Preface

From Quantum Engineering to Applications for Citizens

At the heart of French deep tech, academic and industrial, QUEST-IS 2025, the 1st International Quantum Engineering conference and exhibition, gathered many companies working in the quantum ecosystem from various countries.

It thus addressed technical matters such as quantum computing and algorithms, quantum sensors, quantum communications, crypto and the Internet, including enabling technologies, and obviously quantum engineering targeting applications for citizens.

This multidisciplinary event let attendees from various backgrounds discuss the technical challenges of applied quantum technologies and business opportunities.

We present in this proceedings peer-reviewed papers addressing challenges on quantum engineering.

Quantum Engineering Challenges and Issues for Quantum Computing, Quantum Sensors and Quantum Communication

Quantum engineering denotes the systematic endeavour to transform the abstract formalism of quantum mechanics into practical technologies. It occupies a unique position at the confluence of physics, materials science, control theory, computer science, and systems engineering. The field encompasses three principal technological domains, quantum computing, quantum sensing, and quantum communications, each facing distinct, though deeply interconnected, challenges. Despite their differences, these disciplines share a common difficulty: the engineering of systems that rely upon fragile quantum states, whose coherence can be destroyed by the slightest interaction with their environment.

Quantum computing aspires to create fault-tolerant logical qubits but is constrained by decoherence and the immense overheads of error correction. Quantum sensing seeks ultimate precision, yet must balance coherence preservation with environmental coupling. Quantum communications endeavour to distribute entanglement securely over global distances, but remain limited by photon loss, imperfect memories, and the absence of scalable network architectures.

The future of quantum engineering lies in the emergence of a new discipline, quantum systems engineering, which formalises the design, verification, and integration of quantum devices in the same way that aerospace or microelectronics engineering did for their respective domains. The goal is not merely to understand quantum mechanics but to harness it reliably, reproducibly, and at scale, thereby ushering in the true technological era of the quantum world.

Quantum Computing

In quantum computing, the overarching objective is to construct machines capable of performing computations by manipulating coherent quantum states, superpositions and entanglements, across large numbers of qubits. The central challenge lies in the extreme delicacy of such states. Decoherence, resulting from unwanted coupling between a quantum system and its surroundings, rapidly erases the information encoded in superposition and entanglement. Extending coherence times whilst maintaining precise operational control

constitutes one of the most formidable undertakings in modern engineering. Achieving this requires ultra-pure materials, cryogenic operation at temperatures frequently below tens of millikelvin, and stringent isolation from electromagnetic and vibrational disturbances. Even the smallest imperfections in fabrication or fluctuations in temperature and electromagnetic fields may induce decoherence and computational error.

Given that all physical qubits are inherently noisy, large-scale computation depends critically upon quantum error correction (QEC). Unlike in classical systems, redundancy cannot be implemented by straightforward duplication, since quantum states cannot be cloned. Instead, information must be encoded across entangled collections of physical qubits to form logical qubits that are resilient to noise. Implementing QEC on this scale is immensely demanding: thousands of physical qubits may be required to realise a single fault-tolerant logical qubit. Quantum dissipation engineering is emerging based on quantum feedback for bosonic qubits, to reduce drastically the number of physical qubits and by regulating qubit errors through thermodynamics dissipation (the Lindblad equation). Each qubit must be precisely calibrated, and errors must be identified and corrected in real time, often within microseconds. This imposes stringent requirements upon low-latency control electronics and cryogenic data processing.

A further source of difficulty arises from qubit control and crosstalk. Quantum gates must be executed with extremely high fidelity, typically exceeding 99.9%, if fault-tolerance is to be achieved. The generation of microwave or optical control pulses must therefore be exact, and must not disturb neighbouring qubits. As systems grow in size, the problems of frequency crowding, wiring density, and thermal load increase substantially. One promising avenue involves the development of cryogenic CMOS electronics placed in close proximity to the qubits, thereby reducing latency and electrical noise.

The issue of scalability is equally fundamental. Moving from devices containing a few dozen qubits to those comprising millions necessitates entirely new architectural paradigms. The wiring, signal routing, and thermal management of large cryogenic systems present severe engineering constraints. Modular architectures, wherein several small processors are interconnected by photonic links, may offer a practical path to scale. Another promising approach lies in the integration of heterogeneous qubit technologies, such as superconducting circuits, trapped ions, spin qubits, and photonic systems, into hybrid platforms. Ultimately, progress in this field depends upon a holistic approach known as quantum systems co-design, in which hardware, control electronics, and software are developed in concert, optimising the entire computational stack from materials to algorithms.

Quantum Sensing

Quantum sensing exploits the sensitivity of quantum systems to external perturbations in order to achieve measurements of unprecedented precision. These sensors can detect magnetic and electric fields, gravitational gradients, accelerations, or time intervals with sensitivities far beyond classical limits. Yet the very sensitivity that grants them such power also renders them vulnerable to environmental noise. The fundamental challenge is to preserve coherence long enough to perform a measurement, despite constant interaction with the environment.

Quantum sensors often employ nitrogen–vacancy centres in diamond, trapped ions, or ensembles of cold atoms in interferometric configurations. In every case, the sensor must be sufficiently coupled to the quantity of interest to register a measurable signal, while simultaneously being isolated from unwanted noise. Achieving this delicate equilibrium

requires careful design of electromagnetic shielding, active feedback, and control of the working point so that the sensor operates at regions of reduced susceptibility to fluctuations.

As these technologies mature, miniaturisation and integration have become pressing concerns. Laboratory systems, typically large and fragile, must be transformed into compact, robust devices suitable for deployment in the field. This entails the integration of optical excitation, detection, and control subsystems within constrained physical volumes, as well as effective thermal management. Cold-atom systems, in particular, must balance vacuum integrity, temperature stability, and portability. Long-term calibration and drift correction present further engineering challenges, since even small instabilities in control parameters may degrade sensitivity over time. Autonomous recalibration and machine-learning-based feedback loops are therefore being investigated to ensure sustained performance.

Quantum Communications

Quantum communications seek to employ the principles of quantum mechanics, particularly entanglement and the no-cloning theorem, to enable the secure transmission of information. The best-known application, quantum key distribution (QKD), allows two parties to share cryptographic keys with absolute theoretical security. Beyond this, researchers envisage a quantum internet, capable of distributing entanglement between distant nodes for both secure communication and distributed computation.

The primary obstacle to such a network is photon loss and decoherence during transmission. Photons propagating through optical fibres are absorbed and scattered, restricting the direct distribution of entanglement to a few hundred kilometres. Free-space and satellite channels mitigate this limitation but suffer from atmospheric turbulence, beam divergence, and pointing errors. To overcome these difficulties, the community is developing quantum repeaters, which rely on intermediate nodes equipped with quantum memories to store, purify, and retransmit entanglement. This, however, requires quantum memories with long coherence times, high efficiency, and rapid read–write cycles, capabilities that remain technically elusive.

The development of reliable quantum memories is thus one of the foremost engineering challenges in the field. Candidate systems include ensembles of cold atoms, rare-earth-doped crystals, and spin systems in solid-state matrices. Each exhibits trade-offs between coherence time, storage efficiency, and compatibility with telecommunication wavelengths. The interface between photonic and matter qubits is similarly critical: efficient, low-noise conversion between optical photons and stationary qubits is essential for scalable networking. Approaches such as cavity quantum electrodynamics, optomechanical transduction, and electro-optic conversion are under active investigation, yet none has achieved the required combination of efficiency, fidelity, and scalability.

At a higher level, network synchronisation and control present formidable obstacles. Entanglement distribution and quantum teleportation protocols require extremely precise timing and phase stability across multiple nodes, often at the nanosecond scale. Large-scale quantum networks must integrate quantum channels with classical communication layers for coordination, routing, and error management. The absence of universally accepted standards and protocols compounds these difficulties. Different experimental platforms employ incompatible encodings, operating wavelengths, and modulation schemes. Achieving interoperability across heterogeneous systems will be crucial to realising a practical quantum internet.

Enabling Technologies and System Integration

Across quantum computing, sensing, and communications, certain engineering challenges are universal. Materials science remains a decisive factor: impurities, defects, and surface roughness all shorten coherence times and reduce gate fidelity. Cryogenics presents another formidable constraint; maintaining stable temperatures at the millikelvin level demands complex and expensive refrigeration systems that are difficult to scale industrially. Quantum control theory is assuming an increasingly central role, providing the mathematical framework for designing optimal pulse sequences and feedback algorithms that stabilise quantum dynamics in the presence of noise.

System integration poses further difficulties, since quantum subsystems must interact seamlessly with classical electronics responsible for control, measurement, and feedback. Timing, thermal management, and signal integrity must all be maintained under extreme conditions. Even the task of verification and validation is non-trivial, for measurement inevitably disturbs the quantum state being tested. Benchmarking and certification of quantum devices thus require sophisticated statistical and tomographic techniques. Lastly, the issue of scalability and manufacturability looms large. Many existing devices are artisanal laboratory constructs, far removed from the robust, reproducible systems required for industrial deployment. The transition to mass-manufacturable quantum technologies will necessitate advances in fabrication processes, design automation, and quality assurance.

Reviewing, Support

Again, we present in this proceedings peer-reviewed papers addressing challenges on quantum engineering. We thank the 102 reviewers from 17 countries who provided an average of 3.5 reviews per paper for the submitted 120 reviewed papers, using a single-blind process. Of the reviewed papers, 68 were selected papers for inclusion in these proceedings.

Support for the event came from Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, The Netherlands, Saudi Arabia, Singapore, South Korea, Spain, Sweden, Switzerland, Taiwan, UAE, and the UK, also from the EU, CERN, and Quantum Flagship, and from our Platinum sponsors, Agence Innovation Défense, EDF, IBM, and Thales, and our Gold sponsors, Laboratoire National de Métrologie et d'Essais, CEA, NATO, Alice & Bob, and Quandela.

October 2025

François Gerin
Frédéric Barbaresco

PLATINIUM SPONSOR

AGENCE INNOVATION DÉFENSE

eDF

IBM

THALES

GOLD SPONSOR

LABORATOIRE NATIONAL DE MÉTROLOGIE ET D'ESSAIS LNE

cea

NATO OTAN



ALICE & BOB

QUANDELA

Organization

General Chairs

Frédéric Barbaresco	Thales, France
François Gerin	SEE, France

Local Organizing Chair

Loïc Cantu	SEE, France

Program Committee Chairs

Frédéric Barbaresco	Thales, France
François Gerin	SEE, France

Program Committee

Neil Abroug	Inria, France
Khulud Almutairi	King Saud University, Saudi Arabia
Yasutaka Amagai	AIST, Japan
Nina Amini	CentraleSupélec, France
Dimitris Angelakis	TU Crete and AngelQ Quantum, Greece and CQT, Singapore
Thomas Antoni	CentraleSupélec, France
Sofiane Bahbah	SEDI.ATI, France
Bhashyam Balaji	DRDC Ottawa, Canada
Quentin Barbeau	Danish Embassy, Paris, France
Pascale Bendotti	EDF R&D, France
Mathieu Bertrand	Thales Alenia Space, France
Quentin Bodart	Crystal Q. Computing, France
Nadia Boutabba	IAT, Abu Dhabi, UAE
Philippe Bouyer	Quantum Delta, Netherlands
Harry Buhram	Quantinuum, UK
Andrea Busch	Quobly, France
Marie-Elisabeth Campo	EM Armée air & espace, France
Joan Camps	Riverlane, UK
Victor Canivell	Qilimanjaro, Spain
Patrick Carribault	CEA DAM hpc, France
Adam Connolly	Quantinuum, UK
Franck Correia	DGA, France
Giacomo Corrielli	Ephos/IFN-CNR, Italy
Stefan Creemers	UC Louvain, Belgium
Simon Crispel	GIFAS ALAT, France
Nicolas Daval	Quobly, France
Merouane Debbah	Khalifa University, Abu Dhabi, UAE
Fabrice Debbasch	Sorbonne Université, France
Ivo Pietro Degiovanni	INRIM, Italy
Guillaume De Giovanni	Viqthor, France
Philippe Deniel	CEA, France
Amer Delilbasic	FZ Jülich, Germany

Matthieu Desjardins	C12, France
Alain Dessertaine	La Poste, France
Amanda Diez	CERN, Switzerland
Daniel Dolfi	Thales, France
Christophe Domain	EDF R&D, France
Jens Eisert	Freie Universität Berlin, Germany
Andreas Elben	PSI, Switzerland
Jaap Essing	TNO, Netherlands
Olivier Ezratty	Independent Researcher, France
Nicolas Fabre	Télécom Paris, IPP, QuantEduFrance, France
Thomas Fauvel	Choose Paris Region, France
Benjamin Frisch	CERN, Switzerland
Jacques-Henri Gagnon	Canadian Embassy, Paris, France
Luc Gérardin	University of Sussex, UK
James A. Grieve	TII Abu Dhabi, UAE
Fabrice Guillemin	Orange Labs, France
Stefan Haeussler	Hensoldt, Germany
Pascal Halffmann	ITWM Fraunhofer, Germany
Mélanie Hardman	NPL, UK
David Harvey	Thales, France
Winfried Hensiger	University of Sussex, UK
Mohamed Hibti	EDF R&D, France
Israel Hinostroza	CentraleSupélec/SONDRA, France
Magnus Hoijer	FOI, Sweden
Min-Hsiu Hsieh	Foxconn, Australia
Daniel Huerga	ONERA, France
Toshiyasu Ichioka	RIKEN, Japan
Samy Jousset	Région Ile de France, France
Mathieu Juan	Université de Sherbrooke, Canada
Marc Kaplan	Veriqloud, France
Jian Feng Kong	PIC IHPC, Singapore
Naoya Kono	Secretariat of Science, Technology and Innovation Policy, Japan
Romain Kukla	Naval Group, France
Philippe Lacomme	Université Clermont-Auvergne, France
Jacques-Charles Lafoucrière	CEA, France
Catherine Lambert	Académie des Technologies, France
Aolita Leandro	6G RC Abu Dhabi, UAE
Marc Leconte	F2S, France
Megan Lee	Quantum City, University of Calgary, Canada
Jeremy Leow Kwang Siong	FSTD, Singapore
François-Marie Le Régent	Pasqal, France
Tobias Lindstrom	NPL, UK
Craig Lloyd	NPL, UK
Tony Maindron	CEA, France
Joseph Mikael	EDF R&D, France
Swapan Mandal	Visva Bharati Institute, India
Sabrina Maniscalco	Algorithmiq, Finland
Ulrich Mans	Quantum Delta, Netherlands
Luigi Martiradonna	Riverlane, UK
Francesco Mauro	Sannio University, Italy
Mathias Möller	TU Delft, Netherlands

Mikko Möttönen	Aalto University, Finland
Stéphanie Molin	Thales, France
Liran Naaman	Quantum Delta, Netherlands
Yuichi Nakamura	NEC Corporation, Japan
Hiro Nakata	Jij, Japan
Frank Nielsen	Sony Computer Science Laboratories, Japan
Pierre-Elie Normand	Dassault Aviation, France
Myriam Nouvel	Thales, France
Yasser Omar	PQI (Portuguese Quantum Institute), Portugal
Dimitrios Papadimitriou	Université libre de Bruxelles, Belgium
Cécile Perrault	Alice & Bob, France
Frank Phillipson	TNO NL, OTAN Group, Netherlands
Nico Piatkowski	IAIS Fraunhofer, Germany
Jonathan Pisane	Thales, France
Mathilde Portais	Naval Group, France
Tony Quertier	Orange, France
Setra Rakotomavo	La Poste, France
Matthieu Ratiéville	AID and DGA, France
René Reimann	TII Abu Dhabi, UAE
Florentin Reiter	IAF Fraunhofer, Germany
Hervé Ribot	Minalogic, France
Stéphane Saillant	ONERA/SONDRA, France
Mitsuhisa Sato	RIKEN R-CCS, Japan
Tanguy Sassolas	CEA, France
Hisham Sati	CQTS, NYU, Dubai, UAE
Félicien Schopfer	LNE, France
Jan Schorer	Hensoldt, Germany
André Schrottenloher	Inria, France
Sylvain Schwartz	GIFAS-ONERA, France
Jean Sénellart	Quandela, France
Xavier Sénémaud	SNCF Group, France
Mohammed Serhir	Centrale Supélec/SONDRA, France
Najwa Sidqi	National Quantum Computing Center, UK
Priyank Singh	Aalto University, Finland
Kako Sugiyama	AIST, Japan
Dunlin Tan	Thales Asia, Singapore
Mathieu Taupin	LNE, France
Hiroaki Tezuka	AIST, Japan
Guido Torrese	Thales, Belgium
Malak Trabelsi Loeb	Vernewell Academy, Belgium
Pierfrancesco Ulpiani	Leonardo, Italy
Mathias van den Bossche	Thales, France
Christophe Vasse	Thales, France
Emmanuelle Vergnaud	Teratec, France
Benoît Vermersch	University of Grenoble-Alpes (UGA), LPMMC, and Quobly, France
Daniel Vert	Systematic, France
Stéphane Vialle	CentraleSupélec, Université Paris-Saclay, France
Loïc Viguerie	Cryoconcept, Air Liquide, France
Mark Webber	Universal Quantum, UK
Göran Wendin	RISE, Sweden

Ronin Wu	QunaSys, France
Yu Yamashiro	Jij and Q-STAR, Japan
Jun Ye	A*STAR, Singapore
Su Yi	ED IHPC
Ugo Zanforlin	Leonardo, Italy
Gilles Zalamansky	GIFAS-Dassault Aviation, France
Dominik Zumbuhl	University of Basel, Switzerland

Keynote Talks

The Two Quantum Revolutions: From Concepts to Applications

Alain Aspect [Institut d'Optique - Université Paris-Saclay, France]

The first quantum revolution, based on wave–particle duality, led to the society of information and communication. The second quantum revolution is based on entanglement. Will its applications lead to a new societal revolution?

Quantum Error Correction and Feedback

Pierre Rouchon [Center Automatic and Systems, Mines-Paris, University PSL, Member of Académie des Sciences, France]

Quantum error correction relies on a feedback loop. This feedback generally corresponds to a classical controller. Quantum error correction can also exploit the dissipation associated with the phenomenon of decoherence. Called autonomous correction by physicists, it then uses feedback where the controller is a dissipative quantum auxiliary system. This talk focuses on the development of such quantum controllers to stabilize logical qubits encoded in harmonic oscillators (bosonic code). Two types of encoding will be considered: cat-qubit encoded in two coherent states of opposite phases for which bit-flip errors induced by usual noises can be experimentally almost suppressed; and GKP-qubit encoded in finite energy grid-states approximating position/impulsion Dirac combs where, in principle, both bit-flips and phase-flips could be almost suppressed.

The Interplay Between Quantum Engineering and Quantum Science

Olivier Ezratty [Freelance quantum engineer, mostly known for "Understanding Quantum Technologies"]

Quantum engineering is a relatively new discipline that takes shape as quantum technologies are maturing and turning into commercial products. But what is it exactly? How are science and engineering intermingled in this innovation process? Is the science done, and we are just left with engineering and technology development? What is the engineering scope required for the development of complex quantum systems, particularly fault-tolerant quantum computers? How does quantum engineering connect the dots between the software and hardware stacks? Is the environmental footprint of these emerging technologies integrated in vendors' engineering goals? Are there quantum engineers? How are they trained and how will they be trained?

Quantum Enabling Technologies from Science to Engineering

Richard Versluis [Quantum Enabling Technologies Engineering, TNO / TU Delft, Netherlands]

In this talk Richard will share an overview of the most critical engineering challenges in Quantum Computing that we will face in the upcoming years. He will delve into the transition from academic research to practical engineering, emphasising the complexities and hurdles that need to be overcome. Aspects like scalability, reliability and modularity will be quantified and related to practical design choices and design requirements for future quantum computers and their constituent components.

Experimental Quantum Photonics: From Testing Foundations of Quantum Mechanics to Building Practical Quantum Networks

Djeylan Aktas [Experimental Quantum Communications, Slovak Academy of Sciences, Slovakia]

In this talk we will explore a few seminal experiments in which quantum photonics made it possible to interrogate nature itself by studying the very foundations of quantum theory. From there, we will see how technological progress allowed for scientists to “surf the wave” of the second quantum revolution to actually create practical quantum technologies that can be useful for society by bringing some quantum advantages like enhanced security for modern communication networks, or better sensing capabilities in various domains of applications.

The European Commission’s Vision for Quantum Engineering: Challenges and Opportunities in EU-Funded Projects

Oscar Diez [European Commission Representative, DG CNECT]

In this keynote, I will present the European Commission’s strategy for quantum technologies and its role in building a globally competitive and resilient European quantum ecosystem. The talk will highlight how EU programmes, including Horizon Europe, the Quantum Flagship, the Chips JU and the EuroHPC Joint Undertaking, support the full quantum value chain, from foundational research and engineering to infrastructure, skills and industrial deployment. It will also address current challenges, such as scaling up quantum platforms, strengthening supply chains and fostering innovation across member states, and will outline upcoming opportunities for industry and research communities to engage with EU initiatives in quantum technologies.

Quantum Sensors

Marco Genovese [Director, Quantum Optics Research, INRIM, Italy]

Quantum state coherence and entanglement are very sensitive to interaction with the environment. On the one hand, this represents a problem for developing quantum technologies such as quantum computation or quantum computation, but, on the other hand, it allows for a very high sensitivity to various parameters, leading to the possibility of developing quantum sensors largely surpassing traditional classical sensors. In this talk, I will present this exciting new field, discussing the potentialities of some of the most interesting techniques.

Fair Benchmarking of Quantum Optimisation Applications

Frank Phillipson [Quantum Computing Applications, TNO, Netherlands]

Quantum computing is advancing rapidly, and optimisation is one of its most promising application areas. But measuring progress is not as simple as comparing runtimes or claiming "quantum advantage". In this keynote, I will explore why traditional benchmarking approaches – developed for CPUs and GPUs – fall short when applied to quantum and hybrid systems, and how misleading comparisons risk slowing genuine progress. Building on lessons from classical supercomputing, I will introduce principles for fair benchmarking that emphasise transparency, end-to-end workflows, solution quality, and reproducibility. I will also highlight recent initiatives, from Q-Score to TAQOS, and show how energy use and application-driven metrics are shaping the next generation of evaluation standards. The goal is to provide the community with a practical framework for assessing quantum optimisation responsibly – one that enables researchers, industry, and policymakers to interpret results with confidence and set realistic expectations for the future.

Telecom Integrated QKD Networks: The Madrid QCI Example

Vicente Martin [Quantum Engineering, CSS / Univ. Politécnica de Madrid, Spain]

In the talk I will revise the types of QKD networks deployed in the field, the problems and different choices, concentrating on the solutions used in the Madrid network. This is a multidomain and highly heterogeneous network currently spanning more than 700 km, with 29 nodes hosting 23 QKD pairs from 10 different vendors. Finally several of the use cases implemented will be shown.

Thales' Roadmap and Product Innovations in Cryogenics for Quantum Applications

Christophe Vasse[1], Garmt de Jonge[2], Emilien Durupt[1], Florian Dupe[1], Daniel Willems[2]

[1] Thales LAS France, 26 Av. Jean François Champollion, 31100 Toulouse, France
[2] Thales Cryogenics BV, Hooge Zijde 14, 5626 DC Eindhoven, The Netherlands
christophe.vasse@fr.thalesgroup.com

Abstract. As the quantum technology landscape continues to evolve, the need for advanced cryogenic solutions becomes increasingly crucial. This paper presents Thales innovative products and strategic roadmap aimed at enhancing the capabilities of quantum applications through cryogenics on a range of temperatures (between 2 K to 80 K). We detail Thales state-of-the-art cryogenic technologies, including systems designed for cooling quantum processors, sensors, and other critical components, demonstrating their potential to optimize performance and stability in quantum systems. Furthermore, we outline our vision for the future of cryogenics within quantum technology, highlighting ongoing research initiatives, collaboration efforts, and upcoming product developments. Through a comprehensive analysis of market needs and technological advancements, we emphasize Thales commitment to addressing the challenges of cooling and maintaining quantum coherence.

Keywords: Cryogenics, Sensing, Cryocooler, SWaP, Thales.

1 Introduction

In a rapidly evolving technological landscape, quantum systems are positioned as potential catalysts for transforming the way we process information, ensure the security of communications, and develop innovative solutions across various sectors. Cryogenics plays a vital role in optimizing the performance of quantum technologies by enabling the extremely low temperatures necessary for the manipulation and measurement of quantum sensors or quantum communication. This aspect is clearly identified in several publications [1], [2], [3].

Thales, a global leader in defense, security, and aerospace technologies, is actively engaged in contributing to this quantum revolution. This document presents Thales' roadmap for cryogenics, detailing our vision, objectives, and the necessary steps to develop cutting-edge cryogenic solutions. By aligning our research and development efforts with the needs of quantum applications, we aim to explore the frontiers of innovation while addressing the complex challenges facing the industry.

This roadmap will highlight the approaches and technologies that Thales intends to adopt to create compact, efficient and reliable cryogenic systems for applications between 2 K and 80 K. Through this initiative, Thales aims not only to strengthen its position in the cryogenics domain but also to actively contribute to the emergence of a robust and sustainable quantum ecosystem.

2 Cryogenics and Quantum applications

2.1 Why cryogenics for quantum applications?

Cryogenics is essential for quantum applications primarily due to the need for maintaining extremely low temperatures to enhance the sensitivity and stability of quantum sensors. Quantum systems are inherently delicate, and their operational stability can be compromised by environmental factors such as temperature fluctuations. At cryogenic temperatures, the thermal energy is significantly reduced, which allows qubits—whether they are based on superconducting circuits or trapped ions—to operate with greater fidelity and reduced error rates. Furthermore, many quantum systems leverage superconductivity, where materials exhibit zero electrical resistance and the expulsion of magnetic fields at low temperatures, enabling more efficient quantum computations and information transfer. Therefore, developing advanced cryogenic technologies is crucial for unlocking the potential of quantum computing, enabling breakthroughs in fields such as cryptography, materials science, and complex system simulations. Without effective cryogenic solutions, the performance and scalability of quantum devices would be severely hindered.

At the same time, cryogenics could have a big impact on system size, power consumption, thermal management and maintenance/reliability, etc… Thus cryogenics solutions shall be developed with needs of specific applications in mind.

2.2 Cryogenics for quantum applications

Several applications are targeted for Thales cryocoolers. These applications are mainly linked to quantum sensing and quantum communication. The next table gives examples of applications with their cold temperature range.

Table 1. Quantum applications with their cold temperature range.

Superconducting Quantum Interference Filters (SQIF)	60 – 80 K
RF superconducting filters	20 – 80 K
RF wide range signal treatment with doped crystal	2 – 3 K
Entangled Photon	2 – 3 K
Single Photon counter	2 – 3 K

These applications are described in several reports [2], [3].

These examples address both civil or defense applications. They can be embedded in aircraft or aboard ships as well as be ground based. We can summarize some of the main requirements for cryogenic coolers:

- Low cold temperature and large range of cold temperature (2 K => 80 K)
- Compactness
- Efficiency / Low power consumption
- High availability (24/7 applications)
- No perturbation to the sensor (vibration, noise, electromagnetic compatibility,…)
- Can be easily embedded in aircraft, ships or satellites
- Low constraints regarding export control (dual use or not controlled)

3 Thales roadmap

Thales Cryogenics today is an important supplier of cryocoolers for infrared sensor cooling and applications that require extremely reliable, long life coolers. Such products are ideally suited to be developed into coolers that can be used for quantum sensor cooling. Thales has started a program to develop cryocoolers for quantum sensors along three main axes:

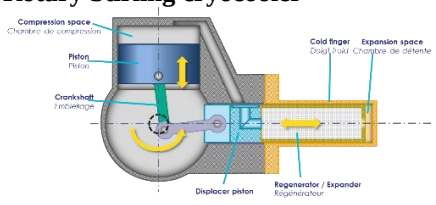
- Optimization of existing cryocoolers regarding efficiency, compactness, and compatibility with sensors,
- Improvement of dewar maturity for quantum sensors,
- Extending our cooling range to 2-4 K .

3.1 40 K - 80 K

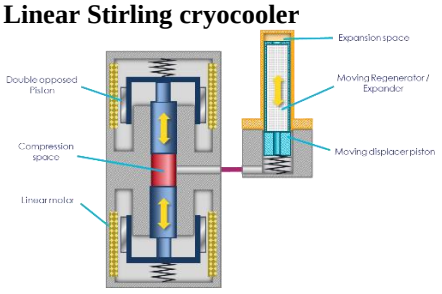
For cold temperature between 40 K and 80 K, Thales Cryogenics intends to use existing COTS Stirling coolers with optimizations/actions:

- Perform Cooler efficiency improvements in order to reach at least 40 K and maybe lower, without a big impact on the SWAP (Size, Weight, and Power),
- Analysis of potential impact of the cooler on quantum sensors (Electro-Magnetic Compliancy or induced vibrations),
- Dewar development for quantum sensors compliant with sensors constraints.

Rotary Stirling cryocooler



Linear Stirling cryocooler





Typical characteristics:
- 60 K ➔ 150 K
- Weight: < 1 kg / Size: < 1 L
- Cooling power: < 1 W @ 80 K
- Electrical consumption: < 6W @ 80 K
- Reliability: > 40 000 h



Typical characteristics:
- 50 K ➔ 150 K
- Weight: < 2.5 kg / Size: < 2 L
- Cooling power: < 2 W @ 80 K
- Electrical consumption: < 50W @ 80 K
- Availability: > 98% after 5 years continuous use

Fig. 1. 40 K-80 K Stirling cryocooler technologies.

3.2 20K - 40K

For the range 20 K - 40 K, Thales Cryogenics will build on its heritage from Space coolers that cover this range, more specific two-stage pulse-tube coolers, by developing non-space /COTS variants.

The next picture shows an example of a two-stage pulse-tube cryocooler for space developed in collaboration with CEA [4]:

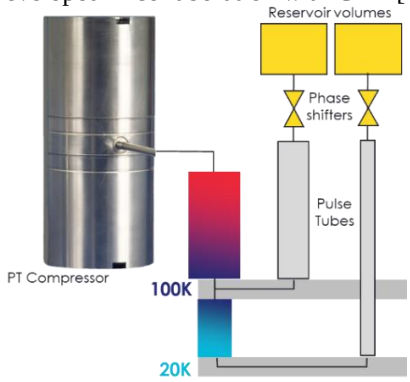




Fig. 2. 20 K - 40 K pulse-tube cryocooler technologies.

Typical characteristics could be (development targets):
- 20 K ➔ 80 K
- Weight: < 15 kg / Size: < 10 L
- Cooling power: > 3 W @ 80 K
  > 200 mW @ 20K
- Electrical consumption: < 200 W
- Availability: > 10 years

These pulse-tube coolers are so-called ‘Stirling-type’ pulse-tube coolers (as opposed to ‘GM-Type pulse-tybe coolers). Stirling-type pulse tube coolers are driven at relatively high frequency by direct compression and expansion of the working fluid. There is no external compressor and no gas lines. Stirling-type pulse-tube coolers are maintenance free, extremely reliable, up to twice as efficient as a comparable GM-type pulse-pulse tube and significantly more compact.

3.3 2 K-4 K

For this lowest temperature range, a combination of several cooler technologies is needed. We aim to combine our strengths in pulse-tube cooler and compressor development and manufacturing with complementary technologies from selected partners. A suitable combination could be the combination of the Thales two-stage pulse-tube cooler mentioned earlier with a closed loop Joule Thomson cooler. We aim to have a portable, rack mountable solution that is highly reliable, scalable and cost effective.

The next schematic explains how a cooler solution could be envisaged

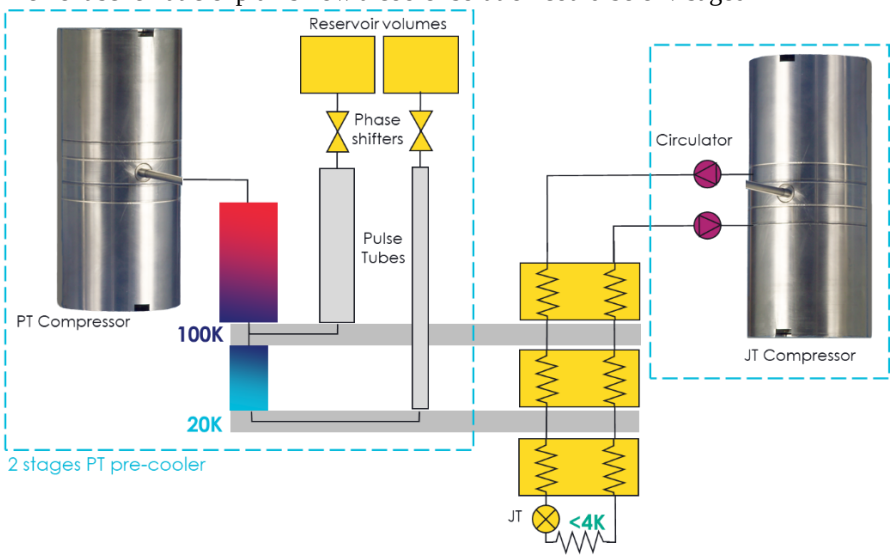


Fig. 3. 2K - 4K cryocooler schematic

Typical characteristics we aim based on feasibility studies are:

- 2 K ➔ 4 K
- Weight: < 20 kg
- 19 inches, rack mountable
- Cooling power: < 3 W of cold @ 80K
  < 10-15 mW of cold @ 2 K
- Electrical consumption: < 300 W
- Availability: > 10 years

4 Conclusion

This paper has outlined Thales' strategic roadmap for advancing cryogenic technologies tailored to the demanding requirements of quantum applications. By leveraging our expertise in low-temperature engineering, thermal management, and integration of scalable cryogenic systems, with our capability to manufacture coolers in large numbers, we aim to enable compact, robust, efficient, and reliable quantum devices operating in the 2 K – 80 K temperature range. Our roadmap addresses critical challenges such as high availability, manufacturability and cost, efficiency, minimizing perturbations, designing dewars, and ensuring system scalability and maintainability, all of which are essential for the transition from laboratory prototypes to industrial quantum solutions. Moving forward, Thales is committed to pushing the boundaries of cryogenics to support the emerging quantum ecosystem, fostering innovation across communication, sensing, and computation domains.

Thales Cryogenics is part of CryoNext, which is the cryogenics part of the French quantum plan.

References

1. Stratégie nationale sur les technologies quantiques – France relance (2021)
2. Rapport sur les technologies quantiques pour les applications Aéronautique-Spatial-Défense, GIFAS (2024)
3. Ezratty, O.; Understanding quantum technologies (2024)
4. Development of a 30–50 K dual-stage pulse tube space cool-er, H. Leenders, G. de Jonge, J. Mullié, T. Prouvé, I. Charles, T. Trollier, J. Tanchon, Cryogenics 88 (2017) 91–100

Preliminary Demonstration of Spin Qubit Control Using the Quantum Instrumentation Control Kit (QICK)

Dana El Hajj[1], Vivien Schmitt[2], Xavier Jehl[2], Benjamin Criton[3], Guillaume De Giovanni[1], Jérémie Theze[1], and Arnaud Perrin[1]

[1] Viqthor, 78100 Saint-Germain-en-Laye, France
[2] Univ. Grenoble Alpes, CEA, Grenoble INP, IRIG, PHELIQS, 38000 Grenoble, France
[3] Université Paris-Saclay, CEA, IRFU, Gif-Sur-Yvette, 91191, France

Abstract. In this paper, we present initial results demonstrating the use of the Quantum Instrumentation Control Kit (QICK) to perform measurements on a spin Qubit. We used the RFSoC to perform energy selective spin readout through a resonator coupled to the qubit. Thanks to of a frequency up-converter we also drive the spin at the Larmor frequency beyond the RFSoC range, typically 15GHz. These early-stage measurements provide evidence that QICK can be adapted for spin qubit control. This work opens a path towards using an Open-Source Hardware platform (QICK) to characterize and control Spin Qubits.

Keywords: QICK , RFSoC , Spin qubits , control electronics.

1 Introduction

Silicon spin qubits are among the leading platforms for scalable quantum information processing, offering long coherence times, high-fidelity operations, and compatibility with industrial fabrication processes [1][2]. Hole spin qubits, in particular, benefit from strong spin–orbit coupling, allowing for all-electrical qubit control via electric-dipole spin resonance (EDSR) [2]. However, implementing such control presents a significant challenge because of the requirement for precise, flexible, and low-latency electronics. In this work, we demonstrate the use of an FPGA-based solution for qubit control and readout using the Quantum Instrumentation Control Kit (QICK)[4].

Originally designed for superconducting qubits, we adapted QICK for the control of a single-hole spin qubit in a quantum dot formed in a natural silicon nanowire. This system performs resonator-based readout, qubit manipulation, and data acquisition. We validated this platform with a sequence of typical spin-qubit experiments performed on a characterized qubit [9] running in a dry dilation refrigerator at 100 mK.

2 Background and Related Work

2.1 Spin Qubits

Hole spin qubits in silicon nanostructures offer fast and electrically driven spin control due to the strong spin-orbit coupling in the valence band [5]. Recent advances have demonstrated high-fidelity[6] and long coherence in nanowire geometries and SiGe/Ge/SiGe heterostructures.[3]

2.2 Quantum Instrumentation Control Kit (QICK)

The Quantum Instrumentation Control Kit (QICK) [4] is an open-source qubit control platform built around a Xilinx RFSoC. The RFSoC integrates digital-to-analog converters (DACs), analog-to-digital converters (ADCs), and programmable logic within a single device, making it an ideal platform for waveform generation, acquisition, and feedback.

The use of an RFSoC significantly reduces latency and allows real-time control logic; this makes it an ideal platform for quantum experiments which require fast and deterministic control. In addition to the architectural advantages of the RFSoC, QICK is mainly made up of a timed processor and a combination of different signal generation and readout blocks, all of which can be modified by the user and tailored to his specific needs. This makes QICK extremely customizable. Not to mention that QICK's software layer is based on Python, a very popular programming language, making it accessible to many users. Not to mention that QICK's software layer can be used by anyone as there are high-level classes and functions provided for ease of us, while also providing the option to write the codes on a register level or through assembly if needed. Moreover, it is possible to integrate QICK with QCoDeS, a widely used Python-based framework for instrument control in quantum research. This makes it straightforward to incorporate QICK into automated measurement routines and data acquisition workflows, which is crucial for reproducible and scalable quantum experiments.

In summary, the QICK platform enables high-resolution pulse generation and digitization on a single FPGA chip with real-time processing. It has been used in superconducting systems[4], but applications to spin qubits remain underexplored.

3 Experimental Setup

3.1 Control Hardware

The required pulses were generated using VIQTHOR's MIMOTHOR platform (**Fig. 1**) which features an RFSoC with QICK installed on it. Through MIMOTHOR, the user can output RF or DC signals, both of which are needed to perform spin qubit

readout and control experiments. Furthermore, the qubit resonance frequency of the device tested was in the 12–18 GHz range which is beyond the output range of the RFSoC. To overcome this limitation and reach the qubit resonance frequency, we relied on VIQTHOR's up converter module (**Fig. 1**). The QICK FPGA generated intermediate-frequency (IF) signals between 1-2GHz, which were mixed with an LO tone to produce the high-frequency pulses needed.

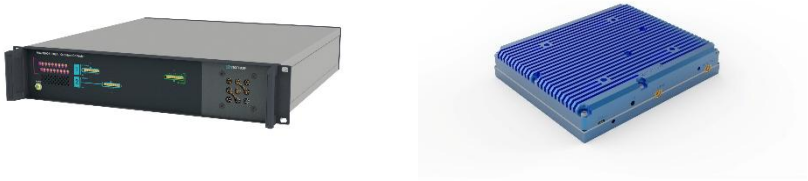


Fig. 1. VIQTHOR's MIMOTHOR Platform (left) and VIQTHOR's Up Converter Module (right) which takes a signal between 1-2GHz and outputs a signal between 2-18GHz

3.2 Spin Qubit

The tests were performed at CEA Grenoble on a single-hole spin qubit in a quantum dot formed in a natural silicon nanowire. The qubit has been thoroughly characterized as reported in [9]. Furthermore, a high-resolution, high-stability iTest DAC system was used to generate static bias voltages.

4 Measurements Performed

4.1 Resonator Spectroscopy

We start by characterizing the RF resonator response by sweeping the probe frequency and measuring the reflected signal. **Fig. 2** shows the reflection spectrum in amplitude, where a clear resonance dip is observed. This resonance shifted in response to charge transitions in the sensing dot. This frequency shift was used to determine the optimal operating point for subsequent readout measurements. Although amplitude is shown here, in practice we primarily use the phase response for high-sensitivity readout measurements, because the phase shows a sharp transition near the resonance.

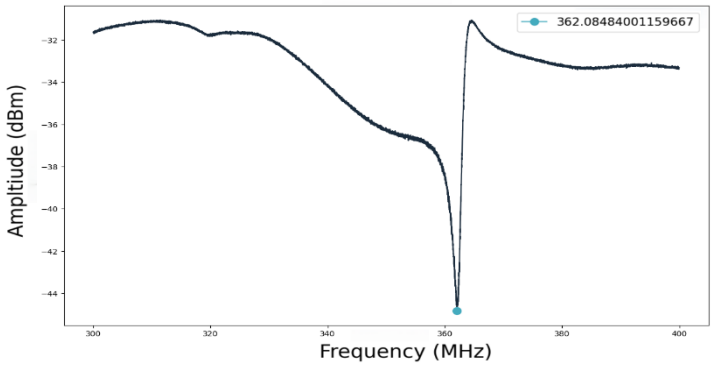


Fig. 2. Resonator Spectroscopy. The sharp dip indicates the resonance frequency of the coupled readout resonator, measured at approximately 362.08 MHz, which is used for subsequent qubit state readout

4.2 Energy-selective spin readout

The so-called ”Elzerman” spin readout scheme relies on connecting the quantum dot to a reservoir and adjust their relative position in energy. When the Fermi level in the contact lies in between the spin up and spin down levels, a spike in current can be observed if the spin was initially up, assuming the tunnel rates are appropriate and the detection fast enough. In contrast if the spin was down, no current ’blip’ is observed [7]. Therefore this readout scheme requires acquiring the resonator’s response for a relatively long time (hundreds of microseconds), keep each trace (see **Fig. 3a**) and accumulate statistics to obtain the spin up probability. This long readout is performed after a series of DC pulses allowing to place the system in any region where we want to manipulate the spin before bringing it back to the readout position (see details in Fig.1 of supplementary materials of [8]).

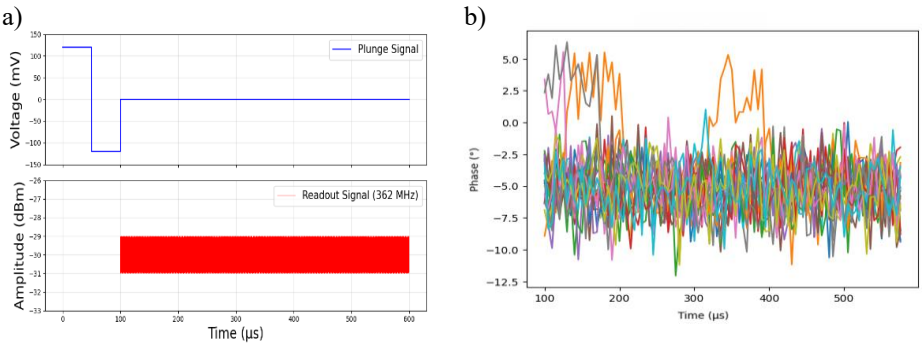


Fig. 3. a) The Plunge and Readout signals needed to perform the Elzerman Readout. b) 20 Elzerman Readout traces, some of them featuring a ’blip’ indicating a spin up

Fig. 3b shows all the individual traces obtained after sending the pulses mentioned

above twenty times. Out of the twenty experiments sent, only some of them will exhibit a phase shift, often called a 'blip' (i.e., a phase value greater than 0). Performing the experiment several times allows us to compute the probability of observing a phase shift, i.e. the probability of having a spin up. Once we are able to detect a phase shift using Elzerman, we can move on to other tests.

4.3 Spin Tail

As shown in **Fig. 4**, we perform a sweep of gate voltage values on T3 (B2 is adjusted accordingly) while measuring the reflected phase signal over 500μs following the initialization (plunge) pulse. The resulting phase map reveals a "tail"-like structure corresponding to the spin-dependent tunneling dynamics. This feature helps us tune the gate voltages for optimal spin readout contrast

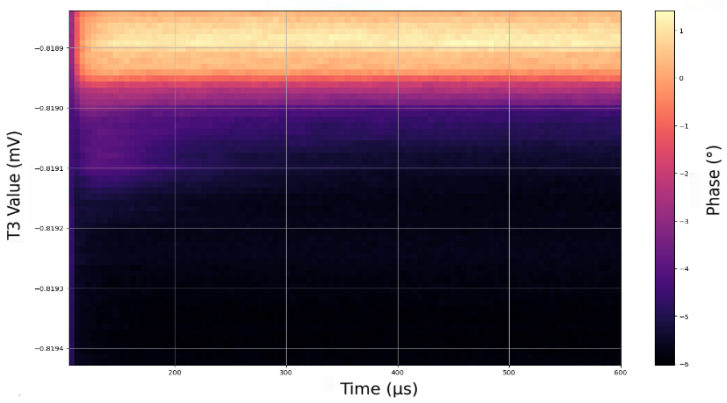


Fig. 4. Phase Reflectometry measurement for various gate voltage values on T3.

The spin tail measurement is critical in order to validate control fidelity and optimizing readout performance when controlling a qubit.

4.4 Qubit Spectroscopy

A qubit spectroscopy is performed to find the qubit's frequency frequency which is determined by the Landé factor g, the Bohr magneton μ_B and the applied static magnetic field which lifts the degeneracy between spins up and down and therefore creates the qubit's 2-level basis: $f = g \times \mu_B \times B$. Since the frequency of a spin qubit is between 12-18GHz, we utilized VIQTHOR's up converter module which takes an input signals between 1 GHz and 2 GHz and shifts them to a frequency within a 2–18 GHz range while maintaining consistent output power and minimal spectral distortion across the full bandwidth. Therefore in our measurements, we use MIMOTHOR to generate an input frequency between 1-2GHz, which will be sent to the up converter module who will output the signal at the user specified frequency. **Fig. 5a** shows the

signals needed to perform the qubit spectroscopy.

Sinusoidal Microwave signals (the drive pulse) are sent to the qubit for a duration of 2µs. The drive frequency is varied. If the drive matches the resonance frequency of the qubit, the hole escapes the dot and triggers a noticeable blip in the reflectometry signal. However, if the drive frequency deviates from the resonance frequency, no changes can be seen in the reflectometry signal as the hole spin remains in its ground state. This means that the resonance frequency is the frequency where we have the highest probability of observing blips, i.e. the highest probability of spin up.

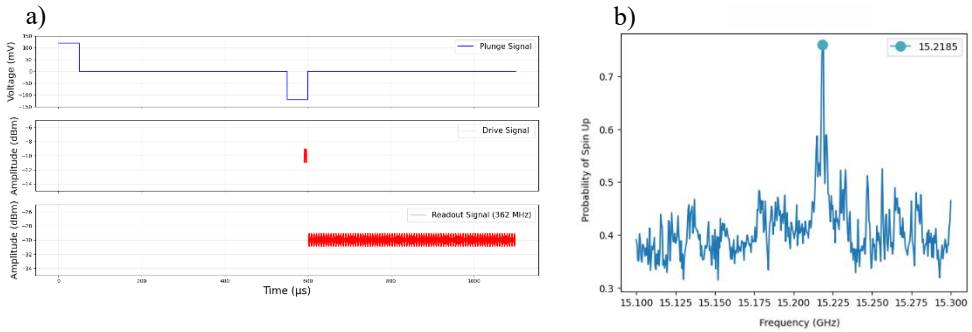


Fig. 5. a) The Plunge, Drive and Readout signals needed to perform Qubit Spectroscopy. b) Output of the Qubit Spectroscopy. The plot shows the probability of spin up as a function of frequency, with a clear resonance peak observed at approximately 15.2185 GHz.

To precisely determine the spin qubit resonance frequency, frequency sweeps with fine step resolution are required. However, when the qubit frequency is initially unknown, performing a high-resolution sweep over a large frequency bandwidth can be time-consuming. To overcome this, we could drive the qubit with a chirp signal. The Chirp signal enables us to rapidly scan a wide frequency range which allows us to identify the approximate resonance region. Once the resonance region is found, we can perform a frequency sweep with fine steps.

4.5 Rabi Oscillations

To characterize the coherent control of the spin qubit, we performed a Rabi oscillation experiment by applying resonant microwave pulses of varying duration. By increasing the pulse length, we notice oscillations in the qubit state population (**Fig. 6**), which correspond to coherent rotations of the spin state on the Bloch sphere. Rabi Oscillations measurements confirm our ability to drive the qubit.

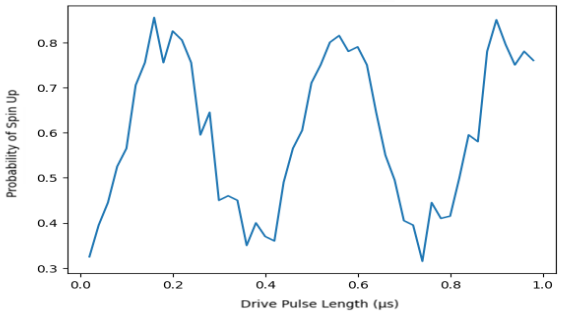


Fig. 6. Rabi Oscillations. The plot shows the probability of spin up as a function of the drive pulse length.

4.6 Rabi Chevron

The Rabi Chevron experiment is done by sweeping both the drive frequency and pulse duration in order to map the coherent Rabi oscillations of the hole spin qubit versus frequency detuning. The resulting 2D plot (**Fig. 7**) displays a chevron-shaped interference pattern, which indicates coherent oscillations occurring near the resonance frequency. The chevron peak corresponds to the qubit's resonance frequency, and the fringe spacing reflects the Rabi frequency of 2.5 MHz as a function of detuning.

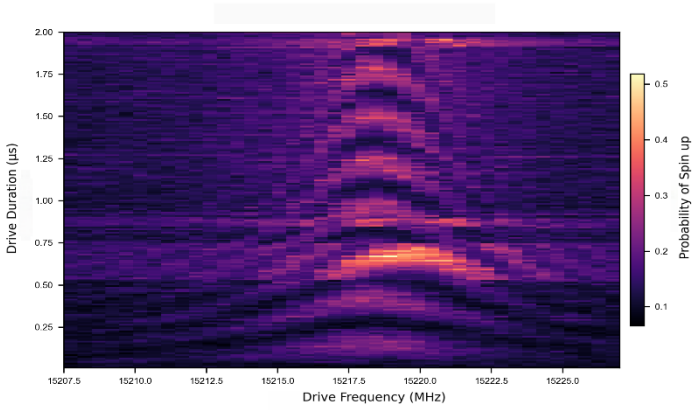


Fig. 7. Chevron plot showing qubit excitation probability as a function of drive frequency and pulse duration. The resonance condition is marked by the peak of the chevron. A shift in the qubit frequency over time is visible.

During this measurement, we observed a gradual shift in the resonance frequency, visible as a horizontal drift in the chevron peak over time. This shift is likely due to environmental or device-level factors such as gate voltage drift due to prolonged

operation or thermal cycling caused by repeated measurements

4.7 Resonance Frequency Drift

The Ramsey Experiment consist of two π/2 pulses separated by a variable delay time. As previously mentioned, during the Rabi Chevron measurement, we observed a shift in the qubit resonance frequency of approximately 1 MHz. Therefore, the subsequent Ramsey Experiment was performed using the pre-drift resonance frequency which resulted in reduced fringe contrast and rapid dephasing, as seen in **Fig. 8**.

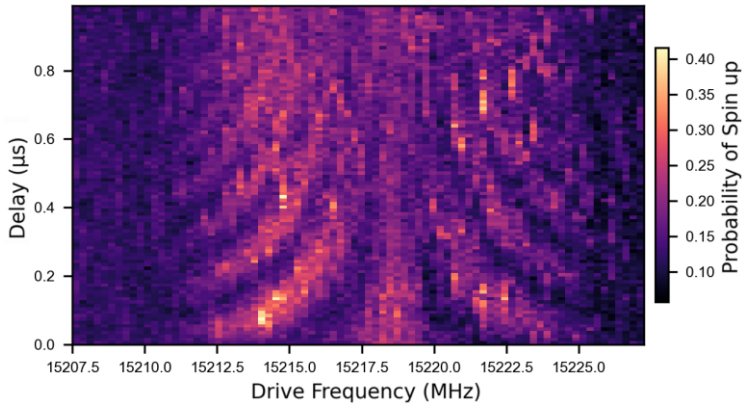


Fig. 8. Ramsey measurement taken after a qubit resonance frequency shift observed during the preceding chevron scan. The drive frequency was not updated to the new resonance after the jump observed in **Fig. 7**, leading to fast decoherence and reduced fringe visibility

To maintain an accurate control over the qubit, resonance calibration must be done frequently in order to account for the shifts in the qubit frequency. The Ramsey experiment's result highlights the necessity of incorporating automated resonance tracking or real-time frequency tuning in future works.

5 Conclusion

In this work, we demonstrated control and readout of a single hole spin qubit in a silicon nanowire using VIQTHOR's MIMOTHOR platform which is based on the QICK platform, as well as VIQTHOR's up converter module. Through a series of characterization measurements which include: resonator spectroscopy, Elzerman readout, spin tail, qubit spectroscopy, Rabi, Rabi Chevron and Ramsey; we were able to validate the effectiveness of the control hardware in generating and delivering the needed pulses accurately.

Additionally, the Ramsey experiment revealed a rapid decay of coherence due to

a frequency drift observed during the rabi chevron experiment. This highlights the sensitivity of spin qubit control to environmental or device-level variations thus emphasizing the need for frequent resonance frequency recalibration.

The results presented in this work show promise of using QICK as a platform for spin qubit control. This work validated the system's ability to generate the needed pulses for all the experiments performed and thus establishes a foundation for scalable and low-cost control architectures for spin qubits.

This work was supported (in part) by the French National program "Programme d'investissement d'avenir, IRT Nanoelec, n° ANR-10-AIRT-05 (Q-Loop)" as well as the CEA PTC program (ERDAQS)

References

1. Stano, P., et al. (2022). Review of performance metrics of spin qubits in gated semiconducting nanostructures. Nature Reviews Physics, 4(10), 672–688. https://doi.org/10.1038/s42254-022-00484-w
2. Burkard, G., et al. (2023). Semiconductor spin qubits. Reviews of Modern Physics, 95(2), 025003. https://doi.org/10.1103/RevModPhys.95.025003
3. R. Maurand et al., "A CMOS silicon spin qubit," Nature Communications, vol. 7, no. 1, Nov. 2016. doi:10.1038/ncomms13575
4. L. Stefanazzi et al., "The Qick (quantum instrumentation control kit): Readout and control for qubits and detectors," Review of Scientific Instruments, vol. 93, no. 4, Apr. 2022. doi:10.1063/5.0076249 arXiv:2111.04249, 2021.
5. B. Voisin et al., "Electrical control of g-factor in a few-hole silicon nanowire MOSFET," Nano Letters, vol. 16, no. 1, pp. 88–92, Dec. 2015. doi:10.1021/acs.nanolett.5b02920
6. Wang, C.-A., et al. (2024). Operating semiconductor quantum processors with hopping spins. Science, 385(6707), 447–452. https://doi.org/10.1126/science.ado5915
7. J. M. Elzerman et al., "Single-shot read-out of an individual electron spin in a quantum dot," Nature News, https://www.nature.com/articles/nature02693
8. B. A. Piot et al., "A CMOS silicon spin qubit demonstrating fast control," Nature Nanotechnology, vol. 17, pp. 1072–1077, 2022. DOI: 10.1038/s41565-022-01196-z. 11
9. Marion Bassi. Tunable resilience to charge noise of a hole spin. Physics [physics]. Universit´e Grenoble Alpes [2020-..], 2024. English. ⟨NNT : 2024GRALY018⟩. ⟨tel-04718717⟩

Microwave interconnectivity as a scalable solution for Quantum Computing: Engineering, Measurement, and Standardization

Laurent PETIT(laurent.petit@radiall.com), Evan ICHIR, Fabrice JANOT, Fleury GRANDJEAN, Guillaume MARION ,Claude BROCHETON, Thierry LE NADAN, Bastien HUON, Jacques MARTINET, Julien LEGRAND

Radiall, 642 Rue Emile ROMANET, Voreppe, France

Abstract. The transition from prototype to large-scale quantum computers requires a new class of microwave interconnect solutions capable of operating reliably at ultra-low temperatures. This paper presents a modular cryogenic interconnect framework designed for compatibility with superconducting quantum processors, addressing critical requirements for signal fidelity, thermal anchoring, and system scalability. The proposed architecture integrates cryogenic attenuators, filters, and switches optimized for operation in dilution refrigerators down to 10 millikelvin. Multiphysics modeling, including electro-thermal co-simulation and mesoscopic heat transport analysis, guides component design. Measurement validation combines calibrated RF methods with qubit-based performance metrics to assess thermal noise impact and signal integrity. Standardized calibration and benchmarking protocols ensure reproducibility across platforms. This work supports ongoing standardization efforts within CEN/CENELEC and paves the way toward scalable, low-noise quantum infrastructures.

Keywords: Cryogenic Microwave Interconnects, Quantum Hardware Scalability, Thermal and Electromagnetic Co-Simulation, Sub-Kelvin Metrology, Standardized RF Component Integration

1 Introduction

1.1 Microwave interconnectivity solution to scale up quantum computer

Quantum computing promises to revolutionize industries ranging from molecular simulation to optimization. However, the practical deployment of such processors demands precise control and readout of qubit states. These processes rely on signal integrity, minimal cross-talk, and thermalization, particularly within cryogenic environments. This paper is structured as follows: Section 2 reviews design functions and constraints. Section 3 details key components. Section 4 covers metrology, and Section 5 discusses standardization efforts. Microwave interconnectivity plays a central role in ensuring these conditions. This domain intersects materials science, RF and microwave engineering, cryogenics, and systems-level integration.

Components such as attenuators, filters, cables, and connectors must preserve their electrical characteristics while operating at temperatures below 100 mK, with tight constraints on electromagnetic compatibility and thermal management.

2 Cryogenic Microwave Interconnects: Functions and Design Requirements

2.1 Functional Roles in Quantum Architectures

Microwave interconnects serve as conduits for qubit control and readout signals for systems scaling from tens to potentially millions of qubits. These include:

- Drive lines that deliver RF pulses to qubits
- Readout chains capturing quantum states via resonators and amplifiers
- DC biasing lines and shielding connections

Each line is subject to frequency-specific losses, phase noise, thermal noise injection, and electromagnetic crosstalk. Control lines typically span DC to 12 GHz, while readout paths extend beyond 20 GHz in dispersive regimes. And the constraints are rising with the "scale up" from tens [1] to several hundred [2] and even millions of qubits that must be precisely controlled and measured [3].

2.2 Multiphysics Design Constraints

Operating in a dilution refrigerator introduces several interdependent constraints:

- Thermal anchoring at each stage (300 K, 3 K, 800 mK, 100 mK, 10 mK)
- Minimized attenuation at low temperatures, preserving signal amplitude
- Avoidance of thermal shorts and vibration sensitivity
- Material selection: superconducting alloys (NbTi), copper-nickel, stainless steel, alumina substrates

3 Component-Level Developments

3.1 Cryogenic Attenuators

Resistive attenuators function as both thermal noise suppressors and signal conditioners, improving qubit readout fidelity and stability. Their design must reconcile the Wiedemann–Franz law and Kapitza resistance, ensuring heat extraction without excessive RF loss. The attenuator structures are based on resistive layers (e.g., TaN or NiCr), embedded in high thermal conductivity ceramics (e.g., AlN). Designs include cartridge-based modules and ganged arrays for high-density integration.

Modeling includes:

- Heat flow simulation using finite element tools
- Mesoscopic modeling of phonon bottlenecks, characterized by ballistic transmission and Kapitza resistance metrics.
- Coupled electromagnetic-thermal co-simulation

Measurements involve both calibrated S-parameters and indirect thermometry based on qubit decoherence and readout fidelity shifts.

Sub kelvin Attenuator Design and Characterization

At sub-Kelvin temperatures, attenuator performance must be characterized not only in terms of RF properties (e.g., insertion loss, VSWR). At millikelvin temperatures, attenuators must exhibit predictable thermal dissipation while maintaining low insertion loss and impedance stability. Mesoscopic modeling captures phonon transmission and heat backflow from resistive zones, and is validated by comparing device-integrated thermometry with indirect readout fidelity degradation in qubits exposed to thermal noise. Such combined approaches enable refining the resistive film design, thermal interface resistance, and ceramic substrate layout.

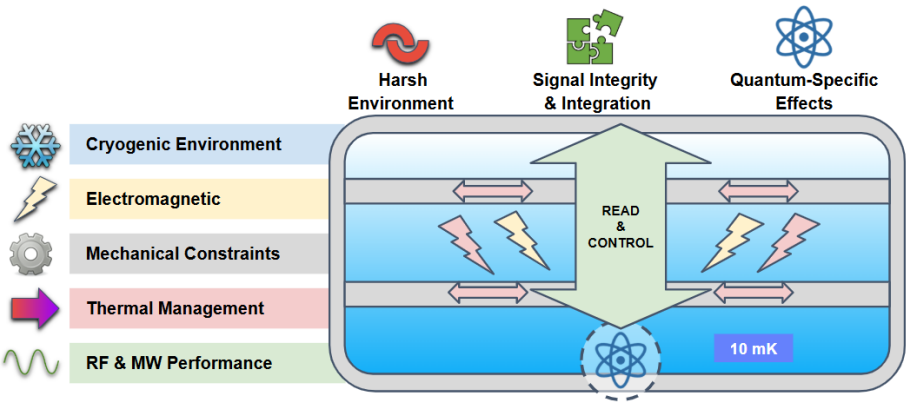


Fig. 1. R&D Constraints for RF&MW Interconnectivity for QT

3.2 Cryo Switches and Multiplexers

Switches are essential for reducing the cable count and enabling reconfigurable experiments. Cryogenic SPDT to SP8T switches are designed to operate with low actuation power (<10 mW), enabling use in 10 mK environments. SMT versions are tested down to 20 mK, showing temperature resilience over multiple actuation cycles.

A key innovation involves packaging with high-conductivity materials, controlled magnetic susceptibility, and hybrid actuation circuits with minimal transient heat load.

3.3 Filters and Infrared Blockers

Filters mitigate spurious modes, reflections, and infrared noise. Their performance in cryogenic conditions depends on:

- Precise impedance matching
- Effective IR absorption (using resin composites or crystalline structures)
- Low-pass filters with cutoff frequencies below 20 GHz

Evaluation involves both network analysis and qubit-based reflectometry techniques monitoring changes in qubit thermal population and parity switching before and after filter integration.

Infrared Filter Thermal Design Validation

Infrared (IR) filters, typically implemented with resin-embedded or multilayer absorbing structures, are essential to block high-energy thermal photons that would otherwise perturb qubit states. Their thermal design effectiveness is indirectly validated by monitoring qubit excited-state populations, parity switching rates, or qubit thermalization dynamics before and after integration of filter modules. These measurements complement electromagnetic simulations and provide system-level evidence of IR filtering efficacy under realistic operating conditions.

4 Integrated Interconnectivity Architectures

Modular sideloaders aggregate over 100 microwave channels across various thermal stages. These systems integrate:

- Solderless, hermetic SMPM clusters
- Bundled semi-rigid or flexible cryogenic cables (CuNi, NbTi)
- Embedded attenuator and filter assemblies

CAD-based mechanical design is complemented by thermal anchoring analysis, and mechanical stress simulations under cooldown cycles. Signal integrity is validated using time-domain reflectometry and noise figure extraction.

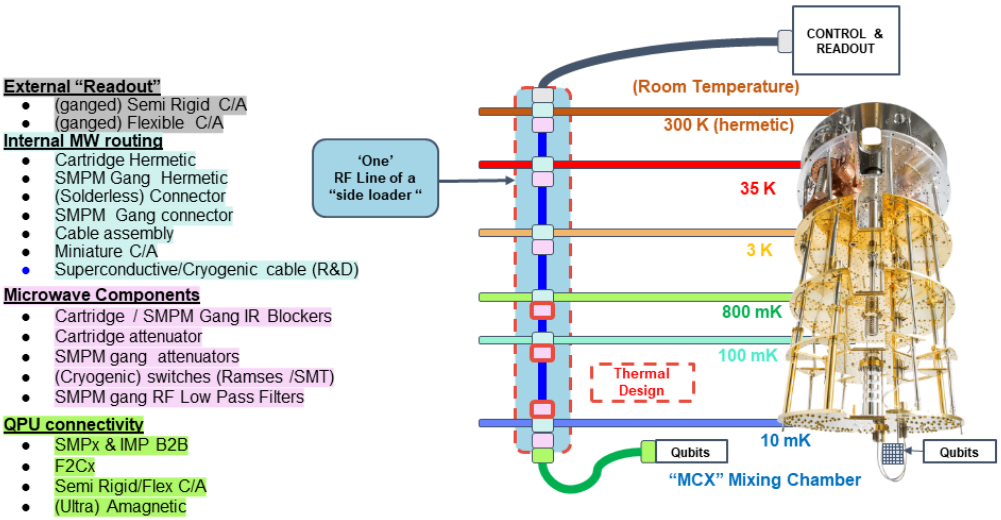


Fig. 2. Integrated Interconnectivity Architectures

5 Cryogenic Metrology and Test Platforms

Precision in RF measurements at cryogenic temperatures is non-trivial. A robust metrological framework includes:

- Cryostat-integrated test benches equipped with a Vector Network Analyzer (VNA)
- TRL (Thru-Reflect-Line), pulse modulation, and arbitrary waveform control
- Temperature-staged calibration using TRL and SOLT methods adapted to cryogenic interfaces
- Local thermometry and thermal imaging for characterizing heat dissipation in components

Both direct methods — such as network analyzer S-parameter measurements, time-domain reflectometry, and resonator Q-factor extraction, and indirect methods such as qubit decoherence signatures, photon statistics, tunneling rates, and parity switching analysis — are employed to assess the behavior and stability of components under cryogenic conditions.

Uncertainty quantification is carried out using both the Guide to the Expression of Uncertainty in Measurement (GUM) and Monte Carlo propagation. The latter estimates uncertainty by simulating numerous outcomes based on probability distributions of input variables, particularly valuable when analytical models are impractical.

In parallel, work is ongoing to improve thermal anchoring reproducibility and to establish benchmarking protocols across laboratory platforms, all coordinated with system-level modeling efforts.

6 Standardization Initiatives

From a measurement and evaluation perspective, standardization aims to:

- Define traceable metrics and test protocols for cryogenic performance
- Harmonize both direct and indirect measurement methods across test sites
- Build consensus on connector and cable specifications to ensure compatibility and integration

These efforts are part of a broader strategy to bridge laboratory R&D and scalable manufacturing by embedding reproducibility, comparability, and traceability into the qualification of quantum interconnects. Standardization is a strategic pillar for the industrialization of quantum technologies. Target parameters include an insertion loss below 3 dB, a VSWR under 1.3, and a heat load below 5 µW per channel at 10 mK.

Radiall coordinates the interconnectivity working group within JTC22 in its scope of "enabling technologies" with the objective of finalization and ongoing adoption within ISO JTC3: Quantum Technologies [5].

7 Modular System Architecture for Cryogenic Quantum Platforms

Building on the above, a comprehensive cryogenic interconnectivity offer has been structured to meet both R&D and industrial scaling needs. This includes:

- Customizable sideloaders with up to 208 coaxial lines using SMPM clusters
- Thermally-optimized attenuator modules for each temperature stage (300K to 10 mK)
- Filter blocks and IR absorbers designed to minimize thermal radiation and EMI
- High-density harnesses composed of flexible and semi-rigid cryogenic cables based on CuNi or NbTi alloy conductors
- Integration support: mechanical CAD, thermal anchoring, signal routing, and EMI modeling

This modular architecture is validated through mK testbeds and aligned with current measurement practices, offering scalable, low-noise, and reproducible signal paths tested with qubit setups showing <2% thermal-induced fidelity loss across 1–20 GHz."

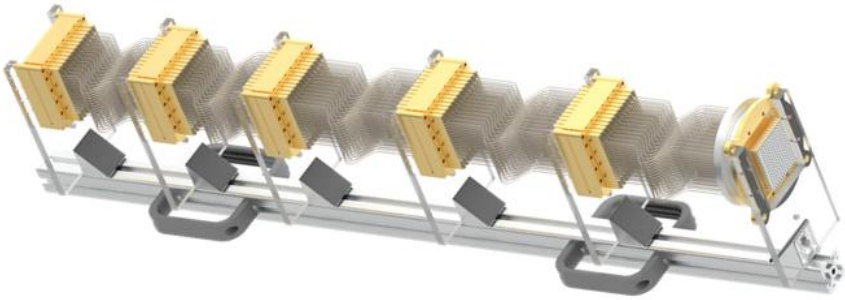


Fig. 3. High Density Sideloader integrating Microwave Cryogenic Interconnectivity [6]

8 Conclusion and Perspective

Microwave cryogenic interconnectivity has emerged as a strategic enabler of scalable quantum computing. Innovations in component design, thermal engineering, and test methodology contribute to reducing noise, improving integration, and enabling high-density scaling.

Future work includes:

- Full qualification of flexible micro-coaxial and FPC-based harnesses
- Deployment of modular sideloaders in production dilution refrigerators
- Expansion of inter-laboratory validation and reference datasets

As quantum processors evolve toward 10k–100k qubit systems, interconnectivity must evolve from a bespoke subsystem into a mature, standardized backbone that integrates RF, thermal, mechanical, and systems engineering into a coherent quantum infrastructure.

References

1. S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinsoo, R. Keller, J. Luetolf, C. Eichler, and A. Wallraff, "Engineering cryogenic setups for 100-qubit scale superconducting circuit systems," *Rev. Sci. Instrum.*, vol. 90, no. 12, pp. 123528, Dec. 2019.

2. N. Masluk, “Cryogenic infrastructure for 400 qubits and beyond,” in *Proc. Cryogenic Engineering Conf./Int. Cryogenic Mater. Conf. (CEC/ICMC)*, Honolulu, HI, USA, Jul. 2023.
3. M. Mohseni et al., “How to build a quantum supercomputer,” *arXiv preprint*, arXiv:2411.10406, 2024. [Online]. Available: https://arxiv.org/abs/2411.10406
4. CENELEC, “Quantum Technologies Committee,” CEN-CENELEC, 2023. [Online]. Available: https://www.cencenelec.eu
5. ISO/IEC JTC 1/SC 42, “Quantum Technologies,” International Organization for Standardization (ISO). [Online]. Available: https://www.iso.org/committee/10138914.html
6. Radiall, “Quantum Technologies — Enabling Scalable Interconnectivity,” *Radiall Industries*, 2023. [Online]. Available: https://www.radiall.com/industries/industrial/quantum-technologies.html?f=0.35382200%201744907020

Scalable Photonic Control System for Large Scale Quantum Computers

Abhilash Amsanpally[1, *], Maeva Franco[2], Arnaud Perrin[1], Jérémie Theze[1], Dana El Hajj[1], Guillaume De Giovanni[1]

[1] VIQTHOR S.A.S, Saint-Germain-en-Laye, France
[2] Lab-STICC, Ecole Nationale d'Ingénieurs de Brest, France
abhilash@Viqthor.com

Abstract. Scaling quantum computers require control systems that are both high-performance and cryogenically efficient. Traditional microwave-based electronics face limitations in such environments due to heat generation, signal loss, and electromagnetic interference. To address these challenges, we present a fully photonic control architecture based on Viqthor's Menthor, a multi-channel RF-over-Fiber (RFoF) transmitter designed for quantum computing applications. Menthor converts microwave signals up to 18 GHz into optical signals, transmitting them over single-mode fiber to minimize thermal load and signal degradation. Paired with IceQube, a cryo-compatible multi-channel microwave converter, the system enables reliable signal recovery directly within dilution refrigerators at temperatures as low as 55 K. Experimental results demonstrate low relative intensity noise (RIN), high dynamic range, and flat frequency response, ensuring high-fidelity control of quantum operations. This platform establishes a scalable, fiber-based approach for high-density qubit control, representing a significant advancement toward next-generation quantum processors.

Keywords: Photonic control systems, Large scale quantum computers, Cryo-photonics, RF-over-Fiber.

1 Introduction

Quantum computing represents a groundbreaking technological advancement with the potential to transform industries such as cryptography, artificial intelligence, drug development, and climate modeling. Despite their promise, realizing large-scale, practical quantum computers face major challenges in scalability, reliability, and operational efficiency. Central to these challenges is the control of qubits, the fundamental units of quantum information, which currently relies predominantly on electronic systems. Traditional electronic control is limited by heat generation, bandwidth constraints, and electromagnetic interference, creating bottlenecks for scaling quantum architectures [1]. In modern systems, superconducting qubits are typically controlled and measured using microwave pulses generated at room temperature and delivered into cryogenic environments via long coaxial cables with multiple stages of attenuation to

suppress thermal noise. While effective for small-scale systems, this method imposes serious obstacles when scaling to large qubit arrays.

To address these limitations, we propose the development of Photonic Control Systems (PCS) for large-scale quantum computing. By adopting a fully photonic approach, this initiative aims to redefine how qubits are manipulated, controlled, and measured, enabling quantum computers that are scalable, efficient, and reliable. The system leverages the inherent advantages of photonics including high speed, precision, and minimal thermal load to establish a fundamentally new paradigm for quantum control.

2 Advantages of Photonic Control Systems

Conventional quantum computing systems often rely on electronic control mechanisms, which, while effective for small-scale systems, encounter significant limitations as the number of superconducting qubits increases. Heat generation, susceptibility to noise, and spatial constraints imposed by electronic circuitry present major obstacles [3]. In contrast, photonic systems offer a fundamentally different control paradigm. Optical signals can interact with qubits with extraordinary precision, operate at the speed of light, and avoid the heat dissipation issues inherent to electronic components. Additionally, integrating photonics with quantum systems leverages mature technologies from the telecommunications and photonics industries, reducing development timelines and costs associated with scaling [4][5].

This work presents a photonic control system (PCS) designed to replace traditional electronic control in quantum computing with advanced optical technologies. By utilizing optical signals, waveguides, and integrated photonics, such as silicon photonics, the PCS addresses the thermal, spatial, and bandwidth constraints that limit current architectures. It supports large-scale qubit arrays with enhanced coherence, lower error rates, and improved scalability. While current electronic approaches struggle with heat, interference, and limited bandwidth, particularly in cryogenic environments [6], our optical approach offers a promising path forward.

3 System Architecture Overview

The PCS is built around two core subsystems: Menthor, a room-temperature RF-over-Fiber transmitter, and IceQube, a cryo-compatible opto-electronic receiver. Menthor, illustrated in Figure 1 with its block diagram in Figure 2, converts RF signals into optical signals using RF-over-Fiber technology, enabling high-frequency signal delivery into cryogenic environments. Operating at wavelengths in the C-band (1525–1555 nm), it employs low-noise DFB lasers to produce stable optical signals spanning 10 MHz to 18 GHz. With low relative intensity noise (RIN), strong harmonic suppression, and an optical output power of 3–10 dBm, Menthor delivers spectrally pure, distortion-free signals for precise qubit operations. Its flat gain and group delay characteristics ensure signal consistency, while a spurious-free dynamic range (SFDR) of approximately 100 dB/Hz$^{2/3}$ and a 1 dB compression point at 18 dBm highlight its high-performance capabilities.

Menthor provides three primary advantages for quantum computing. First, it eliminates heat dissipation associated with traditional RF cables by using fiber optics, preserving the cryogenic environment necessary for qubit coherence. Second, its modular, fiber-based design enables effortless scaling from a few to thousands of qubits without compromising signal quality or thermal stability. Third, it delivers ultra-clean, stable RF signals with low noise and precise shaping, enhancing gate fidelity and reducing operational errors. The current implementation is a compact 2U system with eight independent RF-over-Fiber channels, each equipped with its own modulator and laser for fully parallel, scalable operation. Control is provided via an intuitive Ethernet GUI or SCPI commands, supporting both remote and automated operation. Menthor integrates seamlessly with Viqthor's Mimothor platform, providing a complete quantum control solution with remote monitoring and management capabilities.

IceQube, shown in Figure 3, serves as the modular photonic receiver within the cryogenic environment and is specifically designed for high-density integration. Optical signals generated at room temperature travel through robust, cryo-stable Teflon-jacketed fibers, passing through sealed feedthroughs to maintain vacuum integrity. These microwave-modulated optical signals are delivered to high-speed photodetectors located inside the cryochamber. The fibers, coated with PTFE jackets, are engineered to withstand ultra-low temperatures and minimize outgassing effects [7][8]. Each IceQube module can accommodate hundreds of photodetectors, aligned with cryo-compatible



Fig. 1. Menthor RF Transmitter system to control qubits

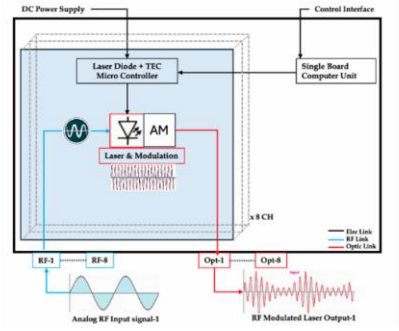


Fig.2. Block diagram of Menthor

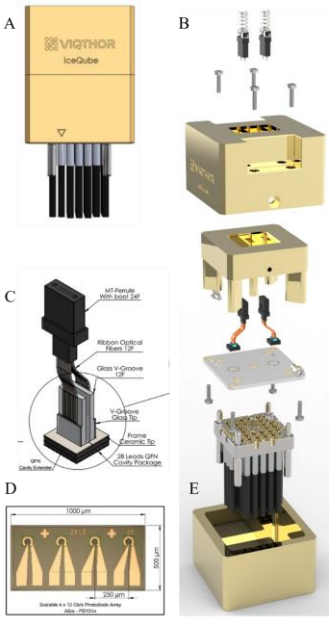


Fig.3. (A) External and (B) internal views of the IceQube module, its internal (C) fiber array integration with (D) photodiode array & (E) High density SMP matrix array.

fiber ribbons containing 12 to 48 fibers. The geometry is carefully optimized, with detector arrays spaced at 250 μm to match standard fiber array connectors. The entire assembly is constructed from cryo-compatible materials, including OFHC copper for thermal anchoring, gold-plated surfaces, and shielded cable interfaces to reduce RF crosstalk and outgassing.

At the cryogenic stage, IceQube's high-speed photodetectors convert the incoming optical signals back into electrical RF, which is routed to the quantum processor through cryo-compatible cabling. This module bridges optical control and quantum processors at millikelvin temperatures using reliable SMP and MMCX interfaces. Its modular design supports flexible stacking inside dilution refrigerators, allowing adaptation to different processor sizes and layouts. By addressing challenges in signal routing and cryogenic stability, IceQube minimizes heat load, eliminates bulky electronics near the qubits, and enhances overall control fidelity. Together with Menthor, IceQube completes the photonic control chain, enabling precise, scalable, and low-noise qubit manipulation.

4 Procedure of Controlling Qubits with The Proposed PCS

The integration of Menthor with IceQube establishes a scalable architecture for delivering precise microwave control signals to qubits inside dilution refrigerators. In this system, classical control electronics at room temperature generate microwave signal up to 18 GHz, corresponding to qubit operations such as single-qubit rotations or two-qubit gates. These signals are converted directly into optical carriers by Menthor modulated lasers, producing optical signals in the 1525–1560 nm range that are transmitted over single-mode fiber. This approach eliminates bulky coaxial cabling and reduces thermal load on the cryostat. The optical signals are carried into the refrigerator through Teflon-jacketed, cryo-stable fibers, which maintain vacuum integrity and support dense multiplexing with fiber ribbons of up to 48 channels for large-scale qubit control. Inside the cryogenic environment, IceQube's high-speed photodetectors convert the optical modulation back into RF signals with minimal distortion, demonstrating stable, linear performance from room temperature down to 55 K, with modulation bandwidths up to 12 GHz. Bias voltages between 0.25 V and 1.0 V optimize carrier extraction, and while responsivity decreases by ~10–14 dB under cryogenic conditions, this can be compensated with bias tuning, increased laser power, or low-noise amplification. The recovered signals are then routed via cryo-compatible coaxial cables to the quantum processor at the millikelvin stage, where they drive transitions and entangling gates with high fidelity. Crucially, the preserved linearity and broad frequency response ensure that control signals arrive undistorted, enabling reliable qubit operations. Finally, the modular IceQube design allows hundreds of photodetectors to be integrated in parallel, providing a pathway to scaling photonic qubit control systems without the thermal and spatial limitations of conventional coaxial wiring.

5 Cryogenic-Photonic Test Results and Analysis

The primary goal of our experimental demonstration was to validate the performance of the Menthor and IceQube system and to determine whether the photonic links remain efficient and stable in the extreme environment of a dilution refrigerator. The experiments were conducted using a comprehensive setup: Menthor was configured to output modulated optical signals in the 1525–1560 nm range, which were transmitted into the cryostat via fiber optics. Inside the cryogenic chamber, IceQube converted the optical signals back into RF and delivered them to measurement equipment, including spectrum analyzers and RF power meters.

Tests were performed at multiple temperature stages, from room temperature down to 55 K, across RF frequencies ranging from 10 MHz to 18 GHz, with input power levels up to +18 dBm. The resulting data formed the basis for the performance analysis presented in the following sections.

5.1 Photocurrent as a function input laser current at 8 GHz modulation.

This test evaluated the photocurrent behavior of high-speed photodetectors (HSPDs) under varying laser currents and temperatures, serving as a critical validation of the proposed photonic control architecture. Figure 4(A) presents curves showing photocurrent output across a range of laser input currents under room temperature (300 K) and cryogenic (55 K) conditions. The analysis was performed for four photodetector bias voltages: 0.25 V, 0.5 V, 0.75 V, and 1.0 V. The laser modulation frequency was fixed at 8 GHz, using a directly modulated laser (DML) at 1550 nm wavelength, consistent with the internal configuration of Menthor. At room temperature (300 K), the photocurrent exhibits a clear linear relationship with increasing laser input current, confirming that the photodetector operates in a highly responsive and stable regime. This behavior aligns with theoretical expectations, as the responsivity of photodiodes under standard conditions typically follows a near-linear dependence on incident optical power [9].

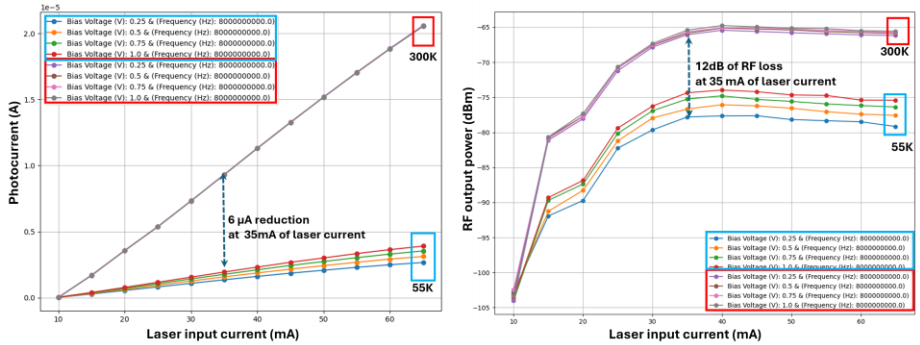


Fig. 4. (A) Photocurrent at 300 K and 55 K at 8 GHz modulation frequency, showing reduced photocurrent of 6micro amps at cryogenic temperatures. (B) RF output power at 300 K and 55 K, highlighting a 12 dB drop under cryogenic conditions.

Under cryogenic conditions (55 K), the photocurrent decreases across all bias voltages. For example, at a laser current of 35 mA, the photocurrent at 55 K is approximately 6 μA lower than at 300 K. Despite this reduction, the linear relationship with laser current is preserved, indicating that the photodetector maintains predictable operation at extremely low temperatures. The reduction in responsivity at cryogenic temperatures can be attributed to intrinsic semiconductor effects, including reduced carrier mobility and changes in the absorption coefficient, which lower the internal quantum efficiency. The retention of linearity is critical, as it ensures that the system can still operate within a predictable regime, allowing calibration and compensation via control or software algorithms. The observed behavior confirms that photonic downlink channels can reliably deliver microwave signals in cryogenic quantum systems. This addresses a key challenge in scaling quantum computers, enabling precise, low-heat, and low-noise control of qubits at millikelvin temperatures. The bias-dependent photocurrent characteristics align with known photodiode performance, demonstrating their suitability for scalable, high-speed quantum control applications.

5.2 RF output power as a function laser input current (300K vs. 55K)

The results in Figure 4(B) compare RF output power as a function of laser input current across bias voltages ranging from 0.25 V to 1.0 V, at a fixed modulation frequency of 8 GHz. Two temperature conditions are shown: room temperature (300 K, red box) and cryogenic temperature (55 K, blue box).

Temperature Impact on RF Output: A significant drop in RF output power is observed at 55 K compared to 300 K. At a laser input current of 35 mA, the RF output power decreases by approximately 12 dB, indicating a strong thermal dependence of the photodetector's conversion efficiency or a potential impedance mismatch under cryogenic conditions. At both temperatures, RF output power increases with laser current up to 35–40 mA, beyond which it begins to saturate or slightly decline. This saturation suggests that the photodetector is approaching its optimal operating range or that thermal and electrical effects start to influence performance.

Bias Voltage Dependence: Across both temperature ranges, higher bias voltages (0.75 V and 1.0 V) consistently produce stronger RF output. This supports the earlier observation that increasing bias improves carrier extraction efficiency, an effect particularly critical in cryogenic operation.

Cryogenic Behavior and Stability: Although overall RF output is lower at 55 K, the response curves remain stable and predictable, confirming that the system continues to function reliably. The observed 12 dB reduction at 35 mA underscores the importance of optimization in cryogenic tuning and impedance matching to maximize system performance.

5.3 Photocurrent response as a function bias voltage at fixed laser current

The experimental investigation shown in Figure 5 characterizes the photocurrent behavior of high-speed photodetectors (HSPDs) under varying bias voltages while maintaining a constant laser current of 35 mA. Measurements were performed at two

temperatures, 300 K and 55 K, across four modulation frequencies: 6 GHz, 8 GHz, 10 GHz, and 12 GHz. These results provide key insights into the effects of temperature and electric field strength on carrier transport efficiency and device responsivity, both critical parameters for photonic control systems in quantum computing environments.

As shown in Figure 5(A), at room temperature, the photodetectors exhibit a non-linear increase in photocurrent with applied bias voltage, eventually saturating at higher voltages. This saturation indicates full depletion of the photodiode's intrinsic region and efficient carrier collection. The maximum observed photocurrent under these conditions was approximately 10 µA. In contrast, at 55 K, the photocurrent is significantly lower, plateauing near 2.5 µA. The reduction is especially pronounced at lower bias voltages; for instance, at 0.25 V and 8 GHz, the photocurrent decreases by about 8 µA compared to room temperature.

At cryogenic temperatures, this reduction is attributed to decreased carrier mobility, altered recombination dynamics, and increased activity of deep-level traps, common effects in semiconductors at low thermal energy. These observations underscore the importance of precise bias tuning in cryogenic photonic systems. Despite the reduced responsivity, the photodetectors maintain a flat frequency response from 6 to 12 GHz, demonstrating a high modulation bandwidth suitable for broadband quantum control. ~~Overall, these results indicate that with proper biasing and thermal calibration, photodetectors remain well-suited for high-speed photonic links in scalable, low-temperature quantum computing architectures.~~

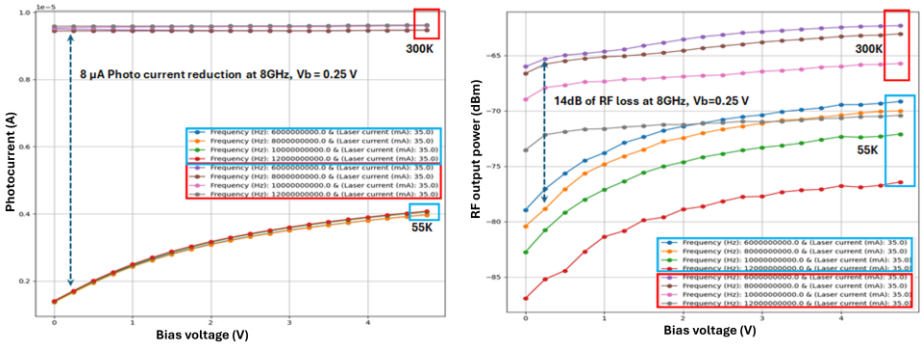


Fig.5- (A) Photocurrent at 300 K and 55 K across 6–12 GHz, showing reduced responsivity at cryogenic temperatures. (B) RF output power at 300 K and 55 K, highlighting a 14 dB drop under cryogenic conditions

5.4 RF output power as a function photodetector bias voltage

This test further evaluates photodetector performance by analyzing RF output power as a function of bias voltage under a fixed laser current of 35 mA and a constant RF modulation frequency of 8 GHz. The RF output was recorded at both 300 K and 55 K, providing a comparative assessment of signal demodulation efficiency in room-temperature versus cryogenic conditions.

As shown in Figure 5(B), at 300 K, the RF output power increases steadily with bias voltage, eventually stabilizing near -65 dBm at higher voltages. This behavior reflects enhanced carrier drift velocities and reduced transit times due to stronger internal electric fields, resulting in efficient optical-to-RF signal conversion. At 55 K, the maximum RF output power drops to approximately -80 dBm, a 14 dB reduction at 0.25 V bias. This attenuation highlights the challenges of signal recovery at cryogenic temperatures, where altered semiconductor properties, such as junction capacitance and carrier dynamics, reduce the photodiode's ability to generate strong RF signals. Despite the lower absolute output, the cryogenic measurements show a consistent, monotonic increase in RF power with bias voltage, mirroring the trend at room temperature. This consistency indicates that the fundamental RF generation mechanisms remain intact, albeit with reduced efficiency, under cryogenic conditions. The maintenance of linearity across the operating range confirms the device's reliability in preserving signal fidelity, which is critical for quantum control operations.

These findings emphasize the importance of electrical bias optimization to mitigate cryogenic-induced performance losses. They also confirm that, while absolute RF power levels are lower at 55 K, the photodetectors retain functional integrity and can operate across the GHz regime with acceptable linearity and frequency response. Collectively, these results validate the feasibility of deploying RF-over-Fiber (RFoF) links in photonic control systems for large-scale quantum computing, where operation inside cryostats at deep cryogenic temperatures is essential. Overall, the data empirically supports the design principles of cryogenic-compatible, high-speed photonic control architectures, demonstrating that strategic bias voltage management and thermal calibration enable photodetectors to meet the stringent performance requirements of next-generation quantum processors.

5.5 RF output power as a function RF input power at 8 GHz modulation.

Test results shown in Figure 6(A) examine the photodetector's linearity and RF gain characteristics by analyzing RF output power as a function of RF input power at a

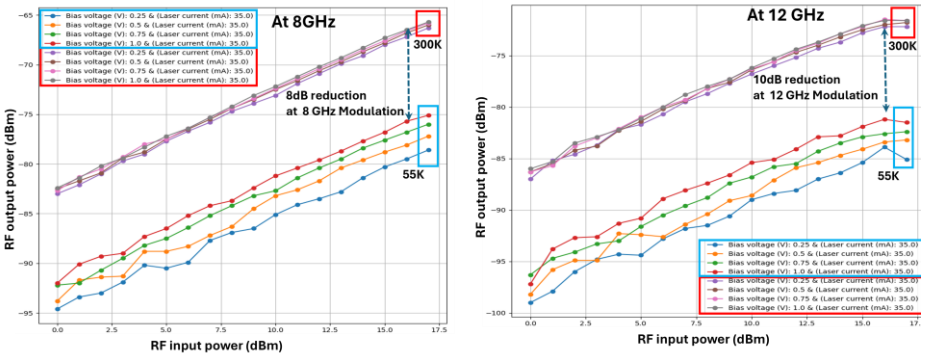


Fig. 6. (A) Linear RF gain across input power range; 8 dB cryogenic loss observed. (B) RF response at 12 GHz with ~10 dB loss at 55 K compared to 300 K.

modulation frequency of 8 GHz. The experiment was conducted at varying bias voltages (0.25 V, 0.5 V, 0.75 V, and 1.0 V) and under two thermal conditions, 300 K and 55 K, with the laser current fixed at 35 mA. This test provides insight into the photodetector's behavior under dynamic power conditions, which is critical for evaluating its suitability in scalable and power-efficient quantum control systems.

At 300 K, the results show a near-linear increase in RF output power with rising RF input power across all bias voltages. The slope of the curves indicates consistent gain and minimal distortion over the tested dynamic range. As expected, higher bias voltages produce stronger RF output, due to improved carrier sweep-out efficiency and reduced transit-time effects. This linear behavior highlights the device's robustness in transmitting modulated signals without introducing nonlinear artifacts, thereby preserving signal fidelity.

At 55 K, the RF output exhibits an ~8 dB reduction compared to room temperature, with losses most pronounced at low input powers, reflecting increased threshold sensitivity and higher minimum power requirements. Despite this attenuation, the output maintains linearity at higher bias voltages, indicating preserved intrinsic gain but reduced efficiency at cryogenic temperatures. This behavior arises from reduced carrier mobility, suppressed recombination, and altered photodiode impedance. While output power decreases, the sustained linear response confirms suitability for cryogenic RF-over-Fiber applications, provided sufficient input levels and optimized biasing, which is critical for deterministic qubit manipulation in photonic quantum computing.

5.6 RF output power as a function of RF input power at 12 GHz modulation.

This section builds upon the findings from Figure 6(A) by reanalyzing RF gain and linearity at a modulation frequency of 12 GHz. This test is critical for assessing the photodetector's bandwidth limitations and its ability to handle high-frequency modulations required for fast qubit gate operations. As in the previous test, measurements were conducted across multiple bias voltages (0.25 V to 1.0 V), with the laser current held constant at 35 mA, under both 300 K and 55 K conditions. As shown in Figure 6(B), at 300 K, the results demonstrate strong RF output power scaling with increasing RF input power across the entire tested range. Like the 8 GHz case, the curves maintain excellent linearity, and higher bias voltages yield improved output strength. These results confirm the detector's ability to operate efficiently at frequencies up to 12 GHz without encountering saturation or bandwidth roll-off, which is essential for high-speed quantum control systems.

Under cryogenic operation at 55 K, the output power exhibits a notable reduction, approximately 10 dB below room-temperature performance at equivalent input levels. This reduction is slightly greater than that observed at 8 GHz, indicating modest frequency-dependent sensitivity when the device is cooled. Nevertheless, the linear trend of RF output versus input power is preserved, and higher bias voltages continue to produce superior output levels. This consistent behavior demonstrates that the photodetector remains effective for high-frequency operation in cryogenic conditions, albeit with slightly reduced gain. The observed 10 dB loss highlights the need for thermal compensation techniques, such as increased laser power, optimized biasing, or post-

demodulation signal amplification. However, the absence of signal compression or nonlinearity at higher input levels confirms that the photodetector preserves signal integrity, a crucial requirement for precision-driven quantum computing systems.

Overall, the results in Figure 6 validate the operational linearity and frequency responsiveness of the photodetector across a broad RF input power range and under extreme temperature conditions. These characteristics are instrumental in enabling scalable, low-noise, and high-fidelity photonic control systems for large-scale quantum computing architectures.

6. Conclusion

The experimental results demonstrate how operating parameters influence the performance of our PCS in controlling solid-state qubits. Optimal operation was achieved with a laser current near 60 mA, consistent with Menthor's default configuration, and a photodetector bias voltage around 0.75 V, balancing efficiency and thermal considerations. The system tolerated RF inputs up to +16 dBm without distortion, highlighting the robustness and linearity of the photonic control platform.

This work validates Viqthor's cutting-edge photonic control system built around Menthor and IceQube, capable of transmitting and demodulating high-frequency RF signals with minimal loss and excellent stability down to cryogenic temperatures as low as 55 K. Beyond simple signal conversion, the PCS provides a modular, scalable, low-noise, and thermally efficient foundation for quantum control. The experimental results establish a practical roadmap for integrating photonic control into next-generation quantum computing systems, representing a significant step toward modular, high-fidelity, and scalable quantum processors.

References

1. Krinner, S., Storz, S., Kurpiers, P. et al., Engineering cryogenic setups for 100-qubit scale superconducting circuit systems. *EPJ Quantum Technol.* 6, 2 (2019).
2. A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Physical Review A 86, 32324 (2012)
3. P. Krantz et al., A quantum engineer's guide to superconducting qubits. *Appl. Phys. Rev*. 6, (2019).
4. Pintus, P., Soltani, M. & Moody, G., Cryogenic optical data link for superconducting circuits. *Nat. Photon.* **18**, 306–308 (2024).
5. S. Joshi and S. Moazeni, Scaling up superconducting quantum computers with cryogenic RF-photonics, *Journal of Lightwave Technology*, vol. 42, no. 1, pp. 166-175, 1 Jan.1, (2024).
6. Lecocq, F. et al. Control and readout of a superconducting qubit using a photonic link. *Nature* 591, 575–579 (2021).
7. Adam J. Christiansen, Characterization of optical fiber at cryogenic temperatures. *SPIE proceedings*, vol.12417, (2023).
8. Xin Lu. Impact of the Fiber Coating on the Temperature Response of Distributed Optical Fiber Sensors at Cryogenic Ranges, *JLT*, vol. 36, No.4, (2018).
9. Eivind Bardalen, Evaluation of InGaAs/InP photodiode for high-speed operation at 4 K, *Int. J. Metrol. Qual. Eng*. **9**, 13 (2018).

Design Concept of NV-Center-Based Wideband Current Comparators

Yasutaka Amagai[1,2†], Hidekazu Muramatsu[1,2], Yuta Kainuma[1,2], Hiromitsu Kato[1,3], Norihiko Sakamoto[2], Nobu-Hisa Kaneko[1,2], Yuji Hatano[4], Mutsuko Hatano[4], and Takayuki Iwasaki[4]

[1] National Institute of Advanced Industrial Science and Technology (AIST), Global Research and Development Center for Business by Quantum-AI technology (G-QuAT) 1-1-1 Umezono Tsukuba, Ibaraki 305-8560, Japan.
[2] National Institute of Advanced Industrial Science and Technology (AIST), National Metrology Institute of Japan (NMIJ) 1-1-1 Umezono Tsukuba, Ibaraki 305-8563, Japan.
[3] National Institute of Advanced Industrial Science and Technology (AIST), Advanced Power Electronics Research Center (ADPERC) 1-1-1 Umezono Tsukuba, Ibaraki 305-8568, Japan.
[4] School of Engineering, Institute of Science Tokyo, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, Japan.

† y-amagai@aist.go.jp

Abstract. Electrical metrology has historically relied on precise current comparators, which play a central role in determining accurate current ratios. Conventional devices, however, depend on inductive pickup coils or cryogenic infrastructures such as SQUIDs, imposing limitations on bandwidth, complexity, and operational environments. To address these challenges, we propose a unified comparator architecture based on nitrogen-vacancy (NV) centers in diamond. The system replaces conventional magnetic pickup or flux feedback mechanisms with a solid-state quantum magnetometer that optically detects magnetic flux in the air-gap of a magnetic core. The NV sensor provides sensitivity to both AC and DC magnetic fields, enabling seamless operation across different metrological regimes, and fully compatible with a modern power electronics requirement. The proposed architecture offers key advantages such as complete electrical isolation, compact system, and potential scalability for portable or embedded applications. Modeling and recent advances in NV magnetometry suggest that the required sensitivity is achievable with current technologies. By addressing both AC and DC current measurement requirements in the field of power industries and fundamental electrical standards, this concept paves the way toward a next-generation current ratio standard.

Keywords: NV center, Diamond, Quantum sensing, metrology, electrical standards, current comparator, current ratio measurement.

1 Introduction

The field of electrical metrology has long relied on quantum-based standards such as the Josephson DC voltage standard [1] and the quantum Hall DC resistance standard [2], along with single-electron tunneling pumps [3,4]. Meanwhile, even with the availability of quantum electrical standards, practical current metrology often requires precise comparison between currents rather than absolute current alone. In several applications such as current transformer calibration, bridge circuits for fundamental resistance measurement, current comparators play a crucial role. These devices allow for precision current scaling, filling a gap between quantum standards and practical instruments.

Traditionally, alternating current (AC) measurements have relied on inductive current transformers and their calibration using current comparator systems that operate at power-line frequencies (50/60 Hz). These devices have achieved impressive accuracy on the order of 10^{-6} [5, 6], thereby underpinning much of the infrastructure required for electric power distribution, grid stability, and industrial metering. In parallel, direct current (DC) ratio measurements have been indispensable for resistance bridge circuits, which ensure the traceability of resistance standards derived from the quantum Hall effect. In this context, SQUID-based [7] current comparators have delivered extreme accuracy down to 10^{-9}, but only at the cost of cryogenic operation [8, 9]. Such low-temperature requirements increase complexity and cost, limiting accessibility to a small number of national metrology institutes. This dichotomy between AC and DC comparators illustrates a broader challenge: while existing technologies are highly specialized, they lack a unified platform that can address both regimes with comparable performance. Moreover, emerging applications in modern power electronics, smart grids, and electric mobility demand measurement solutions that combine high accuracy with compactness, portability, and robustness under real-world conditions.

Recent advances in room-temperature quantum magnetometry using nitrogen-vacancy (NV) centers in diamond offer a promising pathway to address these limitations. NV centers allow for the optical detection of magnetic fields with sub-nanotesla sensitivity and a wide dynamic range, simultaneously covering both AC and DC regimes [10–13]. Unlike conventional pickup coils or SQUIDs, NV sensors operate at ambient conditions, provide inherent electrical isolation, and can be engineered into miniaturized and scalable devices. These features open the possibility of developing a unified current comparator architecture—one that seamlessly supports both AC and DC current ratio measurements within a single, solid-state sensing platform.

In this paper, we present the design concept of such a comparator system based on NV-diamond magnetometers. The proposed approach eliminates the need for complex windings or cryogenic environments, while enabling broadband magnetic field detection through an optically read-out sensor placed in the air-gap of a magnetic core. We focus on the design strategy, sensor requirements, and theoretical feasibility, supported by recent modeling efforts and advances in high-purity diamond engineering. Experimental implementation is currently underway, and results will be reported in a separate publication [14]. By bridging practical metrological requirements with cutting-edge

quantum sensing technology, this work aims to contribute to the next generation of electrical standards and to provide a scalable route toward international standardization.

2 Measurement principles

2.1 Quantum Magnetometry Using NV Centers in Diamond

Nitrogen-vacancy (NV) centers in diamond have gained significant attention as versatile quantum magnetometers and represent one of the most mature solid-state quantum sensing platforms available today. An NV center consists of a substitutional nitrogen atom adjacent to a vacancy in the diamond lattice, forming a point defect with an electronic spin triplet ground state ($S = 1$). The spin sublevels ($m_s = 0$ and $m_s = \pm 1$) experience Zeeman splitting in the presence of an external magnetic field, with the splitting magnitude being directly proportional to the field strength [10].

Optically detected magnetic resonance (ODMR) provides a powerful mechanism to initialize, manipulate, and read out these spin states. When the NV center is illuminated with green laser light, it exhibits spin-dependent fluorescence in the red spectral region. The application of a resonant microwave field modulates the fluorescence intensity, giving rise to characteristic ODMR spectra. These spectra display sharp resonance dips that shift linearly with magnetic field, thereby enabling precise and quantitative magnetic field measurements. By carefully monitoring these shifts, one can resolve both AC and DC magnetic components with remarkable sensitivity.

NV centers offer several advantages compared with other magnetometer technologies. They function robustly at room temperature, and coherent spin control has been demonstrated even near 1000 K [11], highlighting their remarkable thermal stability. They achieve sensitivity levels down to the pico-tesla (10^{-9} T) regime for both AC and DC fields [12, 13], and due to their crystallographic orientation, they can provide vector magnetic field information [15], which is difficult to obtain with many other sensors. Moreover, the solid-state nature of diamond allows compact device integration, long-term stability, and the potential for large-scale sensor arrays, making NV centers scalable for both laboratory and industrial applications [16].

Recent engineering advances have further pushed the performance limits. High-purity, isotopically engineered diamond samples combined with optimized optical and microwave control have demonstrated sensitivities below 10 pT/Hz$^{1/2}$ [17, 18]. Such performance enables the detection of extremely weak residual magnetic fields, including those encountered in precision metrology devices, resistance bridge circuits, and current comparators. Taken together, these attributes position NV centers as one of the most promising platforms for bridging the gap between quantum physics and practical measurement standards.

2.2 Conceptual Integration of NV Sensors into Current Ratio Measurement

The proposed architecture leverages NV-diamond magnetometry to overcome the inherent limitations of conventional current comparators and to provide a unified platform for precision AC and DC measurements. In this scheme, the magnetic core of the comparator is intentionally designed with an air-gap, into which the NV sensor is either

embedded or externally aligned in close proximity. This configuration enables the sensor to optically detect the residual magnetic flux arising from even very small current imbalances. The principle is thus conceptually simple yet powerful: when the primary and secondary currents are perfectly balanced, the net flux through the core vanishes, whereas any deviation produces a detectable magnetic leakage field at the air-gap.

For AC applications, the system functions in a manner closely analogous to traditional inductive comparators. The NV sensor is able to track oscillating magnetic fields at standard power-line frequencies, thereby supporting feedback control schemes that null the flux in real time. The resulting balance condition provides the desired current ratio, while maintaining full electrical isolation between the sensing unit and the current-carrying windings. Importantly, this approach extends naturally beyond 50/60 Hz operation, suggesting applicability to a broader frequency range relevant to modern power electronics.

In the case of DC measurements, where conventional comparators encounter fundamental challenges due to the limited sensitivity of pick-up coils, the NV sensor introduces a qualitatively new detection principle. By directly measuring the static magnetic field generated by current imbalance, the system establishes a null-detection method for resistance bridge circuits and DC ratio standards without recourse to inductive components or cryogenic devices such as SQUIDs. This eliminates many of the operational constraints that have historically restricted DC comparator deployment to specialized laboratories.

Beyond the immediate metrological benefits, the use of NV centers also simplifies both mechanical and magnetic design. The absence of pick-up coils reduces the complexity of winding arrangements and allows the magnetic core to be made smaller, lighter, and more compatible with modular instrumentation. Optical read-out further enhances robustness against electromagnetic interference, while enabling straightforward electrical isolation and safe operation in complex environments such as high-voltage laboratories or industrial substations. These characteristics naturally support miniaturization, portability, and potential integration into embedded systems.

3 Sensor design and Feasibility

3.1 Conceptual Architecture

The proposed current comparator architecture is centered around a magnetic core equipped with multiple windings and an intentionally introduced air-gap, within which an NV center–based diamond magnetometer is carefully positioned. This structural arrangement enables the sensor to directly probe the local magnetic field in the vicinity of the air-gap, allowing for the detection of the net magnetic flux that arises from even a slight imbalance between the currents flowing through the windings. By converting this residual flux into an optically detectable signal, the NV sensor serves as the key element that bridges conventional electromagnetic comparator principles with modern quantum magnetometry.

Figure 1 illustrates the simplified structure of the system. The primary and secondary currents, driven through separate windings, produce opposing magnetomotive forces in a high-permeability core. Under ideal balanced conditions, the net magnetic flux through the core is zero. Any residual flux arising from a mismatch between the currents escapes through the air-gap, where it is detected by the NV sensor. The NV sensor is optically interrogated using a green laser and microwave excitation, with red fluorescence intensity being monitored to extract the local magnetic field via optically detected magnetic resonance (ODMR). This optical readout provides complete electrical isolation from the current-carrying paths, enhancing measurement safety and robustness against electromagnetic interference. To evaluate the feasibility of the proposed approach, we estimate the magnetic field that will be detected by the NV sensor in a typical comparator configuration. Finite element simulation of magnetic flux distribution in a toroidal core with and without an air gap. The flux is significantly decreased due to high magnetic reluctance of the air gap. The presence of the gap causes local flux leakage into the surrounding space, enabling detection by a magnetic sensor such as the NV center. These sensors are thus well positioned to achieve the necessary performance for detecting small residual flux levels in both AC and DC conditions. Moreover, the absence of detection windings significantly simplifies the magnetic design, allowing for reduction in core size and may faster dynamic response. This is especially advantageous for time-sensitive calibration systems. It is also worth noting that even if the sensitivity of the NV magnetometer does not meet the required threshold, longer integration times can reduce random noise. This trade-off between measurement time and precision may allow practical use of metrology.

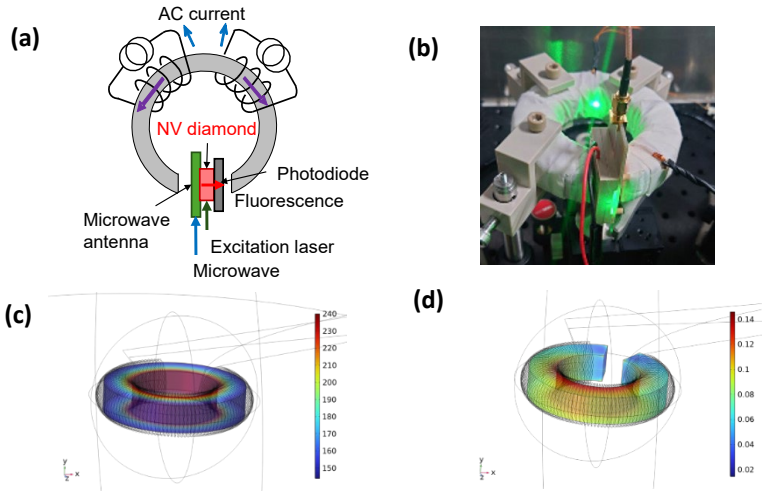


Fig. 1. (a) Schematic drawing of the proposed current comparator architecture using NV centers in diamond as a magnetic flux detector. **(b)** A photograph of the proposed current comparator architecture. **(c)** Simulated magnetic flux density in a toroidal core without any gap. The magnetic field remains uniformly confined within the core material, resulting in negligible external flux leakage. **(d)** Simulated magnetic flux distribution in the same toroidal core after introducing an air gap. The presence of the gap causes local flux leakage into the surrounding space, enabling detection by a magnetic sensor such as the NV center.

3.2 Optical and Electromagnetic Considerations

The NV magnetometer requires direct, line-of-sight optical access to the sensing volume in order to achieve efficient excitation and readout of the spin states. In our design, laser light is introduced laterally through the side of the diamond chip or coupled via an optical fiber, providing flexibility in integration and allowing compact packaging of the sensor head. The resulting red fluorescence is collected either along the same axis or through an orthogonal optical path, then spectrally filtered before reaching a photodiode or avalanche photodiode, ensuring high signal-to-noise detection. The microwave field necessary for spin manipulation is delivered by a coplanar waveguide patterned in close proximity to the sensing region, which ensures homogeneous excitation and minimizes unwanted reflections.

To further enhance sensitivity, the air-gap in the magnetic core is deliberately designed to be narrow and positioned precisely along the anticipated flux leakage path, thereby maximizing the magnetic field concentration at the diamond sensor. Careful attention is paid to the geometry of the sensor head, as a compact form factor is essential to minimize magnetic perturbations and to preserve the intended flux distribution. In addition, multilayer magnetic shielding surrounding the sensor head is considered to suppress environmental magnetic noise and to reduce the influence of mechanical vibrations. Such shielding not only improves measurement stability in laboratory conditions but also enhances robustness for potential deployment in industrial environments, where electromagnetic interference is more severe. Collectively, these optical, microwave, and shielding strategies provide a practical foundation for achieving high-fidelity NV-based current comparator measurements.

3.3 Limitations and Ongoing Work

While initial feasibility studies are promising, several engineering challenges remain to be addressed before the NV-based comparator can be realized as a practical standard. In particular, DC operation may be influenced by leakage flux arising from geometric asymmetries in the magnetic core or from magnetization hysteresis associated with the presence of the air gap. These effects, although less prominent in inductive or SQUID-based comparator architectures, can degrade measurement accuracy if not carefully modeled and compensated. Ongoing efforts therefore focus on accurate finite-element simulations and the development of active compensation techniques to suppress such parasitic flux contributions.

Another critical requirement is the use of homogeneous NV-diamond sensors with minimal device-to-device variation. Inhomogeneities in crystal quality, defect density, or orientation distribution can lead to variability in sensor performance, complicating calibration and long-term reproducibility. Equally important are the engineering aspects of optical alignment, thermal management, and microwave field homogeneity, all of which directly impact the achievable sensitivity and stability. For example, slight drifts in laser coupling or temperature fluctuations can introduce baseline shifts over long measurement times, an issue of particular relevance to metrological applications that demand repeatability at the 10^{-8} level or below.

In addition, practical DC operation often requires the use of auxiliary coils wound around the magnetic core to provide controlled DC offsets for balancing purposes. While such coils are indispensable for fine adjustment, they may themselves act as sources of electromagnetic noise. Careful design of the coil geometry, shielding, and filtering circuitry is therefore essential to suppress these noise contributions and to prevent them from limiting the ultimate sensitivity of the NV read-out.

4 Conclusions

This work presents a conceptual architecture for a novel current comparator system based on nitrogen-vacancy (NV) centers in diamond. The proposed design offers several attractive features, including room-temperature operation, complete electrical isolation, compactness, and compatibility with modern measurement requirements such as portability and broadband functionality demanded by power and energy industries. By eliminating the dependence on bulky pickup coils and cryogenic components, the NV-based approach also simplifies the overall system design, paving the way for instruments that are easier to manufacture, deploy, and maintain. Although the proposed comparator has not yet been fully implemented experimentally, the preliminary design considerations and feasibility analysis strongly suggest that its realization is within reach of currently available NV sensing technologies. Continued improvements in diamond material quality, sensor integration, and optical read-out techniques are expected to further enhance sensitivity and stability, thereby supporting long-term reproducibility essential for metrological applications. Moreover, the architecture is inherently scalable, making it adaptable not only for laboratory-grade standards but also for embedded systems in industrial or grid-level monitoring environments.

Taken together, these advantages indicate that the NV-based current comparator concept represents a promising path toward next-generation current ratio standards. In addition to advancing measurement science, this work may contribute to the broader field of quantum sensing by demonstrating a concrete, application-driven implementation. Experimental validation is currently underway, and results will be presented during the conference presentation.

Acknowledgements

This work was supported by the MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Number JPMXS0118067395, the CSTI, the Cross-ministerial Strategic Innovation Promotion Program (SIP), “Promoting the application of advanced quantum technology platforms to social issues,” and Bridging the gap between R&D and the IDeal society (society 5.0) (BRIDGE) organized by the Council for Science, Technology Innovation (CSTI) (Funding agency: QST). It was also partially supported by the JST ASPIRE program Grant number JPMJAP24C1.

References

1. J. Kohlmann, R. Behr and T. Funkc, “Josephson voltage standards,” *Meas. Sci. Technol.* 14 1216 (2003) https://doi.org/10.1088/0957-0233/14/8/305.
2. W. Poirier, F. Schopfer, “Resistance metrology based on the quantum Hall effect,” *Eur. Phys. J. Spec. Top*. **172**, 207–245 (2009). https://doi.org/10.1140/epjst/e2009-01051-5
3. J. P. Pekola, O.-P. Saira, V. F. Maisi, et al., “Single-electron current sources: Toward a refined definition of the ampere,” *Rev. Mod. Phys.* 85, 1421 (2013) https://doi.org/10.1103/RevModPhys.85.1421.
4. A. Fujiwara, G. Yamahata, N. Johnson, S. Nakamura, N.-H. Kaneko, Silicon Quantum Dot Single-Electron Pumps for the Closure of the Quantum Metrology Triangle, *ECS Transactions* 112, 119 (2023) https://doi.org/10.1149/11201.0119ecst.
5. W. J. M. Moore and P. N. Miljanic: The Current Comparator. UK: Peter Peregrinus Ltd. (1988).
6. N. L. Kusters and W. J. M. Moore, “The compensated current comparator: A new reference standard for current-transformer calibrations in industry,” *IEEE Trans. Instrum. Meas.*, vol. IM-13, pp. 107–114 (1964) https:// doi.org /10.1109/TIM.1964.4313383.
7. Y. S. Greenberg, “Application of superconducting quantum interference devices to nuclear magnetic resonance” *Rev. Mod. Phys*., 70(1) p. 175–222 (1998). https://doi.org/10.1103/RevModPhys.70.175.
8. D. Drung et al., “Improving the stability of cryogenic current comparator setups,” *Supercond. Sci. Technol.*, vol. 22, 114004 (2009) https://doi.org/10.1088/0953-2048/22/11/114004
9. H. Seppä et al., “Thin-film cryogenic current comparator,” *IEEE Trans. Instrum. Meas.*, vol. 48, pp. 365–369 (1999) https:// doi.org/10.1109/19.769602.
10. F. Jelezko and J. Wrachtrup: Single defect centres in diamond: A review. *Phys. Status Solidi A* 203(13), pp. 3207–3225 (2006). https://doi.org/10.1002/pssa.200671403.
11. G.Q. Liu, X. Feng, N. Wang, *et al*. “Coherent quantum control of nitrogen-vacancy center spins near 1000 kelvin,” *Nat Commun* 10, 1344 (2019). https://doi.org/10.1038/s41467-019-09327-2.
12. M. W. Doherty *et al*.: The nitrogen-vacancy color center in diamond. *Phys. Rep.* 528(1), 1–45 (2013). https://doi.org/10.1016/j.physrep.2013.02.001.
13. J. M. Taylor et al.: High-sensitivity diamond magnetometer with nanoscale resolution. *Nat. Phys.* **4**, 810–816 (2008). https://doi.org/10.1038/nphys1075.
14. H. Muramatsu, Y. Kainuma, H. Kato, N. Sakamoto, T. Yamada, C. Urano, H. Abe, S. Onoda, T. Ohshima, Y. Hatano, M. Hatano, N.-H. Kaneko, Y. Amagai, and T. Iwasaki, “Current comparator for both AC and DC ratio measurements with 10^{-8}-level accuracy,” arXiv preprint arXiv:2508.05140, 2025.
15. B. J. Maertz, A. P. Wijnheijmer *et al*., “Vector magnetic field microscopy using nitrogen vacancy centers in diamond,” *Applied Physics Letters*, 96 (9), p. 092504 (2010). https://doi.org/10.1063/1.3364135.
16. J. Pogorzelski, L. Horsthemk *et al*., “Fully Integrated LED Quantum Sensor Based on NV Centers in Diamond,” *Sensors* 24, 743(2024). https://doi.org/10.3390/s24030743.
17. N. Sekiguchi, M. Fushimi, A. Yoshimura *et al*., “Diamond quantum magnetometer with dc sensitivity of sub-10 pT/$Hz^{1/2}$ toward measurement of biomagnetic field,” *Phys. Rev. Applied* 21, 064010 (2024). https://doi.org/10.1103/PhysRevApplied.21.064010.
18. N. Sekiguchi, Y. Kainuma, M. Fushimi, C. Shinei, M. Miyakawa, T. Taniguchi, T. Teraji, H. Abe, S. Onoda, T. Ohshima, M. Hatano, M. Sekino, and T. Iwasaki, “Performance

evaluation of a diamond quantum magnetometer for biomagnetic sensing: A phantom study," *Applied Physics Letters*, vol. 126 194001 (2005) https://doi.org/10.1063/5.0254828.

Detection of grinding burns with NV center in diamond magnetometer

Baptiste Vindolet[1], Guillaume Bourcin[1], Benjamin Ducharne[2, 3], Hoai Nam Nguyen[1], Xavier Mougenot[4], Christophe Gallais[4] and Thomas Hingant[1]

[1] KWAN-TEK, 1 Rue Galilée Espace Innova, 56270 Ploemeur, France
[2] LGEF, INSA-Lyon, UR682, Université Lyon, 8 rue de la Physique, 69621 Villeurbanne, France
[3]ELyTMaX IRL3757, CNRS, Univ Lyon, INSA, Lyon, Central Lyon, Université Claude Bernard Lyon 1, Tohoku Université, Sendai, Japan
[4] Audit et R&T Industriels, SAFRAN TRANSMISSION SYSTEMS, 18 Bd Louis Seguin, 92700 Colombes, France

Abstract. Grinding burns alter the hardness and mechanical properties of steel components and can result in cracks or failure. Nital etching is widely used in the industry to detect grinding burns, but it poses environmental and safety risks and requires visual detection. Alternative techniques such as magnetic Barkhausen noise inspection have been developed, but automation is hindered by calibration drifts and by the sensor dimensions. Here, we introduce a novel non-destructive approach based on nitrogen-vacancy center in diamond to detect leakage magnetic field from grinding burns. This technology enables quantitative vector magnetic field measurement, is contactless and does not require calibration. Grinding burns with different sizes were successfully detected with a high spatial resolution and sensitivity.

Keywords: NV Diamond Quantum Magnetometry, Non-Destructive Testing, High-Resolution Sensing

1 Introduction

Some essential metallurgic elements such as gears or bearings are machined in high quality steel to resist extreme operational requirements [1–3]. To maintain expected tolerances and high surface quality, a rectification stage is performed to guarantee performance, durability and optimal security [4]. This processing step is the trickiest one, potentially leading to a heating of the material. Excessive heat can alter the material microstructure and bring reduced hardness and intrinsic strain formation [5,6]. With time, such defects, called grinding burns (GB), can result in cracks or failures in the material [7].

Detection methods exist to detect GB. These include chemical detection or magnetic non-destructive testing (NDT). Because most ground steel possess ferromagnetic phases, magnetic NDT is well-suited for detecting defects. Destructive methods also

exist offering high precision, but present excessive cost, and are not suitable on production lines due to the invasive inspection [8].

Among chemical methods, nital etching is the most common in industry [9–11]. This method consists of applying a chemical etching component on the sample surface. Material grain (crystalline structure), phase (ferrite, martensite, …), and defect on surface are made visible to the naked eye, becoming darker as usual. While this process is effective and inexpensive, it usually requires an operator to visualize defects, therefore it is difficult to fully automate [12,13] Moreover, it requires the use of toxic chemical products.

Magnetic NDT has advantages of being non-invasive, inexpensive, and make possible integration and automation in a product line, conversely to chemical etching. Among them, the magnetic Barkhausen noise (MBN) analysis can detect grain, texture, phase, or impurities in the material [14–16]. The difficulty is to differentiate all these magnetic phenomena. These can be differentiated by combining MBN with other magnetic methods but require a lot of data for a specific products line and is not easily transposable for another one [17,18].

In this study, a novel approach of magnetic NDT is proposed. This method, based on Nitrogen-Vacancy (NV) center technology, exploits the quantum properties of NV color centers in diamonds to achieve precise and reliable magnetic field measurements [19–21]. Due to its quantum nature, this technology is highly sensitive and does not require calibration, while providing digital measurement with a high resolution. Additionally, the probe is small enough (sub-mm) and is suitable to access difficult geometries, such as gear teeth for instance.

2 Implementation of a stray field measurement setup

The Nitrogen-Vacancy (NV) center is a defect in the crystalline lattice of a diamond, where two adjacent carbon atoms are substituted by a nitrogen atom and a vacant site respectively. Due to its configuration, the NV center diamond absorbs green light, and reemits red fluorescence. The negatively charge NV^- center presents an electronic spin triplet characterized by the projection number $m_s = 0, \pm 1$. Due to spin-spin interaction, the singulet state $m_s = 0$ and doublet states $m_s = \pm 1$ are energetically separated of $D \simeq 2.87$ GHz.Under green light excitation, the NV is pumped into the $m_s = 0$ and the red fluorescence rate from the $m_s = 0$ state is higher than from the $m_s = \pm 1$ states, enabling optically detected magnetic resonance (ODMR), by collecting photoluminescence (PL) over a MW frequency sweep [22]. Furthermore, with the Zeeman effect, the degeneracy of the $m_s = \pm 1$ states is lifted when a magnetic field is applied, and the transition frequencies between the ground state and both excited states are given by [21]:

$$\nu_{\pm 1} = D \pm \frac{\gamma_{\mathrm{NV}}}{2\,\pi} \boldsymbol{B} \cdot \hat{\boldsymbol{u}}_{\mathrm{NV}}, \tag{1}$$

where $\hat{\boldsymbol{u}}_{\mathrm{NV}}$ is the NV center axis, and γ_{NV} is the gyromagnetic ratio of the NV center, with $\frac{\gamma_{\mathrm{NV}}}{2\,\pi} = 28.035$ MHz.mT^{-1}.

Because of the crystal structure, there are four possible orientations of the NV center, allowing vectorial measurements of the magnetic field. These exceptional spin

properties have attracted much interest for the NV center in diamond as a quantum sensor for magnetometry but also electric field sensing and thermometry [19–21].

Inhomogeneous magnetization in ferromagnetic materials results in leakage magnetic field outside the sample. In a magnetized material, these inhomogeneities can occur at cracks, inclusions, corrosions, or grinding burns for instance. The principle of the NV magnetometer measurement is to provide stray field cartography at the sample surface, as depicted in Fig. 1 (a).

The NV diamond is integrated into an endoscopic sensor, glued at the tip of an optical fiber. The diamond is of 500 µm cubic shape, and highly doped with NV center (~ 0.3 ppm). Additionally, a microwave (MW) antenna is localized near to the diamond, to excite the NV magnetic resonances. The antenna is a thin copper coil, connected to an amplified MW synthesizer via an SMA connection. A photograph of the tip of the

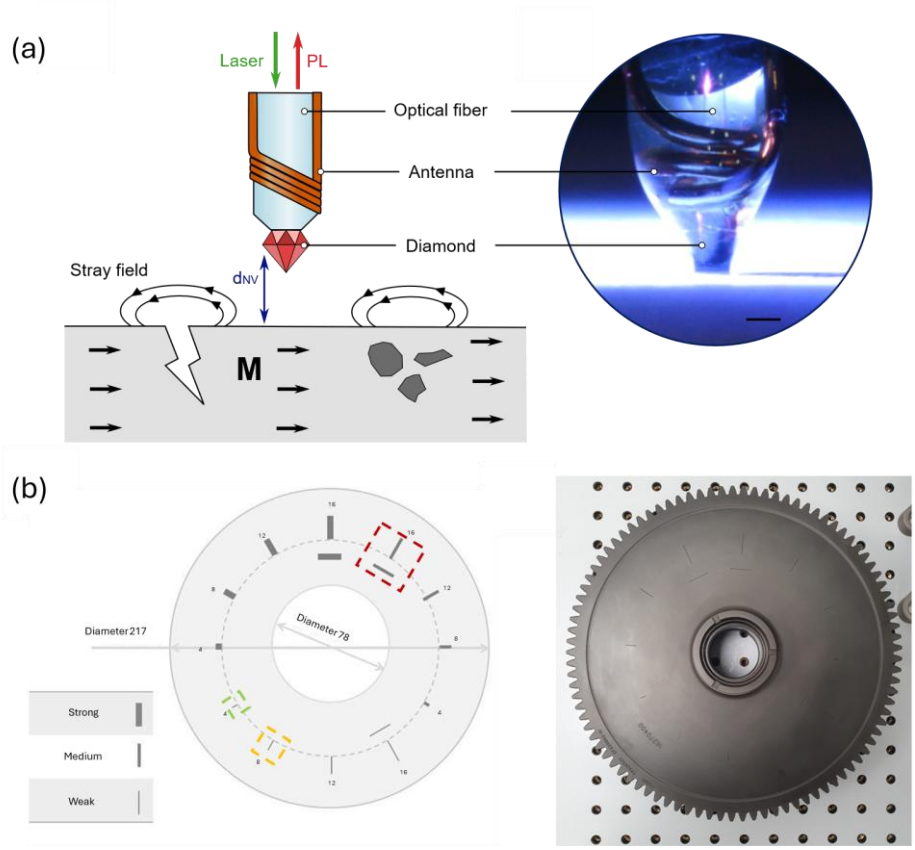


Fig. 1. (a) Schematic representation of inhomogeneous magnetization due to cracks or phase changing in a steel sample resulting in leakage magnetic field. The measurement is performed with a diamond NV center-based endoscope, with a photograph of the tip on the right-hand side. (b) Schematic representation of the gearwheel with grinding burns of different lengths, widths, and depths. Different cartographies of defects presented in this study are emphasized in dashed square regions. A photograph of the gearwheel is provided at the right-hand side.

endoscope is given in Fig. 1 (a). The green laser of 520 nm wavelength couples into the optical fiber from free space. The red PL goes back into the fiber and is separated from the green one thanks to a dichroic mirror. Then, the red light is collected on a photodiode which converts the signal into voltage.

In addition to the endoscope, a static magnetic field is applied - using a permanent magnet - to split the ODMR resonances. The frequency position of one of these resonances is tracked [23]. When an unknown magnetic field occurs from the sample, its projection along the NV center axis shifts the resonance. The tracking is performed using the modulation and demodulation method. The applied MW is typically modulated at the kilohertz range around the center frequency of the resonance ν_0. Hence, the PL signal is also modulated at the MW modulation frequency. After demodulation using a lock-in amplifier, the output signal obtained is similar to a derivative of the resonance curve. This output signal presents a linear region around the center frequency. In this region, the amplitude is directly proportional to the MW frequency, hence to the magnetic field from the defect and cancels at the resonance.

To track the resonance, a proportional-integral-derivative (PID) feedback controller is implemented. The PID adjusts the synthesizer frequency to maintain the demodulated signal at zero. To obtain a cartography of the leakage field, the sample is mounted on a XY translation stage, and the endoscope is fixed above the sample, with a lift-off distance of about 100 µm to prevent physical contact and potential damage. A two-dimensional map is obtained by storing the instantaneous resonance frequency at each pixel. Using the transition frequencies formula given in Eq. (1), the leakage magnetic field is retrieved from the measured instantaneous resonance frequency at each pixel.

In this study, only one resonance is tracked, but it can be generalized to vectorial measurement by tracking simultaneously the four different resonances.

3 Experimental results

In this study, the NV magnetometer is tested on GB detection. For this purpose, a gearwheel made of 16NiCrMo13 with a diameter of 20 cm is used, as illustrated in Fig. 1 (b). Burns were reproduced with laser-treatment and resulting defects are like those due to thermal effects of grinding process. A temper etching process allows to make defects visible to the naked eye.

The defects were previously calibrated with a twin sample using destructive metallographic analysis. The gearwheel presents 15 laser burns of different lengths (4, 8, 12, and 16 mm) along and perpendicular to the radial axis of the gearwheel. The defects are categorized according to either their depths or their widths: the defect depth is either of 60 µm and considered as weak, or of 180 µm and considered as strong; similarly, the width is either of 580 µm (weak) or of 1050 µm (strong). Note that detecting weak burns is not straightforward with standard visual inspection [24].

A ferromagnet has a magnetic history, and the magnetization homogeneity around a defect is not guaranteed. In the case of inhomogeneous magnetization, the leakage magnetic field related to the laser burn would be quasi-inexistant. To overcome this issue, material magnetization preprocessing is necessary. Before each measurement

campaign, the sample is magnetized around the defect under test with a high amplitude in-plane magnetic field using an electromagnetic yoke. Poles of the yoke were separated of 7 cm and placed at each side of the defect, applying a magnetic field up to 30 mT.

3.1 Influence of the magnetization direction

To detect the defect, the pre-processing step of magnetization needs to be performed properly. Indeed, as shown in the following of this study, the defect detection is directly linked to the magnetization direction. Fig. 2 (a) illustrates a cartography of the magnetic field probed with the NV diamond-based endoscope. The range of cartography is 30 mm in X and Y direction, and the pixel size is 200 µm. These directions are quasi equivalent to the azimuthal and radial directions respectively. The cartography area is emphasized by the square red region in Fig 1 (b), which includes two medium defects with a length of 16 mm in the azimuthal (X-axis) and radial (Y-axis) directions. Additionally, the magnetic field is applied along the X-axis, therefore perpendicular to defect along the Y-axis. A smooth magnetic field gradient of about 15 mT originates from sample localization beneath yoke poles throughout the magnetization processing. The defect along the Y-axis is visible with an amplitude contrast of about 65 µT whilst the defect along the X-axis is not. In Fig. 2 (b) the Laplacian of the magnetic field is computed, given by:

$$\Delta B_{\mathrm{NV}}(x,y) = \frac{\partial^2 B_{\mathrm{NV}}(x,y)}{\partial x^2} + \frac{\partial^2 B_{\mathrm{NV}}(x,y)}{\partial y^2} \quad (1)$$

The Laplacian cartography reveals fast spatial magnetic variation, i.e. the default, and suppresses slow variations, i.e. the slow magnetic gradient between the yoke's poles.

In Fig. 2 (c) and (d) the map from the magnetic field and from the corresponding Laplacian are represented, with a magnetization orientation at 45 degrees from the X-axis direction. With this magnetization orientation, both the defects in the azimuthal and radial directions are visible. In the perpendicular direction of a defect, the magnetization divergence occurs in a shorter distance as in its parallel direction. Therefore, the resulting leakage magnetic field is higher when the sample is magnetized along the perpendicular direction to the defect than along its parallel direction.

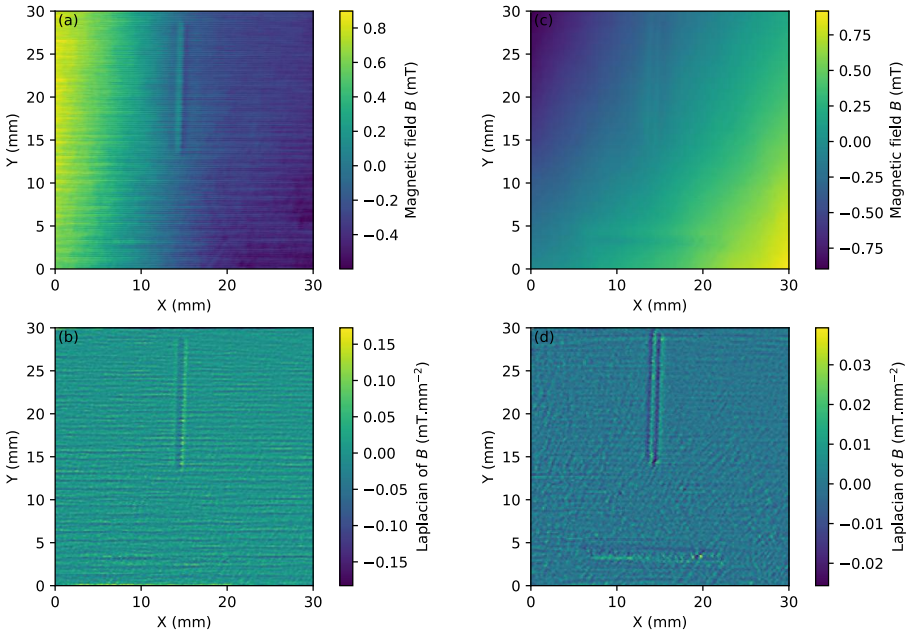


Fig. 2. Measured magnetic field along X and Y axis with a magnetization along X-axis in (a), and at 45° from X-axis in (c). Mapping around two medium defects with a length of 16 mm. In (b) and (d) are depicted the Laplacian of the magnetic field of respectively (a) and (c).

3.2 Weak grinding burns detection

In the previous section, we demonstrated detection with NV magnetometry on medium-intensity GB. In this section, we tried to assess the limit of GB detection with this technology. After performing a setup's parameters optimization with strongest defects to maximize the signal-to-noise ratio, i.e. laser and microwave power, PID loop settings, frequency modulation, scan parameters, and magnetization amplitude, cartographies of weak burns were conducted. Fig. 3 (a) and (b) show the Laplacian map for two different weak burns of respective lengths 4 mm and 8 mm. The pixel size is 125 µm, and the cartography range of 15 and 10 mm, respectively. As mentioned earlier, the magnetization has been performed along the perpendicular direction of the defect to better visualize it. As observed, the NV magnetometer has successfully detected all the weak burns. Due to their dimensions, the weak burns are barely visible to the naked eye.

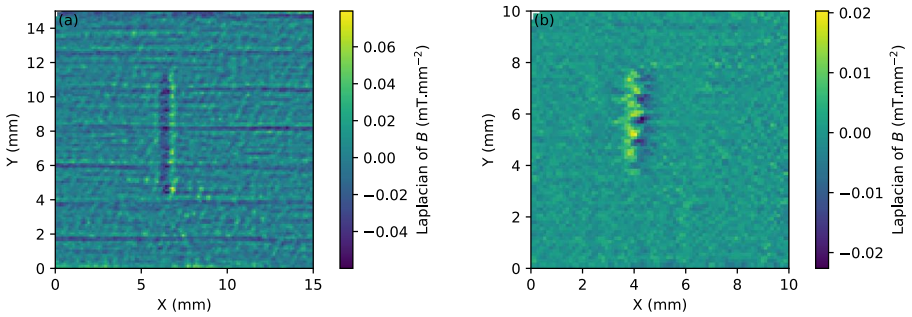


Fig. 3. Measured magnetic field along X and Y axis with a magnetization along X-axis. Maps obtained around a weak defect with a length of 8 mm in (a), and of 4 mm in (b).

4 Conclusion

In this study, we proved the ability of the NV-based magnetometer to detect successfully grinding burns on steel surfaces. With preliminary applying a weak magnetic field of about 30 mT along the perpendicular direction of the defects, all the defects were observed, even the weakest ones. This study demonstrated quantitative measurements of the magnetic flux leakage, with high sensitivity and resolution. More generally, it illustrates the potential interest of diamond quantum sensors for NDT, as a technique combining high spatial resolution, sensitivity and quantitative measurements.

Beyond grinding burns in steel, NV diamond sensors improve existing magnetic NDT, either by increasing resolution or by providing quantitative measurements that do not suffer from calibration drifts.

References

1. R. M'Saoubi, D. Axinte, S. L. Soo, C. Nobel, H. Attia, G. Kappmeyer, S. Engin, and W.-M. Sim, High Performance Cutting of Advanced Aerospace Alloys and Composite Materials, CIRP Ann. **64**, 557 (2015).
2. H. K. Tönshoff, C. Arendt, and R. B. Amor, Cutting of Hardened Steel, CIRP Ann. **49**, 547 (2000).
3. E. V. Zaretsky, Rolling Bearing Steels – a Technical and Historical Perspective, Mater. Sci. Technol. **28**, 58 (2012).
4. K. Wegener, F. Bleicher, P. Krajnik, H.-W. Hoffmeister, and C. Brecher, Recent Developments in Grinding Machines, CIRP Ann. **66**, 779 (2017).
5. S. Malkin and C. Guo, Thermal Analysis of Grinding, CIRP Ann. **56**, 760 (2007).
6. R. I. King and R. S. Hahn, Handbook of Modern Grinding Technology, *Handbook of Modern Grinding Technology* (Springer Science & Business Media, 2012).
7. F. Del Re, M. Dix, and F. Tagliaferri, Grinding Burn on Hardened Steel: Characterization of Onset Mechanisms by Design of Experiments, Int. J. Adv. Manuf. Technol. **101**, 2889 (2019).

8. B. Wang, S. Zhong, T.-L. Lee, K. S. Fancey, J. Mi, Non-Destructive Testing and Evaluation of Composite Materials/Structures: A State-of-the-Art Review (2020).
9. B. He, C. Wei, S. Ding, and Z. Shi, A Survey of Methods for Detecting Metallic Grinding Burn, Measurement **134**, 426 (2019).
10. J. E. Mayer, A. H. Price, G. K. Purushothaman, A. K. Dhayalan, and M. S. Pepi, Specific Grinding Energy Causing Thermal Damage in Helicopter Gear Steel, J. Manuf. Process. **4**, 142 (2002).
11. E. Sauter, M. Winter, and K. Wegener, Analysis of Robustness and Transferability in Feature-Based Grinding Burn Detection, Int. J. Adv. Manuf. Technol. **120**, 2587 (2022).
12. E04 Committee, Test Method for Macroetching Metals and Alloys, (2015).
13. M. W. Seidel, A. Zösch, and K. Härtel, Grinding Burn Inspection: Tools for Supervising and Objectifying of the Testing Process, Forsch. Im Ingenieurwesen **82**, 253 (2018).
14. A. Sorsa, M. Ruusunen, S. Santa-aho, and M. Vippola, Sub-Surface Analysis of Grinding Burns with Barkhausen Noise Measurements, Materials **16**, 159 (2023).
15. A. Thanedar, G. G. Dongre, R. Singh, and S. S. Joshi, Surface Integrity Investigation Including Grinding Burns Using Barkhausen Noise (BNA), J. Manuf. Process. **30**, 226 (2017).
16. B. Ducharne, Non-Destructive Testing of Ferromagnetic Steel Components Based on Their Magnetic Response, in *Non-Destr. Mater. Charact. Methods*, edited by A. Otsuki, S. Jose, M. Mohan, and S. Thomas (Elsevier, 2024), pp. 707–725.
17. P. J. Withers, M. Turski, L. Edwards, P. J. Bouchard, and D. J. Buttle, Recent Advances in Residual Stress Measurement, Int. J. Press. Vessels Pip. **85**, 118 (2008).
18. R. Kern, B. Wolter, I. Altpeter, and G. Dobmann, Industrial Applications of 3MA ? Micromagnetic Multiparameter Microstructure and Stress Analysis, (n.d.).
19. B. Vindolet, B. Ducharne, H. N. Nguyen, X. Mougenot, C. Gallais, and T. Hingant, High-Resolution Non-Destructive Detection of Grinding Burns with NV Diamond Quantum Magnetometer, NDT E Int. **155**, 103439 (2025).
20. C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum Sensing, Rev. Mod. Phys. **89**, 035002 (2017).
21. L. Rondin, J.-P. Tetienne, T. Hingant, J.-F. Roch, P. Maletinsky, and V. Jacques, Magnetometry with Nitrogen-Vacancy Defects in Diamond, Rep. Prog. Phys. **77**, 056503 (2014).
22. M. W. Doherty, N. B. Manson, P. Delaney, F. Jelezko, J. Wrachtrup, and L. C. L. Hollenberg, The Nitrogen-Vacancy Colour Centre in Diamond, (2013).
23. H. Clevenson, L. M. Pham, C. Teale, K. Johnson, D. Englund, and D. Braje, Robust High-Dynamic-Range Vector Magnetometry with Nitrogen-Vacancy Centers in Diamond, Appl. Phys. Lett. **112**, 252406 (2018).
24. M. Inês Silva, E. Malitckii, T. G. Santos, and P. Vilaça, Review of Conventional and Advanced Non-Destructive Testing Techniques for Detection and Characterization of Small-Scale Defects, Prog. Mater. Sci. **138**, 101155 (2023).

Uncut Gem - An Open-Source Hackable Quantum Sensor

Mark Carney[1 [0000−0001−9372−9033]] and Victoria Kumaran[1]

Quantum Village Inc., Delaware, USA and London, UK
{mark,victoria}@quantumvillage.org
https://quantumvillage.org/

Abstract. This work presents an overview of our fully open-source, hackable quantum sensor platform based on nitrogen-vacancy (NV) center diamond magnetometry. This initiative aims to democratize access to quantum sensing by providing a comprehensive, modular, and cost-effective system. The design leverages consumer off-the-shelf (COTS) components in a novel hardware configuration, complemented by open-source firmware written in the Arduino IDE, facilitating portability, ease of customization, and future-proofing the design. By lowering the barriers to entry, our sensor serves as a compact platform for education, research, and innovation in quantum technologies, embodying the ethos of open science and community-driven development.

Keywords: Quantum sensing · Open Source

1 Introduction

Quantum technology needs an 'Apple II' moment. By this, we refer to the importance that the Apple II holds in the history of the desktop/personal computer. From the first spreadsheet, *VisiCalc*, to the first graphical adventure game *Mystery House*, the Apple II was a consumer-grade platform that enabled much of the modern technological world we are now so familiar with [16].

In a similar vein, the release of the Raspberry Pi was viewed as a key enabling technology in the development of the Internet of Things, driving the decentralization of compute in society even further and giving prominence to the Single Board Computer (SBC) [10]. Our work has taken inspiration from this concept and applied this reasoning to Quantum Technologies, and sensing in particular. How can we push the boundary of what is conventionally considered to be possible?

Open Source software has, in the first quarter of the 21st Century, formed the bedrock of much of our technology stacks and infrastructure. Hoffman *et al.* [8] determined that companies would spend around 3.5× more on software annually if open-source options did not exist. With this, the importance of open sourcing the key parts of the quantum technology stack become apparent - it drives adoption upwards whilst driving costs down [8]. This sentiment of increasing

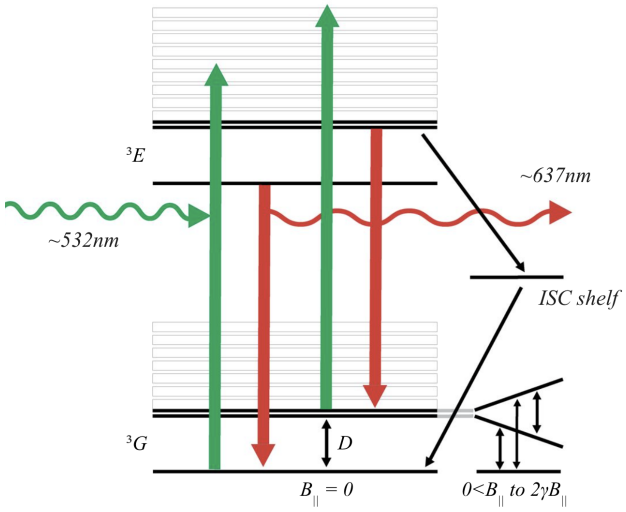


Fig. 1. Energy Level diagram for NV Centre Magnetometry, detailed in section 1.1. **NB** - the intersystem crossing (ISC) shelf involving two singlet states with an infrared (IR) transition occurring at 1042 nm, but this is not relevant to our magnetometry.

openness, interoperability, and institutional momentum is echoed by other open source projects and studies in quantum technology, see [2, 19].

Stegemann *et al.* [20] managed to get the overall cost of an NV Centre diamond experimentation platform down to under €200, utilising COTS parts and a 3D printer. This made our challenge apparent - could we reduce this cost down even further? And could we fill in some of the gaps and create a full-stack device that is fully open-source, moving away from the optics table and towards a hand-held form factor?

We first began designing and developing devices for quantum sensing and magnetometry back in 2023, and this project has been the culmination of 2 years of work towards making at least part of the quantum technology space truly open.

1.1 Brief Overview of NVC Magnetometry

The best surveys of the physics behind Nitrogen-Vacancy (NV) centre diamond magnetometry are [18] for detailed theory, and [7] for a high-level overview.

A Nitrogen Vacancy Centre (NV Centre) is the replacement in a diamond lattice of a single carbon (C) atom with a single Nitrogen (N) atom. In lab-grown diamonds this replacement is typically done through electron bombardment [13]. The mismatch in atomic valencies (4 for C to 3 for N) means that a vacancy is introduced. This defect essentially acts as a virtual atom contained within the lattice structure that has a strong zero-phonon line at 1.945 eV ($\lambda_{ZPL} = 637$nm)

between the excited states ${}^3E$ and ground states ${}^3G$. The hyperfine structures at both these levels yield sub-levels that are split into a lower singlet state $m_s = 0$ and two upper states $m_s = \pm 1$ separated by a difference of $D = 2.87$GHz with no magnetic field present [18]. The vacancy also has no Doppler shift relative to any interacting photons, which is why its effects may be utilized at room temperature.

The ground states Hamiltonian of this, ignoring the hyperfine interactions, is given in [20] and [18] as:

$$\mathcal{H} = hDS_z^2 + hE(S_y^2 - S_x^2) + g\mu_B \mathbf{B} \cdot \mathbf{S} \tag{1}$$

where z is the parallel axis to the NV centre, D and E are the Zero-Field Splitting parameters, S_x^2, S_y^2, and S_z^2 are the spin operators, g is the Landé g-factor, and μ_B is the Bohr magneton. D arises via spin-spin interaction between the two uncoupled electrons in the defect, and E is proportional to the internal strain within the lattice. The resulting resonant frequency ν (see [20]) for the applied microwave field at frequency D is given by

$$\nu = D \pm \sqrt{E^2 + (\gamma \cdot \mathbf{B}_{||})^2} \tag{2}$$

with $\gamma = g\mu_B/h \approx 28\text{GHzT}^{-1}$ [18]. The important takeaway here is that under this simplified regime, provided E is much less than $\mathbf{B}_{||}$, the main variable regarding the splitting is the magnetic field $\mathbf{B}_{||}$ and its orientation to the NV centre axis.

The Energy-level scheme found in figure 1 summarizes the key physical concepts. We can excite these NV centres with green light $\lambda_E \approx 532$nm. The NV centres then relax into the upper excited states ${}^3E$, and will become magnetosensitive under the application of microwaves. In practice, we sweep the microwave spectrum surrounding 2.87GHz, utilizing a similar sweep to the sweep in [20] by starting at 2.614GHz and sweeping up to 3.126GHz in discrete steps of 4MHz.

2 Gap Analysis

Many possible design options were reviewed, including the designs and build characteristics from [4, 5, 9, 12, 17, 20, 21]. The most complete guide was from Stegemann *et al.* [20], which provided a basis for our final design.

Upon analysis, however, we found a number of key things were missing from the available designs; The firmware used in NV Center diamond experiments is generally not available. Likewise, a reliable photodetector design schematic is generally not available, with most examples in the literature not always accurate to best electronic design principles. There are few easily configurable diamond mounts, as most require specialist optical mounting apparatus. Lastly, none of the solutions were designed to be 'hackable' for rapid deployment or prototyping, with little to no reconfigurable electronics and software. Similarly, none of the designs were geared to being ruggedized towards real world deployment.

3 Design Overview

In order to remediate these gaps, we came up with an initial design that made the following key decisions:

1. **Use off the shelf components** - the only specialist part should be the diamond material and any custom PCBs.
2. **Make as many things 'software problems' as possible** - by careful hardware design we can relinquish most control to a single software source. This massively improves debugging, maintenance, and updating a device.
3. **Simplify the design as much as possible.**
4. **Make the device 'hackable'** - by this we mean that the device should be fundamentally and easily reconfigurable and/or extendable in both software and hardware.

3.1 Microcontroller & Firmware

We wrote the source code for this device as Arduino IDE code - this taps into a large user base, with lots of help and built-in support for 26 native Arduino boards and hundreds of derivative boards [1]. This does two things; first, it immediately makes the design much more accessible to more people who are familiar with Arduino, over using an IDE or development tool-chain that is manufacturer specific.

Second, it allows the design to be ported to other chips through Arduino's universal hardware abstraction layer. This future-proofs the design, as any issues affecting the availability of one microcontroller means another more available one may be substituted with ease.

To this end, we used an ESP32 microcontroller - this is very well supported by Arduino IDE, has high quality 12-bit ADC inputs, is cheap and widely available, and has built-in WiFi and Bluetooth Low Energy (BLE) for future connectivity.

The default software libraries for driving an ADF4351 chip (see below) use a short calculation loop to generate the input values for the chip based off the input of a desired frequency. However, the sensitivity of the electronics (see below) means that this interferes with measurements significantly. To mitigate this, we pre-computed all of the desired sweep values and store them in firmware. This only consumes around 30 bytes per frequency at most, and practically eliminates the noise from the microcontroller interfering with the measurement operations.

3.2 Electronics

We made use of off-the-shelf operational amplifiers that are in a standard transimpedence configuration. We use the BPW34 photodiode as the main detection part as it has a large detection area - 3×3mm - and is widely available. This is then amplified by a TL082 dual op-amp, a common low-noise dual operational amplifier.

To minimize the parts required and to only require one 5V power input, we utilize a voltage divider formed from two 1,000 Ohm resistors in series from 5V to Ground. This gives us a floating 2.5V which we utilize as a 'fake ground' to which the non-inverting inputs ('+' pins) of the operational amplifiers are tied. This achieves two things - firstly, it means we can do away with additional inverting voltage regulators, and secondly means this guarantees that that output will be below the 3V limit on the ESP32's ADC input, without needing level shifting.

An ADF4351 PLL signal generator was chosen as a microwave generator as it allows control via SPI bus with a range of 35MHz to 4.4GHz. The output of this is then amplified by two monolithic microwave RF amplifiers, specifically GALI-84+ (or compatible) amplifiers providing $\approx$ 16dB amplification each.

The PCBs were designed to use single-sided soldering and assembly of surface mount components, reducing cost and complexity of production. The PCBs also use top and bottom layer ground planes to reduce parasitics and noise.

Additionally, to make the device hackable beyond just giving full software access, we also incorporated a small prototyping area. These generally consist of 1.0mm copper (Cu) plated holes, spaced 2.54mm apart in a grid array. This allows the incorporation of any device which has output pins with this standard pitch, which is historically derived from the DIN 41612 standard, see [6].

3.3 Light Source

Canonically, 525-532nm light is sourced from a laboratory laser, however these may sometimes prove to be problematic, especially where there is little to no guarantee of adequate laser safety protocols.

One of the main advantages of lasers is the generation of phase coherent photons. This is something we do not utilize for our magnetometry (see equation 1), so it suffices to use any source with the correct frequency output. 'Superbrite' LEDs were sourced that emit the correct frequency, and were found to perform on par with laser sources. This significantly reduces cost and power consumption, whilst also increasing the safety of the device.

3.4 Diamond and Photodiode Encapsulation

We make use of fast-UV curing epoxy to encase the photodetector, a microwave aerial, and NV centre diamond samples. As pointed out in [20], we can omit the need for a dichroic mirror and instead place a square of easily available red lighting filter gel between the photodiode and diamond to act as our primary optical filter.

The epoxy prism is created during a 3-step layering process, corresponding to the diagram in figure 2; Layer 1 is a base layer of epoxy upon which the diamond and microwave aerial are placed after it is cured. A second layer of epoxy is then applied to fix the diamond and aerial in place. A third layer that fixes in place the filter, photodiode, as well as any mechanical fixing parts.

The density of diamond means that it will naturally sink in most epoxies, hence a layering approach is used. Even for fast UV curing epoxy, the excellent

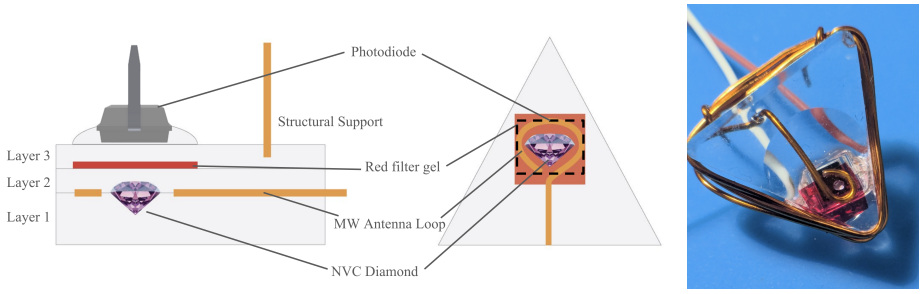


Fig. 2. Left: A diagram depicting the constructions layers of the diamon epoxy prism mount. **Right:** A mounted sample of diamond.

optical properties of clear epoxy are not compromised by the layering technique. It is also easy to reconfigure, as most epoxies only need to be heated to $\sim$ 200C in order to delaminate. The diamond is completely unaffeced by this heating, and so can be retrieved for subsequent encapsulation. In particular, this move away from the optics table towards very low cost epoxy mounting is a key part of the fast iteration and innovation potential for this design.

4 Testing

Although full characterization of the quantum sensor is still yet to be performed, we have managed to get some preliminary checks and estimates of sensitivity. In figure 3 we can see the output from the sensor both with and without the presence of a small magnet placed on top of the diamond prism. These are in line with the measurement of the magnetism of a small bar magnet, $\approx 0.005 - 0.01$ Teslas, found in figure 7 in [20] and figure 3 in [18], by rough approximation.

5 Conclusions and Future Work

On April 14th 2025 we released the full stack as open source on GitHub in [3]. To this end, the schematics, PCBs, parts lists, software code, and build instructions are all fully available, free and open source. We hope to build on this success with future releases and updates to both the hardware and base software to enable more people to learn about and practice with quantum sensing. Table 1 details the Bill of Materials (BoM) for the first release of the device, detailing a total cost that should not exceed £120. In Aug 2025 we released a single-board variant at DEF CON 33 [11, 15] with a lower total price of just under $100. With our subsequent version releases we expect this cost to fall even further.

5.1 Future Work

Firstly, a full and accurate characterization should be performed on multiple diamond samples in order to gather data for both assessing the direct measure-

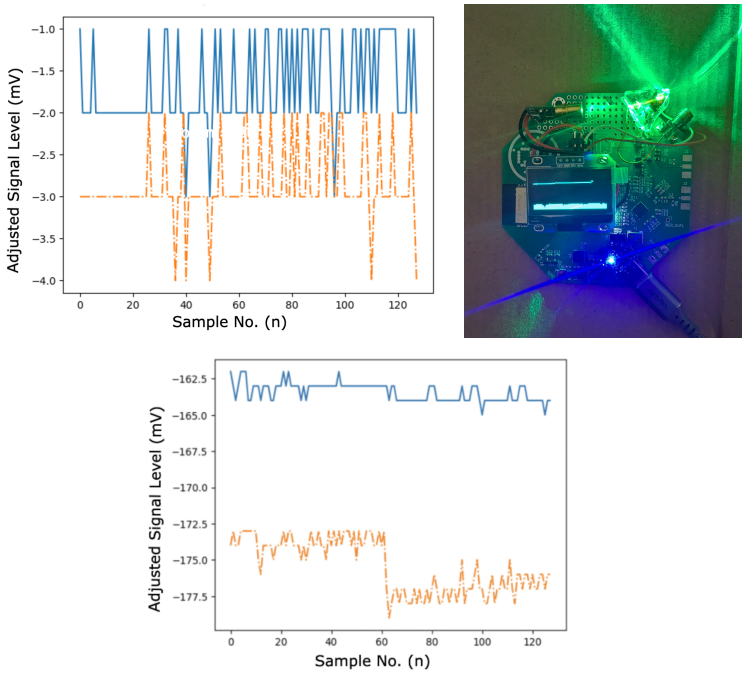


Fig. 3. Upper Left: Plotted readings in noise-adjusted millivolts from the ESP32 via serial over USB. Each point is an average of 6 measurements, with the dashed orange showing measurements with, and blue lines measurements without the presence of a small magnet. The shift in the splitting can be observed both in the drop in levels and in the shift of the valleys moving further apart. The frequency for each measurement is given in MHz from the sample number n by $(2838 + n)$. **Upper Right:** The device during operation. **Lower:** Plotted readings in noise-adjusted millivolts with the sensor in the presence of a large magnet (dashed orange) compared to no magnet (blue). Same scale as other plot, but in a different noise environment, hence lower adjusted readings.

ment and further analysis. There is plenty of scope for improving the output analysis and measurement estimation using edge machine learning and TinyML, something that Homrighausen *et al.* demonstrated in [9], and which we think would be an important step forward in improving the output of these devices.

We would like to see a space-ready version of the device prepared. This would mean changing the form factor to that of a CubeSat, as well as ensuring that all of the materials, soldering, conformal coating, and mechanical properties are suitable for the stresses and strains of space.

There are many use cases that should also be further explored for their suitability to this kind of low-cost quantum sensing setup, including; medical technology use cases like magnetocardiography (MCG), industrial sensing applications, and low-cost magnetometry for Position, Navication and Timing (PNT) use cases, building on work such as Muradoglu *et al.* [14].

Item	Number	Cost
Battery pack	1	Free to £5
OpAmp + electronics	1	up to £5
LED	1	∼£2-£5
ADF4350/1 MW Generator	1	£15-25
MW Gain (40dB) + BNC	1	£10
PCB (unit from run)	1	∼£5
OLED Screen	1	£5
ESP32 board	1	∼£10
NV Center Diamond	1	∼£10
Epoxy and Cu Wire	1 ea.	£10-15
	TOTAL	∼£115

Table 1. Bill of Materials for the first release version sensor.

Acknowledgments. The authors are grateful to the following people for their help during development of the device: Prof. Ben Varcoe, Max Shirokawa-Aalto, David Benoit, Matthew Markham, Rick Altherr, David Bengtson, and Brian McDermott.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Banzi, M.: Getting started with arduino. Make: Projects, O'Reilly Media, 2 edn. (Feb 2009)
2. Bülau, A., Walter, D., Fritz, K.P.: Components for an inexpensive cw-odmr nv-based magnetometer. Magnetism **5**(3), 18 (Aug 2025). https://doi.org/10.3390/magnetism5030018, http://dx.doi.org/10.3390/magnetism5030018
3. Carney, M., Kumaran, V.: Uncut Gem - A Fully Open-Source, Hackable Quantum Sensor (Apr 2025), https://github.com/QuantumVillage/UncutGem
4. Chatzidrosos, G., Rebeirro, J.S., Zheng, H., Omar, M., Brenneis, A., Stürner, F.M., Fuchs, T., Buck, T., Rölver, R., Schneemann, T., Blümler, P., Budker, D., Wickenbrock, A.: Fiberized diamond-based vector magnetometers. Frontiers in Photonics **2** (Aug 2021). https://doi.org/10.3389/fphot.2021.732748
5. Dix, S., Lönard, D., Barbosa, I.C., Gutsche, J., Witzenrath, J., Widera, A.: A miniaturized magnetic field sensor based on nitrogen-vacancy centers (2024). https://doi.org/10.48550/ARXIV.2402.19372
6. Fletcher, A.: Connector industry: A profile of the European Connector Industry - market prospects to 1999. Elsevier (Oct 2013)
7. Frellsen, L.F., Ahmadi, S.: Sensing magnetic fields with diamonds and green laser light. Quantum Views **2**, 10 (Jul 2018). https://doi.org/10.22331/qv-2018-07-17-10
8. Hoffmann, M., Nagle, F., Zhou, Y.: The value of open source software. SSRN Electronic Journal (2024). https://doi.org/10.2139/ssrn.4693148
9. Homrighausen, J., Horsthemke, L., Pogorzelski, J., Trinschek, S., Glösekötter, P., Gregor, M.: Edge-machine-learning-assisted robust magnetometer based on

randomly oriented nv-ensembles in diamond. Sensors **23**(3), 1119 (Jan 2023). https://doi.org/10.3390/s23031119
10. Johnston, S., Cox, S.: The raspberry pi: A technology disrupter, and the enabler of dreams. Electronics **6**(3), 51 (Jul 2017). https://doi.org/10.3390/electronics6030051
11. Kumaran, V., Carney, M.: Diamonds are for hackers - building the first fully open source and hackable quantum sensor. DEF CON 33 (August 2025)
12. Kuwahata, A., Kitaizumi, T., Saichi, K., Sato, T., Igarashi, R., Ohshima, T., Masuyama, Y., Iwasaki, T., Hatano, M., Jelezko, F., Kusakabe, M., Yatsui, T., Sekino, M.: Magnetometer with nitrogen-vacancy center in a bulk diamond for detecting magnetic nanoparticles in biomedical applications. Scientific Reports **10**(1) (Feb 2020). https://doi.org/10.1038/s41598-020-59064-6
13. McLellan, C.A., Myers, B.A., Kraemer, S., Ohno, K., Awschalom, D.D., Bleszynski Jayich, A.C.: Patterned formation of highly coherent nitrogen-vacancy centers using a focused electron irradiation technique. Nano Letters **16**(4), 2450–2454 (Mar 2016). https://doi.org/10.1021/acs.nanolett.5b05304
14. Muradoglu, M., Johnsson, M.T., Wilson, N.M., Cohen, Y., Shin, D., Navickas, T., Pyragius, T., Thomas, D., Thompson, D., Moore, S.I., Rahman, M.T., Walker, A., Dutta, I., Bijjahalli, S., Berlocher, J., Hush, M.R., Anderson, R.P., Szigeti, S.S., Biercuk, M.J.: Quantum-assured magnetic navigation achieves positioning accuracy better than a strategic-grade ins in airborne and ground-based field trials (2025). https://doi.org/10.48550/ARXIV.2504.08167
15. Newman, L.H.: A special diamond is the key to a fully open source quantum sensor. WIRED (Aug 2025), https://www.wired.com/story/fully-open-source-quantum-sensor-uncut-gem/
16. Nooney, L.: The Apple II Age: How the Computer Became Personal. University of Chicago Press (2023). https://doi.org/10.7208/chicago/9780226816531.001.0001
17. Opaluch, O.R., Oshnik, N., Nelz, R., Neu, E.: Optimized planar microwave antenna for nitrogen vacancy center based sensing applications. Nanomaterials **11**(8), 2108 (Aug 2021). https://doi.org/10.3390/nano11082108
18. Rondin, L., Tetienne, J.P., Hingant, T., Roch, J.F., Maletinsky, P., Jacques, V.: Magnetometry with nitrogen-vacancy defects in diamond. Reports on Progress in Physics **77**(5), 056503 (May 2014). https://doi.org/10.1088/0034-4885/77/5/056503
19. Shammah, N., Saha Roy, A., Almudever, C.G., Bourdeauducq, S., Butko, A., Cancelo, G., Clark, S.M., Heinsoo, J., Henriet, L., Huang, G., Jurczak, C., Kotilahti, J., Landra, A., LaRose, R., Mari, A., Nowrouzi, K., Ockeloen-Korppi, C., Prawiroatmodjo, G., Siddiqi, I., Zeng, W.J.: Open hardware solutions in quantum technology. APL Quantum **1**(1) (Feb 2024). https://doi.org/10.1063/5.0180987, http://dx.doi.org/10.1063/5.0180987
20. Stegemann, J., Peters, M., Horsthemke, L., Langels, N., Glösekötter, P., Heusler, S., Gregor, M.: Modular low-cost 3d printed setup for experiments with nv centers in diamond. European Journal of Physics **44**(3), 035402 (Apr 2023). https://doi.org/10.1088/1361-6404/acbe7c
21. Webb, J.L., Troise, L., Hansen, N.W., Achard, J., Brinza, O., Staacke, R., Kieschnick, M., Meijer, J., Perrier, J.F., Berg-Sørensen, K., Huck, A., Andersen, U.L.: Optimization of a diamond nitrogen vacancy centre magnetometer for sensing of biological signals. Frontiers in Physics **8** (Oct 2020). https://doi.org/10.3389/fphy.2020.522536

Molecular Spectroscopy Experiments for Direct Detection of Dark Matter

Florin Lucian Constantin[1]

[1] Laboratoire PhLAM, CNRS UMR 8523, 59655 Villeneuve d'Ascq, France
FL.Constantin@univ-lille1.fr

Abstract. Synchronized spectroscopy measurements for acetylene and molecular iodine reference transitions performed with an ultrastable laser signal may be exploited to constrain jointly couplings of dark matter to many fundamental constants. Enhanced sensitivity is expected from precision spectroscopy with quantum light.

Keywords: Ultralight Bosonic Dark Matter, Optical Spectroscopy, Time Synchronization.

1 Introduction

Dark matter (DM) makes apparently the majority of the matter in our Universe, but its nature is unknown. Precision measurements in atomic, molecular and optical physics were exploited to probe DM formed of 0-spin, sub-eV-mass particles. This ultralight bosonic dark matter (UBDM) [1], that is expected to behave as a classical field coherently oscillating at the corresponding Compton frequency, couples to the Standard Model (SM) particles and yields oscillations of the fundamental constants (FC) [2]. Atomic and molecular spectroscopy [3-5] and atomic clock comparisons [6,7] provided stringent limits on the UBDM field couplings.

This contribution proposes an experiment to probe dark matter fields that is based on precision molecular spectroscopy of reference transitions of acetylene at 1.5 μm and of molecular iodine at 0.5 μm. These transitions enabled high signal to noise rations in spectroscopy and were exploited to build secondary frequency standards [8]. In addition, quantum enhancement was demonstrated in spectroscopy by exploiting squeezed light [9]. Particularly, homodyne detection of squeezed light at 1550 nm enabled noise reduction by 12 dB beyond the standard quantum limit [10] and broad bandwidth [11]. Quantum advantage is already exploited in interferometry for gravitational wave detection [12] and may open the way to a new generation of precision spectroscopy measurements.

2 Probing variations of fundamental constants by optical network spectroscopy

2.1 Principe

The variation of a fundamental constant $\mathcal{F}$ (*e.g.* fine structure constant α, electron or quark masses $\mathrm{m}_i = \{\mathrm{m}_e, \mathrm{m}_u, \mathrm{m}_d, \mathrm{m}_s\}$, quantum chromodynamics energy scale Λ) may be quantified from the coupling of a UBDM field φ to the SM with the relevant coupling coefficient $d_{\mathcal{F}}$:

$$\mathcal{F}(\varphi) = \mathcal{F}(1 + d_{\mathcal{F}}\varphi) \tag{1}$$

The UBDM field oscillates as $\varphi(t) = \varphi_0 \cos(2\pi f_\varphi t)$ in function of the Compton frequency $f_\varphi = m_\varphi c^2/h$, related to the particle's mass, and an amplitude φ_0 depending on the local density of the DM halo [13].

The variations of the nuclear, atomic or molecular transition frequencies and of the resonance frequencies of the Fabry-Perot cavities may be parametrized in function of the variations of the fundamental constants:

$$\frac{\Delta f_X}{f_X} = Q_\alpha^X \frac{\Delta\alpha}{\alpha} + Q_\mu^X \frac{\Delta\mu}{\mu} + Q_q^X \frac{\Delta(\hat{m}/\Lambda)}{\hat{m}/\Lambda} \tag{2}$$

in function of the sensitivity coefficients to the variations of the fine structure constant Q_α^X, proton-to-electron mass ratio ($\mu = \mathrm{m}_p/\mathrm{m}_e$) Q_μ^X, and mean quark mass ($\hat{m} = \mathrm{m}_u/2 + \mathrm{m}_d/2$) Q_q^X. The sensitivity coefficients are calculated using nuclear [14], atomic [15], molecular [16], or solid-state [17] theories.

The fractional variation between the frequency of a Fabry-Perot cavity-stabilized laser f_L and the frequency of a molecular transition f_{mol} is sensitive to the variations of α and μ:

$$\frac{f_L - f_{mol}}{f_L} = (Q_\alpha^L - Q_\alpha^{mol})\frac{\Delta\alpha}{\alpha} + \left(Q_\mu^L - Q_\mu^{mol}\right)\frac{\Delta\mu}{\mu} \tag{3}$$

which are modulated with the differences between the relevant sensitivity coefficients $Q_{\alpha,\mu}^{L,mol}$ of the laser and of the molecular frequency. FC variability is parametrized here conveniently in function of the fine structure constant and the proton-electron mass-ratio, while other small contributions from nuclear structure are neglected. A time variation of the fractional frequency that is higher than the experimental uncertainty indicates detection of the dark matter field. Otherwise, the experimental uncertainty may be translated into constraints of the coupling coefficients of the UBDM field to the SM.

The variations of the fundamental constants are correlated. The proposed approach allows performing many measurements exploiting many molecular transitions with different sensitivity coefficients in order to constrain time variation of each fundamental constant. Such result is necessary to constrain New Physics theories beyond the Standard Model.

2.2 Proposed experimental approach

The proposed experiment is based on a laser at 1542 nm, stabilized on a Fabry-Perot cavity resonance with 10^{-15} level fractional stability at 1 s and continuously referenced against a hydrogen maser at Observatoire de Paris [18]. The laser signal is disseminated to remote geographical locations through REFIMEVE's stabilized optical fiber links [19]. This research infrastructure is exploited here to set up an optical network of molecular spectrometers (Fig. 1) that will perform synchronized laser absorption spectroscopy measurements. At each spectrometer location, the laser signal is amplified and suitably tuned at near-resonance with the molecular line by frequency shifting to probe acetylene by 1542 nm or by frequency tripling to probe molecular iodine by 514 nm. Additionally, each spectrometer may be quantum-enhanced by squeezing of radiation field that probes the molecular absorption (see next section).

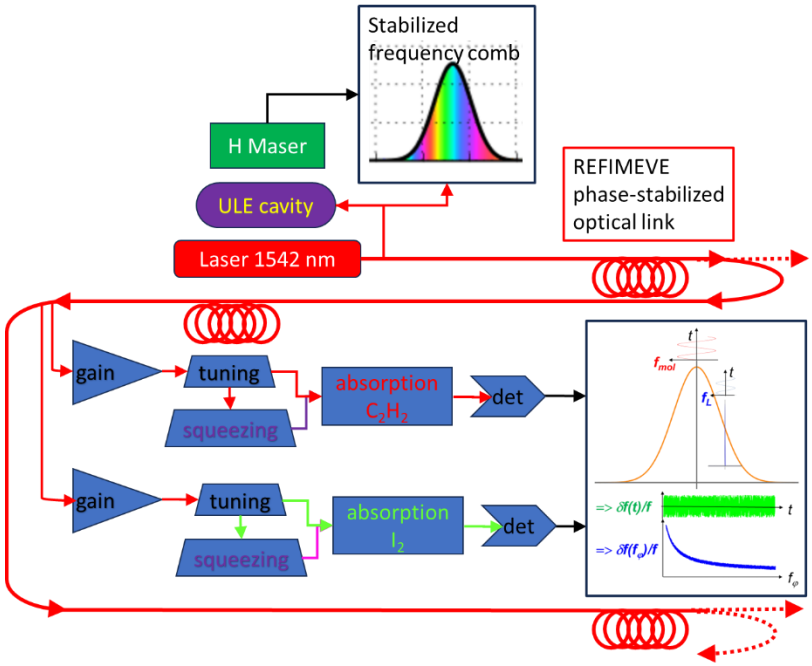


Fig. 1. An optical network of quantum-enhanced spectrometers exploiting the REFIMEVE reference signal at 1542 nm.

The time dependence of each laser absorption signal is recorded and Fourier-transformed. The fractional absorption signal is proportional with the fractional frequency detuning, expressed as:

$$\frac{f_L(t) - f_{mol}(t)}{f_{L(t)}} = \left(Q_\alpha^L h_L\left(f_\varphi\right) - Q_\alpha^{mol} h_{mol}\left(f_\varphi\right)\right)\frac{\Delta\alpha(t)}{\alpha} + \left(Q_\mu^L h_L\left(f_\varphi\right) - Q_\mu^{mol} h_{mol}\left(f_\varphi\right)\right)\frac{\Delta\mu(t)}{\mu} \quad (4)$$

by introducing the response functions $h_{L,mol}\left(f_\varphi\right)$ of the experimental setup at the Compton frequency f_φ of the oscillating fundamental constants. These response functions display low-pass frequency dependencies imposed by several experimental constraints: finite speed of propagation of the acoustic perturbations in the Fabry-Perot cavity spacer or in the laser's cavity and the linewidth of the molecular lines. The response functions have to be determined experimentally by calibrated measurements of the laser absorption dependence on the laser frequency fluctuation. In addition, both linear absorption spectroscopy and pump-probe Doppler-free saturation spectroscopy measurements are required in order to ensure sensitivity to oscillations of fundamental constants over a broad frequency domain. Some experimental setups were already developed for spectroscopy of acetylene [20] and molecular iodine [4].

3 Improved sensitivity in spectroscopy with squeezed light

The sensitivity of the proposed experiment is limited by the amplitude noise of the laser beam that is used for spectroscopy. Differential detection schemes, exploited in previous experimental setups [3,4,20], enable sensitivity limited by the shot noise in photodetection. Here is proposed a dark mater sensing strategy where the molecular absorption is probed with amplitude-squeezed light to improve the sensitivity beyond the standard quantum limit. The absorption signal is monitored using a balanced homodyne detection scheme, that amplifies the amplitude of the signal by beating with a strong local oscillator, while the impact of the associated shot noise is reduced by subtracting the two photodetection signals [21].

The approach, schematically depicted in Fig. 2, exploits a laser emitting in a mode at angular frequency ω and wavevector $\vec{k}$ with the following electric field operator:

$$\hat{E}_i(\vec{r},t) = \hat{\alpha}_i u(\vec{r}) e^{i(\vec{k}\vec{r}-\omega t)} \tag{5}$$

in function of the complex amplitude $\hat{\alpha}_i = \overline{\alpha_\iota} + \delta\hat{\alpha}_i$, expressed in terms of the mean value $\overline{\alpha_\iota}$ and the fluctuation $\delta\hat{\alpha}_i$, and the normalized spatial profile $u(\vec{r})$. $\overline{\alpha_\iota}$ is normalized such as $I_i = |\overline{\alpha_\iota}|^2$ is the intensity of the laser in a transverse section to the laser propagation direction. The laser beam is assumed strong, that is $|\delta\hat{\alpha}_i| \ll \overline{\alpha_\iota}$. The electric field may be also expressed in function of the quadrature operators:

$$\hat{E}_i = \hat{X}_i \cos(\omega t) + \hat{Y}_i \sin(\omega t) \tag{6}$$

that are non-commuting $\left[\hat{X}_i, \hat{Y}_i\right] \neq 0$ and their fluctuations obey to the Heisenberg uncertainty relation $\Delta\hat{X}_i \cdot \Delta\hat{Y}_i \geq |\overline{\alpha_\iota}|^2$. The fluctuations are equal to the standard quantum limit (SQL) $\Delta\hat{X}_i = \Delta\hat{Y}_i = |\overline{\alpha_\iota}|$ for a laser field in a coherent state. The squeezed light has one quadrature with fluctuation compressed below the SQL.

In direct detection of the molecular absorption with a coherent field (Fig. 2(a)), gas absorption, assumed small, is modelled as a beamsplitter with amplitude transmission factor $t = 1 - a/2$ and introduction of a vacuum field $\hat{c}$ with amplitude $\sqrt{1-|t|^2}$:

$$\hat{\alpha}_{out} = t\hat{\alpha}_{in} + \sqrt{1-|t|^2}\,\hat{c} \tag{7}$$

For small absorption and ideal detection and optics, the signal is reduced in proportion with the absorption and the additional noise induced by the absorption may be neglected. The signal to noise ratio in optical power units may be expressed as [22]:

$$\frac{S}{N} = \frac{(\mathrm{Re}(a))^2 \bar{I}_{in}}{2(1+Ex)\Delta\nu} \quad (8)$$

in function of the mean photodetection current $\bar{I}_{in}$ of the laser beam without absorption, the fraction of excess noise Ex beyond the shot noise in the detection of the laser beam, and the detection bandwidth $\Delta\nu$.

In a differential measurement with a coherent field (Fig. 2(b)) an ideal 50:50 beamsplitter is exploited to separate the laser beam in two parts: one beam pass through the molecular gas and is detected with detector 1, the other beam is detected with detector 2. The signal to noise ratio in power units may be expressed as [22]:

$$\frac{S}{N} = \frac{(\mathrm{Re}(a))^2}{4} \frac{\bar{I}_{in}}{2\Delta\nu} \quad (9)$$

in function of the mean photodetection current of the laser beam. A loss of a factor of 4 to the shot-noise limit $Ex = 0$ of eq. (8) is explained by the 50% loss of light intensity by using the beamsplitter.

In this proposal, bright squeezed light is generated for spectroscopy (Fig. 2(c)) by superposing on the 50:50 beamsplitter the coherent laser field $\hat{a}_{in}$ and broadband squeezed vacuum field described with the operator [23]:

$$\hat{a}_s(\omega) = \hat{a}(\omega)\cosh(s) - e^{2i\theta_s}\hat{a}^\dagger(\omega)\sinh(s) \quad (10)$$

where $\hat{a}(\omega)$ is the annihilation operator for the squeezed wave at the same angular frequency that the laser field, $\hat{a}^\dagger(\omega) = [\hat{a}(\omega)]^\dagger$ the creation operator, s the squeezing factor and θ_s the squeezing angle. Squeezed vacuum is generated here using a sub-threshold OPO pumped by the second harmonic of the monochromatic laser, approach that ensures suitable spectral purity and tunability. The emerging bright squeezed beam probes the molecular absorption and the measurement is performed by homodyne detection. A part of the coherent laser beam is phase shifted and used as local oscillator. The difference between the two current intensities is monitored on a spectrum analyzer. To implement saturation spectroscopy measurements, another part of the coherent laser (not shown on Fig. 2(c)) is superposed in the absorption cell on the squeezed beam with an orientation in the opposite direction.

For homodyne detection at the same optical intensities, the signal to noise ratio in optical power units with squeezed light $\left(\frac{S}{N}\right)_{sq}$ is improved to the value $\left(\frac{S}{N}\right)_{cl}$ obtained with classical light. That may be expressed as [24]:

$$\left(\frac{S}{N}\right)_{sq} = \left(\frac{S}{N}\right)_{cl} / \left(1 + \xi\rho S(\Omega, \theta_s)\right) \quad (11)$$

in function of the spectrum of squeezing of the output field of the OPO $S(\Omega, \theta_s)$, a term ρ describing transmission of the OPO field, and a term describing field transmission in the homodyne detection ξ.

At efficient propagation and detection $\xi \to 1$, for large squeezing $S(\Omega, \theta) \to -1$, a large enhancement is expected, that scales $\propto 1/(1 - \mathrm{T}_0)$ in function of the transmission T_0 of the OPO output coupler.

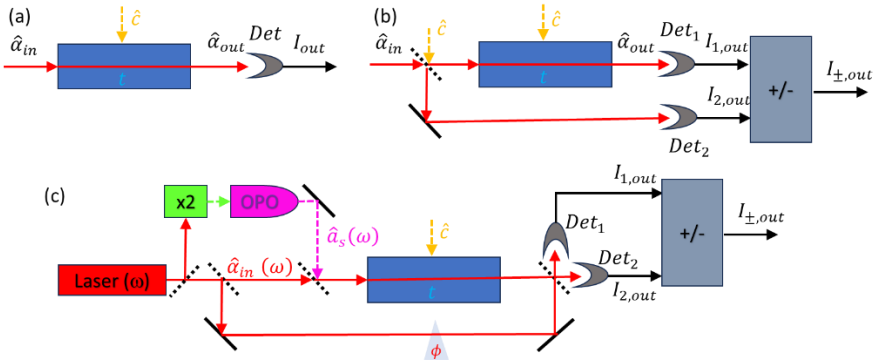


Fig. 2. Detection of a molecular absorption with quantum light. (a) direct detection, (b) differential detection, (c) experimental setup for spectroscopy with squeezed light. Laser emitting at angular frequency ω, x2 second harmonic generation, OPO sub-threshold pumped optical parametric amplifier, ϕ phase-shifter. *Det* detectors providing photodetection currents I_{out}, +/- homodyne detection electronics.

4 Expected performances

First are evaluated the constraints from measurements on molecular iodine and acetylene transitions by saturation absorption spectroscopy using classical light from the cavity-stabilized laser remotely distributed to both experiments.

The proposal aims to perform time-synchronised measurements of the transitions P(27) of the $\nu_1+\nu_3$ band of $^{12}C_2H_2$ at 1542 nm, and R(35) 44-0 of $^{127}I_2$ at 514 nm by suitably tuning the ultrastable optical signal disseminated in the REFIMEVE optical fiber link network. The metrological performances of the molecular clocks [25,26] suggest that these measurements may provide fractional frequency uncertainties of $5 \times 10^{-13}/\sqrt{\tau}$ and $3 \times 10^{-14}/\sqrt{\tau}$, respectively, that are expressed in function of the averaging time τ in seconds. The sensitivity of the molecular transitions to FC variations was described previously [27,28]. The cavity-stabilized laser is mainly sensitive to α-variation [17]. The uncertainties of the spectroscopy measurements are translated subsequently into estimations of the constraints for the coupling coefficients $d_{\alpha,\mu}$ of the UBDM field to α and μ. In doing that, one should take into account high-frequency cutoffs in the experimental response to FC oscillations imposed by the linewidths of the molecular lines and the finite speed of propagation of acoustic perturbations in the laser stabilization cavity. The UBDM field is associated here to the Galactic DM halo on Earth and the estimations for the bounds are shown in Fig. 3.

Squeezing of light enabled measurements with homodyne detection of the reduction of the power spectrum below the shot-noise limit by squeezing factors of 18.2 at 1550 nm [10] and 3.6 at 532 nm [29]. Exploiting squeezed light with these performances in saturated spectroscopy measurements may improve both fractional frequency uncertainties at $3 \times 10^{-14}/\sqrt{\tau}$ for the acetylene clock, and at $8 \times 10^{-15}/\sqrt{\tau}$ for the molecular iodine clock. These estimations are in proportion with the power squeezing factors. The estimations for the bounds for the coupling of the UBDM field

to the SM from saturation spectroscopy with squeezed light are also shown in Fig. 3. Quantum advantage translates here into improved bounds by a factor of 40 for the α-coupling and of 21 for the μ-coupling, respectively.

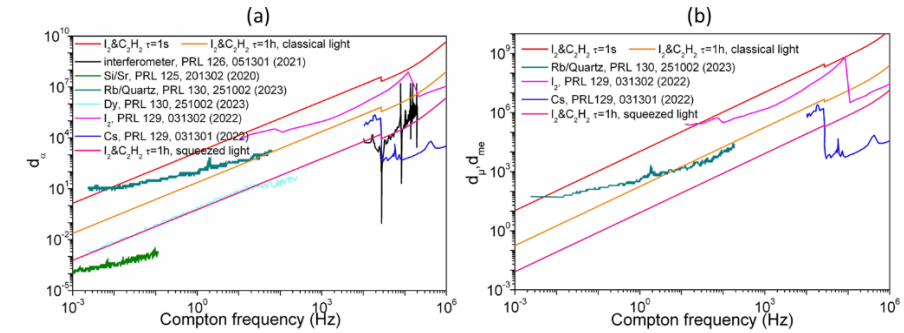


Fig. 3. Joint bounds for UBDM field couplings to (a) fine structure constant α, (b) proton-electron mass ratio $\mu=m_p/m_e$ (averaging time: τ=1 s red line, classical light, τ=1 h orange line, classical light, and τ=1 h pink line, squeezed light). Area above each dependence indicates the excluded range. For comparison, these plots include bounds set on the UBDM couplings to α and m_e from other experiments with classical light.

5 Conclusion

These results indicate that molecular spectroscopy may improve the description of the couplings of an UBDM field to the SM particles by constraining jointly time variations of α and μ. The bounds in the radiofrequency range are competitive with those from previous spectroscopy experiments and enable access to Compton frequencies higher that those were probed in atomic clock comparisons. These bounds may be significantly improved by implementing saturation spectroscopy with squeezed light.

6 Acknowledgement

This work is supported by the Programme National GRAM of CNRS/INSU and by the French ANR agency under contract No. ANR 10-LABX-0048, Labex First-TF. Laboratoire PhLAM is member of the Equipex+ T-REFIMEVE (French contract PIA 3 Equipements Structurants pour la Recherche of the Ministry of Higher Education, Research and Innovation and the General Secretariat for the Investment Plan).

References

1. Jackson Kimball, D. F., van Bibber, K.: The search for ultralight bosonic dark matter. Springer, New York (2023).

2. Arvanitaki, A., Huang, J., van Tilburg, K.: Searching for dilaton dark matter with atomic clocks. Phys. Rev. D 91(1), 015015 (2015).
3. Tretiak, O., *et al*: Improved bounds on ultralight scalar dark matter in the radio-frequency range. Phys. Rev. Lett. 129(3), 031301 (2022).
4. Oswald, R., *et al*: Search for dark-matter-induced oscillations of fundamental constants using molecular spectroscopy. Phys. Rev. Lett. 129(3), 031302 (2022).
5. Zhang, X., *et al*: Search for ultralight dark matter with spectroscopy of radio-frequency atomic transitions. Phys. Rev. Lett. 130(25), 251002 (2023).
6. Filzinger, M., *et al*: Improved limits on the coupling of ultralight bosonic dark matter to photons from optical atomic clock comparisons. Phys. Rev. Lett. 130(25), 253001 (2023).
7. Sherrill, N., *et al*: Analysis of atomic-clock data to constrain variations of fundamental constants. New J. Phys. 25(9), 093012 (2023).
8. BIPM Recommended values of standard frequencies, https://www.bipm.org/en/publications/mises-en-pratique/standard-frequencies, last accessed 2025/06/18.
9. Polzik, E. S., Carri, J., Kimble, H. J.: Spectroscopy with squeezed light. Phys. Rev. Lett. 68(20), 3020–3023 (1992).
10. Shajilal, B., *et al*: 12.6 dB squeezed light at 1550 nm from a bow-tie cavity for long-term high duty cycle operation. Opt. Express 30(21), 37213–37223 (2022).
11. Ast, S., Mehmet, M., Schnabel, R.: High-bandwidth squeezed light at 1550 nm from a compact monolithic PPKTP cavity. Opt. Express 21(11), 13572–13579 (2013).
12. Lough, J., *et al*: First demonstration of 6 dB quantum noise reduction in a kilometer scale gravitational wave observatory. Phys. Rev. Lett. 126(4), 041102 (2021).
13. McMillan, P. J.: Mass models of the Milky Way. Monthly Notices of the Royal Astronomical Society 414(3), 2446–2457 (2011).
14. Flambaum, V. V., *et al*: Limits on variations of the quark masses, QCD scale, and fine structure constant. Phys. Rev. D. 69(11), 115006 (2004).
15. Dzuba, V. A., Flambaum, V. V.: Relativistic corrections to transition frequencies of AgI, DyI, HoI, YbII, YbIII, AuI, and HgII and search for variation of the fine-structure constant. Phys. Rev. A 77(1), 012515 (2008).
16. Jansen, P., Bethlem, H. L., Ubachs, W.: Perspective: Tipping the scales: Search for drifting constants from molecular spectra. J. Chem. Phys. 140(1), 010901 (2014).
17. Pašteka, L. F., *et al*: Material size dependence on fundamental constants. Phys. Rev. Lett. 122(16), 160801 (2019).
18. Xie, X., *et al*: Phase noise characterization of sub-hertz linewidth lasers via digital cross correlation. Opt. Lett. 42(7), 1217–1220 (2017).
19. Cantin, E., *et al*: An accurate and robust metrological network for coherent optical frequency dissemination. New J. Phys. 23(5), 053027 (2021).
20. Oswald, R., Vogt, V., Schiller, S.: Search for oscillating fundamental constants using a paired detector and vibrational spectroscopy. Phys. Rev. D 111(5), 055003 (2025).
21. Yuen, H. P., Chan, V. W. S.: Noise in homodyne and heterodyne detection. Opt. Lett. 8(3), 177–179 (1983).
22. Schwob, C.: Utilisation des faisceaux corrélés au niveau quantique produits par un Oscillateur Paramétrique Optique en spectroscopie de grande sensibilité. Thèse de doctorat. Sorbonne Université, Paris (1997), https://hal.science/tel-00011780/, last accessed 2025/06/18.
23. Loudon, R., Knight, P. L.: Squeezed light. Journal of Modern Optics, 34(6–7), 709–759 (1987).

24. Polzik, E. S., Carri, J., Kimble, H. J.: Atomic spectroscopy with squeezed light for sensitivity beyond the vacuum-state limit. Appl. Phys. B 55(3), 279–290 (1992).
25. Barbarat, J., *et al*: Compact and transportable iodine frequency-stabilized laser. In: Sodnik, Z., Karafolas, N., Cugny, B., (eds.) INTERNATIONAL CONFERENCE ON SPACE OPTICS; 2018, SPIE, vol. 11180, pp. 111800T. SPIE, Bellingham (2019).
26. Hald, J., *et al*: Fiber laser optical frequency standard at 1.54 μm. Opt. Express 19(3), 2052–2063 (2011).
27. Constantin, F. L.: Sensitivity to electron-to-proton mass ratio variation from $^{12}C_2H_2$ rovibrational transitions to $\nu_1+\nu_3$ and $\nu_1+\nu_2+\nu_4+\nu_5$ interacting levels. Vibrational Spectroscopy 85, 228–234 (2016).
28. Constantin, F. L.: Towards probing a variation of fundamental constants with optical clock transitions of $^{127}I_2$. In: Weinstein, D., *et al* (eds.) JOINT CONFERENCE OF THE IEEE IFCS AND ISAF; 2020, IEEE, vol. CFP20ISA-POD, pp. 443–445. IEEE, Piscataway (2020).
29. Baune, C., *et al*: Strongly squeezed states at 532 nm based on frequency up-conversion. Opt. Express 23(12), 16035–16041 (2015).

The Gilles Brassard Laboratory: An open Research Platform for Secure Quantum-Era Communications

Antoine Dierick[1], Lahatra Rakotondrainibe[1], Jonathan Pisane [1]

[1] Thales Belgium, Rue des Frères Taymans, 28, 1480 Tubize, Belgique
antoine.dierick@be.thalesgroup.com

Abstract. The advent of quantum computing poses a significant threat to the security of today's communication networks, making it urgent to protect both current and future data exchanges. While theoretical countermeasures—such as post-quantum cryptography (PQC) and quantum key distribution (QKD)—exist, these technologies still require substantial development before being integrated into real-world infrastructures. The Gilles Brassard Laboratory (GBL) addresses this challenge by providing a research and development platform dedicated to advancing, testing, and validating key components of quantum-resistant communication systems.

The GBL is structured around two main segments: a Quantum Cryptography Infrastructure (QCI) for the implementation and evaluation of PQC and QKD, and a radio communication segment that emulates both terrestrial (LTE/5G/6G) and space-based links (DVB-S, CCSDS, 5G NTN). This dual framework enables researchers and engineers to investigate secure end-to-end communication solutions within a realistic testing environment.

Initial use cases demonstrate proof-of-concept validations, including the implementation of secure end-to-end communications by combining QKD technologies with standardized radio link protocols. Transmissions are encoded in accordance with DVB-S and CCSDS transport layers (broadcast mode only), paving the way toward the deployment of secure, future-proof communication networks.

Keywords: Post Quantum Cryptography (PQC), Quantum Key Distribution (QKD), Quantum Communications infrastructure (QCI)

Introduction

Recent technological advances, particularly in quantum computing, pose significant challenges to the security of current information systems. While robust countermeasures—such as post-quantum cryptography (PQC) and quantum key distribution (QKD)—have been theoretically established, their practical development and deployment remain complex. To address this gap, the Gilles Brassard Laboratory (GBL) was created as a dedicated platform for research and experimentation in quantum cryptography. The laboratory is named after Gilles Brassard, who, together with Charles Bennett, introduced the seminal BB84 protocol [1], recognized as the first QKD scheme and a cornerstone for secure communications based on quantum mechanics.

The purpose of this paper is to present the objectives, architecture, and initial experimental results of the GBL. In particular, we describe how the laboratory enables the integration and validation QKD technologies within realistic communication environments, including satellite-based scenarios. While the platform focuses on QKD; PQC integration is foreseen in future releases.

This paper is structured as follows: Section 1 outlines the main threats faced by contemporary information systems. Section 2 describes the objectives and architectural design of the GBL. Section 3 presents initial experimental results focused on secure satellite communication scenarios using DVB-S (Digital Video Broadcasting – Satellite) and CCSDS (Consultative Committee for Space Data Systems) protocols. Finally, Section 4 concludes the paper and discusses directions for future work.

1 Context

1.1 The Quantum Threat

The security of contemporary information systems relies on a key assumption: that no adversary possesses sufficient computational resources to decrypt protected communications within a reasonable timeframe. Under this paradigm, the security of a system depends both on the adversary's available resources and on the required duration of data confidentiality.

This approach faces a critical limitation. If technological or algorithmic advances occur within the confidentiality period, data can be compromised. An adversary may simply store encrypted communications and later decrypt them once improved capabilities become available. This threat becomes particularly acute with the emergence of quantum computers, which exploit principles of quantum mechanics to solve certain computational problems more efficiently than classical machines.

Two quantum algorithms illustrate this. Grover's algorithm [2] accelerates brute-force search in unstructured spaces, providing a quadratic speedup. More critically, Shor's algorithm [3] efficiently solves the integer factorization problem in polynomial time, undermining the hardness assumptions on which RSA and elliptic-curve cryptography (ECC) are based. Since these schemes are widely deployed in current infrastructures, their security is directly endangered by the rise of large-scale quantum computing [4].

This looming vulnerability highlights the urgency of developing quantum-resistant alternatives to secure communication systems.

1.2 Possible solutions: PQC & QKD

In response to the quantum threat, two major strategies are being pursued. The first, Post-Quantum Cryptography (PQC), seeks to secure communications through cryptographic schemes based on mathematical problems believed to remain intractable even for quantum computers. Prominent candidates include lattice-based, code-based, multivariate, and hash-based constructions [5]. While the security of PQC is conditional, depending on the absence of future algorithmic or technological breakthroughs, its main advantage lies in compatibility: PQC can often be integrated into existing communication infrastructures with minimal hardware changes. This makes it a pragmatic solution for rapidly mitigating "store now, decrypt later" attacks, serving as a near-term safeguard while more advanced technologies mature.

The second strategy, Quantum Key Distribution (QKD)**,** leverages the laws of quantum physics to achieve information-theoretic security. QKD enables two parties to establish a shared secret key by exchanging quantum states, which serve as physical carriers of information. A defining feature of QKD is its inherent ability to detect eavesdropping: any measurement of the exchanged states by a third party induces anomalies, due to the fundamental principle that measurement disturbs a quantum system [6].

While QKD offers unconditional security in theory, practical deployment faces challenges. Quantum states are highly sensitive to environmental disturbances, which severely limits long-distance communication without additional technologies. Although quantum repeaters, whose foundational concept was introduced in 1998 [7], are regarded as the key solution to overcome the distance limitations of optical fibers, this technology has not yet reached maturity. Current quantum communication networks therefore rely on alternative approaches, such as trusted relays [8] or satellite-based links [9].

At the GBL, research efforts are directed toward both strategies, aiming to combine the near-term deployability of PQC with the long-term security guarantees of QKD to build quantum-resilient communication systems**.**

1.3 An Open Quantum Cryptography Laboratory for Researchers

The primary objective of the GBL is to deploy an end-to-end communication network integrating both software and hardware components in order to test, assess, and validate quantum-based cybersecurity and quantum-resilient solutions under realistic conditions. In particular, the laboratory enables:

- **Characterization and assessment of cyber threats**, such as vulnerabilities and risks associated with quantum-enabled attacks, as well as more conventional cybersecurity risks including jamming and spoofing;
- **Deployment and validation of quantum-based security solutions**, evaluated against the performance requirements and characteristics of modern communication networks.

The GBL is conceived as an open and collaborative research facility, accessible to Belgian and European universities, as well as to local and national small and medium-sized enterprises (SMEs) seeking to validate technologies. It also serves as a framework for cooperation with other European laboratories, supporting joint efforts in the context of European Commission initiatives and missions.

The design and management of the GBL have been entrusted to the Thales Group, ensuring professional expertise and operational support. The facility is intended for researchers, engineers, and specialists working on quantum technologies and their integration into secure communication infrastructures.

2 Architectural design of the GBL quantum research platform

The GBL quantum research platform is functionally divided into three complementary segments:

- **Radio communications segment** (terrestrial and space), responsible for emulating end-to-end communication links;
- **User segment**, which provides tools for control, configuration, definition of measurement scenarios, data collection, and visualization of results;
- **Quantum Cryptography Infrastructure (QCI) segment**, dedicated to quantum key distribution and secure key management.

This modular architecture allows the integration of heterogeneous data streams such as quadrature signals (IQ), network traffic (IP), and quantum samples (QC), into a unified environment accessible through a dynamic graphical user interface.

2.1 Radio Communications Segment

The radio segment emulates end-to-end communication links with both terrestrial (LTE/5G/6G) and space-based (DVB-S, CCSDS, 5G NTN) standards. It combines virtualized environments and software-defined radios (SDRs), supported by high-throughput interconnects.

A channel emulator reproduces realistic propagation effects such as delay, Doppler, attenuation, and interference, enabling the evaluation of secure transmissions under degraded conditions.

Signal processing is performed in real time using GNU Radio, supporting both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) configurations.

2.2 Quantum Cryptography Infrastructure (QCI) Segment

The QCI segment ensures secure key distribution between users through a time-bin implementation of the BB84 protocol. It operates as a key provider, delivering quantum-generated keys to consumers via the standard ETSI GS QKD 014 interface. Keys are subsequently used in symmetric encryption modules (e.g., AES), which may be implemented as dedicated devices or integrated directly into SDRs.

A Key Management System (KMS) is integrated at each QKD node to arbitrate key allocation, ensuring flexibility in adding, removing, or transmitting keys across the network. The encryption modules retrieve keys through the ETSI REST interface and apply them to secure the data streams.

The following Figure 1 provides a high-level view of the GBL research platform architecture.

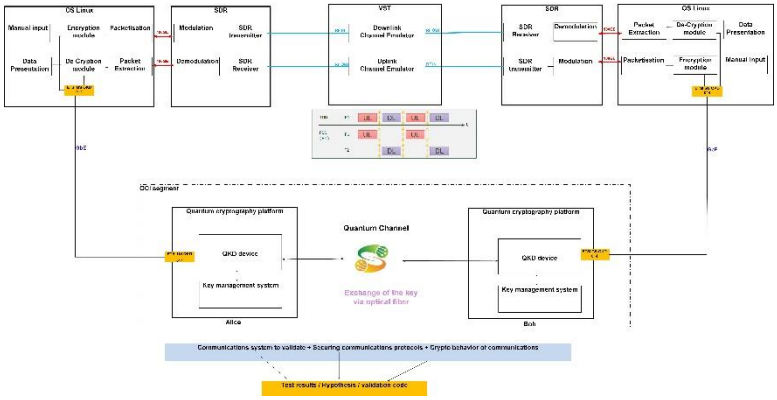


Figure 1. Block diagram of GBL research platform

3 Experimental results

The experimental validation of the GBL platform was carried out by demonstrating secure end-to-end communications combining the radio and QCI segments. Two layers of validation were considered:

- Physical-level validation, assessing the performance of the QKD link through key generation metrics, error rates, and hardware stability.
- System-level validation, focusing on the integration of QKD keys into communication protocols such as DVB-S and CCSDS;

3.1. Physical-Level Validation of the QKD Link

The performance of the quantum channel was assessed using data collected from the Alice and Bob devices, as well as from the FPGA processing stage.

Figure 2 shows the FPGA-reported quantum bit error rate (QBER) measured over time. Values remained within the expected range (<5%), validating the security of the generated keys. Figure 3 presents the FPGA-reported sifted key rate, which confirmed the continuous availability of correlated bits for key delivery. As expected, the final secret key rate (after error correction and privacy amplification) is lower, but still sufficient to support encryption for the DVB-S and CCSDS scenarios.

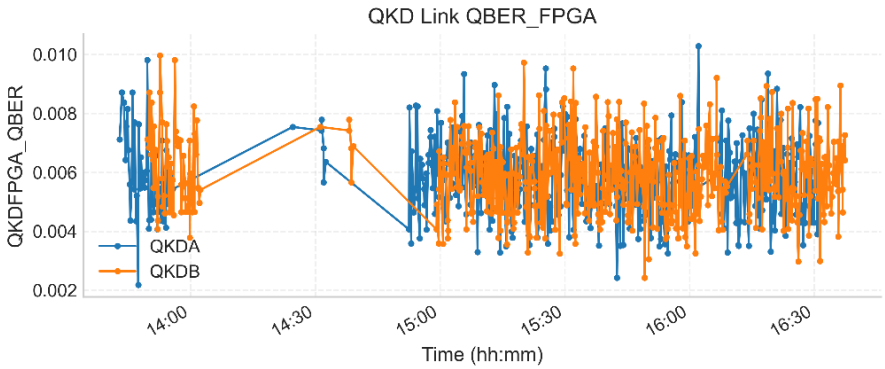


Figure 2. Quantum Bit Error Rate as logged at the FPGA stage (sifted key error, i.e. after basis sifting but before reconciliation and privacy amplification). The plotted series reflects physical impairments (detector dark counts, timing jitter, visibility fluctuations). For BB84 the theoretical one-way security bound is ≈11% [10]; in our implementation a conservative operational threshold of ~5% was adopted to ensure efficient error correction and acceptable finite-key performance.

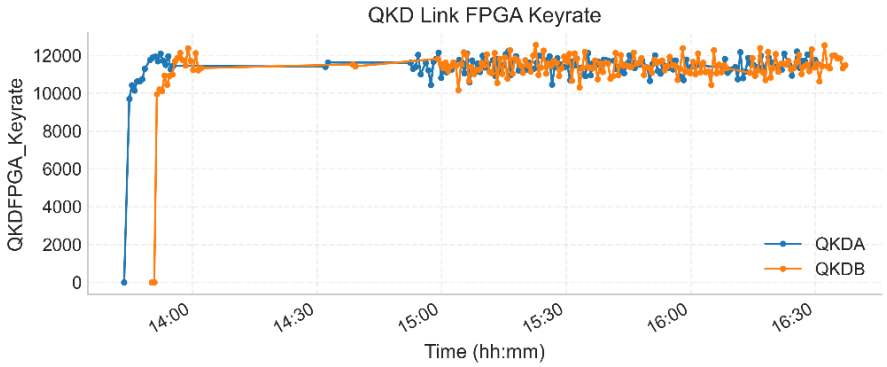


Figure 3. Sifted key rate reported by the FPGA processing stage. The sustained throughput demonstrates the continuous availability of correlated key material for secure encryption, ensuring integration with higher-layer communication protocols.

Additional hardware monitoring confirmed proper operating conditions. Board and detector temperatures remained stable, and optical visibility provided further evidence of link quality. Variations in laser modulation and detector bias were found to correlate with small fluctuations in QBER, but without impacting overall key delivery.

Together, these results demonstrate that the QKD segment operated within expected physical tolerances and successfully generated cryptographic key material suitable for integration into higher-layer communication protocols. The combined analysis of detections, QBER, key rate, and stability parameters provides a comprehensive validation of the quantum link under realistic operating conditions.

3.2 System-Level Validation

Two representative use cases were implemented to illustrate the platform's capabilities:

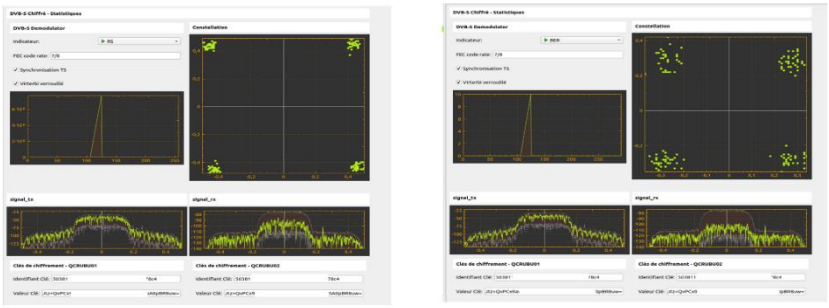
3.2.1 DVB-S scenario

The first use case emulates a satellite video broadcast using the DVB-S (Digital Video Broadcasting – Satellite) standard. A video file was transmitted in real time between two machines, with simultaneous playback at the receiver. Two configurations were tested:

- Unencrypted DVB-S transmission, serving as a baseline scenario;
- Encrypted DVB-S transmission, where the data stream was secured using QKD-generated keys.

Successful validation was demonstrated by receiving uninterrupted video streams in real time (see Figure 4.a). Furthermore, channel impairments were introduced through the VST (e.g., gain loss, delay, Doppler shift, and attenuation) to simulate realistic

conditions. Even under degraded conditions (e.g., 20 dB attenuation), secure transmissions were maintained (see Figure 4.b).



a) b)

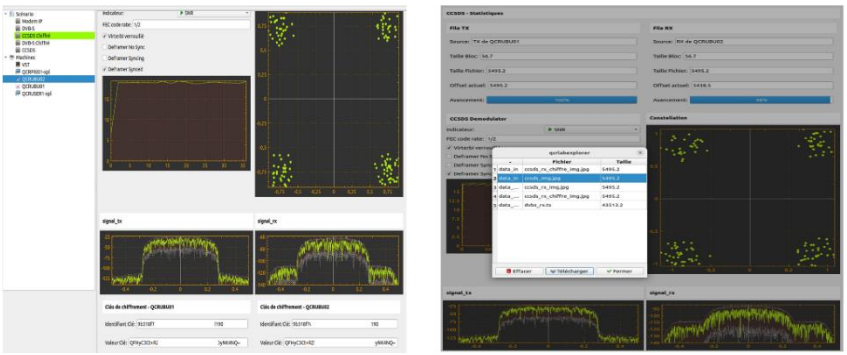
Figure 4. Demonstration of a secure real-time video broadcast over a DVB-S link. The stream was encrypted with QKD-generated keys and successfully received in real time, even under degraded channel conditions (attenuation and Doppler shift) – a) VST no degradation, b) VST with 20 dB attenuation

3.2.2 CCSDS Scenario

The second use case targeted file transfer in space communication contexts, using the CCSDS (Consultative Committee for Space Data Systems) protocol. The setup involved transmitting binary files from a source PC to a target PC under two configurations:

- Standard CCSDS file transfer, without encryption;
- Encrypted CCSDS transfer, where data confidentiality was ensured by QKD-generated keys.

Validation consisted of confirming reliable file delivery without interruption. As with the DVB-S case, simulated channel degradations (e.g., signal-to-noise ratio loss, frequency shift) were applied to test the resilience of the secured link (see Figure 5).



a) b)

Figure 5. Validation of a secure file transfer using the CCSDS protocol. QKD keys were integrated into the link encryption process, ensuring reliable and confidential end-to-end delivery despite channel impairments. – a) VST no degradation, b) VST with 20 dB attenuation/1MHz frequency shift.

4 Conclusion & Future work

This paper has presented the design, implementation, and initial validation of the Gilles Brassard Laboratory (GBL), an open research platform dedicated to quantum cryptography and quantum-resilient communications. The platform combines a radio communications segment, capable of emulating both terrestrial and space-based links, with a Quantum Cryptography Infrastructure (QCI) segment implementing the BB84 protocol and standardized key management interfaces.

The first experimental results demonstrated the feasibility of integrating QKD with classical communication standards. In particular, successful proof-of-concept validations were achieved for secure end-to-end transmissions under DVB-S and CCSDS protocols. In both scenarios, QKD-generated keys were delivered through the ETSI GS QKD 014 interface and used to encrypt real-time video streams or file transfers. Importantly, transmissions remained secure and reliable even when subject to degraded channel conditions such as attenuation, delay, or frequency shifts. These results confirm the capability of the GBL to provide a realistic environment for testing and validating quantum-resilient communication systems.

Looking ahead, several research directions are envisioned:

- Extension to new communication standards: integration of 5G, 6G, and non-terrestrial networks (NTN) within the GBL testbed.
- In future iterations of the platform, PQC implementations will be integrated and validated alongside QKD, enabling comparative studies and hybrid secure communication scenarios.
- Advanced quantum technologies: investigation of quantum repeaters, trusted-node architectures.

- Enhanced resilience testing: characterization of system performance under more complex attack models, including classical cyber threats and combined quantum–classical adversaries.
- Training and education: providing a platform for researchers, engineers, and students to experiment with and learn about next-generation secure communications.

By combining a realistic radio environment with quantum key distribution and post-quantum cryptography research, the GBL paves the way for future secure communication networks that remain resilient in the quantum era.

5 Acknowledgement

The Gilles Brassard quantum research Laboratory is located at the IDELUX site in Transinne, and is funded by the Walloon Region of Belgium.

References

1. Bennett CH, Brassard G (2014) Quantum cryptography: Public key distribution and coin tossing. Theoretical Computer Science 560:7–11. https://doi.org/10.1016/j.tcs.2014.05.025
2. Grover LK (1996) A fast quantum mechanical algorithm for database search. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. Association for Computing Machinery, Philadelphia, Pennsylvania, USA, pp 212–219
3. Shor PW (1997) Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM Journal on Computing 26:1484–1509. https://doi.org/10.1137/s0097539795293172
4. Mosca M (2018) Cybersecurity in an Era with Quantum Computers: Will We Be Ready? IEEE Security & Privacy 16:38–41. https://doi.org/10.1109/MSP.2018.3761723
5. Chen, L., Jordan, S., Liu, Y.K., Moody, D., Peralta, R., Smith-Tone, D., et al. (2016) Report on Post-Quantum Cryptography. US Department of Commerce, National Institute of Standards and Technology, Vol. 12, Gaithersburg. https://doi.org/10.6028/NIST.IR.8105
6. Gisin N, Ribordy G, Tittel W, Zbinden H (2002) Quantum cryptography. Rev Mod Phys 74:145–195. https://doi.org/10.1103/RevModPhys.74.145
7. Briegel H-J, Dür W, Cirac JI, Zoller P (1998) Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication. Phys Rev Lett 81:5932–5935. https://doi.org/10.1103/PhysRevLett.81.5932
8. Sasaki M, Fujiwara M, Ishizuka H, et al. (2011) Field test of quantum key distribution in the Tokyo QKD Network. Opt Express.19(11):10387-10409. https://doi.org/10.1364/OE.19.010387
9. Liao S-K, Cai W-Q, Liu W-Y, Zhang L, Li Y, Ren J-G, Yin J, Shen Q, Cao Y, Li Z-P, Li F-Z, Chen X-W, Sun L-H, Jia J-J, Wu J-C, Jiang X-J, Wang J-F, Huang Y-M, Wang Q, Zhou Y-L, Deng L, Xi T, Ma L, Hu T, Zhang Q, Chen Y-A, Liu N-L, Wang X-B, Zhu Z-C, Lu C-Y, Shu R, Peng C-Z, Wang J-Y, Pan J-W (2017) Satellite-to-ground quantum key distribution. Nature 549:43–47. https://doi.org/10.1038/nature23655
10. Shor PW, Preskill J (2000) Simple Proof of Security of the BB84 Quantum Key Distribution Protocol. Phys Rev Lett 85:441–444. https://doi.org/10.1103/PhysRevLett.85.441

Metrology for real-world QKD

Alice Meda[1], Salvatore Virzì[1], Marco Gramegna[1], Giorgio Brida[1], Marco Genovese[1,2], Ivo Pietro Degiovanni[1]

[1] Istituto Nazionale di Ricerca Metrologica (INRIM), Strada delle Cacce 91, I-10135 Torino, Italy

[2] Istituto Nazionale di Fisica Nucleare (INFN), Sezione Torino, Via Giuria 1, I-10125 Torino, Italy

a.meda@inrim.it

Abstract. Quantum Key Distribution (QKD) is a technology that enables the sharing of secret cryptographic keys between two distant users (Alice and Bob), with intrinsic security guaranteed by the fundamental laws of nature.

QKD has become a mature technology, and in Europe, all 27 member states are collaborating on a European Commission initiative (EuroQCI) to design, develop, and deploy a quantum communication infrastructure. In Italy, the QUID project is responsible for implementing the Italian segment of EuroQCI.

QKD relies on single photons to secure the distribution of the keys and, to become a viable real-world solution, the metrological characterization of optical components and systems is fundamental. To obtain the appropriate security requirements, test and evaluation methods at single-photon level need to be developed; in particular, since the single-photon detectors represent the most vulnerable part of a QKD system, their characterization in terms of operating parameters (quantum efficiency, dead time, jitter, afterpulsing..) is of the utmost importance.

We present the INRIM efforts in the quantum efficiency calibration of single-photon avalanche detectors (SPADs), focusing on QKD application. The detection efficiency is evaluated for a fibre-coupled InGaAs/InP-SPAD and for a free-space Si-SPAD. The calibration is performed using different experimental setups and reference standards with proper traceability chains at the wavelength of 1550 nm and 850 nm respectively. Dependence of detection efficiency on polarization in superconducting nanowire single-photon detectors (SNSPDs) is also reported.

The work is fundamental to align the Italian deployment of QKD, in the framework of QUID, with validation needs, providing test services for the characterization, validation and certification for QKD.

.

Keywords: Quantum Key Distribution, Single-photon Avalanche Detectors, Quantum Efficiency

1 Introduction

1.1 QKD and Single-photon avalanche diodes

Quantum Key Distribution (QKD) is currently the most advanced technology for sharing cryptographic secret keys with a level of security that is fundamentally independent of computational power [1, 2]. Practical consideration of QKD systems and network architectures based on specific use cases has already begun [3, 4], and QKD metropolitan networks have been successfully demonstrated across the globe [5, 6].

The European Union (EU) is actively advancing in this domain, having signed a declaration with all 27 member states to jointly pursue the development of the European Quantum Communication Infrastructure (EuroQCI) [7]. The EuroQCI aims to establish secure quantum communication networks across all EU countries. As part of this initiative, the Italian project QUID (Quantum Italy Deployment) [8] has been launched to initiate the deployment of Italy's contribution to the EU-wide network by implementing QKD systems and building Quantum Metropolitan Area Networks (QMANs) in various cities.

Single-photon avalanche diodes (SPADs) represent the most widely adopted commercial solution for single-photon detection technologies. Among their many applications, QKD relies heavily on SPADs for photon detection [9].

As with all cryptographic systems, practical implementation of QKD demands rigorous investigation of the components involved to ensure their correct operation and integration [10-12]. It is therefore crucial to develop theoretical models that accurately describe the behavior of each system element and establish boundaries on the acceptable error rates during key exchange. Deviations from the ideal device behavior can open side channels or backdoors, which may be exploited by a potential eavesdropper to compromise the communication. In particular, numerous detection-based attack strategies have been documented, including backflash-related attacks [13] and the more severe class of detector control attacks [14].

To maintain QKD security and prevent such tailored attacks, it is essential to model and characterize the SPAD response under various operational conditions, taking into account key parameters such as detection efficiency, dead time, and dark count rate.

Here we focus on the calibration of the detection efficiency of a fibre-coupled InGaAs/InP-SPAD [15, 16, 17] and of a free-space Si-SPAD [18].

2 Detection efficiency characterization

2.1 Fibre-coupled InGaAs/InP-SPAD

The detection efficiency is the probability that a photon of a specific energy (wavelength) incident at the optical input will be detected within a detection gate. The most exploited measurement approach in metrological institutes is based on the traditional substitution method [15], where the detection efficiency is inferred by comparing the output of the single photon detector and the output of a calibrated power meter, illuminated by the same light but with different levels of attenuations.
We develop a facility for the traceable measurement of detection efficiency for the new generation single-photon detectors based on semiconductor (InGaAs/InP SPAD).
The measurement protocol relies on a detection efficiency measurement of detectors that takes advantage of an experimental set-up whose schematic description is reported in Figure 1.

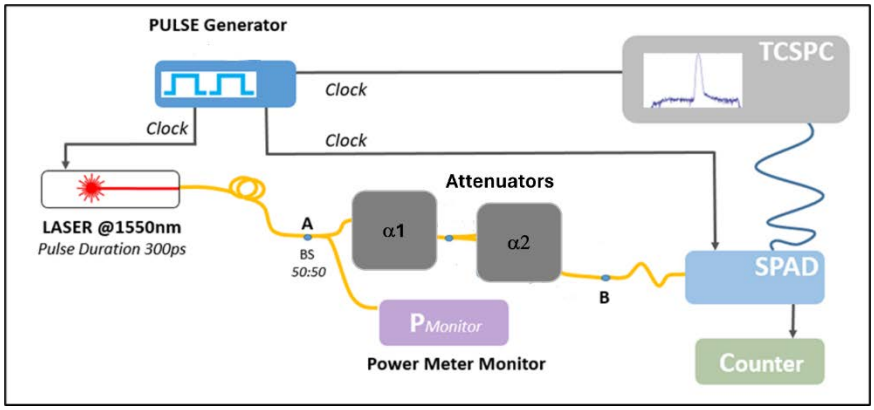


Figure 1 Experimental setup of INRiM InGaAs/InP SPAD detection efficiency calibration facility

A pulsed laser at 1550nm is externally triggered, with a pulse generator, that triggers also a Time-Correlated Single Photon Counting (TCSPC) device, is connected to a first variable attenuator that provides a nominal transmissivity α_1 and a second variable attenuator that provides a nominal transmissivity α_2, for a total transmissivity $\tau = \alpha_1\alpha_2$.
The detection efficiency of the DUT is evaluated according to the formula:

$$\eta = \frac{P_C}{\tau P_0} \quad (1)$$

where P0 is the power in point B and PC is the average power of the effective photons measured by the detector, calculated from the photon rate absorbed by the SPAD

corrected for dark counts, dead time and considering the energy of the photon at λ = 1550 nm.

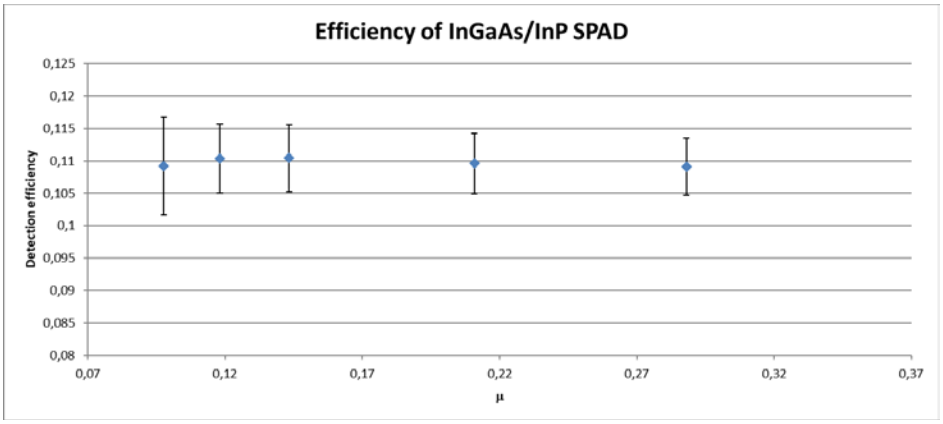


Figure 2 detection efficiency of an InGaAs/InP SPAD

Figure 2 reports detection efficiency as a function of the mean photon number per pulse μ. A proper model for the correction for the effects of dead time on the detection efficiency is applied; details can be found in [15, 16, 17].

The setup dedicated to the detection efficiency measurement is now also suitable for light-polarization-sensitive detectors as SNSPDs, thanks to the introduction of a dedicated polarization measurement and control system, that allows to reconstruct the estimated efficiency distribution across the Poincaré sphere's surface and with a relative uncertainty <5% (k=2) [18].

2.2 Free space Si-SPAD

A second INRiM measurement facility is designed to determine the detection efficiency at a wavelength of 850 nm, where photons are addressed in free space with a spatial distribution diameter smaller than 40 μm.
The calibration setup is shown in Figure 1. The photon source is a Ti:sapphire (Ti:Sa) laser with tunable wavelength, set to λ=850.711±0.006 nm. Optical power is modulated by a polarization-based variable attenuator composed of a half-wave plate and a polarizer, allowing precise control of the single-photon rate.

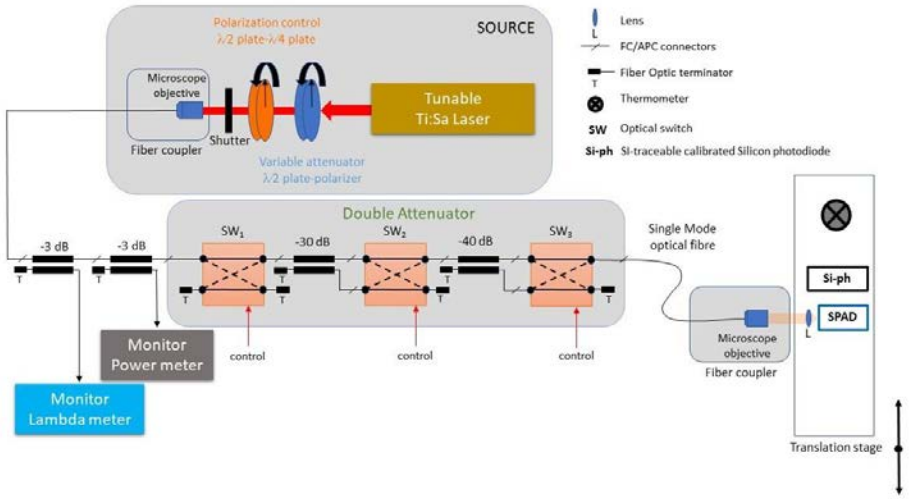


Figure 3 Experimental setup of INRiM Si-SPAD detection efficiency calibration facility

The photon beam subsequently passes through a mechanical shutter and is coupled into a single-mode fiber via a 20× microscope objective. Two 50:50 beamsplitters are employed to continuously monitor both the optical power and the source's wavelength stability. The photons then enter a secondary attenuation stage, composed of two strongly unbalanced beam splitters (999:1 and 9999:1), which introduce fixed attenuations of 30 dB and 40 dB respectively. This double-attenuator setup achieves a total attenuation of 70 dB when the appropriate optical path is selected via three PC-controlled optical switches, matching the total transmissivity τ.

After attenuation, photons are out-coupled and collimated in free space with a Gaussian spatial profile and sent into a shielded enclosure to suppress environmental photon noise. A motorized translation stage positions the selected detector based on the optical power level: a SI-traceable silicon photodiode (Hamamatsu S1337-1010BQ) is used for measuring the input optical power P_0, while the silicon SPAD (DUT) is used to count the detected photons N when the photon flux is reduced to the single-photon regime via the calibrated attenuation τ and to estimate P_c.
Since the typical active area of a free-space silicon SPAD is on the order of hundreds of micrometers, a focusing lens is used to reduce the spatial mode diameter to 40 μm at the detector.

The DUT alignment is optimized using three actuators for translation in the x, y, and z directions. All instruments in the setup are controlled via a LabVIEW interface, and measurement procedures are fully automated.

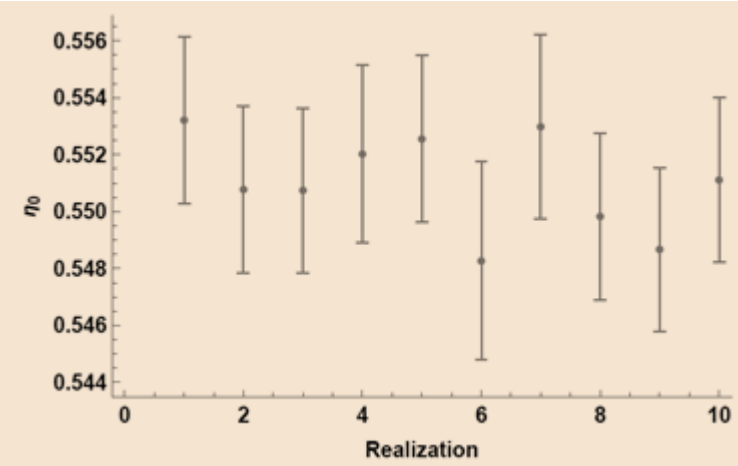


Figure 4 Quantum efficiency measurement of a Perkin-Elmer Si-SPAD

We tested detection efficiency of a Perkin-Elmer Si-SPAD. The results are illustrated in Figure 4. The resulting estimations of efficiency in different runs of measurement exhibit strong agreement. The average value obtained for the detection efficiency is: $\langle \eta_0 \rangle = 0.551 \pm 0.003$ [19].

References

1. C. H. Bennett and G. Brassard, in Proceedings of IEEE International Conference on Computers, Systems and Signal Processing (IEEE, Bangalore, India, 1984).
2. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. 74, 145 (2002).https://doi.org/10.1103/RevModPhys.74.145
3. A. Lewis and M. Travagnin, A Secure Quantum Communications Infrastructure for Europe: Technical background for a policy vision. https://doi.org/10.2760/180945 (2022).
4. A. Meda, A. Mura, S. Virzì, et al. QKD protected fiber-based infrastructure for time dissemination. Sci Rep 15, 13419 (2025). https://doi.org/10.1038/s41598-025-97480-8

5. M. Peev, et al. The SECOQC quantum key distribution network in Vienna. New J. Phys. 11(7), 075001. https://doi.org/10.1088/1367-2630/11/7/075001 (2009).
6. D. Stucki, et al. Long-term performance of the swissquantum quantum key distribution network in a field environment. New J. Phys. 13(12), 123001. https://doi.org/10.1088/1367-2630/13/12/123001 (2011).
7. https://digital-strategy.ec.europa.eu/en/policies/european-quantum-communication-infrastructure-euroqci
8. https://quid-euroqci-italy.eu/it/
9. R. H. Hadfield, Nat. Photonics, 3, 696–705 (2009). https://doi.org/10.1038/nphoton.2009.230
10. Y. Liang and H. P. Zeng, Sci. China: Phys., Mech. Astron. 57, 1218 (2014).https://doi.org/10.1007/s11433-014-5450-0
11. T.-Y. Chen, J. Wang, H. Liang, W.-Y. Liu, Y. Liu, X. Jiang, Y. Wang, X. Wan, W.-Q. Cai, L. Ju, L.-K. Chen, L.-J. Wang, Y. Gao, K. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, Opt. Express 18, 27217 (2010).https://doi.org/10.1364/OE.18.027217
12. D. Stucki, M. Legré, F. Buntschu, B. Clausen, N. Felber, N. Gisin, L. Henzen, P. Junod, G. Litzistorf, P. Monbaron, L. Monat, J.-B. Page, D. Perroud, G. Ribordy, A. Rochas, S. Robyr, J. Tavares, R. Thew, P. Trinkler, S. Ventura, R. Voirol, N. Walenta, and H. Zbinden, New J. Phys. 13, 123001 (2011).https://doi.org/10.1088/1367-2630/13/12/123001
13. A. Meda, I. P. Degiovanni, A. Tosi, Z. Yuan, G. Brida, and M. Genovese, Light: Sci. Appl. 6, e16261 (2017).https://doi.org/10.1038/lsa.2016.261
14. L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics 4, 686 (2010).https://doi.org/10.1038/nphoton.2010.214
15. M. López, A. Meda, G. Porrovecchio, R. A. Starkwood (Kirkwood), M. Genovese, G. Brida, M. Šmid, C. J. Chunnilall, I. P. Degiovanni, and S. Kück, "A study to develop a robust method for measuring the detection efficiency of free-running InGaAs/InP single-photon detectors", EPJ Quantum Technol. 7, 14, (2020).
16. H. Georgieva, A. Meda, H. Hofer, S. M. F. Raupach, M. Gramegna, I. P. Degiovanni, M. Genovese, M. López and S. Kück, "Detection of ultra-weak laser pulses by free-running single-photon detectors: Modeling dead time and dark count effects", Appl. Phys. Lett. 118, 174002, (2021).
17. S. M. F. Raupach, I. P. Degiovanni, H. Georgieva, A. Meda, H. Hofer, M. Gramegna, M. Genovese, S. Kück, and M. López, "Detection rate dependence of the inherent detection efficiency in single-photon detectors based on avalanche diodes", Phys. Rev. A 105, 042615 (2022)
18. Manuscript in preparations.
19. S. Virzì, A. Meda, E. Redolfi, M. Gramegna, G. Brida, M. Genovese, I. P. Degiovanni; Detection efficiency characterization for free-space single-photon detectors: Measurement facility and wavelength-dependence investigation. Appl. Phys. Lett. 25, 125 (22): 221108 (2024).

Development of National Quantum-Safe Network Testbed in Singapore

Hao Qin[1] *, Jing Yan Haw[1] *, Tosanut Rimprongern[1], Matthew Wee[1], Nelly Ng[2], Christian Kurtsiefer[1], Michael Kasper[3], and Alexander Ling[1]

[1] Centre for Quantum Technologies, National University of Singapore
hao.qin@nus.edu.sg, jingyan.haw@nus.edu.sg
[2] School of Physical and Mathematical Sciences, Nanyang Technological University
[3] Fraunhofer Singapore Research Centre@NTU, Nanyang Technological University

Abstract. We present the strategic framework and technical foundations of the National Quantum-Safe Network (NQSN) Testbed — a resilient and fully interoperable quantum-safe network. Built on a star-topology architecture over production-grade fiber infrastructure, the testbed supports multi-protocol quantum key distribution (QKD), post-quantum cryptography (PQC) and accommodates diverse applications from multiple vendors. Interoperability is enabled by a centralized key and network management system, ensuring seamless integration across heterogeneous technologies. We showcase a range of reference applications, including secure data center interconnects, edge computing, hybrid QKD–post-quantum cryptography (PQC) encryption, and multi-layer integration within the OSI stack. These implementations underscore the feasibility and adaptability of deploying quantum-safe technologies in complex, multi-input, multi-output network environments.

Keywords: Quantum key distribution · Quantum key distribution network · Quantum-safe · Quantum-network · Testbed.

1 Introduction

As the quantum computing technology landscape undergoes rapid evolution, there is an urgent need to fortify the communication infrastructure that handles high-value assets or requires long-term protection with quantum-safe security enhancements. In this respect, it is desirable to have a robust quantum-safe network (9; 11; 13; 2; 7) that is highly interoperable, supporting various types of quantum key distribution (QKD) protocol implementations while ensuring seamless compatibility with post-quantum cryptography (PQC).

In this work, we present the strategic approach and delve into technical considerations integral to the construction of a resilient and fully interoperable quantum-safe network (National Quantum-Safe Network, Singapore). Our

* These authors contributed equally to this work.

quantum-safe network testbed employs a star-type network topology, deployed over existing production-grade fiber infrastructure (Fig. 1(a)), with more than 1250 km of dark fibers in total. The testbed features the interoperability of multiple QKD protocols and full-scale Open Systems Interconnection (OSI) layers of applications integration from various vendors.

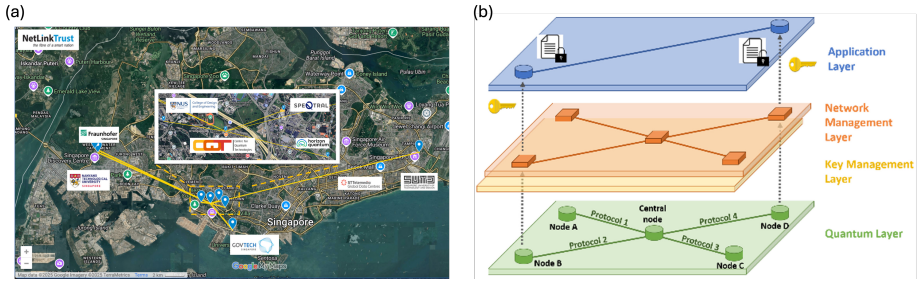


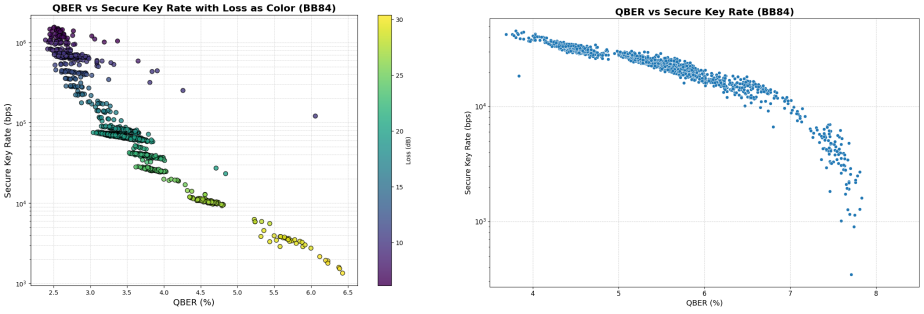
Fig. 1. (a) NQSN Fiber Network, (b) Different layers in NQSN testbed

The systematic construction of the testbed unfolds across several logical layers ([5]; [4]): the quantum layer, key management (KM) layer, QKD network management layer, and application (App) layer (Fig. 1(b)). This layered approach facilitates the configuration and tailoring of the testbed for scalability and interoperability, both horizontally and vertically. Such a structured framework also enables PQC seamless integration at KM and application layer, and also helps to understand the integration and security at each layer of the network architecture. In the following, we introduce some key features of the NQSN testbed.

2 Quantum layer: Multi-QKD protocol

In the Quantum layer, 5 commercially ready QKD devices are deployed to connect the central node with each remote node through the existing production-grade fiber infrastructure. The total loss of all QKD links reaches about 50 dB, with only passive components in between. Commercial QKD devices implementing protocols include BB84, coherent one-way (COW), continuous variable Gaussian modulation coherent state, and entanglement-based BBM92. These QKD devices are evaluated through their functions, performances, and long-term stability over the fiber infrastructure under realistic conditions. Fig. 2 gives an example of the secure key rate performance over the loss changes and Quantum Bit Error Rate (QBER) manipulation of one of the BB84 QKD systems. Fig. 2 (a) measures the secure key rate and QBER behavior with a variable optical attenuator (VOA) connected to the deployed fiber with loss of 15.45 dB. Fig. 2 (b) measures the secure key rate over the QBER manipulation using a dedicated tool to delay the pulses to introduce errors over QKD signal pulses. At

last, the performance of different QKD systems in terms of secure key rate are summarized in the Fig. 3.



(a) Secure key rate vs loss change

(b) Secure key rate vs QBER manipulation

Fig. 2. Secure key rate under different quantum channel conditions.

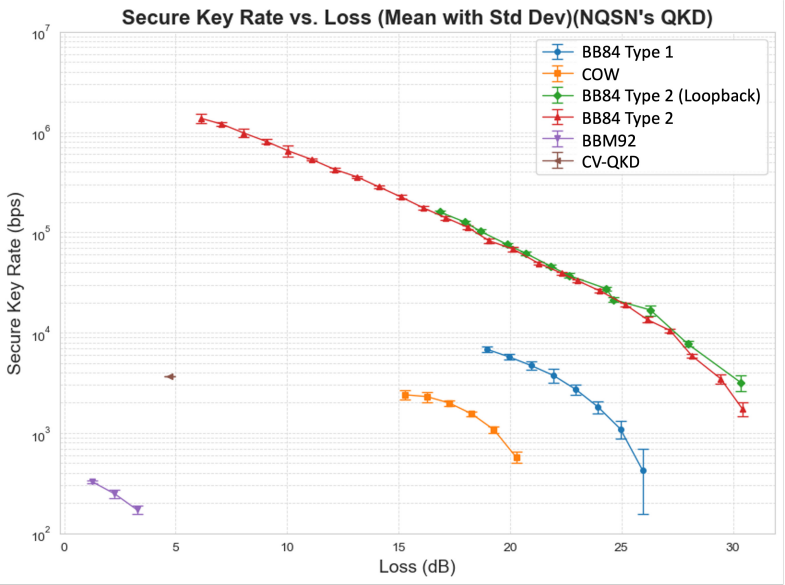


Fig. 3. Secure key rate versus channel loss for multiple QKD systems in NQSN. Error bars indicate measurement variability.

3 KM layer: Interoperable key management services and interfaces

At the heart of our network is an interoperable key management service (KMS) solution developed for the NQSN testbed deployed over fiber infrastructure. The KMS ensures network interoperability, scalability, resilience, and crypto-agility, supporting end-to-end symmetric key applications. It is designed to work with diverse network topologies and to integrate with various QKD devices in the quantum layer and applications in the App layer through standardized interfaces ([6]; [3]).

Within the NQSN testbed, the KMS operated with five commercial QKD systems using distinct protocols from different vendors and demonstrated compatibility with industry-grade symmetric key applications across OSI layers. Key challenges—such as key synchronization, key storage, relaying, interface compatibility, key refresh, and entity authentication—are identified and analyzed. Fig. 4 (Left) demonstrates the key database that stores QKD keys from different QKD devices.

We also evaluated the KMS interfaces, including ETSI QKD GS 014 standards ([3]) between QKD-KMS and KMS-application layers, using metrics such as response time and latency (Fig. 4 (Right)). Interoperability issues arising from implementation variations were also assessed to enhance compatibility and performance.

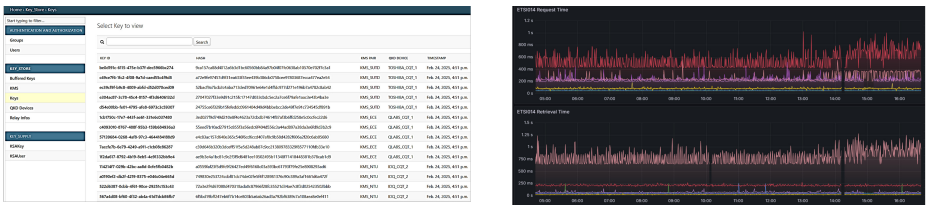


Fig. 4. (Left) NQSN KMS Key Database (Right) ETSI GS QKD 014 Request Time and Retrieval Time

We further showcased a commercial KMS solution (evolutionQ) with four types of QKD systems (plus a QKD simulator) in a linear topology spanning five QKD nodes, validated by an end-to-end QKD virtual private network (VPN) application (Fig. 5, top). Advanced approaches, including Distributed Symmetric Key Establishment based on secret sharing (Quantum Bridge), were also demonstrated for cross-border connectivity and QKD integration. (Fig. 5, bottom)

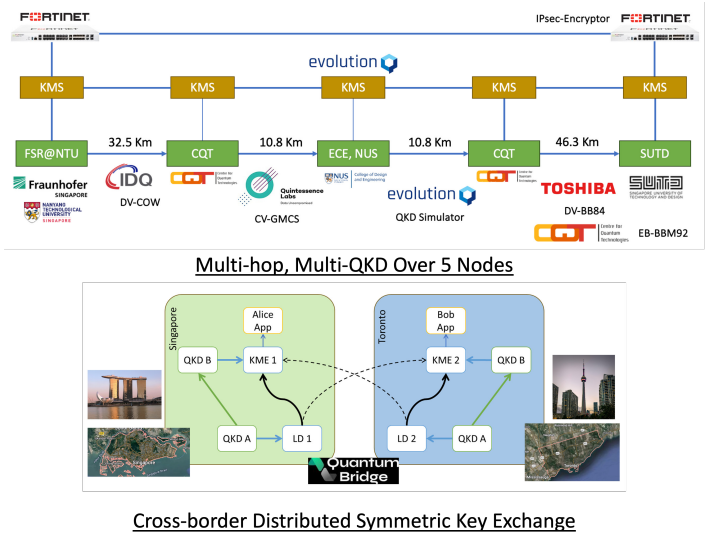


Fig. 5. Key management layer use cases

4 Application layer: Versatile Reference Use Cases & Hybrid QKD-PQC Integration

In the application layer, various quantum-secured use cases, spanning data center applications, edge computing, hybrid QKD-PQC encryption, and integration at different OSI layers, are explored from a technical perspective, providing critical insights into the deployment in QKDN with the multi-input/multi-output KM architecture. We focus on practical deployments within the NQSN, showcasing key integration scenarios that highlight the synergy between QKD and PQC. We demonstrated QKD-secured applications across various OSI layers, including physical, data link and network layers, illustrating how secure keys are provisioned and utilized in various network environments.

For the evaluation of quantum-safe technologies in critical infrastructure, we performed two representative reference deployments:

1. QKD-secured Data Centre Interconnect (10)
   ID Quantique Cerberis XGR QKD system (COW Protocol) is deployed over NetLink Trust commercial fibre between STTelemedia-Global Data Centres sites, producing >2 Gbit of QKD-derived key material and enabling AES-256 QKD-VPN secure data transfers.
2. QKD-secured Edge Cloud Computing (8)
   In this implementation, field-deployed fiber optics are utilized to connect AWS Edge Compute devices with on-premises Fortinet devices. QKD generates symmetric keys, which are used to establish a quantum-secured VPN that protects edge cloud workloads.

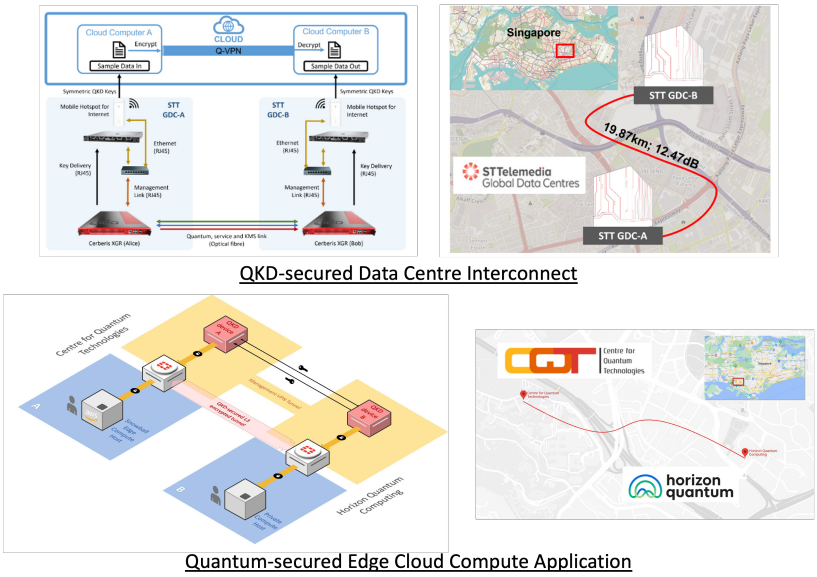


Fig. 6. Quantum-secured infrastructure use cases.

For OSI layer use cases, below are some examples organized within the NQSN Testbed (Fig.8):

1. QKD-secured video communications
   We deployed Fortinet FortiGate for OSI Layer-3 routing/firewalling and Hitachi Energy MPLS-TP encryption cards for Layer-2.5 transport encryption (11). Video traffic is protected using symmetric keys from a QKD system over 32.48 km (11.52 dB of loss) of fibre links between Centre for Quantum Technologies (CQT) and Fraunhofer Singapore Research Centre@NTU (FSR@NTU).
2. Three-node QKD-encrypted communications
   We deployed an end-to-end Layer-2 encryption across three sites (CQT, FSR@NTU, GovTech) using ST Engineering OSI Layer-2 encryptors. Symmetric keys are distributed and relayed by the ST Engineering Key Management System (KMS) and are sourced from two QKD different systems (BB84 and COW Protocol).
3. QKD-secured streaming over 5G backbone
   We setup streaming traffic between CQT and Singapore University of Technology and Design (SUTD), which is protected end-to-end using QKD as the primary source of symmetric keys with a PQC fallback, delivered to Layer-2 encryptors over a 5G backbone infrastructure. The deployment uses BB84 QKD System over 46.6km link (>19dB loss)and Thales CN6010 encryptor performing AES-256 encryption with keys constantly refreshed.

4. Quantum Encryption at Optical Transport Layer
   We are organsing physical-layer encryption using Ciena L1 hardware encryptors, consuming QKD keys via the NQSN KMS.

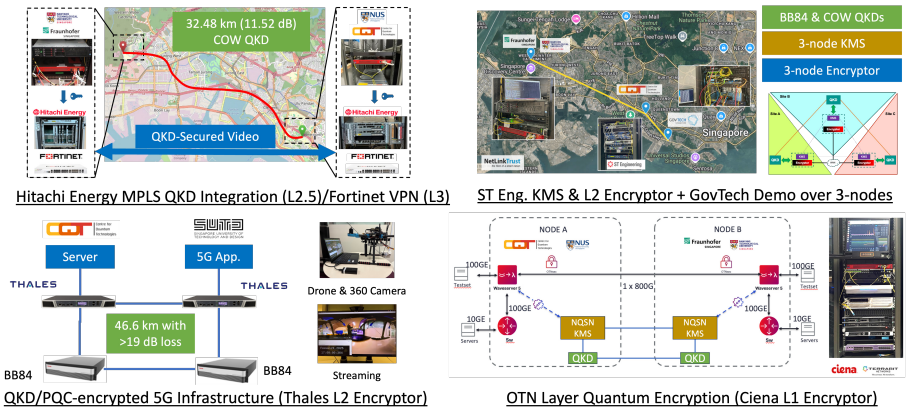


Fig. 7. Quantum-secured OSI layers use cases.

These demonstrations reveal how these quantum-safe technologies can be integrated into modern communication infrastructures.

Additionally, we explored hybrid QKD-PQC implementations, which combine the strengths of both quantum and classical cryptographic techniques to enhance security in various applications. By integrating QKD's ability to securely distribute encryption keys with PQC's resilience against quantum attacks, these implementations provide tailored solutions for a diverse range of security use cases. Several hybrid QKD-PQC use cases evaluated on the NQSN includes

1. QKD-PQC Hybrid VPNs
   We deployed a three-site IP-Sec virtual private networks (VPNs) link using Fortinet appliances integrated with a QKD key source and Check Point appliances using PQC keys. Test plan covers speed tests and file transfer validation across the three nodes.
2. Hybrid QKD-PQC Quantum-Safe application
   We developed an application with layered security model, combining quantum-secure primitives (QKD, OTP) plus Open-SSH (with PQC option from Open Quantum Safe (12)) and conventional symmetric ciphers (e.g. AES). This comprehensive approach ensures information-theoretic confidentiality for short and high-value messages.
3. QKD-PQC Defense in depth
   We deployed Fortinet Fortigate L3 VPN with NQSN-KMS and Quintessence-Labs CV-QKD (qOptica™ 100). Additionally, we implement the Viavi TeraVM PQC Key Exchange Mechanism (KEM) to secure HTTPS traffic on top of this infrastructure.

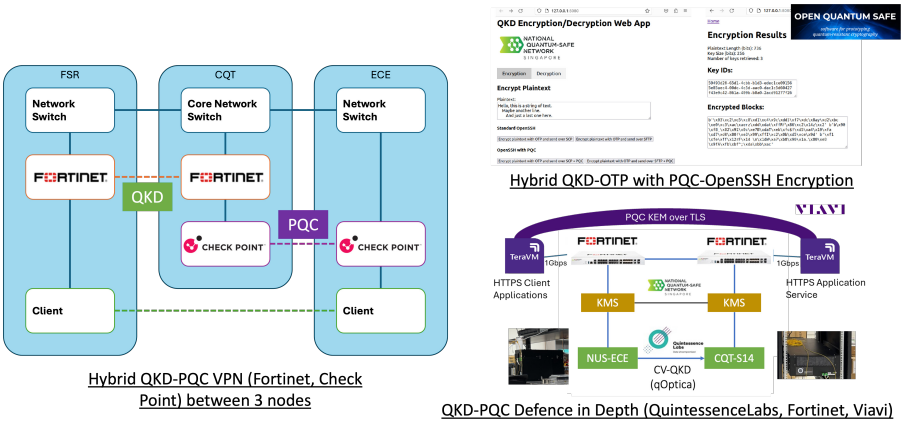


Fig. 8. Hybrid QKD-PQC use cases.

While NQSN hosts QKD technologies sourced from a diverse range of vendors, which form the secure backbone of the network, the inclusion of PQC further complements QKD by offering cryptographic algorithms designed to resist quantum attacks. This dual-layered approach enhances the network's security, serving as a last-mile connectivity solution, a resiliency mechanism, or an additional security layer, depending on the specific application requirements.

5 Conclusion

In this work, our objective is to discuss the challenges encountered during deployment of a practical quantum-safe network, such as interoperability issues, scalability constraints, and integration complexities. Lessons learned from these experiences will provide valuable insights into optimizing hybrid architectures for future applications. Ultimately, these insights underscore the potential of hybrid QKD-PQC architectures in ensuring resilient and scalable quantum-safe networks, paving the way for secure communication in the quantum era.

Acknowledgments. We acknowledge NQSN partners (www.nqsn.sg) for their support. We acknowledge funding support from A*STAR via National Quantum Office (NQO) under its Quantum Engineering Programme 3.0 (National Quantum-Safe Network Testbed, W24Q3D0001), National Research Foundation, Singapore and A*STAR under its Quantum Engineering Programme (National Quantum-Safe Network, NRF2021-QEP2-04-P01), Infocomm Media Development Authority (National Quantum-Safe Network Plus) and start-up grant for Nanyang Assistant Professorship awarded to Ng NHY of Nanyang Technological University, Singapore.

Disclosure of Interests. The authors declare no competing interests.

Bibliography

[1] Bächli, R., Shoaie, M.A., Floeter, R., Palangadan, V., Foery, A.: Quantum key distribution for mpls-tp traffic encryption. In: CSE N°36 — CIGRE. CIGRE — CSE (Feb 2025)

[2] Dynes, J., Wonfor, A., Tam, W.S., Sharpe, A., Takahashi, R., Lucamarini, M., Plews, A., Yuan, Z., Dixon, A., Cho, J., et al.: Cambridge quantum network. npj Quantum Information **5**(1), 101 (2019)

[3] European Telecommunications Standards Institute: ETSI GS QKD 014, Quantum key distribution; Protocol and data format of key delivery API to Applications (2019)

[4] Infocomm Media Development Authority: IMDA Reference Specification Quantum Key Distribution Network (Singapore) (2023)

[5] International Telecommunication Union: ITU-T Recommendation Y.3800, Overview on networks supporting quantum key distribution, plus Corrigendum 1 (2020)

[6] International Telecommunication Union: ITU-T Recommendation Y.3803, Quantum key distribution networks – Key management (2020)

[7] Martin, V., Brito, J.P., Ortíz, L., Mendez, R., Buruaga, J., Vicente, R., Sebastian-Lombrana, A., Rincon, D., Perez, F., Sanchez, C., et al.: Madqci: a heterogeneous and scalable sdn-qkd network deployed in production facilities. npj Quantum Information **10**(1), 80 (2024)

[8] Moreno, J., Proctor, C.: Implementing a quantum-secured network in a metropolitan area (Mar 2023), `https://aws.amazon.com/blogs/quantum-computing/implementing-a-quantum-secured-network-in-a-metropolitan-area/`, aWS Quantum Technologies Blog

[9] Peev, M., Pacher, C., Alléaume, R., Barreiro, C., Bouda, J., Boxleitner, W., Debuisschert, T., Diamanti, E., Dianati, M., Dynes, J.F., et al.: The secoqc quantum key distribution network in vienna. New journal of physics **11**(7), 075001 (2009)

[10] Qiu, K., Haw, J.Y., Qin, H., Ng, N.H.Y., Kasper, M., Ling, A.: Quantum-secured data centre interconnect in a field environment. Journal of Surveillance, Security and Safety **5**(3) (2024)

[11] Sasaki, M., Fujiwara, M., Ishizuka, H., Klaus, W., Wakui, K., Takeoka, M., Miki, S., Yamashita, T., Wang, Z., Tanaka, A., et al.: Field test of quantum key distribution in the tokyo qkd network. Optics express **19**(11), 10387–10409 (2011)

[12] Stebila, D., Mosca, M.: Post-quantum key exchange for the internet and the open quantum safe project. In: Avanzi, R., Heys, H. (eds.) Selected Areas in Cryptography (SAC) 2016, Lecture Notes in Computer Science, vol. 10532, pp. 1–24. Springer (October 2017), `https://openquantumsafe.org`

[13] Wang, S., Chen, W., Yin, Z.Q., Li, H.W., He, D.Y., Li, Y.H., Zhou, Z., Song, X.T., Li, F.Y., Wang, D., et al.: Field and long-term demonstration

of a wide area quantum key distribution network. Optics express **22**(18), 21739–21756 (2014)

QuaNTUM: A Modular Quantum Communication Testbed for Scalable Fiber and Satellite Integration

Julien Chénedé[1,2 [0009−0000−9721−8249]], Tjorben Matthes[1,2 [0000−0003−2798−9843]], Josefine Krause[3] Asli Cakan[1,2 [0000−0002−1920−2804]] and Tobias Vogl[1,2 [0000−0002−0993−0648]]

[1] Technical University of Munich , Arcisstraße 21, 80333 Munich, Germany https://www.tum.de/

[2] Munich Center for Quantum Science and Technology, Schellingstrße. 4, D-80799 Munich, Germany https://www.mcqst.de/

[3] Friedrich-Schiller-Universität Jena, Albert-Einstein-Straße 5, 07745 Jena, Germany

Abstract. Secure communication is a cornerstone of modern society, underpinning everything from financial transactions to critical infrastructure. As classical encryption faces growing threats from advancing computational power, quantum communication offers a fundamentally secure alternative, leveraging the laws of physics to protect data.
We introduce QuaNTUM (Quantum Network at the Technical University of Munich), a modular and extensible quantum communication testbed designed to enable scalable, flexible, and secure quantum communication across fiber-based campus networks and satellite-ground links. QuaNTUM integrates early deployments of solid-state quantum emitters in small satellites, bridging terrestrial and free-space channels. As an open-access platform, it supports experimental quantum communication protocols, quantum device benchmarking, and hybrid network integration.
The terrestrial network connects universities and research institutes across the high-tech campus in Garching near Munich, using single-mode fibers in a star-shaped topology. Each node features polarization-maintaining components, multiplexers, and time-synchronized analysis modules, with a central switching hub for dynamic routing. Active polarization control and real-time feedback ensure low error rates and stable qubit transmission, enabling high-fidelity quantum key distribution (QKD) and entanglement distribution.
A central focus of QuaNTUM is its use of deterministic single-photon sources based on optically active defects, such as in hexagonal boron nitride (hBN) or excited erbium atoms. These sources are currently being deployed in orbit, marking one of the first demonstrations of a solid-state quantum emitter in space. By uniting fiber and free-space links with scalable hardware and open protocols, QuaNTUM offers a basis that could adapt to future hybrid quantum networks, supporting both current testbed research and the long-term vision of a global quantum internet.

Keywords: quantum communication · hBN single-photon emitter · CubeSat · fiber-based testbed · quantum network architecture

1 Introduction

The arrival of quantum computer threatens today's encrypted networks by efficiently solving complex mathematical problems (like factoring large integers), on which nowadays most of our secured communication system are relying on [1]. One replacement candidate is quantum communication which guarantees a safe communication method against any eavesdropper or even the operator of the network. Several projects were done during the past years proving the feasibility of such links using long distance communication via satellites [2] and shorter ones via fibers [3]. Before quantum repeaters become technically feasible and enable a full "quantum internet", hybrid networks can be created using both fiber and satellite links, with satellites serving as trusted nodes. Ultimately, QuaNTUM aims to enable long-distance quantum networking, linking remote quantum processors and supporting reliable quantum information exchange.

Quantum key distribution (QKD) is a remarkable technology: it uses the fundamental laws of physics to secure communication. By encoding keys onto single photons, it makes eavesdropping impossible, thanks to the no-cloning theorem [1] and the fact that any attempt to measure the photons disturbs them. Yet, a limitation remains: when we rely only on fiber optics, QKD hits a wall after a few hundred kilometers, or even less for some implementations. The signal weakens exponentially, and unlike classical signals, amplification alone is insufficient.

Realizing a true "quantum internet" relies on quantum repeaters, which are still nowadays at an early stage [3]. In the meantime, there is a workaround: a hybrid approach that combines the best of both worlds. Satellites can bridge vast distances through free space, where signal loss is almost negligible above 10 kilometers, while fiber networks handle the last kilometers, connecting buildings, campuses, or even cities.

Despite these advancements, a critical gap remains: the absence of a large-scale, real-world open testbed that integrates these technologies into a cohesive framework. The QuaNTUM project addresses this integration challenge by establishing a star topology fiber network across the Garching campus near Munich. This infrastructure interconnects various institutes and laboratories, enabling shared access to quantum resources. Unlike trusted-node QKD networks that have already been demonstrated in Boston, Vienna, Tokyo, and China, QuaNTUM is designed as an open research testbed that directly connects laboratories for end-to-end quantum experiments without relying on intermediate trusted relays. Researchers can thus conduct experiments involving single-photon emission, for instance, from hBN [4] color centers or erbium-doped materials [5], optical quantum memories [6,7], and diverse QKD protocols [3], as well as investigate quantum entanglement distribution [8].

In a subsequent phase, the network will be extended to Munich's city center and to a ground station to support optical satellite links. Preparatory work is already underway with the low-Earth-orbit CubeSat QUICK^3 [9], which is performing foundational quantum optics experiments in microgravity. Upon completion of the commissioning phase, the QuaNTUM network will be made accessible to the Munich Center for Quantum Science and Technology (MCQST),

the Munich Quantum Valley (MQV), and external partners. This will facilitate collaborative research in areas such as secure quantum communication, quantum sensing, and the interfacing of quantum memories.

2 QuaNTUM Testbed Architecture

2.1 Fiber-Optic Backbone

The QuaNTUM fiber-optic backbone is implemented as a start topology single-mode network centered at the TUM-MI node (Departments of Mathematics and Computer Science), roughly located in the center of the campus. Some standard Telecom-Band SMF-28 fibers form the primary campus backbone and interconnect the principal sites described in **Fig. 1** with typical link lengths of 1 km to 2 km. At the central node a quantum reconfigurable add–drop multiplexer (q-ROADM) is specified to provide wavelength de/multiplexing, low-loss optical switching and a consolidated point for polarization control hardware and timing distribution. The network design preserves the option of direct fiber-to-fiber links to minimize insertion loss for critical quantum channels.

Cable composition and connectorization are chosen to balance compatibility with telecom infrastructure and the specific wavelength requirements of quantum systems. Each main cable contains multiple SMF-28 fibers (optimized for telecom O-/C-bands at 1330 nm and 1550 nm) together with dedicated short-wavelength fibers (780-HP and 1060-XP) to support key quantum technologies [10]:

- Solid-state qubits: Color centers in diamond (e.g. NV centers at 637 nm [11]) and silicon carbide (e.g. silicon vacancies at 737 nm [12]) require near-visible wavelengths, but fiber attenuation below 780 nm is prohibitive for kilometer-scale links—hence the 780 nm lower bound in our design.
- Quantum memories: Rare-earth-doped crystals (e.g. Pr^{3+}:Y_2SiO_5 at 606 nm or Tm^{3+}:Y_2SiO_5 at 793 nm [13]) and alkali vapors (e.g. Rb D2 transitions at 780 nm [10]) often operate in the 700–800 nm range, motivating the inclusion of 780-HP fibers.
- Telecom compatibility: Long-distance quantum communication leverages the near-infrared (NIR) free-space window and telecom O-/C-bands (1330/1550 nm).

E2000 APC simplex terminations are specified for quantum ports to minimize back-reflection and allow individual fiber access. Loss budgeting and noise mitigation are central design constraints. The end-to-end optical loss model includes fiber attenuation, splice loss, connector mated loss, and insertion loss of passive/active elements (Wavelength Division Multiplexing [14], switches, polarizers). To protect single-photon channels from classical noise, high-power classical signals are separated spectrally or routed on separate fibers; APC connectors are used to reduce back-scatter and Raman-induced background. For polarization-encoded protocols the backbone incorporates motorized fiber polarization controllers (FPCs) and software feedback to maintain state fidelity under environmental perturbations. Market-ready devices devices can provide polarization compensation such as silicon-based decoder [15], Pockels cell [16] and fiber squeezing devices [17].

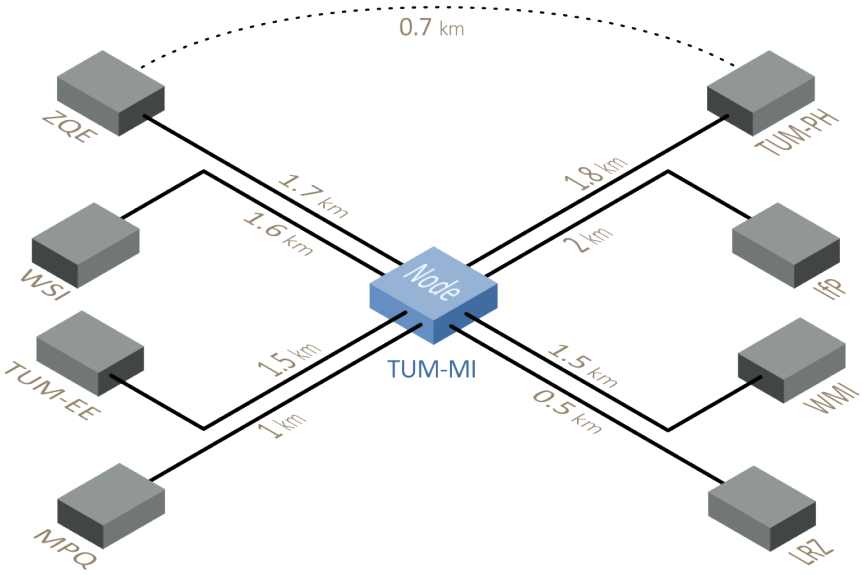


Fig. 1. Topology of the QuaNTUM testbed on the Garching campus near Munich: star-shaped fiber connections using SMF-28 ultra fibers, 780-HP and 1060-XP between the departement of physics (TUM-PH), Electrical Engineering (TUM-EE), the Max Planck Institute of Quantum Optics (MPQ), the Leibniz Supercomputing Centre (LRZ), the Walter Schottky Institut (WSI), the Walther-Meißner-Institut (WMI) and with the central node in School of Computation, Information and Technology (TUM-MI). An additional special cable with only SMF-28 ultra fibers is connecting directly ZQE and TUM-PH (dotted-line). – For geographic reference, see the map: https://v.bayern.de/N74hf – ©Bayerische Vermessungsverwaltung 2025

2.2 Timing and Control

Synchronization and deterministic control are fundamental to all QuaNTUM experiments: they enable reliable coincidence detection, accurate protocol timing, and automated feedback for polarization and routing. The timing architecture therefore combines a dedicated classical reference channel, local high-resolution time-tagging at each measurement node, and a coordinated control plane that drives active hardware and manages experiment scheduling.

A low-jitter classical reference channel is provisioned alongside the quantum wavelengths to distribute a 10 MHz clock and a 1 pulse per second (PPS) marker. This provides a stable, common time base and avoids many limitations of satellite-based references for short-range campus networks. Each polarization analysis module (PAM) is paired with a synchronized time-tagger capable of sub-100-ps resolution for arrival-time stamping; these local time stamps are es-

sential for coincidence analysis, quantum bit error ratio (QBER) estimation and post-processing.

For deterministic, low-latency distribution and scalable network timing, the testbed will evaluate precision time protocols such as White Rabbit [18] because they provide sub-nanosecond synchronization and known latency bounds over fiber links.

The control plane is implemented as a hybrid of FPGA/embedded control at the hardware layer and a centralized scheduler at the management layer. FPGA units or microcontrollers co-located with the q-ROADM and at user nodes perform deterministic control of optical switches, electro-optic modulators and motorized FPCs with microsecond actuation times. Timing-critical loops, for example, live polarization compensation, run locally at each node (fast perturbation, local evaluation, accept/reject), while higher-level policies (when to perform a full reference recalibration, scheduling of satellite passes) are managed centrally.

Practical implementation requires explicit calibration and monitoring: per-link fixed delays must be measured and recorded (optical time-domain reflectometer and two-way time transfer), time-tagger offsets and cable dispersive effects characterized, and ongoing telemetry (detector counts, QBER, temperature) integrated into the control dashboard. These elements together ensure that QuaNTUM achieves the temporal precision and operational automation necessary for reliable QKD and entanglement distribution experiments across a campus-scale hybrid fiber network.

2.3 Entanglement and QKD Layers

From a security standpoint, protocols divide into device-independent (theoretically secure against all attacks, with no hardware assumptions) and device-dependent (relying on trusted measurement tools but compatible with off-the-shelf components). In practical terms, implementations split between discrete-variable (e.g. polarized photons) and continuous-variable (e.g. coherent states) approaches [19]. QKD protocols fall into three primary types:

Table 1. QKD protocol classification by structure

Type	**Key Mechanism**
Prepare-and-Measure (PM)	Alice prepares quantum states; Bob measures them to generate a shared key.
Entanglement-Based	Alice and Bob exploit entangled particle correlations for key distribution.
Measurement-Device-Independent	Security remains intact even if measurement devices are untrusted.

QuaNTUM is implementing two complementary quantum channel layers on the same physical backbone: a wavelength-multiplexed entanglement distribution layer for multi-user and a PM-QKD layer that supports standard schemes.

For entanglement distribution, a broadband spontaneous-parametric-down-conversion (SPDC) source in a Sagnac configuration [20] is specified as the primary photon-pair generator. The SPDC output is spectrally partitioned by a demultiplexer; the q-ROADM in the central node assigns channel pairs to endpoint links, producing a reconfigurable, wavelength-multiplexed entanglement fabric. At each user endpoint, polarization analysis modules (PAMs) implement random-basis selection and perform polarization projection and detection. Time-stamping of events with sub-100-ps resolution enables coincidence identification and correlation analysis for entanglement verification. The architecture is designed to support multi-hop protocols [18] by routing idler photons to intermediate Bell-state measurement stations when required, enabling tests of networked entanglement distribution.

The PM layer will support different standard QKD implementations (for example B92, BB84, SARG04 [19]). Deterministic single-photon emitters based on localized defects in hBN (at 575 nm) are primarily intended for short-reach or on-chip experiments, as their visible-band emission is incompatible with long-distance transmission over telecom-band fibers. For integration with the testbed's backbone, a frequency conversion strategy is employed: Visible photons are converted to telecom bands via nonlinear optics [21] before entering the network.

For long-distance operation, the testbed provisions alternative sources (e.g. telecom-compatible Spontaneous Parametric Down-Conversion pairs [22] or InGaAs-based single-photon emitters[23]). Real-time operational control closes the loop for secure operation: measured QBER and coincidence statistics drive local feedback (motorized fiber polarization controllers and software routines) for live polarization compensation and calibration.

3 Single-Photon Emitters in hBN

3.1 Deterministic Fabrication

In QuaNTUM we implement deterministic fabrication by localized electron-beam irradiation of multilayer hBN using a standard scanning electron microscope (SEM) [24]. Exposures are targeted to sub-micron coordinates to create emitter arrays with high lateral precision; process parameters (beam energy, current, dose and spot dwell time) are optimized to maximize yield and spectral reproducibility. The produced emitters show a reproducible zero-phonon line (ZPL) near 575 nm and are consistent with carbon-related defect complexes identified by density-functional calculations. This approach is compatible with wafer-scale patterning workflows and with subsequent on-chip or micro-optical integration because the emitter position is known *a priori*. Scalability hinges on automation of exposure patterning, batch handling of hBN flakes or films, and yield optimization to reduce per-device screening. For production of fiber-coupled modules or on-chip arrays, deterministic sites are pre-aligned to lithographically defined photonic structures (waveguides, cavities, grating couplers) so that subsequent assembly and coupling are mainly mechanical rather than optical alignment tasks.

3.2 Optical Performance

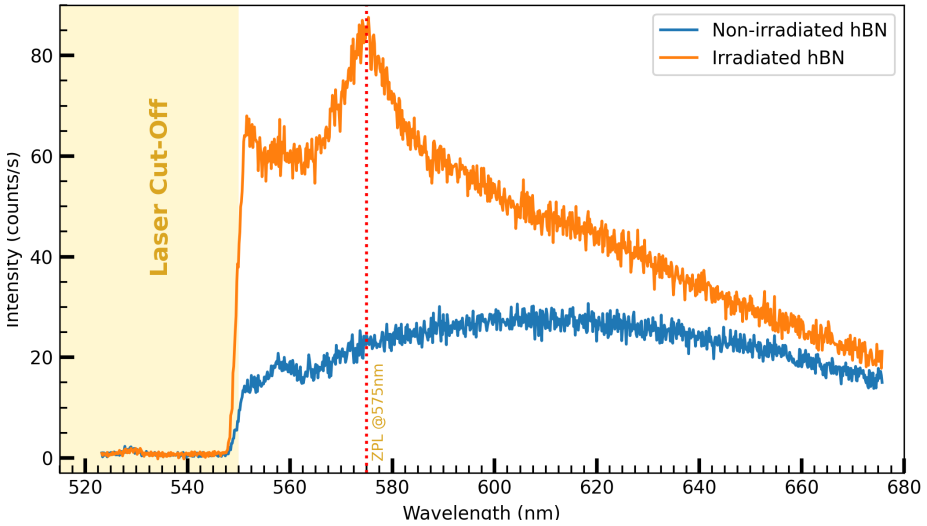


Fig. 2. Room-temperature photoluminescence spectrum of a single hBN emitter fabricated by localized electron irradiation, featuring a ZPL near 575 nm (orange line). Comparison with a non-irradiated hBN (blue line). Laser excitation at 530 nm and long-pass filter at 550 nm.

hBN hosts a diverse family of quantum emitters with zero-phonon lines (ZPLs) spanning UV to NIR wavelengths [25], enabling compatibility with a wide range of quantum systems. For this work, we focus on defects produced by localized irradiation that, under 530 nm femtosecond laser excitation, exhibit a ZPL near 575 nm with a typical linewidth (FWHM) on the order of tens of nanometers at room temperature (**Fig. 2**). The second-order correlation function $g^{(2)}(0) < 0.1$ confirms single-photon emission [24], while the reported excited-state lifetime of 3.83(1) ns supports high repetition-rate operation [24]. Long-term photostability has been observed in many sites even 3 years post-irradiation, though device-level qualification—including performance under variable temperature and optical pump conditions—remains necessary for network deployment or space applications. Visible-band ZPLs and room-temperature linewidths are advantageous for bright, high-rate operation but pose compatibility challenges with standard SMF-28 telecom infrastructure and with narrowband quantum memories. To address these issues, two engineering strategies are relevant: (1) spectral narrowing and tuning via photonic engineering (microcavities, Purcell enhancement, strain or electric-field tuning) to reduce the effective linewidth and stabilize the ZPL; and (2) quantum frequency conversion (QFC) to translate emitted photons to telecom wavelengths for low-loss fiber transmission or to

memory-matched atomic transitions. Microcavity coupling can also increase the collection efficiency and reduce the required source brightness for a target key rate or coincidence rate, while QFC preserves quantum statistics if conversion noise is sufficiently suppressed.

3.3 Integrated photonics

Integration of hBN emitters with photonic platforms (e.g. fiber integrated platforms [26]) addresses three objectives: improve collection/coupling efficiency into single-mode fibers, control emission dynamics (Purcell effect) and enable on-chip routing and switching for compact, stable modules [4]. In practice, deterministic emitter sites are aligned to on-chip waveguides, microcavities or grating couplers using lithographic markers and high-precision assembly. Finite-difference time-domain method (FDTD) simulations guide the resonator and waveguide design to maximize coupling efficiency and to position cavity resonances relative to the emitter ZPL [27]. Thanks to photonic integration, stricter constraints are typically designed for satellite deployment. The QUICK^3 payload strategy uses a miniaturized version of a different emitter platform based on $MoSe_2$ [9]: the chip is mounted within a compact, thermally managed optical bench that includes the pump laser, stabilization optics, fiber-coupling stage and pointing interface. Packaging choices emphasize mechanical resilience (vibration and shock tolerance), thermal control over the expected orbital temperature range, radiation tolerance of active components, and minimal moving parts.

4 Conclusion

In conclusion, the QuaNTUM project addresses a critical gap in the field of quantum communication by providing a large-scale, real-world open testbed that integrates fiber and satellite links into a cohesive framework. This platform, akin to a campus-wide hybrid network, offers researchers and engineers a controlled yet dynamic environment to experimentally validate, refine, and benchmark system performance under realistic conditions. By leveraging advanced technologies such as solid-state quantum emitters and polarization control, QuaNTUM aims to overcome the limitations of current quantum communication systems. The successful integration of these technologies could pave the way for a global quantum internet, enabling secure communication and advanced quantum applications. Future work will focus on expanding the network, improving the performance of quantum emitters, and exploring new applications in quantum communication and sensing.

Acknowledgments. This research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy- EXC-2111-390814868 (MCQST). The authors acknowledge support from the Federal Ministry of Research, Technology and Space (BMFTR) under grant number 13N16292 (ATOMIQS).

Disclosure of Interests. The authors declare that they have no competing interests relevant to the content of this article.

References

1. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, Mar 2002.
2. C.-Y. Lu, Y. Cao, C.-Z. Peng, and J.-W. Pan, "Micius quantum experiments in space," *Rev. Mod. Phys.*, vol. 94, p. 035001, Jul 2022.
3. G. Granados, W. Velasquez, R. Cajo, and M. Antonieta-Alvarez, "Quantum key distribution in multiple fiber networks and its application in urban communications: A comprehensive review," *IEEE Access*, vol. 13, pp. 100446–100461, 2025.
4. T. Vogl, R. Lecamwasam, B. C. Buchler, Y. Lu, and P. K. Lam, "Compact cavity-enhanced single-photon generation with hexagonal boron nitride," *ACS Photonics*, vol. 6, no. 8, pp. 1955–1962, 2019.
5. A. Gritsch, L. Weiss, J. Früh, S. Rinner, and A. Reiserer, "Narrow optical transitions in erbium-implanted silicon waveguides," *Phys. Rev. X*, vol. 12, p. 041009, Oct 2022.
6. A. I. Lvovsky, B. C. Sanders, and W. Tittel, "Optical quantum memory," *Nature Photonics*, vol. 3, pp. 706–714, Dec 2009.
7. A. Çakan, C. Cholsuk, A. Gale, M. Kianinia, S. Paçal, S. Ateş, I. Aharonovich, M. Toth, and T. Vogl, "Quantum optics applications of hexagonal boron nitride defects," *Advanced Optical Materials*, vol. 13, no. 7, p. 2402508, 2025.
8. S. P. Neumann, A. Buchner, L. Bulla, M. Bohmann, and R. Ursin, "Continuous entanglement distribution over a transnational 248 km fiber link," *Nature Communications*, vol. 13, p. 6134, Oct 2022.
9. N. Ahmadi, S. Schwertfeger, P. Werner, L. Wiese, J. Lester, E. Da Ros, J. Krause, S. Ritter, M. Abasifard, C. Cholsuk, R. G. Krämer, S. Atzeni, M. Gündoğan, S. Sachidananda, D. Pardo, S. Nolte, A. Lohrmann, A. Ling, J. Bartholomäus, G. Corrielli, M. Krutzik, and T. Vogl, "Quick^3 - design of a satellite-based quantum light source for quantum communication and extended physical theory tests in space," *Advanced Quantum Technologies*, vol. 7, no. 4, p. 2300343, 2024.
10. C. Cholsuk, S. Suwanna, and T. Vogl, "Tailoring the emission wavelength of color centers in hexagonal boron nitride for quantum applications," *Nanomaterials*, vol. 12, no. 14, 2022.
11. C. Bradac, W. Gao, J. Forneris, M. E. Trusheim, and I. Aharonovich, "Quantum nanophotonics with group iv defects in diamond," *Nature Communications*, vol. 10, p. 5625, Dec 2019.
12. S. Castelletto, "Silicon carbide single-photon sources: challenges and prospects," *Materials for Quantum Technology*, vol. 1, p. 023001, mar 2021.
13. C. Simon, M. Afzelius, J. Appel, A. Boyer de la Giroday, S. J. Dewhurst, N. Gisin, C. Y. Hu, F. Jelezko, S. Kröll, J. H. Müller, J. Nunn, E. S. Polzik, J. G. Rarity, H. De Riedmatten, W. Rosenfeld, A. J. Shields, N. Sköld, R. M. Stevenson, R. Thew, I. A. Walmsley, M. C. Weber, H. Weinfurter, J. Wrachtrup, and R. J. Young, "Quantum memories," *The European Physical Journal D*, vol. 58, pp. 1–22, May 2010.
14. M. Bathaee, M. Rezai, and J. A. Salehi, "Quantum wavelength-division-multiplexing and multiple-access communication systems and networks: Global and unified approach," *Phys. Rev. A*, vol. 107, p. 012613, Jan 2023.

15. Y. Du, X. Zhu, X. Hua, Z. Zhao, X. Hu, Y. Qian, X. Xiao, and K. Wei, "Silicon-based decoder for polarization-encoding quantum key distribution," *Chip*, vol. 2, no. 1, p. 100039, 2023.
16. C. Caputo, M. Simoni, G. A. Cirillo, G. Turvani, and M. Zamboni, "A simulator of optical coherent-state evolution in quantum key distribution systems," *Optical and Quantum Electronics*, vol. 54, p. 689, Sep 2022.
17. F. Kaiser, B. Fedrici, A. Zavatta, V. D'Auria, and S. Tanzilli, "A fully guided-wave squeezing experiment for fiber quantum networks," *Optica*, vol. 3, pp. 362–365, Apr 2016.
18. M. Alshowkan, J. M. Lukens, H.-H. Lu, and N. A. Peters, "Resilient entanglement distribution in a multihop quantum network," *Journal of Lightwave Technology*, pp. 1–8, 2025.
19. M. Kumar and B. Mondal, "A brief review on quantum key distribution protocols," *Multimedia Tools and Applications*, vol. 84, pp. 33267–33306, Aug 2025.
20. T. Strömberg, P. Schiansky, R. W. Peterson, M. T. Quintino, and P. Walther, "Demonstration of a quantum switch in a sagnac configuration," *Phys. Rev. Lett.*, vol. 131, p. 060803, Aug 2023.
21. P. Fisher, R. Cernansky, B. Haylock, and M. Lobino, "Single photon frequency conversion for frequency multiplexed quantum networks in the telecom band," *Phys. Rev. Lett.*, vol. 127, p. 023602, Jul 2021.
22. R. Ikuta, Y. Kusaka, T. Kitano, H. Kato, T. Yamamoto, M. Koashi, and N. Imoto, "Wide-band quantum interface for visible-to-telecommunication wavelength conversion," *Nature Communications*, vol. 2, p. 537, Nov 2011.
23. J. Zhang, M. A. Itzler, H. Zbinden, and J.-W. Pan, "Advances in ingaas/inp single-photon detector systems for quantum communication," *Light: Science & Applications*, vol. 4, pp. e286–e286, May 2015.
24. A. Kumar, C. Cholsuk, A. Zand, M. N. Mishuk, T. Matthes, F. Eilenberger, S. Suwanna, and T. Vogl, "Localized creation of yellow single photon emitting carbon complexes in hexagonal boron nitride," *APL Materials*, vol. 11, p. 071108, Jul 2023.
25. C. Cholsuk, A. Zand, A. Çakan, and T. Vogl, "The hbn defects database: A theoretical compilation of color centers in hexagonal boron nitride," *The Journal of Physical Chemistry C*, vol. 128, no. 30, pp. 12716–12725, 2024.
26. T. Vogl, Y. Lu, and P. Koy Lam, "Room temperature single photon source using fiber-integrated hexagonal boron nitride," *Journal of Physics D: Applied Physics*, vol. 50, p. 295101, Jun 2017.
27. P. Ji, C. Qian, J. J. Finley, and S. Yang, "Thickness insensitive nanocavities for 2d heterostructures using photonic molecules," *Nanophotonics*, vol. 12, no. 17, pp. 3501–3510, 2023.

Network-wide Quantum Key Distribution with Onion Routing Relay

Pedro Otero-García, David Pérez-Castro, Manuel Fernández-Veiga, and Ana Fernández-Vilas

atlanTTic - research center, Universidade de Vigo
{pedro.otero,dperezcastro,mveiga,avilas}@det.uvigo.es

Abstract. The advancement of quantum computing threatens classical cryptographic methods, necessitating the development of secure quantum key distribution (QKD) solutions for QKD Networks (QKDN). In this paper, a novel key distribution protocol, Onion Routing Relay (ORR), that integrates onion routing (OR) with post-quantum cryptography (PQC) in a key-relay (KR) model is evaluated for QKDNs. This approach increases the security by enhancing confidentiality, integrity, authenticity (CIA principles), and anonymity in quantum-secure communications. By employing PQC-based encapsulation, ORR aims to avoid the security risks posed by intermediate malicious nodes and ensures end-to-end security. Our results show a competitive performance of the basic ORR model, against current KR and trusted-node (TN) approaches, demonstrating its feasibility and applicability in high-security environments maintaining a consistent Quality of Service (QoS). The results also show that while basic ORR incurs higher encryption overhead, it provides substantial security improvements without significantly impacting the overall key distribution time. Nevertheless, the introduction of an end-to-end authentication extension (ORR-Ext) has a significant impact on the Quality of Service (QoS), thereby limiting its suitability to applications with stringent security requirements.

1 Introduction

The development of quantum computing is rapidly evolving [1], with recent advancements in quantum computation that brings algorithms [2] like Shor's and Grover's closer to reality. These advances pose a potential risk to existing cryptographic systems [3], as a quantum adversary could eventually break the security of classical cryptography, especially asymmetric algorithms.

The scientific community has focused on two main approaches to address quantum threats: quantum cryptography (QC), which uses quantum systems for security, and post-quantum cryptography (PQC), which seeks algorithms resilient to quantum attacks. QC leverages quantum mechanics to ensure security, with Quantum Key Distribution (QKD) being a prominent solution [4], relying on quantum superposition or entanglement [5]. However, QKD faces implementation challenges [6]: The lack of commercial quantum repeaters or scalability constrains, both of which limit its range of applicability.

In contrast, PQC does not require specialized hardware and offers a more practical, cost-effective solution, although it is based on complex mathematical problems that could eventually be broken by sufficiently powerful quantum computers [7]. NIST is leading the effort to standardize PQC algorithms, including CRYSTALS-Kyber, CRYSTALS-Dilithium, HQC or FALCON.

Despite the hardware limitations of QC, QKD remains the only solution proven to be unconditionally secure [8], proving to be safe even against quantum attacks. However, scalability remains a challenge, with models like key-relay (KR) [9] and trusted-node (TN) [10] not ensuring end-to-end confidentiality across intermediate nodes. Thus, a hybrid QC and PQC approach is required for comprehensive security in quantum key distribution networks (QKDNs).

This paper evaluates the feasibility of our previous work [11] —a secure key distribution approach based on the key-relay (KR) model, enhanced with Onion Routing (OR) [12] and PQC techniques to ensure CIA principles and anonymity— by comparing it with the main existing alternatives: key-relay and trusted-node models. The remainder of the document is structured as follows: Section 2 reviews existing QKDN security models, Section 3 outlines the technical background, Section 4 introduces the ORR model, Section 5 presents the experimental setup, and Section 6 provides a comparative performance analysis. Finally, Section 7 summarizes the findings, highlighting ORR(-Ext)'s advantages and drawbacks in security and QoS ending with the future lines of work to follow.

2 Related Work

In QKDNs without commercial quantum repeaters, the primary approaches to overcome distance limitations are KR and TN models. Both require fully trusted nodes, introducing security vulnerabilities [13]. Rass *et al.* [14] emphasize that network security is only as strong as the least trusted intermediate node. De Santis *et al.* [15] propose satellite-based configurations to reduce dependence on multiple intermediaries, assuming only one or two trusted nodes. Calsi *et al.* [16] introduce a KR enhancement where nodes establish QKD links with both nearest and next-nearest neighbors, improving resilience but remaining susceptible to multiple-node attacks. In TN models, Vyas *et al.* [17] suggest trust-level segmentation to relax security assumptions at intermediate nodes, though this requires additional infrastructure for trust verification.

To address malicious node threats, post-quantum cryptography (PQC) has been integrated into classical protocols [18]. Rios *et al.* [19] show that combining Kyber and Dilithium enhances classical crypto performance at high-security levels, albeit with increased traffic overhead. Hybrid QKD–PQC solutions are increasingly favored for critical infrastructure in the NISQ [1] era [20], as they

[1] Noisy Intermediate-Scale Quantum (NISQ) refers to nowadays quantum devices with tens to hundreds of qubits. They are powerful enough to surpass some classical simulations but remain limited by noise and lack of full error correction, making them a transitional stage before fault-tolerant quantum computers.

offer both long and short-term protection. One approach uses QKD to encrypt PQC-based key exchanges [21].

Building on this trend, we adopt a hybrid method combining KR-based QKD with OR and PQC, enabling secure key distribution between distant nodes, as introduced in our earlier work [11].

3 Technological Background

QKD and PQC derive security from fundamentally different principles. QKD offers unconditional security based on quantum mechanics, while PQC achieves computational security against quantum attacks. While QKD ensures point-to-point confidentiality, PQC excels in speed and scalability. By combining both, one can build resilient systems for both short- and long-term security. Our method incorporates OR and PQC-encrypted keys into a KR-based QKDN, achieving layered protection and preserving end-to-end confidentiality.

3.1 Key-Relay Quantum Key Distribution Networks

In KR networks, a sender and receiver share a secret key S via a path of intermediate QKD-enabled nodes. The sender generates S using a QRNG, encrypts it with the shared quantum key, and transmits it through the network. Each node decrypts and re-encrypts S for the next hop. Although simple and effective, this model exposes S to every intermediate node, compromising confidentiality. KR is also vulnerable to spoofing on classical channels. Although intercepting a quantum key is unlikely without compromising a QKD node, such attacks can still disrupt network operations.

3.2 Onion Routing

OR is a layered encryption technique designed to anonymize communication paths. A message is encrypted in successive layers, starting with the public key of the final node and moving backward to the first one. Each node decrypts one layer, learns only the next hop, and forwards the message, preserving sender-receiver anonymity. OR enhances privacy but introduces higher latency and exposes the exit node to potential attacks. Some extensions [22] improve security by verifying message integrity end-to-end, defending against message tampering and overload-based denial-of-service (DoS) attacks on onion routers, however, increasing the computation and the size of the message.

4 Model Principles

The proposed model, Onion Routing Relay (ORR), guarantees unconditionally secure key distribution within QKDN. It builds on the KR model for QKDN and integrates an OR extension that ensures end-to-end authentication. This

approach upholds CIA principles and anonymity of the destination node. Figure 1 illustrates the basic operation of ORR.

Since classical OR protocols remain vulnerable to quantum attacks, the model replaces insecure classical cryptography —particularly asymmetric cryptography— with PQC. Rather than using public-key cryptography to encrypt the onion layers, ORR uses PQC Key Encapsulation Mechanism (PQC-KEM) algorithms to distribute keys among nodes. It then applies symmetric encryption algorithms, such as AES,[2] to create the encrypted layers (onions). The ORR model is detailed described in our previous work [11].

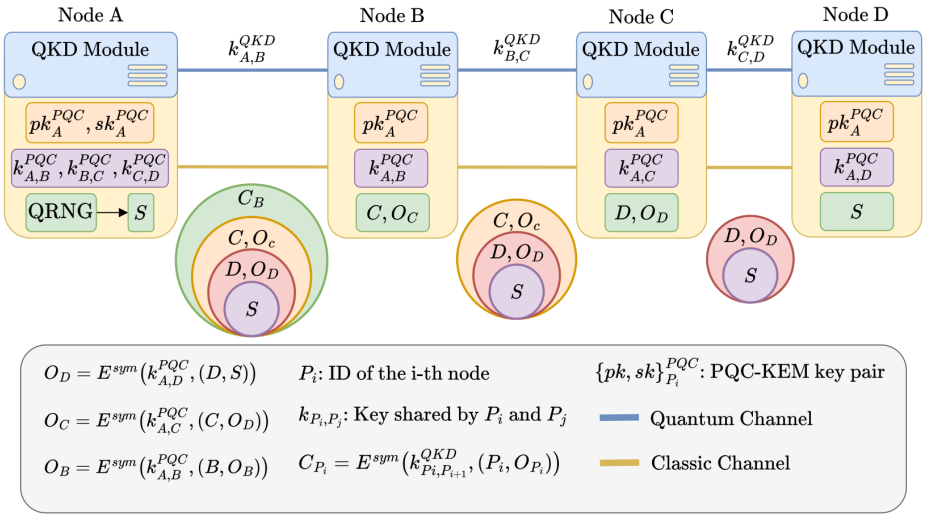


Fig. 1. Simplified example of the Onion Routing Relay key distribution model. In this scheme, it can be seen a circuit formed by three nodes (B, C and D) in addition to the initiator node (A). Node A possesses a KEM key-pair (pk^{PQC}_A, sk^{PQC}_A), which is used to compute shared keys with the nodes in the circuit ($k^{PQC}_{A,B}, k^{PQC}_{A,C}, k^{PQC}_{A,D}$). Using these keys, along with the QKD key shared with the first intermediate node B ($k^{QKD}_{A,B}$) A is able to create the onion ciphertext that then sends to B, thereby the ORR key distribution.

5 Test scenario & Implementation

To compare the different models, a series of scripts were developed in C to simulate their behavior. A GitHub repository[3] is available to readers, providing

[2] Due to Grover's algorithm halving the security of symmetric cryptography, AES-256 should be used in a post-quantum world to maintain the same security level as AES-128 today.

[3] https://github.com/pedrotega/ORRvsTNvsKR

Table 1. Cryptographic Algorithms and Their Implementations

Category	Implementation	Category	Implementation
Symmetric Encryption	AES-256-CBC (OpenSSL)	PQC-KEM	Kyber-768 (LibOQS)
PRNG	RAND_bytes() (OpenSSL)	XOR	Directly in C

the source code and execution instructions. It is important to note that due to the performance differences between the ORR model and its extended version with end-to-end authentication (ORR-Ext), it has been decided to evaluate each approach separately.

For the experiment, only a pair of QKD nodes was available –specifically the Cerberis QKD XGR [23] developed by ID Quantique– which provided the quantum functionalities for the simulations. Each simulated node in the scripts requests QKD keys from one of the real nodes, which works as a key-managment-system, and retrieves the key by its identifier from the other real node. All the request follows a RESTful API format specified in ETSI-014 [24].

In the code, each node is executed in a different thread that allows the execution of code concurrently or sequentially, as appropriate, at any given time. Since a QRNG was not available for this project, a PRNG (Pseudo-Random Number Generator) was used as a replacement. Table 1 lists the algorithms and functions employed in the implementation, along with the respective software suppliers.

The parameters measured in the tests include the encryption time and the key distribution time, which is calculated from the moment the secret is generated at the initiator node until it reaches the destination node. For the encryption time, the measurement varies depending on the model: In ORR the longest encryption time in this is spent encrypting the initial onion at the initiator node. In the version with authentication end-to-end (ORR-Ext) the time to create the authentication extension is included. For the TN model, the encryption time is measured during the calculation of the ciphertext to be sent to the destination node at the trusted node. Finally, in the KR model it is only necessary to measure the time of XOR operation when encrypting the secret in the initiator node.

For the implementation of ORR-Ext, the HMAC-256 algorithm was employed as signature scheme. This choice is motivated by the fact that it is still considered secure even in the presence of Grover's algorithm, offering post-quantum resistance without the overhead associated with PQC alternatives. In contrast to HMAC, PQC signature schemes are asymmetric –which necessitates the inclusion of the public key in each message– significantly increasing the overall message size. Table 2 presents a comparison of the message lengths in both ORR and ORR-Ext models under different signature algorithms, highlighting the trade-offs in communication efficiency introduced by PQC signature methods.

Table 2. Ciphertext length (Bytes) for a circuit of 5 nodes

Variant	Onion	Public Key	Signature	Ciphertext
ORR	416	0	0	432
ORR-Ext-HMAC-256	416	0	32	1472
ORR-Ext-Falcon-1024	416	1793	1280	16612
ORR-Ext-Dilithium3	416	1952	3293	27537

6 Results and Discussion

The results presented in this section have been obtained by averaging 100 iterations with each simulated model for circuits[4] with 3, 5, 7, 9, and 11 nodes. Figure 2 illustrates the average time required for the encryption procedure in each model. The KR model exhibits a highly consistent and minimal encryption time, ranging from 1.5 to 1.93 μs, remaining largely unaffected by the number of nodes due to its simple XOR-based encryption using QKD keys. The TN model shows a moderate increase, starting at 4.76 μs for a 3-node circuit and reaching 29.7 μs at 11 nodes, as encryption requires handling more ciphertexts with each additional node. The ORR model, which implements layered encryption, maintains moderate growth in encryption time, starting at 33.4 μs and increasing up to 47.25 μs with 11 nodes. Meanwhile, the ORR-Ext model introduces a significant overhead due to the additional operations required to ensure end-to-end authentication. It starts at 238.78 μs for 3 nodes and escalates sharply to 5327.31 μs for 11 nodes.

Regarding the key distribution time, represented in Figure 3, the growth trend across models reflects the increasing communication and coordination overhead as the number of nodes rises. The KR model again stands out as the fastest, beginning at 167.7 μs for 3 nodes and reaching 272.31 μs for the largest circuit. The TN model, which starts slightly above KR at 224.24 μs, experiences more dramatic growth, requiring 862.55 μs at 11 nodes. The ORR model, while initially slower than KR and TN (227.62 μs), maintains a steadier increase, finishing at 389.02 μs. In contrast, the ORR-Ext model incurs the highest key distribution times due to the authentication processes involved and the major length of their ciphertexts, with the average time rising from 465.8 μs for 3 nodes to 6571.89 μs for the 11-node configuration.

From the results obtained in the previous section, it can be concluded that the inclusion of layered encryption and authentication mechanisms results in significantly different performance profiles among the four models. As expected, the ORR-Ext model introduces the most substantial encryption overhead, requiring several orders of magnitude more time compared to the TN and KR models. Specifically, the encryption time in ORR-Ext can be more than 1800 times higher than in KR and over 150 times higher than in TN for the 11-node circuit. This is due to the added authentication extension operations needed to guarantee end-to-end security, which accumulate across layers.

[4] A circuit is formed by the intermediates and destination nodes.

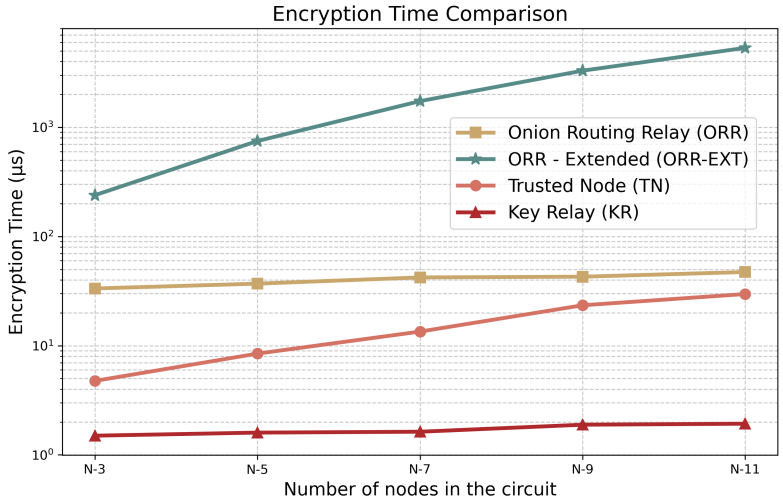


Fig. 2. Encryption time time comparison of the different models

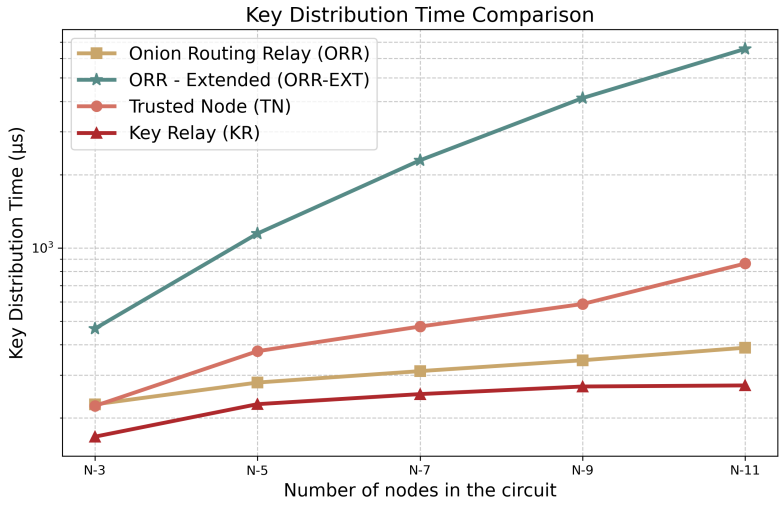


Fig. 3. Key distribution time comparison of the different models

The TN model shows a moderate but clear growth in encryption time as the number of nodes increases, stemming from the need to aggregate and process ciphertexts at the trusted node. This processing overhead is absent in the KR model, which maintains consistently minimal encryption times, as its XOR operation does not depend on the circuit size. The basic ORR model without the authentication extension, while implementing layered encryption, shows a more controlled increase, demonstrating its balance between added security and computational efficiency compared to ORR-Ext.

Nevertheless, the most impactful distinction across models lies in the key distribution time, which directly affects the system's QoS. The ORR-Ext model, in particular, suffers from very high distribution times due to both the authentication-related message exchanges and the context switching between threads simulating node-to-node communication. These results suggest that in practical deployments, where network delays and message propagation times are even more significant, the additional burden introduced by ORR-Ext could represent a limiting factor in terms of responsiveness. TN also exhibits a steep rise in key distribution time, particularly for larger circuits, primarily due to the additional communication step required to forward the key to the final destination.

In contrast, the KR and basic ORR models display a more favorable trend, especially at higher node counts. Notably, while basic ORR's encryption time is higher than that of KR, its key distribution time remains relatively competitive. For example, in the 11-node case, the difference between ORR and KR in key distribution is approximately 117 μs, which is consistent with the expected communication overhead introduced by layered routing, yet still acceptable within practical bounds.

Therefore, it can be concluded that although ORR-Ext significantly enhances security guarantees through authentication, it does so at the cost of scalability and performance. In realistic environments where encryption times become negligible compared to network latency, models like ORR—-despite their layered encryption—-can achieve a competitive QoS while offering enhanced protection against malicious intermediaries, presenting a viable compromise between efficiency and security.

7 Conclusions and Future work

In this paper, we have proposed and tested a key relaying solution based on the principles of onion routing, to ensure the anonymity of end-to-end keys and the sequence of intermediate nodes. Our evaluation shows that the cost of nested PQC encryption with an ORR scheme is high but affordable if the extra level of security in relaying must be achieved. Therefore, ORR and ORR-Ext both strengthen the security of key relaying in QKD networks at the cost of some extra bandwidth only. Future work will focus on implementing these models in a real QKDN to evaluate whether the proposed solutions can meet the requirements of practical deployment scenarios. In particular, the objective is to determine

whether ORR-Ext can sustain a minimum key distribution rate that would make it suitable for highly security-sensitive applications.

Acknowledgments. This work was supported under the grants TED–2021–130369B–C31 funded by MICIU/AEI/ 10.13039/501100011033 by the "European Union NextGenerationEU/PRTR"; PID2023-148716OB-C31 funded by MICIU/AEI/10.13039/501100011033; grant ED431B 2024/41 (GPC) by the Galician Regional Government; and the regional agreement for laboratories and demonstration centers in cybersecurity within the RETECH program.

References

1. D. Aasen *et al.*, "Roadmap to fault tolerant quantum computation using topological qubit arrays," 2025. [Online]. Available: https://arxiv.org/abs/2502.12252
2. S. P. Jordan and Y.-K. Liu, "Quantum cryptanalysis: shor, grover, and beyond," *IEEE Security & Privacy*, vol. 16, no. 5, pp. 14–21, 2018.
3. S. A. Syed, "The quantum threat: Preparing for the impending impact on cyber security," *International Journal of Engineering Technology Research & Management (IJETRM)*, vol. 7, no. 03, 2023.
4. C. H. Bennett and G. Brassard, "An update on quantum cryptography," *Springer Berlin Heidelberg*, 1984.
5. A. K. Ekert, "Quantum cryptography based on bell's theorem," *Physical review letters*, vol. 67, no. 6, p. 661, 1991.
6. e. a. Stanley, "Recent progress in quantum key distribution network deployments and standards," in *Journal of Physics: Conference Series*, 2022.
7. e. a. Alagic, "Status report on the fourth round of the nist post-quantum cryptography standardization process," National Institute of Standards and Technology, Tech. Rep. NIST IR 8545, 2025.
8. V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Reviews of modern physics*, vol. 81, no. 3, pp. 1301–1350, 2009.
9. C. Elliott, "Building the quantum network," *New Journal of Physics*, vol. 4, no. 1, p. 46, 2002.
10. International Telecommunication Union, "Quantum key distribution networks – Key management," ITU-T, Tech. Rep. Y.3803, Dec. 2020.
11. P. Otero-García and et al., "Onion routing key distribution for qkdn," *arXiv preprint arXiv:2502.06657*, 2025.
12. D. Goldschlag, M. Reed, and P. Syverson, "Onion routing," *Communications of the ACM*, vol. 42, no. 2, pp. 39–41, 1999.
13. e. a. Huttner, "Long-range qkd without trusted nodes is not possible with current technology," *npj Quantum Information*, 2022.
14. S. Rass, M. Mehic, M. Voznak, and S. König, "Hacking the least trusted node: Indirect eavesdropping in quantum networks," *IEEE Access*, 2024.
15. e. a. De Santis, "Parallel trusted node approach for satellite quantum key distribution," *arXiv preprint arXiv:2406.08562*, 2024.
16. e. a. Davide Li Calsi, "End-to-end qkd network with non-localized trust," 2025. [Online]. Available: https://arxiv.org/abs/2411.17547
17. N. Vyas and P. Mendes, "Relaxing trust assumptions on quantum key distribution networks," *arXiv preprint arXiv:2402.13136*, 2024.

18. H. Shim, B. Kang, H. Im, D. Jeon, and S. Kim, "qtrustnet virtual private network (vpn): Enhancing security in the quantum era," *IEEE Access*, 2025.
19. R. Rios, J. A. Montenegro, A. Muñoz, and D. Ferraris, "Towards the quantum-safe web: Benchmarking post-quantum tls," *IEEE Network*, 2025.
20. e. a. Zeng, "Towards efficient and secure quantum-classical communication networks," in *2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA)*. IEEE, 2024, pp. 520–523.
21. I. B. Djordjevic, "Joint qkd-post-quantum cryptosystems," *IEEE Access*, vol. 8, pp. 154 708–154 712, 2020.
22. C. Kuhn, M. Beck, and T. Strufe, "Breaking and (partially) fixing provably secure onion routing," 2019. [Online]. Available: https://arxiv.org/abs/1910.13772
23. ID Quantique, "Cerberis xgr qkd system," 2021. [Online]. Available: https://www.idquantique.com/quantum-safe-security/products/cerberis-xgr-qkd-system/
24. ETSI, "Quantum Key Distribution (QKD); Protocol and data format of REST-based key delivery API," European Telecommunications Standards Institute, Tech. Rep., 2019.

Complete centralized architecture for QKD networks

Anne de Chastenet[1], Eric Augé[2], Gonzague Reydet[2], Maxime Coursimault[2], Philippe Legros[1], Stephanie Molin[3]

[1] Thales, Thales Systèmes d'information et communication sécurisés, 4 avenue des Louvresses, 92230 Gennevilliers, France
[2] Thales, Thales Systèmes d'information et communication sécurisés, 110 avenue Leclerc, 49309 Cholet, France
[3] Thales, Thales cortAIx Labs,1 avenue Augustin Fresnel, 91767 Palaiseau, France
anne.dechastenet@thalesgroup.com

Abstract. We propose a complete architecture for a QKD (Quantum Key Distribution) network, from the physical layer up to the network controller. Our design is based on a centralized approach for Quantum Keys Management to mitigate security risks on the trusted relays.
This paper explores the essential components, design principles, and architectural models of QKD networks, which are critical in enabling secure information exchange amid rising cybersecurity concerns. Security within QKD networks is paramount, encompassing measures to address various threats. This includes the establishment of quantum-safe classical networks, rigorous node authentication, and dynamic eavesdropping countermeasures. Moreover, our architecture accommodates scalability and interoperability to integrate new nodes and technologies seamlessly. All these features are addressed in this paper, with a focus on the centralized Key Management System (KMS) we designed to drive QKD.

Keywords: QKD, Trusted Node, centralized, XOR, security, ITU-T Y.3803.

1 Introduction

Quantum Key Distribution (QKD) elementary links on optical fiber are currently limited to a distance of few tens of km due to fiber propagation losses. Therefore, QKD Networks (QKDN) use Trusted Nodes (TN) to digitally relay the shared secret key. QKD keys are established on each elementary link and stored in the TN; a final key (KSA key) is then elaborated end-to-end using the QKD keys generated along the path. Key Management (KM) is therefore at the heart of QKDN. Several acknowledged schemes regarding the key transfer technique are provided by Rec. ITU-T Y.3803 [1].

The first two usual schemes, "classical key transfer" or "key forwarding", consist of a "decentralized approach". The key is generated at one end-node (User Node, UN) or chosen to be the QKD key shared with its Access Node to the QKDN (AN) and forwarded hop-by-hop using the One-Time-Pad (OTP) encryption technique (Information-Theoretic-Secure process) to the other end node. In the decentralized approach,

the key is thus accessible inside each node along the path, raising the need for duly and expensively protected trusted relay nodes.

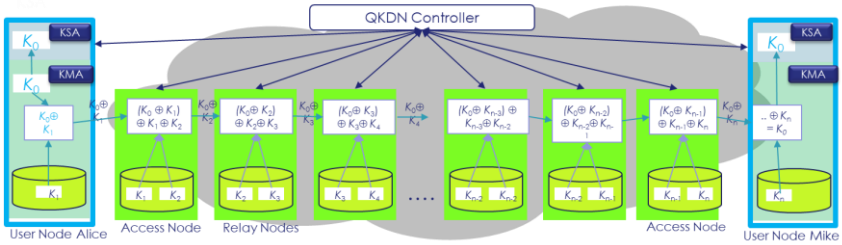


Fig. 1. One classical approach of ITU for key transfer (as per Figure 6 of ITU-T Y.3803). Each Node includes a local Key Manager, to store KMA-keys and transfer/deliver KSA-keys.

In alternative schemes (figures 7 and 8 of Rec. ITU-T Y.3803), the key to share is not forwarded to nodes along the path. Instead, QKD keys generated along the elementary links are used to provide an end-to-end key. The second of those 2 schemes is based on a centralized architecture, with an additional Key Management Agent, part of the KMS.

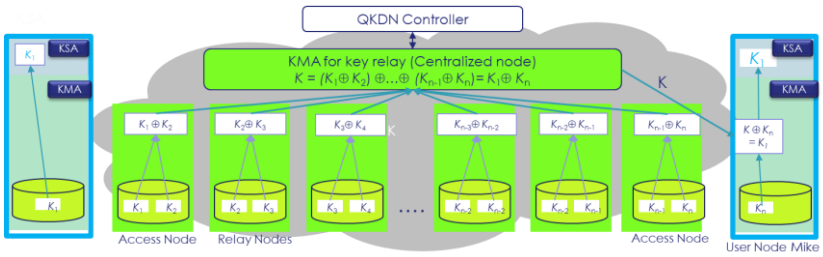


Fig. 2. Centralized approach of ITU for key transfer (as per Figure 8 of ITU-T Y.3803).

In this configuration, only the "access node" of user A knows directly the key, drastically reducing security weaknesses of the whole network (and the AN of user B could deduce the key, not directly but if it can intercept the contribution sent by user B).

No full and mature standard is available for the moment for QKDN, especially for the centralized approach of key transfer (Figure 2) which is auspicious to enhancements. Characteristics of this new approach need to be grasped in new perspectives. In the following, we sketch design concepts for an enhanced centralized QKD Network and call it "centralized XOR node" approach.

2 Main components of Key Management in centralized approach

Key Management System (KMS) is the KM layer, a specificity of QKDN (Fig 2). In the "centralized XOR node" approach, it includes local and centralized instances:
Local KMS (LKMS):

Each Trusted Node hosts a LKMS, including a Key Management Agent (KMA) and a Key Supply Agent (KSA). The KMA formats and stores as KMA-keys the keys generated by QKD on the quantum links with its adjacent nodes, whereas the KSA interfaces with the key consumer and delivers final keys, KSA-keys.

A Centralized KMA for key relay:

The centralized entity in Figure 2 is not explicitly described by the ITU. We chose to implement it in two distinct instances: the Central Key Management Core (CKMC) and the "XOR node". The CKMC is a key stone within the centralized XOR architecture, behaving like a central hub between all modules (QKDN Manager, QKDN Controller, LKMS). Its primary functions are to trigger the transfer of XORed keys from LKMSs to the centralized "XOR node" and elaborate the metadata of the key shared by user node (Alice) with another (Mike). The CKMC actuates also the centralized "XOR node" in charge of the global XOR computation and its delivery to Mike.

Management and Control functions are similar to the ones of classical networks.

The QKDN Manager is in charge of the configuration and constant monitoring of all network components. It ensures the network's health and stability.

As well, it initiates and automates the configuration of the KMS, i.e. the CKMC, the central XOR node and the LKMS. The QKDN Manager provides the static network topology to the QKDN Controller for the key routing path computation and allowances according to user's SLA and is in charge of the billing function.

The QKDN Controller computes end-to-end route across the QKDN in response to a route request, issued by the CKMC, which should be the optimal path to trigger.

For this computation, the QKDN Controller requires the complete view of the QKDN topology (static and dynamic, like link failure for instance) as well as key resource availabilities per node (key volume in real time, per link to be more specific).

3 Enhancements of the "centralized XOR node" approach

In our approach, the CKMC and "XOR node" could be distinct entities, installed in the Network Operation Centre (NOC). Collocating the NOC with a QKD node is beneficial for authentication and encryption of its exchanges.

The CKMC does not host keys but is in charge of gathering key metadata from KMA storages in the various nodes of the network (metadata, incl. key IDs, key hash and key length). It is also in charge of checking consistency of storages of all nodes and of providing synthetical information to the QKDN Controller and Manager.

The "XOR node" only receives individual XOR computation from each LKMS and computes a resulting XOR bit string from which Mike will deduce Alice's key.

In this design, the numerous inter-node KMA links (in meshed networks) are replaced by only one KMA link per node connected to the central one. In addition, KSA links between adjacent nodes are useless. Our approach drastically reduces the complexity, optimizing the number of links, especially for architecture going at scale.

Unlike forwarding schemes, this centralized scheme is prone to enhancements:

Simultaneous actuation of all nodes; Once the path has been provided to the CKMC by the QKDN Controller, the CKMC can immediately and simultaneously ask each node on the path to send the XOR of chosen KMA keys with its 2 adjacent nodes.

Such parallel process minimizes the transfer delay, which is not increasing with the number of nodes of the path, but also the probability of flop, as failure of a node before it could transfer the key is much more unlikely than in classical hop-by-hop transfers, and as, all KMA keys of the path being reserved simultaneously, no conflict between various requests for keys in the network will occur.

Symmetrisation of process, for a doubled capacity of transfer;

The XOR node delivers a resulting string to Mike, which is the XOR of Alice's key and Mike's key. Our XOR node sends this same string to Alice, so that she can get Mike's key: they share 2 keys instead of 1, sparing transport keys.

Centralization of storage and elaboration of metadata, for a better system check and reduction of metadata transfer and divulgation;

Full metadata of the transferred KSA key are sensitive: they give attackers too much information on the transfer process (between which addressees it is shared, which nodes and KMA keys were used...). In classical key transfer, they usually are elaborated and transferred progressively with the key, which is a vulnerability.

The CKMC will easily elaborate and archive full metadata of transferred keys and provide addressees with only non-sensitive synthetical metadata (key Id, key length, computed key age, computed end-to-end security parameter).

Limitation of information on a need-to-know basis, for a mitigation of risks due to attacks from a specific entity in the QKD network that would turn out to be malicious;

For instance, the QKDN Controller is asked by the CKMC to find a path between access nodes and does not know about the concerned users' nodes.

Relay nodes do not know about the concerned users' nodes: they are only requested by the CKMC to send the XOR of KMA key with Id xx and KMA key with Id yy, with tag MsgId to the XOR Node. The XOR node knows which nodes send him contributions but does not know about their role in the path: access or user or pure relay nodes.

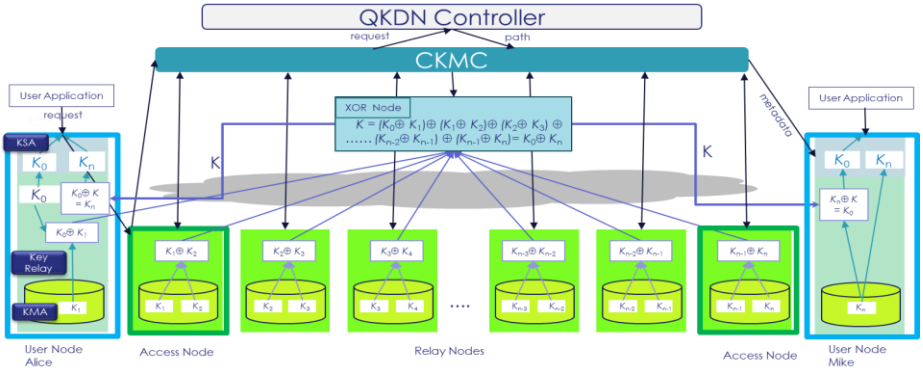


Fig. 3. Simplified process and exchanges of a key request

Mitigating the risk on relay node compromission;

Security of Users' and Access Nodes is paramount. They correspond to the definition of fully Trusted Nodes. Notably, their exchanges with CKMC and XOR node will be encrypted and authenticated.

Accessing a relay node does not give enough information, the attacker must also identify the other nodes involved in the key transfer, intercept their communications with the XOR node, be able to correlate them with the actual transfer, deduce each KMA key involved to retrieve the transferred key. Such an attack would be all the more complicated since XORs sent by the access (and users) nodes are encrypted.

Possible additional mechanisms, as a 3rd centralized entity for parallel and independent security features, for further in-depth protection (Patented by Thales);

Ease of hiding the transfer process to external attackers;

The CKMC affects different various Message Identifiers for the responses of the various nodes, to make traffic observers unable to correlate these responses to a same key transfer. Only the XOR node has the adequate list of these MsgIds.

Use of a QRNG (Quantum Random Number Generator) key generated at Alice's node;

It can be used as KSA key instead of a KMA key shared with her access node so that the access node is, in this case, unaware of the final KSA key (the interest of this choice has to be evaluated in some implementations where users' nodes are chosen to be unknown to the XOR node) or combined with quantum keys to resist to attacks on integrity (patented by Thales).

4 Risk mitigation in a centralized QKD network

Note that we consider Quantum Communication Infrastructures which deal with one unique sensitivity level, as mixing various sensitivity levels would induce potential security risks.

One of the security loopholes in the non-centralized architecture is the key transportation. Indeed, the confidentiality of the QKD key exchanged between two neighbours is preserved. It is no longer the case when a third-party node is located in the middle of the path. The key to share goes through this node in plain text or at best encrypted with a local key. So, the third party has access to the key. It becomes a greater problem when the third party is located in a foreign country. The centralized architecture prevents this need-to-know issue. All parties involved in the key transportation know only the key exchanged with its neighbourhood (which is intrinsically the case in a QKDN).

4.1 Authentication

The security brought by QKD technology is based on two principles:

- The key material exchange performed by the quantum layer;
- The authentication of the classical link for key negotiation.

The key material exchange is recognized to have Information-Theoretic Security (ITS). In order to keep this property, the authentication should get it too. ITS can be obtained by pre-shared keys (for the very first time) renewed by reserving exchanged QKD keys for it or by loading new shared key by a trusted-courier.

European National Cyber-Security Agency wrote a common paper [2] saying that to preserve ITS property, the Service Application should use ITS algorithm when it uses

the QKD keys, One-Time-Pad for encryption etc. This will be possible only for very sensitive communication with low volume of transmitted data.

Therefore, another axis for authentication may be the use of Post-Quantum Cryptography (PQC). Indeed, authentication is not concerned by the threat Store Now and Decrypt Later (SNDL) and PQC covers Quantum Computer threats. PQC implementation can provide additional security by encrypting the exchanges too.

If some elements of the QKDN cannot be connected with a quantum layer; QKDN Manager, QKDN Controller, CKMC etc; PQC will have to be used for the protection of their exchanges (authentication and/or encryption).

4.2 Complexification of attacks

The security of the nominal centralized architecture can be easily improved.

Limiting information on a need-to-know basis: e.g.

- Nodes do not know for whom they are transferring a key, the CKMC only knows the service applications involved and asks for a route between the nodes where the applications are connected without telling the subscriber Id. This improves the confidentiality in case of remote spying on the network.
- msgIds are all different, when the CKMC sends the XOR request, it uses a unique msgId for each XOR request/answer and gathers them on the same internal ID. Indeed, in case of one node compromission, the XOR sent from the compromised node to the CKMC can be decrypted with the local keys and the traffic interception of other involved nodes can help to get the final transferred key (getting one key after the other from the intercepted messages). Without common Id, the correlation is harder to be done, the XOR messages can be considered as a part of a whole exchange from external view.

Adding layers: e.g.

- Addition of a “Secret Layer” on top of the described design can also be easily implemented in a centralized way, as an additional security service (Thales patent) and a way to provide protection against errors inserted by malicious or failing QKDN nodes or XOR node (Thales patent).
- Splitting centralized KMA into CKMC and XOR node and limiting their exchanges on a need-to-know basis will prevent a malicious XOR node to collaborate with a corrupted access node for attacks

Dummy messages: in order to disturb traffic interception, dummy requests can be sent from CKMC to LKMS. The fake XOR sent by LKMS is not used by CKMC but can corrupt the result of the spy.

5 Results and Discussions

5.1 Security assessment

A ground-breaking aspect of QKD Security is that it is mastered via a security parameter, **ε-security**, as an upper bound for the probability of the QKD key to be wicked,

(either too much information available to Eve or errors in key deduced by Bob). This notion has to be extended when several QKD links intervene for the transfer. Let's note:

- **P** the probability of a transferred key to be wicked (errors or compromised)
- $\mathbf{n_j}$ the probability for node j to be hacked (all data and operations performed by this node are known by the attacker or can be corrupted)
- $\mathbf{p_j}$ the probability for a KMA-key of link j to be wicked (including ε-security parameter of QKD but also probability of compromission in storage).
- $\mathbf{q_j}$ the probability that transmissions from node j can be understood by a hacker (including the ability to correlate this message with a given key transfer)
- $\mathbf{a_j}$ the probability of transmissions from node j being altered (by hackers or because of transmission errors)
- Nodes 0 and n are the Users' Nodes (UN), Nodes 1 and (n-1) are the Access Nodes (AN), Nodes 2 to (n-2) are the relay nodes on the path
- K_0 is the KSA key to be transferred, and generated at the user's node

In the case of classical transfer, the KSA key is not wicked if and only if none of the steps in the transfer is wicked (notably, one hacked node is enough for that). Thus,

$$P = 1 - (1- n_n).\ \prod_{j=0}^{n-1}\ (1- n_j).\ (1 - a_j).\ (1- p_j .q_j)$$

P increases with the number of nodes.

In the case of the centralized design, the KSA key is wicked when one transfer KMA key is wicked and the attacker also has knowledge of the nodes in the path and accesses all k consecutive XORs sent to the XOR node, deduces all other keys to get K_n or K_1 and is able to get either the XOR sent by Alice or the one received by Mike to get K_0. Thus, the probability is bounded by:

$$P = 1 - (1- n_n).\ (1- a_n).\ \prod_{j=0}^{n-1}\ (1- n_j).\ (1- a_j).\ (1- p_j.\ r_j)$$

$$r_k = \prod_{j=0}^{k}\ q_j\ +\ \prod_{j=k+1}^{n}\ q_j\ -\ \prod_{j=0}^{n}\ q_j \qquad \text{for } 0 \le k < n$$

The r_k coefficients moderate the increase of P with the number of nodes.

Note: this probability P is an upper limit. With the centralized approach,

- a wicked relay node would not be enough to get the KSA key: the nj values for relay nodes should be decreased to consider only corruption of data.
- the knowledge of the nodes in the path is all the more difficult to get for an attacker as no full metadata transit and spying the exchanges of CKMC or XOR node can be fruitless if those exchanges are subtle (e.g. with different msgIds). A probability for such a knowledge should also be set (instead of 1 in the above formula).

As it will be seen, this upper bound is already adequate for the comparison.

The probability of a transferred KSA key to be wicked highly depends on assumptions on the elementary probabilities, as can be shown on the following figures.

For the first case, we assumed 10^{-6} for n_j, 10^{-4} for p_j, 10^{-4} for a_j, and 10^{-1} for q_j. And, when vulnerable relay nodes, 10^{-5} for n_j, 10^{-3} for p_j, and 1 for q_j.

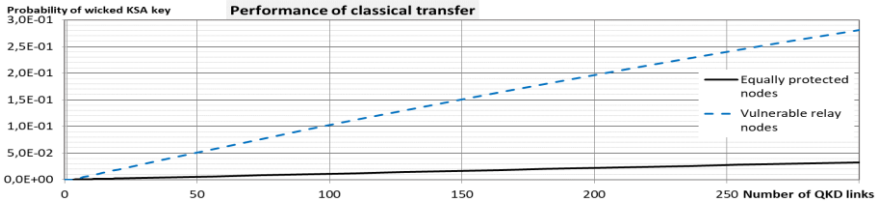


Fig. 4. Performances of classical scheme

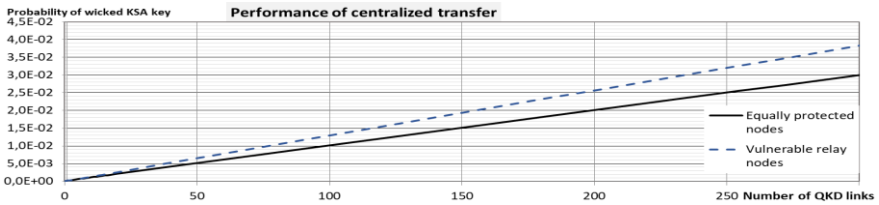


Fig. 5. Performances of the Centralized XOR node approach

For the second case, we assumed 10^{-5} for n_j, 10^{-3} for p_j, 10^{-4} for a_j, and 5.10^{-1} for q_j. And, when vulnerable relay nodes, 10^{-4} for n_j, 3.10^{-2} for p_j, and 1 for q_j.

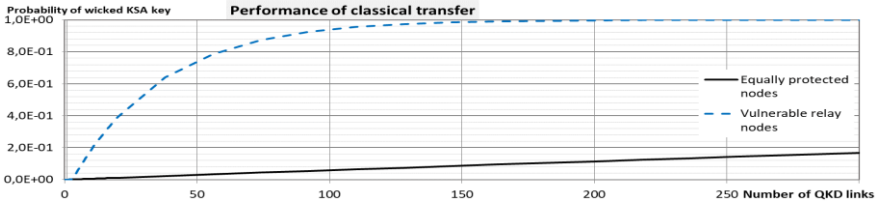


Fig. 6. Performances of classical scheme with less protection

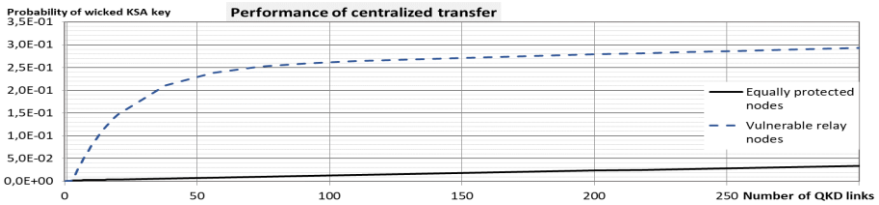


Fig. 7. Performances of centralized scheme with less protection

In a situation of all nodes having same protection characteristics, the probability of a wicked KSA key is always decreased with the centralized approach (factor respectively 1.2 and 5 for the two cases here above) with respect to classical key forwarding.

In cases where relay nodes are less protected than other nodes (user and access nodes), degradation in performances of classical transfer as a function of total number of nodes is drastic whereas security of centralized transfer is far less affected.

The comparison of performances of classical and centralized schemes shows drastic advantage of the basic centralized scheme. We did not include yet the impact of the enhancements proposed with our approach, that would increase the advantage.

5.2 Discussion

About the security evaluation;

The theoretical evaluations here above show how security is impacted by the hypothesis of vulnerability of the various nodes or of their communication means. This view demonstrates the advantage of the "Centralized XOR node" approach.

The values of the various p_j, n_j, q_j and a_j have to be assumed. They correspond to the probability that vulnerabilities in the system can be exploited, which may be hard to quantify: for instance, p_j and n_j also depend on the confidence we have in the system provider, the system certification or in the system administrators and q_j and a_j depend on the confidence about the links protection.

Assumptions have to be made, depending on the protection level that is considered, for physical access, certification or sophistication of encryption and authentication processes. Then, theoretical computations can be refined to determine impact of enhancements (not quantified yet) and trade-offs between security elements design such as:

- Generation of key with a user QRNG
- Hiding identity of Alice and Mike to the XOR node.

About foreseen drawbacks and advantages of the centralized solution;

Load of the CKMC or the XOR node in case of high traffic requests may be a question and should be checked with a simulation of the complete QKD Network. However, this load should be lower than the load of the QKDN Controller that performs path computations. It is to be noted that the principles we described here are also applicable whenever these functions are distributed to a cluster of CKMC or XOR nodes (for instance, considering a geographic or role repartition), that would allow for load balancing and redundancy, provided the inclusion of adequate protocols to run this distribution.

Authentication of CKMC and XOR node exchanges is mandatory (QKDN Controller exchanges should also be authenticated, in any design) to avoid attacks and data alteration. That is what the XOR node and the CKMC entities are preferably users of the QKD network, so that their authentication to each user node can be easily updated with KSA keys.

Duration of key transfer, which is a very important system performance, is minimal and does not increase with the number of nodes, with the centralized approach.

Management of metadata is also eased and more secure, as no transfer of full metadata is necessary. Users will be sent only synthetized metadata whereas full metadata will be centralized managed and archived for potential revocation actions.

About perspectives;

A cluster of "centralized" entities instead of one, with back-up functions, is also a way of dealing with availability and of avoiding single points of failure and shall be considered in future operational QKD Networks.

Splitting responsibilities and knowledge of information can also be helpful for security.

Extending the concept of "centralized XOR node" for the interconnection of various QKD networks, each one with its own design, can lead to attractive solutions.

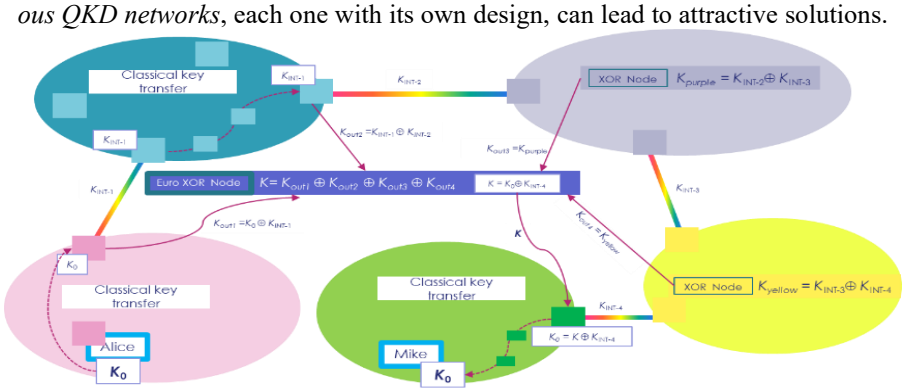


Fig. 8. Example of interconnection of heterogeneous QKD domains with centralized approach

6 Conclusion

We have designed an enhanced centralized XOR node approach of a global QKD network and analysed its performances.

We showed how the security of the network is impacted by the hypothesis of vulnerability of the various nodes or of their transmission means.

Our results show the advantage of the "Centralized XOR node" approach compared to the classical transfer approach in terms of security.

We consider this approach as a good solution also for the interconnection of various QKD networks, that may not use the same scheme and will thus avoid having to trust all QKD networks on the path.

Acknowledgements. The authors thank the European Commission for supporting part of this work within EuroQCI projects (e.g. PETRUS and FranceQCI)

References

1. Rec. ITU-T Y.3803. SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS, NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES, "Quantum key distribution networks – Key management", Dec. 2020.
2. French Cybersecurity Agency (ANSSI), Federal Office for Information Security (BSI), Netherlands National Communications Security Agency (NLNCSA), Swedish National Communications Security Authority, Swedish Armed Forces, "Position Paper on Quantum Key Distribution" (26.01.2024).

QKD-Encryption of Air Traffic Control Data with Negligible Extra Latency

Abraham Sotomayor[1][0000-0003-3847-3703], Paulette Gavignet[1][0009-0001-5935-2807], Patrice Robert[1], Sandrine Gravant[1], Luca Calderaro[2][0000-0002-1758-6760], Alessandro Emanuele[2], Costantino Agnesi[2][0000-0003-0830-2057], Gonzague Reydet[3], Olivier Ferrouillat[3], Stephanie Molin[4], Nicolas Courtel[5], Sylvain Fournier-Bazoune[5], Alexandre Tremon[5], Thierry Andre[5]

[1] Orange, 2 Av. Pierre Marzin, 22300 Lannion, France
[2] ThinkQuantum S.r.l., Via della Tecnica, 85, IT-36030 Sarcedo (VI), Italy
[3] Thales,110 avenue du Maréchal Leclerc, 49300 Cholet, France
[4] Thales, 1 avenue Augustin Fresnel, 91767 Palaiseau, France
[5] French State, Ministère chargé des Transports, DGAC/DSNA/DTI, 1 avenue du Dr-Maurice-Grynfogel 31035 Toulouse Cedex 1, France
abraham.sotomayor@orange.com

Abstract. This study explores the implementation of a Quantum Key Distribution (QKD) system using a Trusted Node (TN) and an optical switch. The TN is used for transmitting the end-to-end encryption key from one QKD link to the next, while the optical switch allows to reduce the number of required Alice/Bob pairs, thereby lowering costs. A testbed was developed within the FranceQCI project to evaluate the performance of the optical switch-based QKD system in the context of Air Traffic Control (ATC) data transmission. The setup involved two polarization-based QKD transmitters, the optical switch, a single receiver, and two encryptors. Results show that the optical switch effectively allows the transmission of end-to-end encryption keys. Furthermore, the availability of encryption keys remains sufficient even during physical link interruptions, ensuring data availability and integrity. User-centric evaluations demonstrated that the QKD-based encryption process did not significantly impact data latency, aligning with operational requirements.

Keywords: QKD, Trusted Node, Optical Switch, VPN Encryption.

1 Introduction

The deployment of Quantum Key Distribution (QKD) systems using commercial optical fibers is currently limited by distance due to fiber attenuation, which reduces the number of useful keys after the post-processing step [1]. One approach to extend the reach of a QKD-based key generation is to utilize Trusted Nodes (TN), where quantum keys are established section by section, and a global encryption key is transmitted end-to-end using One-Time-Pad (OTP) encryption with the various quantum keys [2,3]. However, a significant drawback is the requirement for additional

pairs of Alice/Bob, one for each additional TN, which increases the overall cost. A proposed strategy, as discussed in [3-7], involves using an optical switch to alternate between different QKD links. The selection of active QKD links is orchestrated by a Key Manager System (KMS) according to a defined configuration.

The effectiveness of the optical switch-based QKD must be evaluated from a user perspective to determine if the complete cryptographic system, comprising QKD, the optical switch, and encryptors, meets user requirements at the application layer, such as throughput, latency, and the availability of encryption keys in the event of a failure or attack on the QKD system.

In this context, within the framework of the FranceQCI [8] European collaborative project, an institutional end-user, the "Direction des services de la navigation aérienne" (DSNA), hosted a test campaign to assess the performance of QKD interconnected to a conventional terrestrial network. The objective was to compare, from a user perspective, both direct unencrypted and encrypted transmissions using QKD of simulated realistic Air Traffic Control (ATC) data.

2 Setup Configuration

2.1 User Platform

The end-user is the Direction générale de l'aviation civile / Direction des services de la navigation aérienne (DGAC/DSNA), the French Air Navigation Service Provider. Air Traffic Control (ATC) data are characterized by low flow rates, currently peaking at a maximum of 100 Mbps in worst-case scenarios. This data primarily consists of real-time information transmitted between remote sites, including sensors, local and central sites, and dedicated networks. The main threat to ATC data is an attack on data integrity, while data availability is a critical factor essential for ensuring public traffic security and confidentiality is less critical. Table 1 provides a general description of ATC data characteristics.

Table 1. Air Traffic Control (ATC) data characteristics.

Traffic data on a site (en-route ATC center, i.e. Area Control Center)	Radar sensor data	~ 20 kbps × 6 radars per zone => 120 to 160 kbps
	Radar Tracks	TMA4 (Terminal Manoeuvring Area) flow (broadcast to Airport Controls): ~ 50 to 100/150 kbps
		TMA8 flows: ~ 20 kbps only
	Flight Plans[1]	~ 11 kbps

[1] In certain emergency situations, all flight plans may need to be sent simultaneously, with a maximum of approximately 900 flight plans, resulting in a data rate of about 10 Mbps.

	Digital Voice for Air Controllers	~ 120 kbps (vs. 64 kbps classically) × simultaneous communications (permanent radio channels); for the largest site: 8400 kbps (telephone) + 70 000 kbps (radio).
Maximal latency	Data for Surveillance	~ 150 ms
	Voice	~ 50 ms (jitter < 5 ms)
Availability	> 99,999 %	
Data or integrity loss	$< 10^{-5}$	
Various priorities to be handled	Radio, voice and radar	
	Flight plans and ATFCM (Air Traffic Flow and Capacity Management)	
	Other data only in best effort.	

2.2 Testbed Implementation

The test campaign was designed to evaluate the encryption of simulated realistic ATC data. To achieve this, the platform was divided into two Virtual Local Area Network (VLAN) architectures for comparison: one with encryption and one without. This setup facilitated the same data exchanges, as illustrated in Fig. 1.

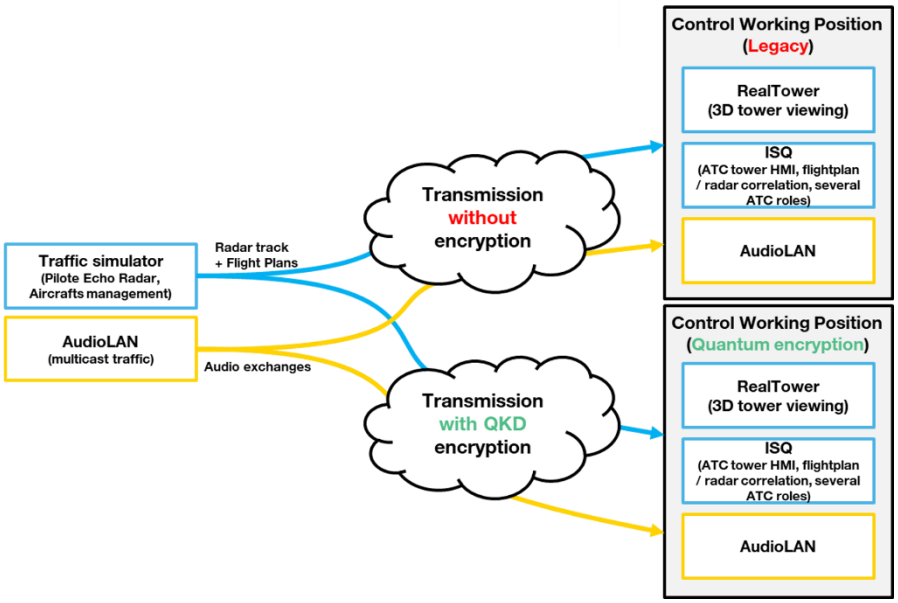


Fig. 1. Testbed implemented at the Direction des Services de la Navigation Aérienne (DSNA).

The transmission system with QKD encryption is illustrated in Fig. 2. The QKD system consists of two polarization-based QKD transmitters, named Alice1 and Alice2, and a single QKD receiver, named Bob, forming two links or channels using G.652 optical fibers. Each transmitter connects to the output ports of the optical switch via

the quantum channel port, while its corresponding input connects to Bob's quantum channel port. The optical switch selects one active channel at a time. The QKD transmitters, Alice1 and Alice2, provide secure keys to the encryptors, which then encrypt the data from the traffic simulator.

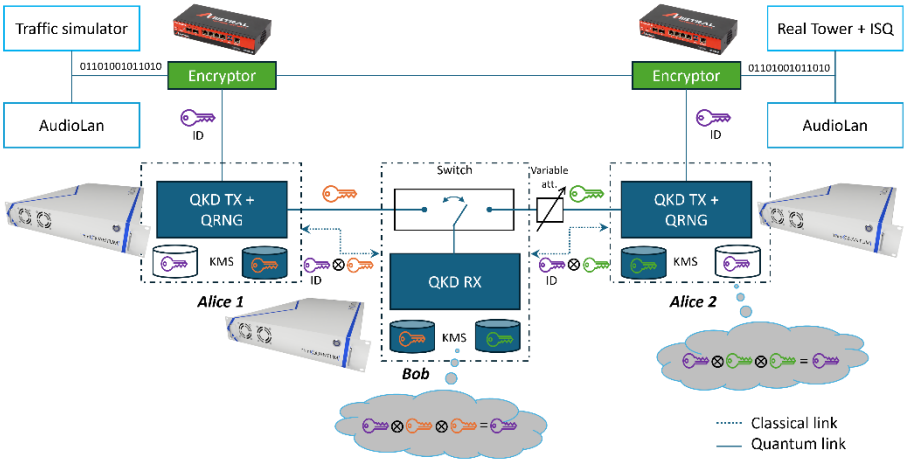


Fig. 2. Encryption system comprising QKD devices, an optical switch, and two encryptors.

The QKD devices used were the commercial Quky developed by ThinkQuantum S.r.l., while the optical switch was sourced from Agiltron. The encryptors employed on the quantum network were the Mistral IP9001 from Thales. In this work, we utilized the standard configuration of the Quky devices, which includes one fiber port for the quantum channel in the C-band and an RJ45 port for the classical channel.

If no switching is employed, the number of QKD transmitter/receiver (Tx/Rx) pairs must match the number of physical links. In a network with n nodes, this leads to a maximum of $\frac{n*(n-1)}{2}$ Tx/Rx pairs. By contrast, the use of optical switching can significantly reduce the number of required QKD devices, since nodes may share common transmitters or receivers through time-multiplexing across multiple physical connections. Furthermore, by employing virtual connections, i.e., trusted-node hops, a full mesh topology can be realized with only n QKD devices. The drawback of this approach is a reduced secret key generation rate, as resources are shared, and link-activation overheads limit the effective time available for key distribution. Such overheads can, however, be mitigated by minimizing the number of switching operations. Ultimately, this introduces a trade-off between system performance and deployment cost that must be assessed with respect to the network requirements. In the present use case, the analysis favored the multiplexed approach.

2.3 QKD System

The QuKy devices implement the efficient three-state BB84 protocol using the polarization degree of freedom of the photon to encode the quantum states. Furthermore, the system implements the one decoy intensity strategy to mitigate Photon-Number-Splitting (PNS) attacks. This simplified protocol guarantees the same level of security of the original BB84 protocol, whereas the one-decoy scheme can provide higher rates in the finite-key scenario compared to the two-decoy scheme.

The system can be seamlessly integrated in a data center thanks to the 19" rack-mount enclosure and interface with standard network equipment. The quantum transmission only requires one simplex fiber, since the temporal synchronization is performed via the Qubits4Sync algorithm which uses the qubits themselves as a frequency and absolute time reference [9]. This implies that this QKD system does not requires any synchronization subsystem, which is usually implemented with a second fiber or channel wavelength multiplexing to share the time reference.

The wavelength used for the transmission of the quantum signals is centered at 1550 nm, which is the most convenient when using a dedicated fiber, as this ensures the minimum losses per km of fiber. The transmitter is equipped with a polarization encoder, iPOGNAC, that guarantees stable, low-error, and calibration-free polarization encoding, generating the horizontal and vertical states for the key generation basis and the diagonal polarization for the control basis [10]. Furthermore, these QKD devices perform phase randomization via laser gain switching, and randomness is provided by a source-device-independent Quantum Random Number Generator (QRNG) integrated within the QKD transmitter. Both features are crucial for the secure implementation of the QKD protocol.

At the receiver, the quantum states are projected into two mutually unbiased bases, which are randomly and passively selected by a beam splitter. The bases are aligned before each QKD link is activated; this base agreement procedure takes some minutes to complete. After the projection, the single-photons are detected using InGaAs/InP single-photon avalanche detectors (SPAD). The polarization drift induced by the fiber is compensated by an all-fiber polarization controller. The correction signal applied to this polarization controller is given by a set of public quantum states which are framed between the qubits used for the QKD protocol. These states are uncorrelated with the qubits used for the QKD protocol and therefore do not represent a security threat. The QKD systems perform a real-time post-processing to generate the secure key using an auxiliary classical communication channel.

The auxiliary channel is authenticated using the Message Authentication Code (MAC) Wegman-Carter algorithm. Before the first secure key is generated via QKD, the Pre-shared Key (PSK) is used for authentication. Once secure key is available, the Wegman Carter algorithm is updated with a portion of the QKD key.

2.4 Generation of end-to-end encryption key

In this section, we explain the process for delivering end-to-end encryption keys between Alice1 and Alice2 using Bob as a TN (as depicted in Fig. 2). Using the Quky devices, the KMS, which is software embedded in each Quky device, determines, through a global configuration file, the rules to generate end-to-end encryption keys.

- Alice1 generates a secret key for end-to-end encryption (the purple key in Fig. 2) through the QRNG embedded in the QKD transmitter. This key must be securely transmitted to Alice2.
- First, Alice1 and Bob, with the optical switch connecting Bob to Alice1, establish a QKD key (the orange key in Fig. 2).
- Alice1 then encrypts the purple key using the orange key with OTP encryption and sends it to Bob via a classical link. This classical link does not need to be secure (but authenticated) as long as the purple key is fully encrypted, and Bob acts as a trusted node.
- Bob retrieves the purple key using the orange key.
- To securely send the purple key from Bob to Alice2, the switching mechanism should be activated. According to the KMS configuration set in this testbed, the switching mechanism is based on the available key collected in the buffers of each QKD link[2]. When a QKD link has produced a minimum of two key blocks, the system checks the ratio between the amount of key collected in the two physical links and compares it with the ratio of the relative weights defined in the KMS configuration file. Depending on this comparison, the system decides to keep the current active QKD link or to switch and activate the other one. When a QKD link's available keys reach 98% of its buffer size, priority is given to the other link regardless of the weight's comparison. This is to avoid waste of secret keys. If both links have their buffers more than 98% full, the link with the least stored key is activated when the decision is made.
- Therefore, once one of these conditions is met, and Bob and Alice2 are physically connected, they can establish a QKD key (the green key in Fig. 2).
- Then, Bob uses this key to encrypt the purple key with OTP and sends it to Alice2, who retrieves the purple key using the previously established green key.
- Finally, the purple keys are delivered to the encryptors.

2.5 Cryptographic Application

The Mistral solution from Thales enables the establishment of secure IPsec Virtual Private Networks (VPN) between network enclaves to be protected. The Mistral

[2] All QKD devices have implemented a key buffer storing secret keys (end-to-end encryption keys) and QKD keys. While these buffers are used in the switching mechanism, they also serve to avoid interruptions in the encryption process due to a lack of secret keys.

IP9001 encryption device allows for throughput of up to 2 × 1 Gbps Full Duplex and low latency of less than 1 ms.

For the needs of QKD networks and EuroQCI projects, a specific firmware has been developed for the Mistral IP9001 device to add ETSI 004 and ETSI 014 interfaces, allowing the retrieval of quantum keys from any external Key Management Entity (KME) system compliant with one of these standards. Quantum keys are utilized for the Advanced Encryption Standard (AES)-256 ciphering of VPN data traffic. In this setup, quantum keys replace classical keys obtained through the Diffie-Hellmann algorithm during the Internet Key Exchange (IKE) negotiation protocol. For the DSNA testbed experimentation, the two Mistral IP9001 encryptors were configured to request and use new quantum keys after processing 700 MB of ciphered data.

3 Results and Discussions

We conducted several tests to measure the stability of the QKD system over extended periods, focusing on the Secret Key Rate (SKR) and Quantum Bit Error Ratio (QBER). Additionally, we assessed the impact on user data traffic by specifically measuring the latency for encrypted data and comparing it with the latency for non-encrypted data.

3.1 Measurements of SKR and QBER

In our experiments, we considered a scenario where the Alice1-Bob and Alice2-Bob links each have a fiber length of 5 km, with a variable optical attenuator installed on the Alice2-Bob link to simulate different span losses. The Alice1-Bob link exhibited a total loss of approximately 4 dB (including the optical switch), while the attenuator for the Alice2-Bob link was set to 5 dB, plus an insertion loss of 4 dB, resulting in a total loss of approximately 13 dB. Fig. 3 and 4 show the recovered SKR and QBER for each link, respectively.

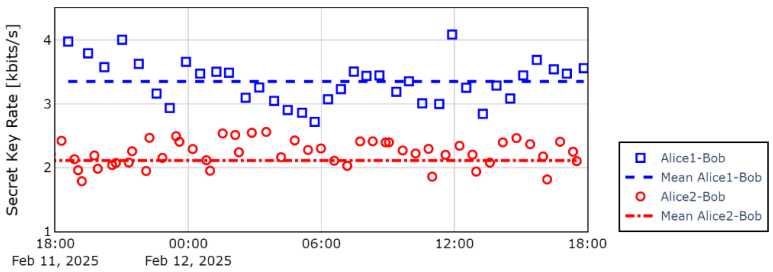


Fig. 3. Secret Key Rate (SKR) measured in Alice1-Bob with total loss ~4 dB (blue squares) and Alice2-Bob with total loss ~13 dB (red circles). The dashed lines are the mean SKR over 24 hours of measurement for Alice1-Bob: ~3.35 kbits/s and for Alice2-Bob: ~2.11 kbits/s.

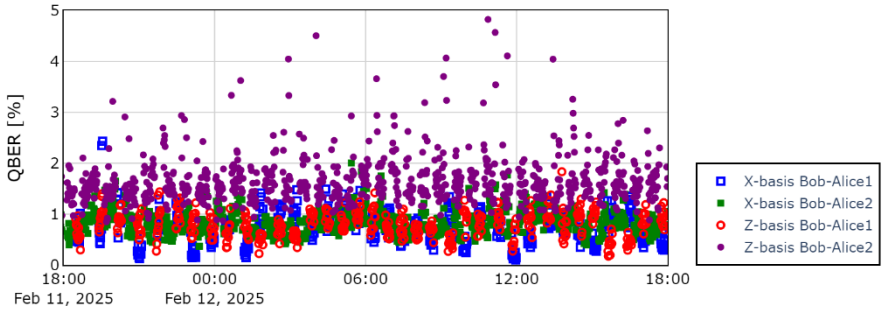


Fig. 4. Quantum Bit Error Ratio (QBER) measured in Alice1-Bob with total loss ~4 dB and Alice2-Bob with total loss ~13 dB for X and Z basis.

3.2 Interruption of quantum links to check key consumption

To evaluate the use of global keys by the encryptors, we observed the number of available keys in Alice1 and Alice2. During normal ATC operation, the available keys are maintained at their maximum of ~ 10 MB, as the QKD system continuously generates encryption keys. To assess the impact on available keys during link failures or events such as Denial of Service (DoS) attacks, we cut the two physical links (Alice1-Bob and Alice2-Bob) and monitored the decrease in available keys in the buffers. Given the low data rate of ATC, we observed a minimal decrease in the available keys, as shown in the left-side and right-side insets of Fig. 5. To challenge the encryption system, we transmitted 50 GB of data within a 10-minute timeslot, resulting in a data rate of ~ 85 MBps. The Mistral encryptors in this setup have a key renewal threshold of 700 Mb, triggering an AES-256 key renewal every second. After 40 hours, ~ 170,000 bytes of keys were utilized out of the 10 million available (not shown in Fig. 5). This indicates that data can still be encrypted for an extended period, even during link failures or DoS attacks.

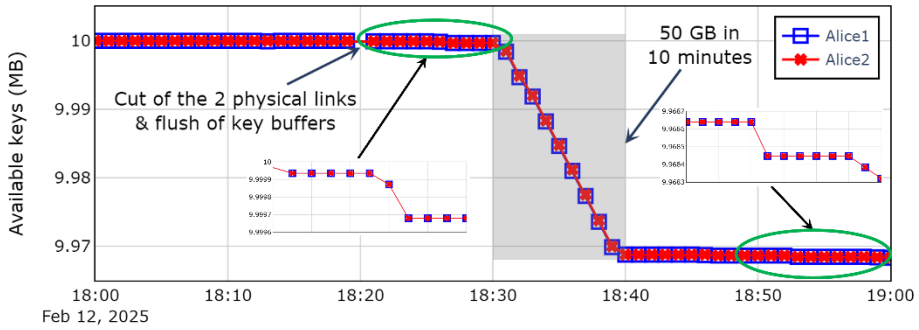


Fig. 5. Evolution of the use of available end-to-end encryption keys.

3.3 User Results

From the user's perspective, maintaining consistent latency was crucial. Therefore, the test focused on measuring the latency of the encrypted data flow and comparing it to that of non-encrypted data. The data used in this test consisted of simulated surveillance and audiolan data. Fig. 6 presents these measurements.

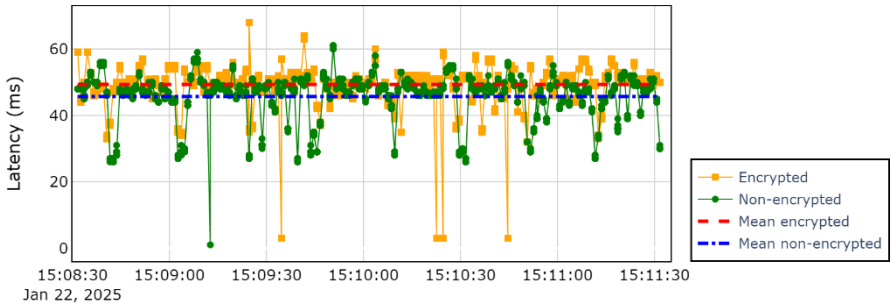


Fig. 6. Comparison of latency between encrypted and non-encrypted data flows.

As observed, there is no visual or auditory difference in latency between the encrypted and non-encrypted data, with no impact on the user application.

4 Conclusions

The suitability of QKD-based encryption was tested from the user's perspective to ensure it meets requirements such as consistent data latency and the availability of end-to-end encryption keys. An optical switch was introduced to allow for a single receiver. After confirming acceptable values for the SKR and a low QBER, we established that the available keys are sufficient to continue encrypting the data flow, even during QKD link failures or DoS attacks. Latency measurements in the user application confirmed that the time required to encrypt messages with QKD and Mistral encryptors is nearly negligible, making this solution suitable for user needs. It is worth mentioning that optical switching reduces the SKR in a single QKD link, but in this case, this was not an issue due to the ample quantity of accumulated encryption keys and the low traffic rate to be encrypted. Data availability is a critical aspect for this use case. In this regard, implementing a full (or partial) redundancy scheme, accounting for other technologies such as post-quantum cryptography (PQC) or classical Diffie-Hellman key exchange, could be necessary. Additionally, a secure key storage and usage policy could be considered, given the significant amount of generated secure keys.

Acknowledgements

This research was co-funded by the European Union under the Digital Europe Program Grant Agreement No. 101091675 (FranceQCI).

References

1. M. Dušek, N. Lütkenhaus, and M. Hendrych, “Chapter 5 - Quantum cryptography”, vol. 49, E. Wolf, Ed. Elsevier, 2006, pp. 381–454.
2. E. Pincemin, P. Gavignet et al., “400-Gbps Coherent Transmission of 100-Gbps QKD-Secured Data Stream Over 184-km of Standard Single Mode Fiber Through Three QKD Links and Two Trusted Nodes”, Journal of Lightwave Technology, Vol. 42, No. 12, June 15, 2024
3. L. Salvail, M. Peev, E. Diamanti, R. Alléaume, N. Lütkenhaus, and T. Länger, “Security of trusted repeater quantum key distribution networks”, J. Comput. Secur., vol. 18, no. 1, pp. 61–87, Jan. 2010.
4. Picciariello, F., Karakosta-Amarantidou, I., Rossi, E. et al., “Intermodal quantum key distribution field trial with active switching between fiber and free-space channels”. EPJ Quantum Technol. 12, 6 (2025). https://doi.org/10.1140/epjqt/s40507-025-00306-9
5. A. Tayduganov, V. Rodimin, E. Kiktenko, V. Kurochkin, E. Krivoshein, S. Khanenkov, V. Usova, L. Stefanenko, Y. Kurochkin, and A. Fedorov, "Optimizing the deployment of quantum key distribution switch-based networks," Opt. Express 29, 24884-24898 (2021).
6. V. Karavias, A. Lord and M. C. Payne, "Switching in quantum networks: an optimization investigation," in Journal of Optical Communications and Networking, vol. 16, no. 3, pp. 404-418, March 2024, doi: 10.1364/JOCN.513637.
7. Brunner, H.H., Fung, CH.F., Peev, M. et al., “Demonstration of a switched CV-QKD network”. EPJ Quantum Technol. 10, 38 (2023). https://doi.org/10.1140/epjqt/s40507-023-00194-x
8. Petrus Homepage, https://petrus-euroqci.eu/, last accessed 2025/04/18
9. Calderaro L, Stanco A, Agnesi C, Avesani M, Dequal D, Villoresi P, Vallone G (2020), “Fast and Simple Qubit-Based Synchronization for Quantum Key Distribution”. Phys Rev Appl 13:054041. https://doi.org/10.1103/PhysRevApplied.13.054041
10. M. Avesani, C. Agnesi, A. Stanco, G. Vallone, and P. Villoresi, "Stable, low-error, and calibration-free polarization encoder for free-space quantum communication," Opt. Lett. 45, 4706-4709 (2020).

Multi-Level Parameter Estimation for Receiver Noise in CV-QKD

Guillaume Ricard*[1], Yves Jaouën[1], and Romain Alléaume[1]

Telecom Paris - Institut Polytechnique de Paris[1], 19 Pl. Marguerite Perey, 91120 Palaiseau, France
Ricard@telecom-paris.fr*

Abstract. We present an experimental method for continuous-variable quantum key distribution (CV-QKD) that removes the need for conventional receiver noise calibration. By exploiting Multi-Level Parameter Estimation (MLPE) with controlled attenuation at the receiver, our approach directly estimates the total receiver noise, including contributions from the local oscillator, while simultaneously extracting channel-induced excess noise. This calibration-free technique allows continuous operation without halting quantum transmissions, increasing experimental throughput and robustness. We implement the method in a dual-polarization, local-local oscillator CV-QKD system multiplexed with a classical communication channel, demonstrating precise receiver noise estimation. Our results highlight MLPE as a practical tool for scalable, real-time CV-QKD deployment.

Keywords: Continuous-variable quantum key distribution (CV-QKD)· parameter estimation · multi-level parameter estimation (MLPE) · Receiver noise

1 Introduction

Continuous Variable Quantum Key Distribution (CV-QKD) has seen significant progress in recent years, with successful field trials [14] and demonstrations over impressive distances [6]. Among the different families of QKD protocols, the continuous variable approach stands out as a particularly promising option for metropolitan-scale deployment, thanks to its cost-effectiveness, compatibility with standard telecommunication hardware, and its ability to coexist with classical optical communication systems [9],[17].

In the early development of CV-QKD, system performance was heavily constrained by excess noise, much of which originated from the receiver's electronic components. To mitigate this, researchers treated this electronic noise as trusted [13], an assumption justified by the controlled access to the receiver's laboratory and precise calibration of this noise source. This strategy enabled CV-QKD implementations to extend to much longer distances.

However, electronic noise is not the sole contributor to receiver-induced noise. Technical impairments of the Local Oscillator (LO) also impact the overall excess noise. These LO-related contributions are typically overlooked or implicitly

absorbed either into untrusted noise terms or even in the shot noise estimator in conventional security models.

In this work, we propose a practical method that alleviates the need for conventional receiver calibration by leveraging Multi-Level Parameter Estimation (MLPE). By modulating the amplitude at the receiver using controlled attenuators, MLPE provides independent measurements of both total receiver noise and channel-induced excess noise. This approach allows continuous operation without interrupting key distribution, improving experimental throughput, and enabling real-time noise monitoring.

We implement this method in a dual-polarization CV-QKD system operating with a local-local oscillator (LLO) configuration and multiplexed classical communication channels. Our results demonstrate that MLPE accurately estimates the total receiver noise and separates it from channel contributions, paving the way for full receiver noise calibration-free CV-QKD experiments that maintain high performance over long distances and in realistic experimental conditions.

2 Multi-Level Parameter Estimation (MLPE)

We propose a novel approach for parameter estimation in CV-QKD that leverages an amplitude modulator placed at Bob's side. While amplitude modulation at Bob has previously been explored as a countermeasure against side-channel attacks [11, 15], our scheme uses it to refine the estimation of key channel parameters, thereby enhancing security and performance.

2.1 Standard parameter estimation

In conventional parameter estimation, Alice reveals a subset of her transmitted symbols, allowing Bob to compute statistics over his received signal. With discrete modulation, such as 64-QAM probabilistic constellation shaping (PCS), Bob can compute not only the overall variance of the received signal (V_B), but also the conditional variance for each symbol sent by Alice ($V_{B|A}$). This conditional variance effectively removes the contribution of the modulation, isolating the channel noise. Incorporating this information provides an additional equation that can be exploited for more accurate parameter estimation.

In standard CV-QKD, the receiver noise is typically modeled as the sum of shot noise and electronic noise:

$$\nu_{\text{rec}} = 1 + \nu_{\text{el}}. \tag{1}$$

For heterodyne detection, the total and conditional variances at Bob are:

$$V_{B|A} = \frac{\eta T}{2}\xi_A + \nu_{\text{rec}} \tag{2}$$

$$V_B = \frac{\eta T}{2}V_A + \frac{\eta T}{2}\xi_A + \nu_{\text{rec}} \tag{3}$$

Excess noise and transmittance can then be estimated via:

$$\xi_A = \frac{2}{\eta T}(V_{B|A} - \nu_{\text{rec}}) \tag{4}$$

$$T = \frac{2(V_B - V_{B|A})}{\eta V_A} \tag{5}$$

2.2 MLPE with controlled attenuation

In our MLPE scheme, Bob introduces controlled optical attenuations η_1 and η_2 using his amplitude modulator. These known attenuations effectively adjust the full channel transmittance, providing two additional equations that allow a more precise determination of channel parameters.

For example, in practice, ν_{rec} also contains optical contributions from the local oscillator, which are not fully captured by electronic noise and shot noise alone:

$$\nu_{\text{rec}} = 1 + \nu_{\text{el}} + \nu_{\text{LO}} \tag{6}$$

Accurate estimation can therefore be achieved with controlled attenuations at Bob. With two known attenuations η_1 and η_2, the variances become:

$$V_{B,\eta_i} = \frac{\eta_i T}{2} V_A + \frac{\eta_i T}{2} \xi_A + \nu_{\text{rec}} \tag{7}$$

$$V_{B|A,\eta_i} = \frac{\eta_i T}{2} \xi_A + \nu_{\text{rec}} \tag{8}$$

This leads to direct MLPE estimators:

$$\nu_{\text{rec}} = \frac{V_{B,\eta_1} - R_{\text{att}} V_{B,\eta_2}}{1 - R_{\text{att}}} \tag{9}$$

$$\xi_{A,\eta_i} = \frac{2}{\eta_i T}(V_{B|A,\eta_i} - \nu_{\text{rec}}) \tag{10}$$

where $R_{\text{att}} = \eta_1/\eta_2$

And finally, if we know the electronic noise variance $\nu_{\mathbf{el}}$, we can estimate the local oscillator contribution to receiver noise that is currently wrongly attributed to shot noise:

$$\nu_{\text{LO}} = \frac{V_{B,\eta_1} - R_{\text{att}} V_{B,\eta_2}}{1 - R_{\text{att}}} - 1 - \nu_{\mathbf{el}} \tag{11}$$

3 Shot-Noise Unit (SNU) and Calibration Cycle

In practical CV-QKD, the shot-noise unit must be calibrated. The standard procedure typically involves two steps:

1. Measure the electronic noise with both the LO and signal off:

$$\widehat{\sigma}^2_{\text{calib},1} = \sigma^2_{\text{elec}} \tag{12}$$

2. Measure the full receiver noise with the LO on and signal off:

$$\widehat{\sigma}^2_{\text{calib},2} = \sigma^2_{\text{rec}} \tag{13}$$

From these measurements, the shot-noise variance is estimated as

$$\widehat{N_0} = \widehat{\sigma}^2_{\text{calib},2} - \widehat{\sigma}^2_{\text{calib},1} \tag{14}$$

allowing all subsequent data to be expressed in SNU.

This conventional estimator neglects the contribution from the local oscillator, ν_{LO}, which means MLPE cannot separate it accurately unless the shot noise is already known. Explicitly accounting for the LO yields

$$\widehat{N_0} = N_0 + \sigma^2_{\text{LO}} \tag{15}$$

In this framework, the LO contribution is incorrectly attributed to shot noise and cannot be distinguished using only the standard calibration equations: electronic noise calibration (no LO contribution), full receiver noise calibration, and the measured variances V_B and $V_{B|A}$, where both shot noise and LO contributions appear identically.

Nevertheless, MLPE remains valuable. Electronic noise is typically small and stable, requiring infrequent calibration. By contrast, total receiver noise often fluctuates significantly, forcing frequent calibrations in conventional setups and imposing a duty cycle on quantum transmission: time must be spent alternating between calibration, collecting sufficient statistics, and actual quantum communication. MLPE mitigates this limitation by directly estimating ν_{rec} from controlled attenuated signals, eliminating the need to halt transmission for frequent calibration. This enables near-continuous monitoring of receiver noise and improves overall experimental throughput.

4 Experimental system design

We have designed a coherent optical communication system that can jointly perform classical and shot-noise-limited quantum communications. Both coherent communications channels are multiplexed over the same fiber and operated in the so-called local local oscillator (LLO) [12] configuration. This allows in particular to perform CV-QKD while closing the security loophole related to calibration attacks on transmitted local oscillator [7]. Our approach notably focuses on recovering the carrier phase using a "symbiotic" classical communication channel [2],[16], rather than relying on pilot tones [10]. While more complex to perform than sole CV-QKD, this approach has the potential to increase quantum communications (and CV-QKD) pervasiveness by minimizing the resources necessary to deploy them but also by pushing further the conditions under which QKD and classical communications can coexist over the same fiber.

At Alice's side (the emitter), a 100 Hz linewidth laser at telecom wavelength is modulated using a dual-polarization IQ modulator. The classical channel is

digitally shifted from the carrier by 4 GHz, and similarly, the quantum channel is shifted by 1 GHz, as illustrated in the spectrum of Figure 1. It's crucial that the classical signal is accurately pulse-shaped to prevent its frequencies from interfering with the quantum signal. Such filtering is not needed on the quantum signal however, since the DA-LMS algorithm (at reception, as discussed in the next section) naturally converges to an RC-like filter, while also accounting for non-ideal response of electrical devices. Additionally to being spectrally separated, classical and quantum channels are also modulated on orthogonal polarization states and the quantum signal is selectively attenuated until it reaches the desired modulation variance V_A.

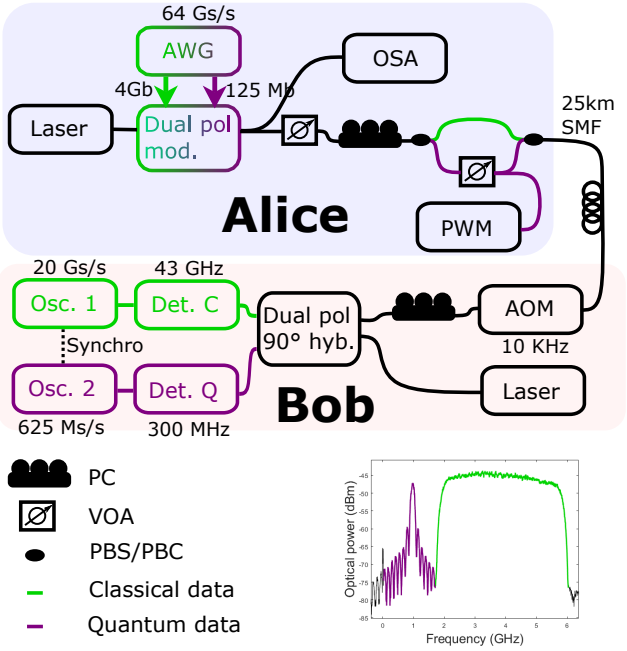


Fig. 1: Block diagram of the experimental setup and optical spectrum read on the OSA.

After being recombined, both classical and quantum signals travel through a 25 km single-mode fiber (SMF) link before reaching Bob's lab (the receiver). Just after the channel, the light encounters an acousto-optic modulator (AOM) driven by a 10 kHz square analog electrical signal. This AOM allows to perform CV-QKD acquisitions at different total transmission levels, hence allowing a more refined parameter estimation procedure.

Finally, light enters our dual polarization 90° hybrid along with the LO, whose optical frequency is tuned to detect the quantum signal in baseband. Coherent detectors measure both quadratures of the resulting optical signals, on both

polarizations. The signals are digitally sampled using two distinct oscilloscopes that share the same clock reference and trigger for accurate time synchronisation.

The balanced receivers used for quantum signal detection exhibit a 13dB clearance over a 350 MHz bandwidth. The losses induced within the detection apparatus consist in the detectors efficiency (0.5 dB), losses in the hybrid (2dB), and losses in the AOM (3dB). Those induced by our 25 km SMF channel are 5 dB. Shot noise calibration is made by cutting incoming light with LO still on [3]. In our case, this is done before each block acquisition of $1.8 \cdot 10^5$ symbols.

5 Digital signal processing (DSP)

We developed a joint digital DSP routine leveraging the high signal-to-noise ratio (SNR) estimates from classical signals to precisely correct in frequency and phase the quantum signals.

The classical signal is generated as a Quadrature Phase Shift Keying (QPSK) signal at 4 Gb, pulse-shaped using a Root Raised Cosine (RRC) filter with a roll-off of 0.1, facilitating sharp frequency cutoff and minimizing leakage across quantum signal frequencies. Meanwhile, the quantum signal is modulated as a 64 states Quadrature Amplitude Modulation (64-QAM) at 125 Mb with Probabilistic Constellation Shaping (PCS) [1] [4].

Upon reception, the classical signal undergoes a series of fairly standard processing steps [8], with a customized two-convergence radius (to account for Bob's amplitude modulation) version of the Constant Modulus Algorithm (CMA). We typically observe error-free classical transmission. For the quantum signal, corrections in both frequency and phase are applied using estimations derived from the classical signal. Subsequently, a data-aided least mean squares algorithm (DA-LMS) [5] is employed to optimize a Finite Impulse Response (FIR) filter that will account for imperfections and optimally downsample the QKD signal to a 1 sample per symbol regime, at the cost of sacrificing 4000 quantum symbols (out of $1.8 \cdot 10^5$) for convergence. We carefully tune the learning rate of the LMS algorithm to maximize performance. Any remaining frequency offset is then corrected during parameter estimation by iteratively minimizing excess noise.

6 Results

6.1 Attenuation accuracy

Calibration of the attenuation ratio R_{att} is central to the success of MLPE. This calibration can be performed once optically, after which the ratio and its stability can be verified from the data itself. In practice, we use either a classical signal that experiences the same attenuation as the quantum channel, or high-intensity quantum data. As shown in Table 1, the ratio of variances between the signals at attenuation levels η_1 and η_2 provides a precise and stable estimate of R_{att}, with tight 95% confidence intervals over several independent data blocks. The values

obtained from classical and quantum signals are consistent, confirming both the calibration and the long-term stability of the attenuation settings.

The raw measurement statistics further illustrate this effect. Figure 2(a) shows time-domain quadrature traces under three different conditions: full extinction (receiver noise only), intermediate attenuation η_2, and higher transmission η_1. The corresponding probability density functions are displayed in Fig. 2(b). While a Gaussian model accurately fits the receiver noise (extinction case), it no longer provides a good description of the modulated quadrature distributions at finite transmission, where clear deviations from Gaussianity appear. These observations underline the reliability of the attenuation calibration as well as the statistical distinctness of the different attenuation regimes.

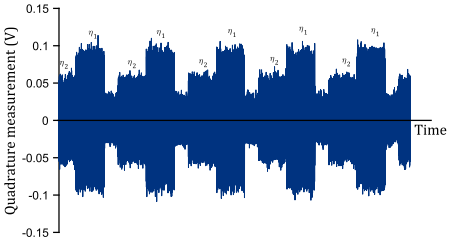
Table 1: Ratio of variances between signals in η_1 and η_2 attenuation configurations with 95% confidence intervals, used to verify attenuation stability over four data blocks, for both classical and (high-intensity) quantum data.

Signal	**Ratio of transmittances η_1/η_2**	**CI 95%**	**Samples per block**
Classical	3.2636	0.0032	5 000 000
Classical	3.2513	0.0032	5 000 000
Classical	3.2455	0.0034	5 000 000
Classical	3.2455	0.0033	5 000 000
Quantum	3.3136	0.0255	156 250
Quantum	3.2874	0.0242	156 250
Quantum	3.3042	0.0253	156 250
Quantum	3.2454	0.0247	156 250

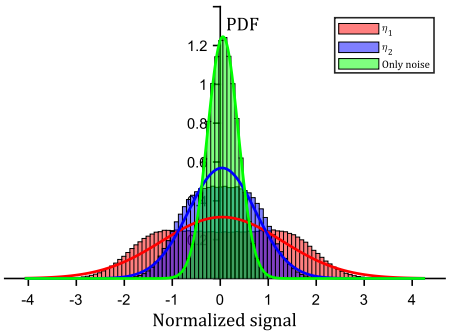
6.2 Noise estimation

The receiver noise can be calibrated independently, and MLPE provides a direct way to cross-check this calibration. Figure 3(a) shows the estimated excess noise as a function of the modulation variance V_A for both attenuation configurations η_1 and η_2. A linear fit with intercept constrained to zero yields $\xi_{\eta_1} = 0.0043V_A$ and $\xi_{\eta_2} = 0.0013V_A$, fully consistent with the measured attenuation ratios and confirming that the excess noise scales as expected with the input modulation.

In addition, Figure 3(b) presents the estimate of the receiver noise variance ν_{rec} obtained from MLPE. The result is in excellent agreement with the independently calibrated value, demonstrating consistency between experimental measurements and theoretical expectations. This agreement validates both the accuracy of the attenuation calibration and the robustness of the MLPE-based noise estimation procedure.

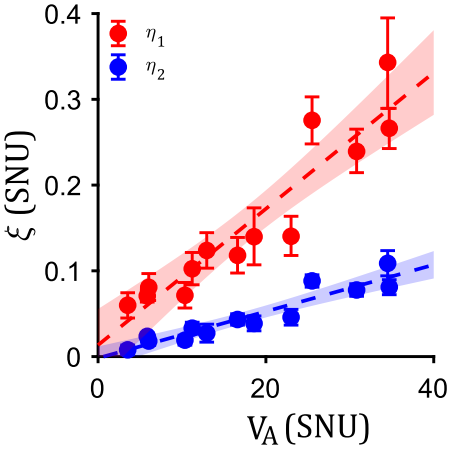


(a) Time-domain quadrature measurements in volts vs. time, showing three distinct attenuation levels: complete extinction (receiver noise only), attenuation η_2, and attenuation η_1.

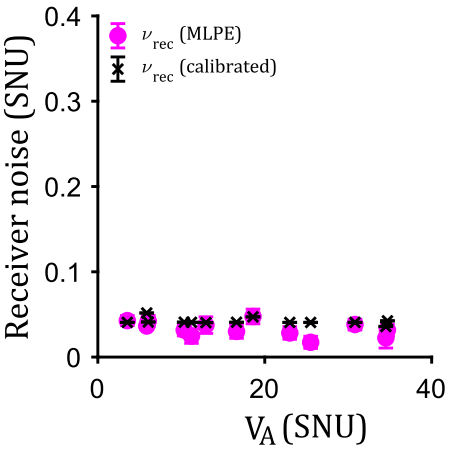


(b) Histograms of quadrature distributions for η_1, η_2, and noise only. Solid lines: Gaussian fits (good for noise only, poor for modulated signals).

Fig. 2: Quadrature measurements illustrating the effect of attenuation. (a) Time-domain traces show three distinct attenuation regimes. (b) Corresponding PDFs with Gaussian fits highlight that while noise is well described by a Gaussian model, modulated signals deviate significantly.



(a) Excess noise ξ vs. modulation variance V_A for the two attenuation settings, with linear fits and uncertainty.



(b) Receiver noise estimate ν_{rec} from MLPE compared with independent calibration, showing excellent agreement.

Fig. 3: Noise estimation using MLPE. (a) Excess noise scales linearly with V_A and matches the expected attenuation ratios. (b) Receiver noise estimate agrees with the independently calibrated value, validating the approach.

6.3 Performance

The fitted noise model allows us to predict the asymptotic secret key rate in the trusted-receiver scenario. Figure 4 shows the expected performance, with the modulation variance V_A optimized at each distance. The computation relies on the trusted excess noise model obtained by fitting the dependence of ξ on V_A, and the key rate is evaluated using the asymptotic proof of [4]. Our experimental point at 25 km is compatible with a secret key rate of 3.73 Mbps.

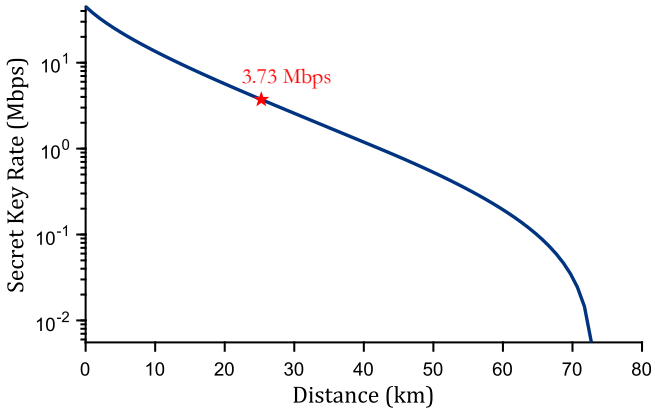


Fig. 4: Asymptotic secret key rate versus distance in the trusted-receiver scenario, with modulation variance V_A optimized at each point.

7 Conclusion and discussion

We have introduced and experimentally validated the MLPE calibration method, which enables direct estimation of receiver noise from quadrature measurements without requiring additional calibration steps. The method was shown to be robust, providing stable attenuation ratios and noise estimates consistent with conventional calibration procedures. We further demonstrated that the excess noise inferred from MLPE scales linearly with the modulation variance. This model allowed us to predict the asymptotic secret key rate with optimized modulation variance, and our experimental data at 25 km confirmed compatibility with a secret key rate of 3.73 Mbps.

These results establish MLPE as both a practical calibration tool and a means to improve throughput in prepare-and-measure CV-QKD systems, including standard implementations relying on pilot tones for phase recovery. Future work will focus on extending MLPE to the finite-size regime, where calibration accuracy and stability are critical for practical deployment.

8 Acknowledgements

This work has been supported by the European Union's Horizon Europe research and innovation program, under the project QSNP (grant agreement 101114043), by Digital Europe under the project FranceQCI (grant agreement 101091675) and by French National Research Agency under France 2030 program, project QCommTestbed (grant agreement ANR-22-PETQ-011).

Guillaume Ricard gratefully acknowledges the support of the Paris Region PhD program.

References

[1] R. Aymeric et al. "Quantum key distribution and classical communication coherent deployment with shared hardware and joint digital signal processing". In: *Proc. SPIE Int. Soc. Opt. Eng.* 12133 (2022), pp. 22–32.

[2] Raphaël Aymeric et al. "Symbiotic joint operation of quantum and classical coherent communications". en. In: *Optical Fiber Communication Conference (OFC) 2022*. San Diego, California: Optica Publishing Group, 2022, W2A.37. ISBN: 978-1-55752-466-9.

[3] Hans H. Brunner et al. "Precise Noise Calibration for CV-QKD". In: *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. ISSN: 2161-2064. July 2020, pp. 1–4.

[4] Aurélie Denys, Peter Brown, and Anthony Leverrier. "Explicit asymptotic secret key rate of continuous-variable quantum key distribution with an arbitrary modulation". en. In: *Quantum* 5 (Sept. 2021). arXiv:2103.13945 [quant-ph], p. 540. ISSN: 2521-327X.

[5] Md. Saifuddin Faruk and Seb J. Savory. "Digital Signal Processing for Coherent Transceivers Employing Multilevel Formats". en. In: *Journal of Lightwave Technology* 35.5 (Mar. 2017), pp. 1125–1141. ISSN: 0733-8724, 1558-2213.

[6] Adnan A. E. Hajomer et al. "Long-distance continuous-variable quantum key distribution over 100-km fiber with local local oscillator". In: *Science Advances* 10.1 (Jan. 2024). Publisher: American Association for the Advancement of Science, eadi9474.

[7] Paul Jouguet, Sébastien Kunz-Jacques, and Eleni Diamanti. "Preventing calibration attacks on the local oscillator in continuous-variable quantum key distribution". In: *Physical Review A* 87.6 (June 2013). Publisher: American Physical Society, p. 062313.

[8] Kazuro Kikuchi. "Fundamentals of Coherent Optical Fiber Communications". en. In: *Journal of Lightwave Technology* 34.1 (Jan. 2016), pp. 157–179. ISSN: 0733-8724, 1558-2213.

[9] Rupesh Kumar, Hao Qin, and Romain Alléaume. "Coexistence of continuous variable QKD with intense DWDM classical channels". en. In: *New Journal of Physics* 17.4 (Apr. 2015). Publisher: IOP Publishing, p. 043027. ISSN: 1367-2630.

[10] Fabian Laudenbach et al. "Pilot-assisted intradyne reception for high-speed continuous-variable quantum key distribution with true local oscillator". In: *Quantum* 3 (Oct. 2019). arXiv:1712.10242 [quant-ph], p. 193. ISSN: 2521-327X.

[11] Marco Lucamarini et al. "Implementation Security of Quantum Cryptography-Introduction, challenges, solutions| ETSI White Paper No. 27". In: (2018).

[12] Bing Qi et al. "Generating the local oscillator "locally" in continuous-variable quantum key distribution based on coherent detection". In: *Physical Review X* 5.4 (Oct. 2015). arXiv:1503.00662 [quant-ph], p. 041009. ISSN: 2160-3308.

[13] Vladyslav C. Usenko and Radim Filip. "Feasibility of continuous-variable quantum key distribution with noisy coherent states". In: *Physical Review A* 81.2 (Feb. 2010). arXiv:0904.1694 [quant-ph], p. 022318. ISSN: 1050-2947, 1094-1622.

[14] Brian P. Williams et al. "Field test of continuous-variable quantum key distribution with a true local oscillator". In: *Physical Review Applied* 21.1 (Jan. 2024). Publisher: American Physical Society, p. 014056.

[15] Shengjie Xu et al. "Counteracting a saturation attack in continuous-variable quantum key distribution using an adjustable optical filter embedded in homodyne detector". In: *Entropy* 24.3 (2022), p. 383.

[16] Yuehan Xu et al. "Simultaneous continuous-variable quantum key distribution and classical optical communication over a shared infrastructure". In: *Applied Physics Letters* 123.15 (2023).

[17] Hai Zhong et al. "Continuous-variable quantum key distribution coexisting with classical signals on few-mode fiber". EN. In: *Optics Express* 29.10 (May 2021). Publisher: Optica Publishing Group, pp. 14486–14504. ISSN: 1094-4087.

Crystalline waveguides for efficient integrated quantum memories

M. Chan[1,2,3], A. Ferrier[4,5], A. Talneau[3], P. Berger[2], L. Morvan[2], S.Welinski[2]

[1]Thales SIX GTS France, Gennevilliers, 92230, France;
[2]Thales Research and Technology, Palaiseau, 91767, France;
[3]Centre de nanosciences et de nanotechnologies, Université Paris-Saclay, CNRS, Palaiseau, 91120, France ;
[4]Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie, Paris, 75005, France ;
[5]Faculté des Sciences et Ingénierie, Sorbonne Université, UFR933, 75005 Paris, France
mickael.chan@thalesgroup.com

Abstract. This paper presents progress in the development of crystalline waveguides to enable efficient and integrated quantum memories for long-distance quantum communication. The primary aim is to tackle the difficulties presented by the weak light-matter interaction strength observed in rare-earth ion (REI) doped crystals, specifically in Yttrium Orthosilicate (YSO), which limits optical storage efficiencies. In contrast to other approaches for photonic integration, our method utilizes traditional lithography and dry-etching techniques. This approach maintains the beneficial properties of bulk crystals while offering design adaptability. We refined processes such as bonding, lapping, polishing, and dry-etching to achieve the optimal YSO waveguide structure, as shown in our simulations. This study paves the way for completely integrated quantum repeaters and also improves the integration of other quantum technologies based on REI-doped crystals, including quantum computing, single-photon generation, RF spectral analysis, and microwave-to-optical quantum transduction.
Keywords: Quantum communications, quantum memories, rare-earth-doped crystals, integrated photonics

1 Introduction

Quantum communications aims at the development of the Quantum Internet [1], establishing the secure transfer of quantum information (QI) through a global network composed of quantum computers, sensors, processors and end-users. In this architecture, photons serve as the primary carriers of information, functioning as "flying qubits" that experience minimal interference from their environment. Although fiber optics represent a natural medium for these networks, the transmission of quantum information over long distances is hampered by scattering and absorption losses inherent in fibers. Furthermore, due to the non-cloning theorem[2], the application of classical amplification methods is barred, necessitating the use of a specific protocol and system called Quantum Repeaters (QRs) to achieve long-range communication[3]. The QR protocol employs a segmented approach, connecting shorter distance links through intermediate

nodes that utilize entanglement swapping to establish communication [4]. Integral to this approach is Quantum Memory (QM), which allows the storage of entangled photonic states at these nodes while waiting to successfully establish entanglement with adjacent nodes. Research groups are increasingly focused on optimizing QM systems for various quantum applications, including distributed quantum computing and quantum key distribution. A critical parameter for effective QM [5]is the storage time, which must exceed the duration required for the entanglement swapping process. Promising candidates for QM include atomic vapors [6]and color centers in diamonds [7], but this paper's focus is on rare-earth ion (REI) doped crystals, which demonstrate significant potential for long-distance quantum repeater applications [8]. In magnetically quiet host materials like Yttrium Orthosilicate (YSO), certain REIs exhibit long optical and hyperfine coherence times at cryogenic temperatures [9]. Another distinguishing feature of REI-doped crystals for network applications is multiplexing, in particular frequency multiplexing where the same device can store multiple photons at different frequencies. This is made possible by the large inhomogeneous linewidths (typically GHz) compared to homogeneous linewidths (typically kHz), enabling the establishment of multiple memory channels within a single crystal absorption line

Recent advancements have successfully demonstrated the optical storage of single photons in REI-doped crystals [10], [11] alongside several promising network applications[12], [13]. Despite these successes, many existing systems struggle with moderate storage efficiencies due to various challenges such as low optical depth and suboptimal light-matter interactions. The weak oscillator strength of REIs limits absorption, which cannot be enhanced simply by increasing ion concentration or crystal length. Indeed, constraints imposed by spin-spin interaction processes necessitate maintaining low ion concentrations—often in the parts per million (ppm) range—to preserve coherence times. Additionally, the compromise between beam divergence and laser power density restricts overall device length. To address these challenges, integrated photonics offers a pathway to enhance light-REI interactions and push towards fully integrated quantum repeater devices[14]. Various approaches are being explored, including the utilization of $LiNbO_3$-based waveguides[15], silicon photonics[16], and femtosecond laser micromachining[17]. However, many options demonstrate degraded coherence performance or struggle to integrate complex functionalities like couplers, curvature, or cavities effectively. In this work, we introduce a novel fabrication methodology (Fig.1) for single-crystalline YSO waveguides, created from bulk crystals through conventional lithography and etching techniques. This approach aims to maintain the intrinsic properties of the bulk material while allowing for significant design flexibility. The fabrication process leverages our advancements in crystal bonding to a dedicated substrate, involves thinning YSO layers to micrometric dimensions, and employs dry etching of YSO membranes. Finite Difference Eigenmode (FDE) simulations have been carried out to identify the optimal geometries and dimensions for the REI waveguide structures.

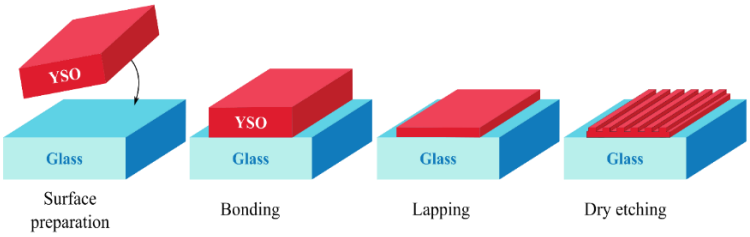


Fig. 1. Fabrication process of crystalline waveguides based on dry etching and UV lithography. Glass substrate is used as a low refractive index media.

2 Simulation considerations

The preliminary phase of waveguide simulation involves selecting an appropriate fiber-to-structure coupling method suitable for cryogenic conditions. A practical technique is direct edge coupling, utilizing UV-cured polymers to bond the optical fiber to the waveguides. Our optimization focuses on two primary criteria: modal overlap between the fundamental modes of the fiber and waveguide, and the effective indices of both. For convenience, we will denote the product of these two parameters as coupling efficiency in the subsequent analysis. Furthermore, we'll examine the spatial mode distribution and transmission losses attributed to curvature, which will provide insights into the design versatility of our fabrication approach. The device architecture is predicated on YSO bonded to borosilicate SiO_2 glass, designed in either a ridge or shallow ridge configuration. This specific glass was selected due to its thermal expansion coefficient, which is closely aligned with that of YSO. All simulations are conducted at a single operational wavelength of 1536 nm, corresponding to the ($^4I_{15/2}$-$^4I_{13/2}$) transition in site 1 of Er^{3+}: YSO. At this wavelength, considering YSO's birefringence, the refractive indices of YSO and borosilicate glass are recorded at 1.7892 and 1.4514, respectively[18]. The waveguide's cross-section is examined within the (b, D_2) plane, while the D_1 axis is oriented parallel to the light propagation direction. This particular choice of axis is constrained by the crystal growth method. The optical fiber used in the simulation is a standard SMF28 Corning fiber with a core/cladding radius of 8/125 µm.

2.1 Ridge case

YSO/SiO_2 ridges were examined first as the most straightforward approach to our objectives. As detailed in Section 3, we consider a maximum achievable etch depth of 1.5 µm. Preliminary simulations reveal that ridges with a vertical thickness (e) exceeding 1 µm always demonstrate multimode behavior regardless of width, while no viable modes are present when e is below 700 nm. Consequently, we have fixed the thickness at 1 µm to ensure single-mode operation. Figure 2a illustrates the coupling efficiency

across ridge widths ranging from 1 µm to 30 µm. Given the waveguide's reduced dimensions, an additional SiO_2 top layer is added to improve modal overlap with the optical fiber. The optimal coupling efficiency was determined to be 23% at a width of 14 µm, coupled with a top SiO_2 layer thickness of 1 µm. At these dimensions, the waveguide also exhibits multimode behavior, necessitating modal filtering through a tapered geometry to achieve single-mode conditions. This approach involves an adiabatic reduction of the width from 14 µm down to 1 µm, as explained n[19]. Despite facing challenges related to the relatively low coupling efficiency and the mechanical difficulty of thinning a YSO membrane to a thickness of 1 µm, the structure demonstrates considerable resistance to curvature. As shown in Figure 2b, with bending radii exceeding 60 µm, losses attributed solely to bending were calculated to remain below 3 dB/cm. This resistance to curvature can be easily understood considering the high differences in refractive indices.

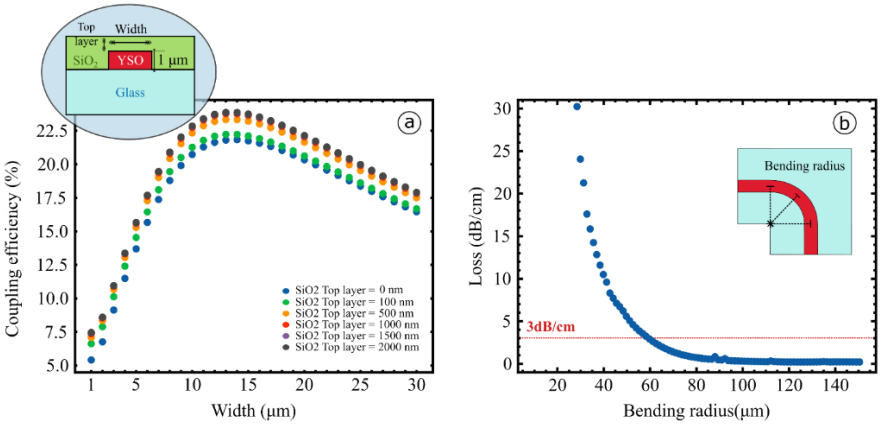


Fig. 2. a) Coupling efficiency (Modal overlap x ratio of effective indices) to ridge width in the case of YSO-SiO_2 1µm-thick ridge, with different SiO_2 top layer thicknesses with single mode fiber; b) Bending loss in the case of YSO-SiO_2 1x1 µm ridge

2.2 Shallow ridge case

Utilizing shallow ridge structures significantly lightens the manufacturing constraints associated with membrane fabrication. Our analysis indicates that maintaining a total membrane thickness of less than 10 µm, coupled with a shallow etch depth of 1.5 µm, is sufficient for the propagation of eigenmodes. As illustrated in Fig. 3a, the 2D map displays coupling efficiencies with the x-axis representing total membrane thickness (inclusive of the 1.5 µm shallow ridge depth) and the y-axis showing ridge widths ranging from 1 to 30 µm. An optimal coupling efficiency of 75% is achieved at a ridge width of 11 µm, corresponding to a total YSO layer thickness of 8 µm, which is significantly more feasible compared to previous ridge configurations. This improvement is primarily due to the enhanced total thickness, which contributes to a more symmetrical waveguide structure and increases modal overlap. Moreover, Fig. 3b confirms the waveguide's spatially monomodal nature for these dimensions, supporting only the

TE00 and TM00 modes. This monomode behavior arises from the minimal effective index differences associated with shallow etch depths, enabling the confinement and propagation of fundamental modes while effectively suppressing the existence of higher-order modes.

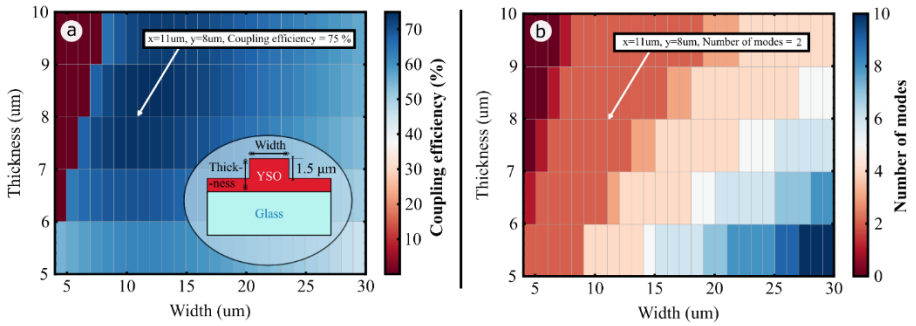


Fig. 3.: a) Coupling efficiency for YSO-SiO_2 Shallow ridge; b) Number of modes (TE & TM)

The shallow ridge design of the waveguides facilitates fabrication but significantly constrains their resistance to curvature. This limitation arises because the membrane can support membrane modes with effective indices that closely match those of the modes in the ridge section.. To address this, we opted to initially simulate the transmission characteristics of an S-bend, a commonly used configuration in integrated photonics that allows for efficient spatial separation between input and output and simplifies the injection process. Of the various S-bend designs, the Euler bend is the most prevalent, primarily due to its effectiveness in minimizing curvature losses [20].

For simulation efficiency, we developed an algorithm leveraging the commercial ANSYS Lumerical EME (Eigenmode expansion) solver. This algorithm calculates the curvature radius along the length of the waveguide and segments the device into sections where the lengths is chosen based on the rate of curvature change. The greater the curvature variation, the shorter the section. The number of segments was determined through convergence testing. Key parameters for the simulation include input-output spatial offset, etch depth, and total membrane thickness, while maintaining a waveguide width of 11 µm based on previous findings regarding coupling efficiency (Fig. 4).

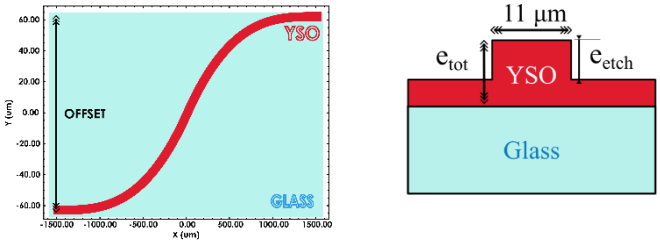


Fig.4 Simulation parameters for Euler S-bend.

Figure 5 illustrates the transmission characteristics of the S-bend as a function of input-output offset across various etch depths. We conducted simulations for total

membrane thicknesses ranging from 1 to 8 µm. For a typical offset of 125 µm with an etch depth limited to 1.5 µm, the ideal waveguide for mode matching presented in Figure 3 (8 µm total membrane thickness, 11 µm width) exhibits a transmission efficiency of less than 1%. This significant loss can be attributed to the minimal difference in effective indices between membrane modes and ridge modes, where the curvature allows mode coupling not only with waveguide modes but also with multiple membrane modes, resulting in considerable bend losses. Notably, deeper etching enhances transmission efficiency, causing the waveguide to approximate the behavior of a ridge waveguide. Similarly, reducing the membrane thickness yields similar results. Thus, a technological optimization decision becomes imperative. If fabrication constraints restrict etch depth, as we will discuss further, then achieving effective S-bend performance necessitates additional membrane thinning. Conversely, if further thinning is unfeasible, prioritizing etch depth and recipe optimization is crucial for improving transmission performance.

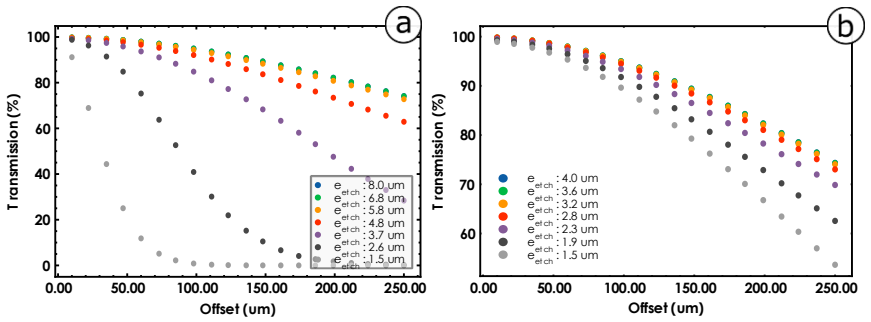


Fig.5 Euler S-bend transmission for two total membrane thicknesses (a: 8 µm; b: 4 µm) as a function of input-output offset, with varying etch depth

3 FABRICATION PROCESS: BONDING, THINNING, ETCHING

3.1 Bonding

The following paragraph addresses the bonding aspects of YSO. The initial technique explored is direct molecular bonding, as it is supposed to maximize light interaction with the REI-doped crystal and is compatible with cryogenic conditions. As demonstrated in [21], we have successfully achieved direct bonding of a commercial 10 x 5 mm YSO crystal to borosilicate glass at an annealing temperature of 520 °C. However, cracks (Fig.6a) can be observed in the crystal, which we attribute to the slight difference in YSO and glass flatness, causing increased mechanical strain on the bonded stack. Multiple temperature sweeps to 7 K indicated no significant changes in the assembly's structural integrity. While this technique has shown to be promising, it presents substantial fabrication challenges that must be addressed to ensure reliable

outcomes. Key constraints include surface cleanliness, roughness, and flatness. Specifically, achieving superior surface quality on YSO has been a persistent challenge due to its hard and brittle nature, compounded by the absence of established polishing protocols in the literature. To address this issue, we have utilized an intermediate layer of benzocyclobutene (BCB) as an intermediate bonding medium. BCB is a widely used adhesive in semiconductor applications and provides a more manageable approach. However, this intermediate layer must remain minimal in thickness (< 300 nm) to preserve the confining properties of the waveguide. Although this solution addresses the roughness concerns, it still necessitates good surface flatness, requiring at least λ/3 flatness for an 8 x 8 mm square (taking the 633 nm of HeNe laser as λ reference). This standard, while less stringent than the λ/10 flatness required for molecular bonding, still presents significant difficulty for fabrication. An example of such bonding is displayed in Fig. 6b).

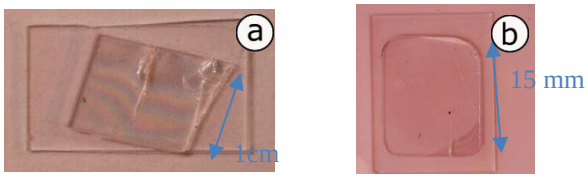


Fig. 6.: a) Direct bonding of YSO on Glass. b) BCB bonding of YSO on glass. Interference patterns observed on the lower right-hand corner suggest air pockets that lead to poor bonding. We attribute this effect to flatness differences between the YSO slab and glass

3.2 Lapping and Polishing

Results were obtained with a Logitech PM5 lapping machine, with sample flatness measured via a Fizeau interferometry system (Mahr Type FI 1040 Z). This setup enables direct monitoring of samples on the fixed polishing head, mounted on a jig. In accordance with previously established requirements, the initial polishing phase is focused on achieving optimal flatness for effective bonding. A flat, slightly convex profile is critical, particularly with 80% of the central surface exhibiting a Peak-To-Valley (PV) value below 300 nm. During the polishing process, we continuously monitored sample shape and plateau condition, adjusting parameters such as pressure, plateau spinning speed, and jig positioning every 2 to 3 minutes. As a result, we achieved a surface flatness of λ/6, as depicted in Fig. 7a. It is noteworthy that this flatness was recorded while the samples were still on the jig. After removal, an increase in the PV value was observed, likely attributed to thermal effects from the gluing process, which may have induced significant deformation in the samples' profiles (see Fig. 7b).

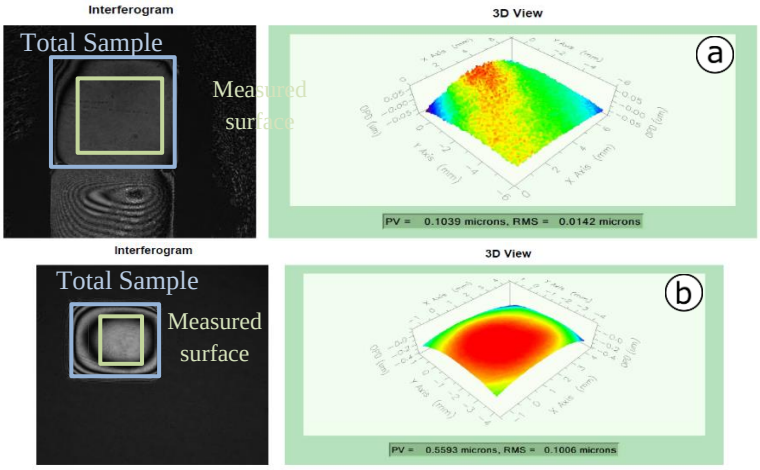
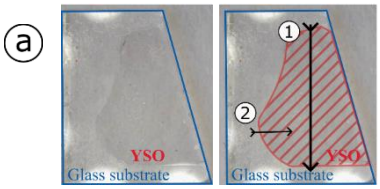


Fig. 7. Fizeau interferometry measurements of a commercial YSO sample under surface preparation polishing. a) Before jig take off PV = 0.104 μm. b) after JIG take off PV = 0.559 μm. The significant PV value difference is attributed to the heating and gluing process that could have caused material deformation. Peak-to-Valley (PV)

Following the bonding procedure, a lapping process is undertaken to achieve micrometer-level thickness in the crystal membrane. The development of this protocol is beneficial not only for applications involving ridges and shallow ridges of YSO but also essential for single rare-earth ion detection, where a micrometer-thin membrane of YSO is utilized within a Fabry-Perot cavity [22]. The lapping process is performed on a cast iron plateau using progressively finer Al_2O_3 grain size solutions. Throughout this procedure, the thickness is continuously monitored using a precision indicator, with final verifications conducted through a Thetametrics white light reflectance spectrometer for micrometer thickness samples and a Dektak XT. Stylus mechanical profilometer for slightly thicker ones. As illustrated in Fig. 8, we achieved an 8 μm-thick membrane that maintained commendable surface homogeneity. Notably, the membrane exhibited no cracks or displacements during the lapping process, with the exception of initially unbonded regions, as shown by the preliminary interference pattern (Fig. 8b). This highlights the critical role of the bonding process in the overall fabrication protocol, as it has a direct impact on the final flatness and effective surface area of the membrane.



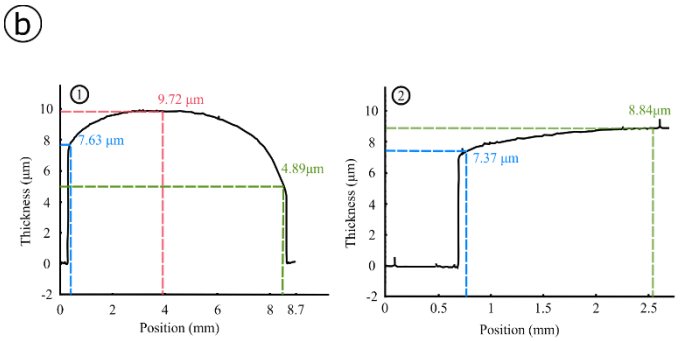


Fig. 8. a) Lapped YSO membrane from the bonded assembly in Fig. 4b) with a delimitation of YSO and glass on the right picture; b) Two stylus profilometry following paths 1 and 2 are shown as an example, indicating a convex shape and a mean thickness of around 8 µm.

3.3 Etching

After completing the polishing steps, we will present our findings regarding the ICP (Inductively Coupled Plasma) Reactive Ion Etching (RIE) of YSO, with further details available in reference [23]. The etching mechanism involves both physical etching, driven by accelerated Argon (Ar) ions, and chemical etching relying on the reactive plasma environment to enhance material displacement and etch rates. The interplay between these two contributions plays a critical role in defining the quality, morphology of the etched surface, etch rate, and selectivity. ICP-RIE is a well-established etching method utilized extensively for various oxides and glasses. Notably, JB Bradley's research [24] evaluated a spectrum of chemistries, including Fluorine, Chlorine, and Bromide-based systems, particularly focusing on Al_2O_3 and Y_2O_3. This study highlighted Chlorine-based etching as superior, achieving the highest etch rates for Y_2O_3, thereby establishing a solid foundation for defining etching parameters for YSO. Several ICP chlorine-based etching processes have been investigated to fabricate shallow ridge waveguides in YSO. The plasma compositions explored include Cl_2/Ar, Cl_2/N_2, and Ar-only configurations. This variation allows for a thorough examination of the underlying etching mechanisms. Dielectric masks, such as SiO_2 or SiN, were chosen based on their compatibility with the respective plasma compositions. Experimental results indicate that the etching process is predominantly ion-driven. The Ar-only plasma configuration produced the highest YSO etch rate at 80 nm/min, with a selectivity ratio of 0.5 using a SiN mask under high ICP power (800 W) and a considerable RF bias (350 V); however, this setup led to trenching issues. A decrease in RF bias to -100 V reduced trenching significantly, although it also resulted in a lower etch rate, but increased selectivity, both of which can be attributed to the reduced energy of the Ar ions. In contrast, both Cl_2/Ar and Cl_2/N_2 plasma configurations, utilizing chlorine ions, yielded lower etching rates and selectivities against a SiO_2 mask but resulted in smoother sidewalls and eliminated trenching. SEM images of the etched YSO under these conditions are shown in Fig. 9. Practically, the etch depth achievable for YSO is constrained by the thickness of the dielectric mask; given the selectivity values obtained, a 3 µm mask thickness is required to etch down to 1.5 µm of YSO. Utilizing higher dielectric thicknesses would necessitate stricter lithography requirements to sustain vertical etch walls. With the current etching parameters, achieving a 1.5 µm YSO etch depth with vertical walls for an

11-µm wide shallow ridge geometry, as discussed in Section 2, remains feasible.

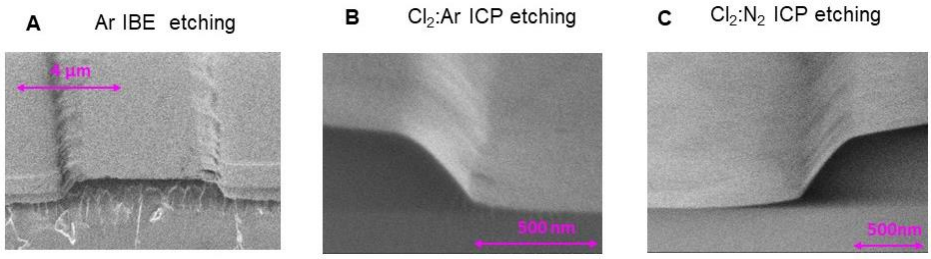


Fig. 9. SEM images of etched YSO with different plasma compositions. Data and discussions will be presented in detail in [23]

4 Conclusion

REI-doped materials are emerging as strong candidates for quantum repeaters and memory applications due to their long optical and spin coherence times, especially when doped into low magnetic noise matrices, such as YSO. Nevertheless, the intrinsic weak light-matter interaction in bulk samples significantly limits optical storage efficiency. This can be addressed through integrated photonics, which allows for the confinement of light within a reduced modal volume. In this study, we detailed improvements to our dry etching fabrication process, highlighting key procedures such as bonding, lapping, and etching. These advancements pave the way for the reproducible and scalable fabrication of YSO waveguides directly from bulk substrates, ensuring the retention of their unique spectroscopic features while offering design versatility. Future investigations will focus on performing waveguide characterization under cryogenic conditions. Additionally, we will undertake spectroscopic measurements to determine any potential degradation in ion characteristics and investigate light-matter interactions, possibly through Rabi frequency measurements, to evaluate enhancements in efficiency and overall performance.

References

[1] H. J. Kimble, "The quantum internet," *Nature*, vol. 453 (7198), 1023–1030, (2008)

[2] D. Dieks, "Communication by EPR devices," *Physics Letters A*, vol. 92 (6), 271–272 (1982)

[3] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, "Quantum repeaters based on atomic ensembles and linear optics," *Rev. Mod. Phys.* 83 (1), 33-80 (2011).

[4] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, "Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication," *Phys Rev Lett* 81 (26), 5932–5935 (1998)

[5] P. Goldner, A. Ferrier, and O. Guillot-Noël, "Rare Earth-Doped Crystals for Quantum Information Processing," in *Handbook on the Physics and Chemistry of Rare Earths*, vol. 46, Elsevier (2015)

[6] M. Hosseini, B. M. Sparkes, G. Campbell, P. K. Lam, and B. C. Buchler, "High efficiency coherent optical memory with warm rubidium vapour," *Nature Communications*, vol. 2(1) ,174 (2011)
[7] E. Togan *et al.*, "Quantum entanglement between an optical photon and a solid-state spin qubit," *Nature*, vol. 466 (7307), 730–734 (2010).
[8] T. Böttger, C. W. Thiel, R. L. Cone, and Y. Sun, "Effects of magnetic field orientation on optical decoherence in Y_2SiO_5," *Phys. Rev. B*, vol. 79(11), 115104, (2009)
[9] M. Zhong *et al.*, "Optically addressable nuclear spins in a solid with a six-hour coherence time," *Nature*, vol. 517 (7533), 177–180 (2015).
[10] A. Ortu, A. Holzäpfel, J. Etesse, and M. Afzelius, "Storage of photonic time-bin qubits for up to 20 ms in a rare-earth doped crystal," *npj Quantum Information*, vol. 8(1), 29 (2022)
[11] J. V. Rakonjac, D. Lago-Rivera, A. Seri, M. Mazzera, S. Grandi, and H. de Riedmatten, "Entanglement between a Telecom Photon and an On-Demand Multimode Solid-State Quantum Memory," *Phys. Rev. Lett.*, vol. 127(21), 210502 (2021).
[12] J. V. Rakonjac, S. Grandi, S. Wengerowsky, D. Lago-Rivera, F. Appas, and H. de Riedmatten, "Transmission of light-matter entanglement over a metropolitan network," *Optica quantum* vol 1(2), 94-102 (2023)
[13] T.-X. Zhu *et al.*, "A Metropolitan-scale Multiplexed Quantum Repeater with Bell Nonlocality," *arXiv preprint arXiv:2508.17940*, 2025.
[14] E. Saglamyurek *et al.*, "A multiplexed light-matter interface for fibre-based quantum networks," *Nature Communications*, vol. 7(1) (2016)
[15] T. Chakraborty *et al.*, "Towards a spectrally multiplexed quantum repeater," *npj Quantum Information*, vol. 11(1), 3 (2025)
[16] I. Craiciu, M. Lei, J. Rochman, J. G. Bartholomew, and A. Faraon, "Multifunctional on-chip storage at telecommunication wavelength for quantum networks," *Optica*, vol. 8 (1),114–121 (2021)
[17] T.-X. Zhu *et al.*, "Integrated spin-wave quantum memory," *National Science Review*, vol. 11(11), nwae161 (2024)
[18] T. Böttger, Y. Sun, C. W. Thiel, and R. L. Cone, "Spectroscopy and dynamics of Y_2SiO_5 at 1.5μm," *Phys. Rev. B*, vol. 74(7), 075107 (2006)
[19] J. Van Asch *et al.*, "A methodical approach to design adiabatic waveguide couplers for heterogeneous integrated photonics," *Journal of Physics: Photonics*, vol. 6(2024)
[20] L. Zhang *et al.*, "Low-loss, ultracompact n-adjustable waveguide bends for photonic integrated circuits," *Opt. Express*, vol. 31(2), 2792–2806 (2023)
[21] A. Talneau, S. Welinski, and A. Ferrier, "Rare-Earth-Doped Y2SiO5 Crystal Directly Bonded on Glass for Efficient Optical Quantum Technologies Platform," *physica status solidi (a)*, 221: 2300718 (2024)
[22] A. Ulanowski, J. Früh, F. Salamon, A. Holzäpfel, and A. Reiserer, "Spectral Multiplexing of Rare-Earth Emitters in a Co-Doped Crystalline Membrane," *Advanced Optical Materials*, vol. 12(15), 2302897 (2024)
[23] A. Talneau *et al.*, "High surface quality Y2SiO5 silicate-crystal waveguides etched by chlorine-based inductive coupled plasma reactive ion etching," *Journal of Vacuum Science & Technology A*, vol. 43(4), 043002 (2025)

[24] J. Bradley, F. Ay, K. Worhoff, and M. Pollnau, “Fabrication of low-loss channel waveguides in Al2O3 and Y2O3 layers by inductively coupled plasma reactive ion etching,” *Applied Physics B*, vol. 89, 311–318 (2007).

Entanglement routing in large-scale quantum networks with ground-space interconnection

Agathe Blaise[1⋆], Luca Paccard[2], Fabrice Arnal[2], Laurent de Forges de Parny[2], and Stéphanie Molin[1]

[1] Thales SIX GTS, 4 avenue des Louvresses, 92230 Gennevilliers, France
[2] Thales Alenia Space, 26 avenue Jean-François Champollion - BP 33787, 31037 Toulouse-Cedex 1, France

Abstract. Quantum Information networks (QIN) aim to interconnect quantum systems via quantum repeaters through end-to-end entanglement of qubits. The integration of both satellites and ground segments is essential to provide global connectivity, even over large distances. In this paper, we study the problem of entanglement routing in large-scale quantum networks – i.e., with a large number of nodes, with a hybrid space-ground architecture. We decompose the routing problem into two subcomponents: request scheduling and path selection, and structure our approach around two phases: a proactive and a reactive step. We design and implement an entanglement routing strategy in an emulated European quantum network, first in a custom Python-based simulator for performance forecasting, and then into the NetSquid quantum network simulator. Our work thus provides a tool allowing for a comparison between terrestrial and satellite segments in terms of successful end-to-end entanglement in a large-scale network and shows that both segments are required for optimal service. Our results indicate that, within the emulated European topology, terrestrial networks can support end-to-end entanglement for approximately 35% of user pairs — the remaining pairs are too distant and suffer from prohibitive losses. In contrast, the satellite-based network achieves connectivity for up to 91% of user pairs, although some pairs remain unreachable due to a lack of simultaneous satellite visibility. However, while satellites can outperform terrestrial networks in total photon distribution, they may not be suitable for low-latency requests if out of visibility, making ground networks essential. This highlights the complementary roles of space and terrestrial segments in quantum networking.

Keywords: QIN · Quantum repeaters · Entanglement routing

1 Introduction

The future Quantum Internet will enable quantum communication between local quantum processors located anywhere on Earth. Such an infrastructure will support a wide range of applications, including secure communications, distributed quantum computing, and high-precision clock synchronization.

⋆ Agathe Blaise and Luca Paccard contributed equally to this work.

The fundamental resource of a QIN is *entanglement distribution*. One of its key building blocks is quantum repeater, which allows the transfer of quantum states over long distances through *entanglement swapping* – as optical amplification is prohibited in QIN. We consider the establishment of entanglement between two distant quantum nodes – A and C – via an intermediate repeater node, B. Quantum sources generate entangled Einstein-Podolsky-Rosen (EPR) pairs locally and send them to both the repeater and the endpoint nodes. Upon receiving one entangled particle from each source, B performs a projective measurement and transmits appropriate instructions to C over a conventional communication channel. C can then apply a corrective operation on her qubit, which creates an end-to-end (E2E) entanglement with the qubit received by A. The entanglement done locally at every repeater for every link toward an adjacent node can fail with a non-negligible probability, but it is heralded by some technology means such that the outcome is known to the repeater itself. We then assume that there is a network element that collects all these outcomes and, therefore, has perfect knowledge of which links between quantum computers can be used to transfer the status of qubits. Thus, the controller tries to satisfy a set of demands from the upper layer applications by performing an ephemeral **routing** of multiple requests, which is only valid until the next round of local link entanglements.

The integration of **satellite segments** with terrestrial networks aims to provide global connectivity, even over large distances. A global quantum Internet will therefore consist of both space segments, with satellites serving ground stations, and terrestrial segments. Satellites are expected to operate in Low or Medium Earth Orbit (LEO or MEO), as these altitudes allow for a better photon reception rate on the ground.

These systems would be composed of multiple satellites organized in constellations, to enable the deployment of terminals anywhere within the coverage area and move toward near-real-time service.

In this work, we focus on the problem of **entanglement routing** in large-scale quantum networks that include both satellite and terrestrial components. We provide a tool allowing for a comparison between terrestrial and satellite segments in terms of successful E2E entanglement and show that both segments are required for optimal service. We also propose and evaluate several strategies that address request scheduling and path selection. The remainder of the paper is organized as follows. §2 introduces the related work on entanglement routing and large-scale quantum networks. In §3, we present the architecture of quantum networks and detail our methodologies for entanglement routing in large-scale quantum networks. §4 presents our results obtained via extensive simulations. Finally, §5 concludes the paper.

2 Related work

Many studies have addressed the problem of entanglement routing in QIN. Some challenges are similar to those found in classical networks, such as handling

multiple simultaneous connection requests between different user pairs, request scheduling and resource allocation, multi-path management, network reconfiguration, and link capacity estimation. However, there are also several challenges arise due to the specific to the nature of quantum networks, such as: 1) the stochastic nature of links, due to the low success probability of entanglement distribution, 2) the very short lifetime of quantum memories, 3) the handling of concurrent entanglement requests, which can lead to congestion on quantum links as each link can only handle one request at a time unlike packet switching in classical networks. Several major strategies for entanglement routing have emerged in the literature.

The first major distinction is between **proactive** (continuous) and **reactive** (on-demand) routing strategies. Proactive strategies involve continuously generating entangled pairs between neighboring nodes, allowing fast E2E entanglement establishment through entanglement swapping once a request arrives. Reactive strategies, by contrast, initiate entanglement generation only upon receiving a request. They begin by selecting an appropriate path, then sequentially establish entanglement between neighboring nodes and perform entanglement swapping as needed – reducing idle occupancy and decoherence risks.

A second distinction involves **centralized** vs. **distributed** routing approaches. Centralized approaches [1, 13, 10, 5, 8] rely on global knowledge of the network and use classical algorithms like Dijkstra to compute paths based on physical characteristics (link transmittance, EPR pair generation rate). These assume either each node or a controller maintains up-to-date knowledge of the entire network state and produces globally optimal routing decisions, maximizing network throughput and minimizing latency. On the other hand, distributed approaches [2, 11] operate using only local information shared among neighboring nodes, making them more suitable for dynamic and large-scale networks.

Another critical aspect of routing is the management of **concurrent** entanglement requests, which can compete for limited quantum resources (e.g., memories or channels). For instance, [7] proposes a centralized scheduler that determines paths, orchestrates request timing, and disseminates decisions across the network. Finally, **multi-path** routing [6, 8, 11] supports robustness and load balancing by precomputing multiple routes for a single user pair. If one route becomes unavailable, due to an entanglement swap failure or expired links, alternative paths can be quickly utilized.

New challenges to the entanglement routing problem arise with the integration of satellites into large-scale QIN: time-varying topologies due to satellite motion, atmospheric losses and weather sensitivity affecting link reliability, and short-time communication windows. The work in [4] defines a high-level architecture of a generic QIN, then focuses on the architecture of the space segment.

Specific characteristics must be considered when selecting a routing algorithm. Proactive strategies can be particularly useful for atmospheric links, which are unstable and affected by external factors such as weather. Efficient routing in networks combining terrestrial and satellite components requires hybrid-aware algorithms that can exploit satellites for long-distance connectivity. This entails

dynamic path planning based on factors such as inter-city traffic demand, meteorological conditions, and dynamic satellite visibility maps. Leveraging satellite links in this way can relieve congestion in terrestrial networks and enable global entanglement distribution.

3 Large-scale entanglement routing

3.1 Quantum network architecture

Fig. 1 illustrates the reference architecture of the global quantum Internet network under study, composed of multiple metropolitan networks interconnected via both space segments (satellites serving ground stations) and terrestrial segments (when the physical distance between access stations is reasonable, i.e., less than a hundred kilometers). The satellites are equipped with an entangled photon source and handle the distribution of the photon pairs to the ground – we do not consider Inter-Satellite Links in this scenario.

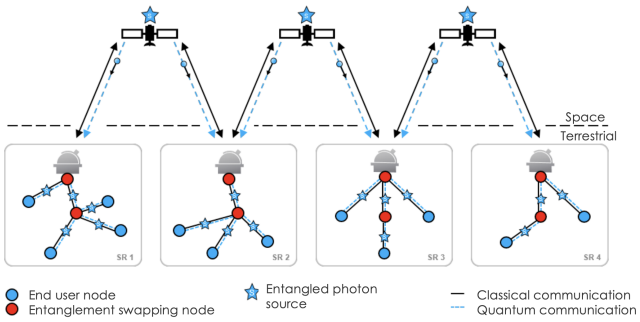


Fig. 1: Reference architecture.

3.2 Problem definition

Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a static graph physical defining the network so that:

- Each node in $\mathcal{V}$ (with $N = |\mathcal{V}|$) hosts a quantum repeater;
- Each edge in $\mathcal{E}$ (with $M = |\mathcal{E}|$) is a lossy optical channel of length L_e (km) and power transmittivity $\eta_e = e^{-\alpha L_e}$, where α depends on the propagation media of the quantum channel. Each supports $S(e) \in \mathbb{Z}^+$ parallel channels(e.g., spatial, spectral, polarization modes);
- The network serves K source-destination pairs (A_j, B_j), for $1 \leq j \leq K$, with $A_j, B_j \in \mathcal{V}$.

The time is slotted, with the assumption that each memory can hold a qubit perfectly for $T \geq 1$ time slots. At the link level, entanglement generation is

attempted repeatedly over the $S(e)$ parallel channels available on each edge e. Each attempt succeeds with probability $p_0(e) = e^{-\alpha L_e}$.

Two probabilities are defined throughout the whole process:

- Probability p (related to the external phase) that is the success probability of local entanglement between two quantum repeaters;
- Probability q (related to the internal phase) that is the success probability of entanglement swapping (internal-link attempts).

4 Simulations and experiments

In this section, we describe the continental European quantum network we simulated and the results of our forecast and operational simulations. We note that it sometimes rely on simplifying assumptions and parameters, like the lifetime and heralding capabilities of quantum memories. Additionally, our geographic model is limited to continental Europe: we do not include overseas territories not to introduce too much complexity to the problem.

In this work, we focus on polarization encoding, which remains one of the most widely used schemes in free-space quantum communication. Polarization qubits are straightforward to generate and measure, and they have already been successfully employed in satellite-based experiments. While alternative encodings such as time-bin or orbital angular momentum (OAM) may offer advantages in certain scenarios, polarization is currently the most practical choice for long-distance free-space links. We note, however, that polarization encoding is also subject to logical errors due to atmospheric turbulence and misalignment, which must be compensated in a full implementation.

4.1 Emulated network

We define a hypothetical European network composed of 84 cities, including capitals and big cities of Western European countries. Each city is connected to its three nearest neighbors. This results in the interconnection graph shown in Fig. 2. Two approaches are considered to connect two end users located in different cities:

1. **Terrestrial**: Entanglement is distributed between the end users by relaying through other cities, using a hop-by-hop approach. The selection of intermediate nodes is made using Dijkstra to determine the shortest path, taking into account network constraints.
2. **Satellite**: Another means of connecting users is through a satellite. We focus on low Earth orbit (LEO) satellites, as they offer shorter propagation distances compared to medium (MEO) and geostationary (GEO) satellites, reducing transmission losses and increasing the photon detection probability.

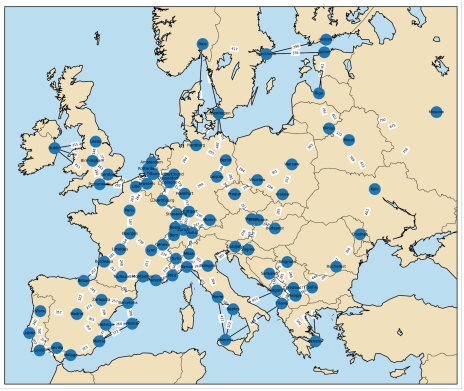


Fig. 2: Interconnection graph of a possible European quantum network.

4.2 Forecast performance simulation

The objective is to study and compare the predictive performance of entanglement distribution on a continental Europe scale using a terrestrial network or a satellite link. The aim is to determine macroscopically the most efficient path to connect two users located in Europe.

Elementary link We model the quantum network as a chain of elementary links. An elementary link is composed of an entangled photon source (EPS), two propagation channels, and two quantum memories with N storage slots (or modes) each. Quantum memories are assumed to be heralded, meaning that writing a quantum state onto the memory outputs a signal indicating that the memory is storing a quantum state. We assume multimode quantum memories with a capacity of up to 50 *temporal* storage modes. Temporal multiplexing is particularly relevant for quantum repeater architectures, as it allows multiple entanglement attempts within the coherence time of the memory. Other types of multiplexing, such as spatial or orbital angular momentum modes, are more vulnerable to degradation during long-distance free-space propagation and are therefore not considered in this work.

The elementary links are connected through a Bell State Measurement (BSM) module as highlighted in Fig. 3.

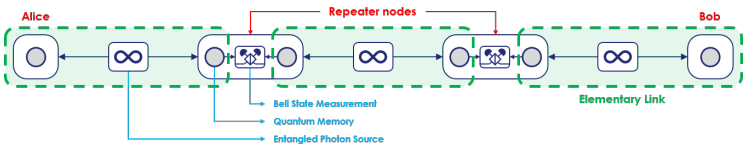


Fig. 3: Chain of quantum repeaters.

We aim to take advantage of the multi-modal capability of the quantum memories, as they reduce the impact of losses in the propagation channels. In a chain of two elementary links (Fig. 4), if a photon is lost between a source and a node on link A, but at the same time another photon is received from another pair, then the latter can be stored in a memory until a photon from another pair of link B is received. We can individually model each elementary link and connect them through a BSM. Our objective is then to determine the probability of successfully transmitting at least one pair of entangled photons between the two nodes after N time slots.

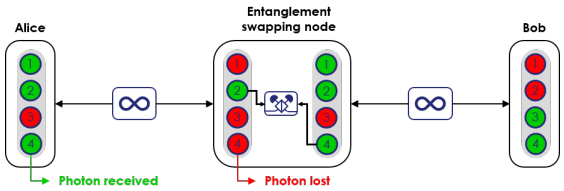


Fig. 4: Two elementary links of a chain of quantum repeaters.

We find that the probability of successfully transmitting at least one EPR pair between the two end-user nodes, on a chain of M quantum repeaters is:

$$p^{chain}_{(success\geq 1)} = \frac{\left[1 - \prod_{t=1}^{N}(1 - \eta_{src} \cdot \eta_{ch}^2 \cdot \eta_{QM}^2)\right]^{M+1} \times \left[\eta_{BSM} \cdot \eta_{det}^2\right]^{M}}{N} \quad (1)$$

The development of the model can be found in §A.

Performances of terrestrial and space paths We compare the performance of the entanglement distribution between end-users with a terrestrial path and a satellite path. For the terrestrial path, we consider the network proposed in Fig. 2 and connect the end-users with the different nodes in the network using a shortest path algorithm. For the space path, we consider a LEO satellite with an altitude of 800 km and an inclination of 60°. We run a 4-month simulation using STK (Systems Tool Kit) and compute the dual link budget between the satellite and the pairs of ground users. The total dual visibility is also considered to take into account the availability of the satellite during the simulation period. Details of these data can be found in §B.

Fig. 5 illustrates the number of EPR pairs distributed over the simulation period of 4 months. Each point corresponds to a pair of users. For the ground network, the link budget between end-users is computed using Eq. 7. To obtain the number of pairs distributed over 4 months, we consider that the network is running 24 hours a day and 7 days a week during the period.

The satellite link is considered as an elementary link with a source (the satellite) and two quantum memories (the end-users). Its performance is thus computed using Eq. 5. We observe that a satellite link becomes more interesting than the ground network when the distance between the end-users exceeds 400 km. For lower distances, the ground network is naturally more effective due to lower losses and greater availability of the service. However, even when considering its limited availability, the satellite allows the distribution of a greater number of EPR pairs than the ground network. Beyond a limit distance of around 2200 km, we observe a cut-off in the number of entangled photons pairs distributed by the satellite because, beyond this distance, two nodes cannot be simultaneously visible to the satellite according to the parameters chosen for the simulation (altitude, minimum elevation, etc).

We note that these results depend on the assumptions and parameters used in the simulation. Our simulations are designed to capture the global behavior of entanglement distribution over the two network segments, enabling qualitative analysis of routing strategies and architectural trade-offs.

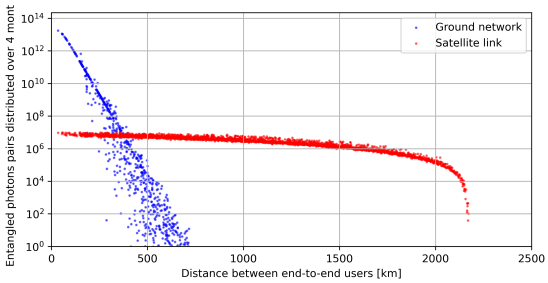


Fig. 5: Number of EPR pairs distributed over 4 months.

Satellite, while outperforming ground networks for distant users, has a significant drawback in terms of availability. Users needing low-latency entanglement resources might find the satellite on the opposite side of the planet, making terrestrial networks the only viable option.

For satellite-based quantum communication, additional effects such as Doppler shifts and imperfect mode matching must be taken into account. Doppler shifts can lead to temporal misalignment between entangled photons, while mode mismatch reduces interference visibility and therefore the overall success probability of entanglement swapping. Both effects can be modeled as an effective reduction in detection efficiency, meaning that the optimistic 99% efficiency assumed in our baseline analysis should be interpreted as an upper bound.

4.3 Operational performance simulation

The objective is to evaluate the routing performance as implemented in the NetSquid simulator. NetSquid [9] is a Python-based library designed to simulate

a wide range of components in a quantum network, with a strong emphasis on real-time network behavior simulation.

Implementation We developed a simulation of E2E entanglement routing in a quantum network using NetSquid. Our work builds upon a source code implementing routing [3], extends it with space (satellites, ground stations) and ground (metropolitan nodes) components, and implementing the topology of a European network. Quantum links encompass both fiber and free-space links. Each quantum node also contains a quantum processor (e.g., for entanglement swapping) and a quantum memory. More details can be found in §C.

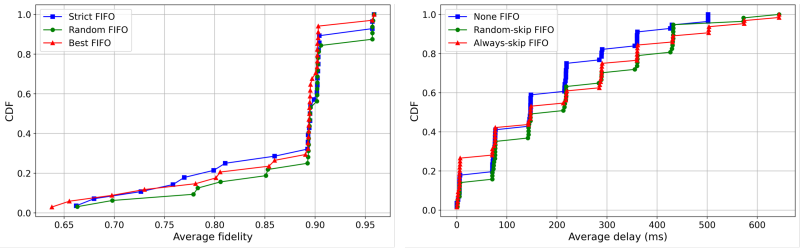
The selected path for connecting two stations is computed with the shortest path that maximizes the E2E available capacity, as described in the previous simulation. After finding the shortest path for each pair, we must determine which pairs will be processed and in what order, in cases where not all pairs can be handled at the current time slot. We implemented and compared several scheduling policies:

- **FIFO**: Pairs are processed in order of arrival without considering path costs, ensuring strong fairness.
- **Best Efficiency**: A path is skipped if its cost exceeds the average cost of all candidate paths, ensuring overall efficiency.
- **Weighted FIFO**: A path is skipped with a probability $P_{skip} = max\{0, 1 - \frac{\sigma}{E[c]}\}$, with σ the path cost and $E[c]$ the average path cost over all paths.

For each simulation, we select random pairs of cities to connect over T time slots. The tuning of parameters in the simulation is described in D.

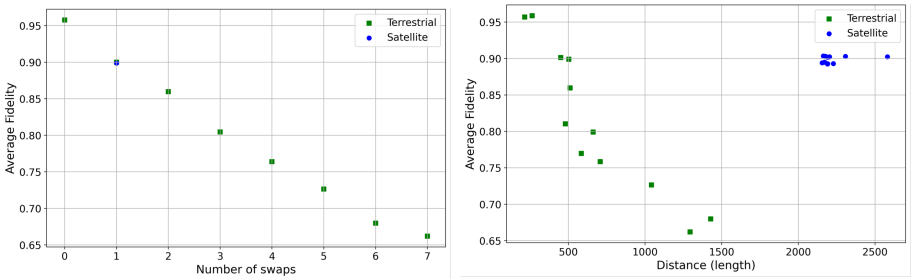
Fig. 6a illustrates the cumulative distribution function (CDF) of the average fidelities for the three selected scheduling policies. On average, higher fidelity is achieved with Best Efficiency and Weighted FIFO. This is because, by avoiding the processing of requests whose path cost is significantly above the average, network resources are used more efficiently, favoring entangled pairs whose nodes are not too distant. Fig. 6b shows the CDF of the average delays for the three scheduling strategies. Both Weighted FIFO and Best Efficiency outperform FIFO, as they achieve lower average E2E entanglement distribution latency. However, we note that Best Efficiency and Weighted FIFO tend to favor shorter paths, which penalize other requests systematically.

We compare the average fidelity depending on the number of intermediate swaps in Fig. 7a, to evaluate the performance of the space and ground segments. We see that the configuration with two metropolitan nodes connected by a fiber link yields the highest fidelity. When a single intermediate swap is required, using either a satellite or a ground station results in roughly equivalent fidelity. However, beyond one swap, the fidelity is consistently higher using the satellite. Fig. 7b further shows the fidelity as a function of the total distance traversed. We observe that the distribution of photons via the space segment results in higher fidelity, even though the physical distance is significantly greater.



(a) CDF of the average fidelity. (b) CDF of the average delay.

Fig. 6: CDF for E2E routing between 2 nodes for different scheduling strategies.



(a) Depending on the number of swaps. (b) Depending on the distance (kms).

Fig. 7: Fidelity values for the terrestrial and satellite scenarios.

5 Conclusion

This paper investigated the challenges and potential of a QIN composed of both terrestrial and satellite-based components. We focused on the problem of establishing E2E entanglement between distant quantum nodes. We modeled an emulated European-scale quantum network, incorporating optimistic physical parameters for quantum memory, decoherence, link loss, and processing delays. Our simulations explored both forecast and real-time operational scenarios, allowing us to evaluate the performance of different routing mechanisms under various network conditions. We implemented our routing strategies within a simulation framework built on NetSquid, extended with a space-ground model including both fiber and free-space quantum links. Our results highlight the complementary roles of space and terrestrial segments in quantum networking.

References

1. Caleffi, M.: Optimal routing for quantum networks. IEEE Access **5** (2017)
2. Chakraborty, K., et al.: Distributed routing in a quantum internet (2019)
3. Claudio Cicconetti: Quantum networking experiments with NetSquid. https://github.com/ccicconetti/netsquid (2025)

4. de Forges, L., et al.: Satellite-based quantum information networks: Use cases, architecture, and roadmap (2022)
5. Gyongyosi, L., Imre, S.: Entanglement-gradient routing for quantum networks. Scientific Reports **7** (2017)
6. Kozlowski, W., et al.: Designing a quantum network protocol. In: CoNEXT. ACM (2020)
7. Li, C., et al.: Effective routing design for remote entanglement generation on quantum networks. npj Quantum Information **7** (2021)
8. Pant, M., et al.: Routing entanglement in the quantum internet. npj Quantum Information **5** (2017)
9. QuTech: NetSquid: The Network Simulator for Quantum Information using Discrete Events. https://netsquid.org/ (2025)
10. Schoute, E., et al.: Shortcuts to quantum network routing (2016)
11. Shi, S., et al.: Concurrent entanglement routing for quantum networks: Model and designs. IEEE/ACM Transactions on Networking **PP** (2024)
12. Shibata, H., et al.: Detecting telecom single photons with 99.5% system detection efficiency. APL Photonics **6**(3), 036114 (2021)
13. Van Meter, R., et al.: Path selection for quantum repeater networks. Networking Science **3** (2012)

A Model for forecast performance simulation

In this appendix, we will develop the model for the forecast performance simulations presented in Section 4.2.

The parameters involved are the following:

- η_{src}: efficiency of the EPS;
- η_{ch}: efficiency of the propagation channel (either free-space or optical fiber);
- η_{QM}: efficiency of the quantum memory (we include both writing and reading efficiencies in this parameter);
- N: number of storage modes of the quantum memory (i.e., multiplexing capabilities in the quantum memory).

The probability of failing to transmit entanglement between two nodes at time t is:

$$p_{fail}(t) = 1 - \eta_{src} \cdot \eta_{ch}^2 \cdot \eta_{QM}^2 \tag{2}$$

For the description of the model, the channel efficiencies will be considered equal, but could be tuned to reflect the real losses of the network.

The probability of failing to transmit entanglement between two nodes after N time slots is then:

$$p_{fail}^{all} = \prod_{t=1}^{N} (1 - \eta_{src} \cdot \eta_{ch}^2 \cdot \eta_{QM}^2) \tag{3}$$

Thus, the probability of transmitting at least one pair of entangled photons on an elementary link after N time slots is:

$$p^{N}_{(success\geq 1)} = 1 - \prod_{t=1}^{N}(1 - \eta_{src} \cdot \eta^2_{ch} \cdot \eta^2_{QM}) \tag{4}$$

Generally speaking, when we consider a link budget, we look at the losses or reception probabilities for a photon (or a pair of photons in our case) on a single time slot. Here we need to divide the probability in Eq. 4 by the number of time slots considered to be able to compare the performances of links with quantum memories of various storage modes:

$$p_{(success\geq 1)} = \frac{1 - \prod_{t=1}^{N}(1 - \eta_{src} \cdot \eta^2_{ch} \cdot \eta^2_{QM})}{N} \tag{5}$$

A chain of quantum repeaters is simply a succession of elementary links, to which BSM modules are added as shown in Fig. 3. A BSM is characterized by two elements:

- η_{BSM}: intricate probability of success of the BSM (i.e., the projected state is either $|\psi^+\rangle$ or $|\psi^-\rangle$);
- η_{det}: efficiency of the single photon detectors in the BSM module (two detectors must click during a successful BSM).

Thus, entanglement swapping has the following efficiency:

$$\eta_{swapping} = \eta_{BSM} \cdot \eta^2_{det} \tag{6}$$

From Fig. 3, we see that in a repeater chain composed of two repeater nodes (i.e.,nodes performing the BSM), there are three elementary links. By generalizing to M repeater nodes, we count:

- M repeater nodes;
- $M + 1$ elementary links

From Eq. 5 and Eq. 6, the probability of successfully transmitting at least one pair of entangled photons between the two end-user nodes, on a chain of M quantum repeaters is:

$$p^{chain}_{(success\geq 1)} = \frac{\left[1 - \prod_{t=1}^{N}(1 - \eta_{src} \cdot \eta^2_{ch} \cdot \eta^2_{QM})\right]^{M+1} \times \left[\eta_{BSM} \cdot \eta^2_{det}\right]^{M}}{N} \tag{7}$$

B Parameters for forecast performance simulations

Tab. 1 presents the parameters of the forecast performance simulation. Some are based on optimistic assumptions, as they reflect the expected performance of future hardware rather than the current state of the art. Recent advances

in superconducting nanowire single-photon detectors (SNSPDs) have demonstrated system detection efficiencies as high as 99.5% at telecom wavelengths [12]. While such performance is not yet standard in laboratory environments, we use this value as an upper-bound benchmark to assess the ultimate performance of satellite-based quantum communication scenarios.

Parameter	**Symbol**	**Value**
Wavelength	λ	1550 nm
Source rate	R_{src}	100 MHz
Source efficiency	η_{src}	0.25
Quantum memory efficiency	η_{QM}	0.9
Number of storage modes per quantum memory	N	50
BSM efficiency	η_{BSM}	0.5
Detector efficiency	η_{det}	0.99
Fiber attenuation factor	α	0.2 dB/km
Satellite telescopes diameter	D_{TX}	50 cm
Ground station telescope diameter	D_{RX}	80 cm

Table 1: List of the parameters of the forecast performance simulation.

Tab. 2 illustrates the satellite data for the pairs of users Paris-Toulouse and Madrid-Amsterdam. The number of EPR pairs distributed is lower for Madrid-Amsterdam because the two users are further apart, hence reducing the mean elevation angle and increasing the mean slant range, as well as reducing the total duration of dual visibility.

User pair	**Distance** (km)	**Mean slant range** (km)	**Mean elevation angle** (°)	**Dual visibility** [hh:mm:ss]
Paris	588	1097	46.9	18:56:15
Toulouse		1063	49.2	
Madrid	1481	1187	39.6	05:49:28
Amsterdam		1199	39.0	

Table 2: Example of satellite simulation data for two pairs of users.

C Details of the implementation within NetSquid

As highlighted in Fig. 8, the three nodes Alice, Bob, and Carol are interconnected by both quantum and classical links, each of equal length. Two types of nodes were implemented: a `TerrestrialNode` class for metropolitan nodes, which also serve as ground stations, and a `Satellite` class for the satellites. Both classes take as input the node's position (latitude, longitude, and altitude). Similarly, two types of quantum links are distinguished: fiber links on the terrestrial segment connecting two `TerrestrialNode` instances, and free-space links between a `TerrestrialNode` and a `Satellite` node. A quantum source is placed at the midpoint of each fiber link and co-located with each satellite, emitting EPR pairs toward the connected nodes.

The quantum channel introduces a constant propagation delay $\delta = d/c'$ through a link of length d (in kilometers), where c' is the propagation speed (set to 200,000 km/s for fiber, and 318,000 km/s in free space in the simulations). Additionally, the quantum channel may cause the loss of a qubit with a certain probability. Different loss models are used depending on the type of link. NetSquid includes a fiber loss model, `FibreLossModel`, where the probability of photon loss is given by: $p_{loss} = 1 - (1 - p_{init}) \cdot 10^{\frac{-\eta d}{10}}$, where p_{init} is the probability of qubit loss immediately after generation, and η is the attenuation factor along the fiber (in dB/km).

Fig. 8 also illustrates the structure of a quantum node as implemented in the NetSquid simulations. At each node, every flying qubit received from a quantum source is immediately absorbed into its designated slot in the quantum memory. The node then determines whether the local entanglement attempt with the paired qubit was successful. Qubits stored in memory undergo decoherence due to dephasing noise, modeled by stochastically applying a Pauli-Z gate with probability: $p_{dephasing} = 1 - e^{-\Delta t \cdot R_{dephasing}}$, where $R_{dephasing}$ is the dephasing rate (in Hz), and Δt is the time elapsed since the qubit was absorbed into memory.

The quantum processor acts on the stored qubits to perform entanglement swaps, implemented via BSM on pairs of qubits, and applies correction operations using X and Z gates (based on received correction bits). Each operation requires a fixed time of 10 ns. Moreover, the X and Z gate instructions introduce noise modeled as depolarizing noise, by stochastically applying Pauli X, Y, and Z gates with probability: $p_{depol} = 1 - e^{-\Delta T \cdot R_{depol}}$, where R_{depol} is the depolarization rate (in Hz), and ΔT is the time required to perform the operation.

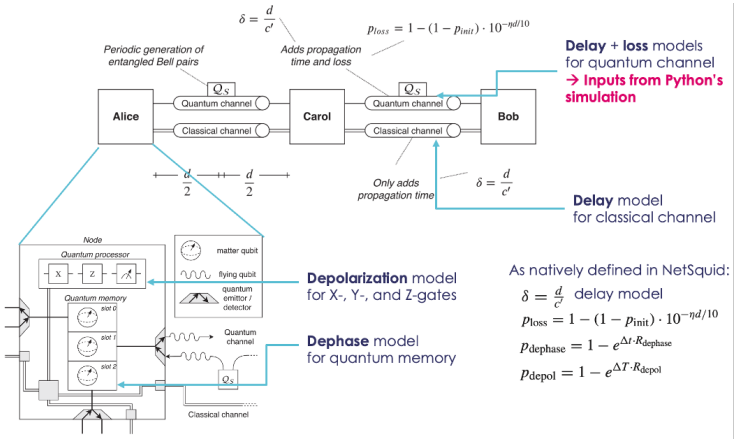
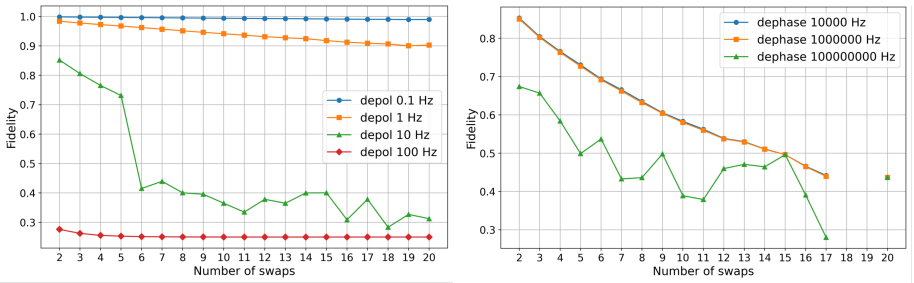


Fig. 8: Illustration of the nodes and channels as implemented in NetSquid.

D Tuning of the parameters for the NetSquid simulation

The first step is to determine suitable parameters for the simulation. We select values for the depolarization rate $R_{depol} \in \{0.1, 1, 10, 100\}$ Hz and $R_{dephase} \in \{0.01, 1, 100, 10000\}$ Mhz. Fig. 9a shows the fidelity of the distributed quantum state between two node pairs as a function of the number of intermediate quantum repeaters, for various depolarization rates. For $R_{depol} \leq 1$ Hz, the number of entanglement swapping operations has little influence on the E2E fidelity, which appears unrealistic. For $R_{depol} = 100$ Hz, the fidelity drops below 0.3 after only two entanglement swaps, which is too low. We then select an intermediate and realistic value: $R_{depol} = 10$ Hz. Fig. 9b shows the fidelity of the distributed quantum state as a function of the number of repeaters, for various $R_{dephase}$ values, with $R_{depol} = 10$ Hz. We observe that for $R_{dephase} = 1$ MHz, the curves are nearly identical, suggesting that dephasing has no significant impact in that range. Starting from $R_{dephase} = 100$ MHz, a clear influence of the number of swapping operations is visible. We thus select the combination $R_{depol} = 10$ Hz and $R_{dephase} = 100$ MHz for our simulations.



(a) For different R_{depol} values.

(b) For different $R_{dephase}$ values.

Fig. 9: Fidelity depending on the number of swaps.

Paving the road towards quantum-safe communications for the Alliance

Joanna Sliwa, Konrad Wrona, Mikolaj Pietruczuk

NATO Communications and Information Agency, The Hague, Netherlands
{joanna.sliwa, konrad.wrona, mikolaj.pietruczuk}@ncia.nato.int

Abstract. This paper presents an overview of experimental studies focusing on quantum-safe communications in military applications, conducted within the NATO Innovation Continuum. The experiments explore a range of topics, including cryptographic agility, secure software supply chains, and quantum-resilient friend-or-foe identification for unmanned aerial vehicles (UAVs). Collectively, these studies highlight challenges and design considerations for military communication and information systems (CIS) architectures that prioritize resilience to emerging quantum threats.

Keywords: cryptographic agility, quantum-safe satellite communications, key management systems, quantum key distribution (QKD), post-quantum cryptography (PQC).

1 Introduction

Governmental and defense organizations worldwide increasingly recognize the need for cryptographic agility in their systems, required for securing their communications against the growing threat posed by quantum-enabled adversaries. While post-quantum cryptography (PQC) appears to be a promising short-term mitigation against cryptographically relevant quantum computers, it may prove vulnerable to new cryptanalytic attacks. Quantum security technologies, such as quantum key distribution (QKD), offer the promise of enabling secure communication channels that can withstand attackers with virtually unlimited computational resources. However, current QKD solutions also present limitations, including a short effective range and vulnerability to side-channel and availability-based attacks on the quantum channel. In such an uncertain environment, designing systems with built-in cryptographic agility is the most future-resilient approach.

The experiments planned to be conducted within the NATO Innovation Continuum (IC25) in 2025 are aimed at exploring ways to achieve quantum-secure communication channels in key military use cases over various transmission media, as well as to investigate the benefits, limitations, and performance of cryptographically agile systems' architecture. Moreover, as shown by extensive industry and deep tech participation in the experiments, NATO Innovation Continuum fosters cooperation with companies and national research centers developing dual-use solutions for the

Alliance. Thanks to this approach, we aim to accelerate innovation, reduce costs, and enhance the agility of supply chains in the defense industry.

2 Motivation

Some of the top concerns in modern warfare are staying at the forefront of innovation and delivering capabilities that meet the needs of operators and provide an advantage over potential adversaries. These concerns underscore the importance of investing in innovative activities and planning for the application of emerging and disruptive technologies in military contexts.

The NATO Communications and Information Agency (NCIA) recently published its Technology Strategy Towards 2030 [11], which addresses the need for delivering high-value, modern solutions for the Alliance and thus advancing the Technology, People, and Processes pillars of the NATO Digital Transformation Vision [12]. One of the key themes of the NCIA Technology Strategy is Quantum Technology (QT), broken down into four strategic objectives: mitigating the quantum threat, performing technology watch, innovating, and creating quantum-literate personnel.

Following its Technology Strategy, NCIA is actively engaged in short-term and long-term activities aimed at satisfying the objectives for the QT Theme. The most immediate ones relate to the approaching Q-day[1] and the necessity of implementing secure and resilient communication patterns. These efforts encompass engagement in the NATO STO Research Task Group on Quantum Technologies Vulnerabilities, which explores means to potentially disrupt the operation of quantum sensors, computers, and QKD devices with different attack vectors. The path from the laboratory to the field for QT must demonstrate that any new technology does not introduce additional security gaps in the system, providing a reliable and highly available capability to the warfighter. Another activity addresses long-term solutions that would make the communications infrastructure independent from the cryptographic primitives protecting these information flows, supporting cryptographic agility and resilience of the CIS. For this purpose, NCIA, together with its partners, prepares a series of experiments for the NATO Innovation Continuum exercise in 2025, organized by the NATO Allied Command Transformation (ACT).

The remainder of the paper provides details of these experiments, with a particular focus on the research covered by each of them.

3 Background

In the context of national security and defense, the implementation of cryptographically agile systems is of critical importance, mainly due to the extended operational lifespan of such infrastructures. These systems implement complex and rigid architectures that require fulfilling strict requirements for resilience and availability, making rapid cryptographic transitions particularly challenging. Additionally, the classified information they store often has a prolonged confidentiality

[1] Approaching moment of building Cryptographically Relevant Quantum Computer able to break current encryption schemes.

requirement, significantly increasing the threat posed by "Harvest-Now, Decrypt-Later" attacks. Mosca's Theorem [10] highlights this urgency, demonstrating that if the required confidentiality duration of data exceeds the time it will take to break current cryptographic algorithms, then measures towards securing such systems must be taken now. In military systems, where sensitive information may need to remain secure for decades, this timeline often extends well beyond current estimates for the introduction of Cryptographically Relevant Quantum Computers. The combination of long-term data sensitivity and low cryptographic agility renders such systems particularly vulnerable to emerging quantum threats.

Simultaneously, mobile communication platforms such as unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) are playing an increasingly important role in tactical environments due to their agility and real-time responsiveness on the battlefield. However, they also introduce new risks, and if compromised, the consequences may be even loss of life. This potentially very high threat impact highlights the urgency of developing quantum-resilient communication mechanisms customized to mobile tactical platforms. Post-quantum cryptography (PQC) offers a practical and scalable solution for several key applications, including platform authentication and secure key exchange for encrypting real-time data, such as video streams transmitted over radio frequencies. Its compatibility with existing infrastructure makes it a more immediately viable approach. Nonetheless, quantum key distribution (QKD) performed via UAVs remains a valuable fallback, potentially useful in establishing encrypted ad-hoc networks in contested or infrastructure-less environments while reducing reliance on centralized key distribution. Exploring both technologies ensures greater resilience for future battlefield communications.

Cryptographic agility is understood as a capability for flexible addition of new cryptographic scheme implementations in a software or hardware system, to reduce the time required to transition and to allow for seamless updates to future cryptographic standards. The latest NIST Cybersecurity White Paper on Considerations for Achieving Crypto Agility [13] defines crypto agility in 3 dimensions, as the ability for machines to select their security algorithms in real time and based on their combined security functions, add new cryptographic features or algorithms to existing hardware or software, resulting in new, stronger security features, and the ability to gracefully retire cryptographic systems that have become either vulnerable or obsolete. Crypto agility can be used on a system, protocol and application level. In our experimental work, we focused on the ability of a system to swiftly and securely transition between cryptographic primitives as threats evolve or standards change over time.

If we consider that certain vulnerabilities can cause cryptographic algorithms be compromised and trigger their update, this approach in a long term also promotes cyber resilience, which is the system's ability to resist, minimize, and mitigate a degradation caused by a successful cyber-attack on a system or network of computing and communicating devices [9]. Cyber resilience by design means that such an ability is intrinsic to the system; the resilience mechanism seamlessly integrates with the primary functional elements and other modules of the system. Cyber resilience by design differs from cyber resilience by intervention, where the intervention of an external actor that "resides" outside of the system is required to provide resilience.

4 State of the art

The National Institute of Standards and Technology (NIST) has been at the forefront of the effort to standardize post-quantum cryptography. In August 2024, NIST released initial PQC standards, including key encapsulation mechanism FIPS 203 (derived from CRYSTALS-Kyber) and lattice-based digital signature algorithm FIPS 204 (derived from CRYSTALS-Dilithium), crucial for quantum-resistant encryption [14]. Simultaneously, the ITU (International Telecommunication Union) is developing standards for quantum-safe networks (in the FG-QIT4N working group [15]), while ETSI (European Telecommunications Standards Institute) actively researches and promotes quantum-safe cryptography and Quantum Key Distribution (QKD) through dedicated working groups and conferences [16]. Migration plans promote creation of cryptographic inventories, emphasize cryptographic agility and hybrid key approaches [3] [4].

As already emphasized, cryptographic agility can be analysed in scope of the system, application and protocol. In this article we focus on the out of band key distribution and agility in selection of cryptographic key providers. Several interfaces define the state of the art in key supply. These are crucial for designing a cryptographically agile system as they effectively allow for decoupling the key management layer from the underlying transmission layer, which is the consumer of keys. Due to this characteristic, the application that is ingesting cryptographic keys is agnostic to the key establishment mechanism.

ETSI has played a central role in defining how applications receive keys. Especially important are two standards [16]: ETSI GS QKD 014, specifying a simple, widely adopted REST API that allows applications to request keys on demand and ETSI GS QKD 004, session-based interface that supports Quality of Service (QoS) parameters, such as guaranteed key rate and jitter. Based on QKD 004, a dedicated commercial key supplier interface has been proposed by CISCO. SKIP (Secure Key Integration Protocol) can provide Post-Quantum Preshared Keys (PPKs) for quantum-safe encryption in IKEv2 and IPsec VPNs. It enhances the standard IKEv2 protocol [8] by mixing traditional Elliptic Curve Diffie-Hellman (ECDH) keys with PPKs during the session key derivation process. Functioning as a REST API over HTTPS, SKIP forwards these PPKs from an external key source into switching devices (e.g. routers). This source, which can be a QKD device, a Key Management System (KMS), or another service, must implement the SKIP Server and use an out-of-band mechanism to deliver the identical PPK to both communicating peers.

5 Experiments

When seeking technical solutions for implementing cryptographic agility, we can assume that different key providers offer varying levels of key management security. This variation allows for selecting the optimal provider based on the transmission source's requirements (e.g., national security policy regulations) or the data type's classification (e.g., security marking of information). Additionally we focus on less dynamic scenarios: when the cryptographic algorithms implemented become

compromised, and having the possibility of changing the key provider without necessity to redesign the system brings time and cost savings.

In our experiments we focus on the out of band key distribution and agility in selection of cryptographic key providers. We defined several experiments that will be carried out in October 2025 during the NATO Innovation Continuum (IC25) exercises, organized by the NATO Allied Command Transformation.

5.1 **Design for cryptographic agility & cyber resilience**

The first experiment focuses on the crypto agility by design and is aimed to explore the following research question: How does the decoupling of the key management layer from the transmission layer facilitate the creation of quantum-secure systems through cryptographic agility and cyber resilience in military communication systems? There are 3 objectives for experiment 1: evaluate the performance and integration feasibility of ML-KEM and QKD (MDI-QKD[2], DV QKD and BBM92 QKD) for secure key exchange in military communications.; implement a modular architecture that decouples the key management layer from the transmission layer to support independent cryptographic algorithm updates; demonstrate cryptographic agility and cyber resilience by testing the ability to swap key exchange mechanisms with minimal disruption to the underlying transmission channel.

The scenario assumes that military data centers serve as the backbone for tactical, operational, and strategic communications, storing critical data, running intelligence analytics, and hosting Command and Control (C2) systems. These data centers typically feature a high-throughput network environment with large volumes of data traffic, including encrypted communications between data centers themselves. Current widely used network security protocols, such as IPSec, are not well-suited for the high throughput demands set by data centers. MACSec, being a Layer 2 protocol, should allow for significantly higher throughput.

In our system, the decoupling of key management from the transmission channel allows for the interchangeability of key exchange mechanisms, providing support for both quantum and post-quantum key exchange schemes without altering the transmission layer (see Fig. 1). Such an abstraction enables us to achieve crypto agility, as the transmission layer does not need to know whether the key originates from QKD or PQC, nor does it require significant changes to itself to accommodate a key from a different source. Subsequently, this innate crypto agility equips our system with cyber resilience by design. If an incident occurs, such as a failure of QKD or the compromise of a specific cryptographic primitive, the system can rapidly respond by substituting the affected cryptographic component without requiring a complete system redesign. The architectural decoupling of the key management layer from the transmission layer ensures that such failures are isolated, preventing cascading effects across the system. This, in turn, leads to minimizing the impact of cryptographic failures and maintaining the system's operational continuity.

[2] MDI – Measurement Device Independent, DV – Discrete Variable, BBM92 – Standard developed in 1992 by Charles H. Bennett, Gilles Brassard, and N. David Mermin

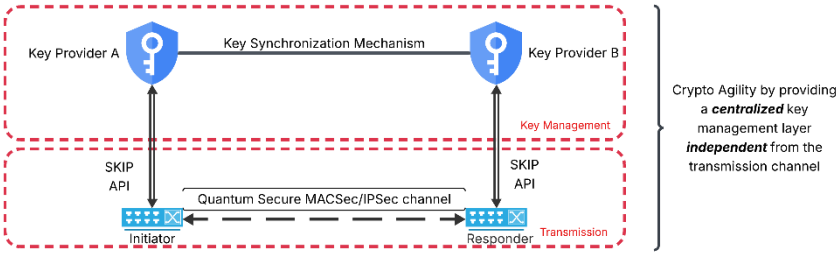


Fig. 1 Architectural decoupling of key management from the underlying transmission

The system will be tested against crypto agility and cyber resiliency criteria. For the first one, the system's ability to integrate and switch between PQC and QKD mechanisms without disrupting communication. Moreover, the ease of implementation will be tested, with an emphasis on the scalability of incorporating new cryptographic mechanisms into the key management layer. For the latter, we will focus on the emulation of system's behavior when one of the cryptographic primitives fails (e.g., a QKD device failure, deprecation of one of the key exchange algorithms, or a successful attack on it), with an emphasis on the system's ability and ease of switching to an alternative method. We are interested in measuring the performance metrics, including the key rate, key failure rate, and number of successful key derivations that will be logged and analyzed to evaluate the feasibility of integrating PQC and QKD into systems with real-world data throughput. For the QKD we are also interested in verification of the resilience to atmospheric interference. The key performance metrics will be logged for an open-space quantum channel under non-optimal and potentially disruptive atmospheric conditions. Subsequently, they will be compared against the same metrics under optimal atmospheric conditions to test the resilience of the open-space QKD channel.

5.2 Secure mobile platforms

The second set of experiments focus on the performance of the PQC standards in dynamic tactical scenarios. Mobile platforms are an integral part in modern military scenarios, providing real-time intelligence and communication support, both of which ultimately lead to increased situational awareness. Secure Identification Friend or Foe (IFF) can effectively mitigate adversarial spoofing attempts or unauthorized access to friendly networks. PQC offers a viable solution for implementing IFF and securing high-value video streaming. Nonetheless, PQC integration poses challenges, including computational and data overhead, key size constraints and energy limitations on UAVs.

This experiment is aimed at verification to what extent post-quantum cryptography mechanisms can be integrated into mobile platforms for secure authentication and real-time, bandwidth-constrained data transmission, such as UAV identification and low-latency 1080p video over tactical radio. It has 2 main objectives related to 2 scenarios: (1) to implement a system that recognizes friendly UAVs using quantum-secure public key infrastructure systems, evaluate the performance of successful authentication attempts and define limitations of such an approach; (2) to showcase the feasibility of

integration and evaluate the performance of post-quantum cryptography in high-bandwidth video encryption over radio frequency.

The relevance of these two scenarios relate to the fact that in highly critical tactical environments, UAVs often operate in close coordination with ground forces and other airborne assets. In such scenarios, the ability to reliably authenticate friendly drones and securely provide live video feed can be vital to the success and protection of friendly forces. Adversarial attempts to spoof or disrupt these systems can compromise operations and endanger human lives. Evaluating PQC under these operational pressures ensures it can meet the demands of real-world deployments.

Two systems will be prepared for (1) performing UAV authentication and (2) encrypting video stream over tactical radio. The former will include three UAVs equipped with a PQC capable FPGA accelerators and directional antennas. The latter will consist of a mobile platform equipped with a camera, a radio transmitter (also capable of encoding video), a radio receiver (capable of decoding the video) and an operation workstation.

Quantum secure IFF scenario will be evaluated based on timing metrics, gathered for both the initiator and the responder. These will include total authentication time, data link latency, data link throughput, computation time for both parties, authentication throughput. After that the protocol disruption will be analyzed based on bit loss ratio, e.g. how many times the protocol should be repeated to perform a successful authentication. Quantum secure video over tactical radio will use multiple metrics collected from the encrypted and unencrypted video feed and then compared against themselves. The metrics will include encoder, encryption and decryption latency.

5.3 **Quantum Key Distribution Analysis**

There are only a handful of scenarios where QKD offers a promising solution for the use in military applications. The IC25 exercises focus on dynamic operational scenarios that could provide high resiliency of key provision for the deployed forces. As the first step in these considerations, the use of the free-space entanglement - based QKD (BBM92) has been proposed to identify challenges of this technology implementation for key distribution from space (e.g. HAPS[3] platforms, satellites). As an emulation of the flying platform, we assume to use a UAV (drone) that will be stabilized to provide entangled photons to two base stations.

This scenario poses several challenges. Maintaining precise laser alignment between ground stations and the possibly moving platform is difficult due to the motion, vibrations and limited stability of drones in flight, which can be also further degraded by windy conditions. Additionally, free-space optical QKD is highly sensitive to atmospheric conditions such as fog, rain or even intensive sunlight. These factors render reliable key exchange under real-world conditions a complex and demanding task. Nonetheless, QKD is analysed as an important fallback option which, as opposed to PQC, is unlikely to be affected by developments in cryptanalysis or increasing computing resources available to adversaries[4].

[3] High-Altitude Platform Stations

[4] Note that QKD may also be subject to additional Side – Channel Attacks that would significantly reduce its overall security [17].

The third experiment aims to verify to what extent are Quantum Key Distribution systems viable for mobile secure communication networks, and what are their potential vulnerabilities. It is planned around 3 objectives: Evaluate the performance and atmospheric resilience of an open-space QKD channel to provide quantum secure video call encryption (*Exp3.scenario 1*); perform a successful side-channel attack (SCA) on a UAV-based mobile QKD system, effectively extracting a quantum-generated key (*Exp3.scenario 2*); evaluate mitigations for such SCA (*Exp3. Scenario2*).

Ad-hoc mobile networks (MANETs) secured with QKD can be relevant in environments where conventional infrastructure is unavailable or compromised. By using high altitude flying platform to establish secure optical links between ground nodes[5], encryption keys can be distributed dynamically and on-demand across mobile units. That can aid in tactical flexibility for units in dynamic situations that still require encrypted communications, for example over Radio Frequency (RF).

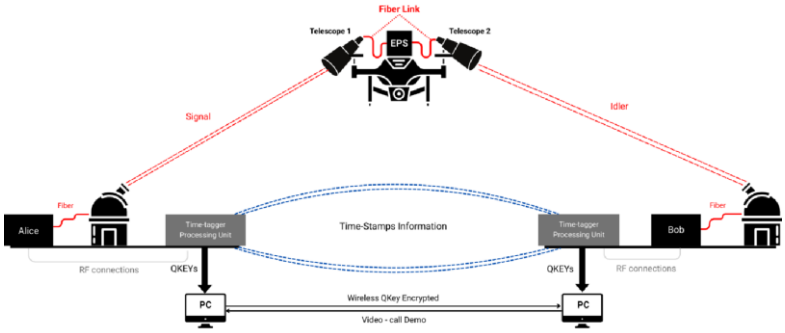


Fig 2. Proposed system design for experiment 3

The system for *Exp3.scenario1* consists of a drone, hovering in one place, equipped with an entangled photon source (EPS), two ground stations (Alice and Bob), the time tagging processing unit and two clients consuming the keys to encrypt a video channel using AES-256. Additionally it is assumed that communications between the time-tagging units will be encrypted with the FPGA – based PQC (see Fig 2). For *Exp3.sceanario2* the adversarial setup for performing an SCA consists of n antennas, each dedicated to a single photon detector (SPD), an analog-to-digital converter (ADC) and a reinforcement learning algorithm specifically trained with previous ADC outputs.

It is planned to measure free-space QKD performance with quantum key generation rate, Quantum Bit Error Rate (QBER), and link stability over the duration of the drone's flight. Coincidence windows between Alice and Bob's detection events will be evaluated to optimize temporal filtering and maximize the secure key yield. QBER will be tracked over time to identify any degradation due to optical misalignment, atmospheric turbulence, or external disturbance. The encryption performance will be verified through the continuity and integrity of the real-time video call. In cases of protocol failure (e.g. high QBER or depleted key supply), the time and cause of the failure will be logged.

For *Exp3.scenario 2* we will focus on QKD side channel vulnerability analysis and possibility to infer keys exchanged by the endpoints. RF traces will be analyzed to

[5] Even though the scenario covers mobile units, it is assumed that ground station is not moving during the key transfer.

detect repeatable patterns corresponding to detector clicks. Machine learning models trained on labeled RF data will be used to classify pulses and assign them to specific SPDs. The inferred time tags will be then compared against the legitimate QKD data to assess accuracy.

6 Summary

The article focuses on exploring technical implementation of different quantum-resistant solutions supporting communications, with great emphasis on dynamic, tactical scenarios and challenges of both – PQC and QKD. Experimenting with cryptographic agility and out-of-band key distribution is critical for future-proofing security. Agile key provider exchange allows organizations to switch providers based on security needs or policy changes, ensuring optimal key management. Hybrid scenarios (combining classical and PQC) offer a crucial transition bridge, mitigating risks from unknown PQC vulnerabilities. However, deployment of any solution needs to take into account performance issues resulting e.g. from larger key sizes and higher computational overhead (for current PQC algorithms), sensitivity to atmospheric conditions, jamming and vulnerability to side-channel attacks (for QKD). These pose significant challenges and performance problems in resource-constrained tactical environments.

The results of experiments will be used to inform NATO ACT about the efficiency of evaluated technologies and prepare recommendations for the NATO CIS quantum threat mitigation plan.

7 References

[1] K.-S. Shim at al., "Design and Validation of Quantum Key Management System for Construction of KREONET Quantum Cryptography Communication," *Journal of Web Engineering*, Jul. 2022, doi: https://doi.org/10.13052/jwe1540-9589.2151.

[2] P. James, S. Laschet, S. Ramacher, and Luca Torresetti, "Key Management Systems for Large-Scale Quantum Key Distribution Networks," Aug. 2023, doi: https://doi.org/10.1145/3600160.3605050.

[3] Barker, E., et al. Considerations for Achieving Crypto Agility. NIST CSWP 39. https://doi.org/10.6028/NIST.CSWP.39.ipd

[4] Timelines for migration to post-quantum cryptography, UK NCSC, March 2025, https://www.ncsc.gov.uk/guidance/pqc-migration-timelines

[5] ITU-T Y.3800 (2019) Corrigendum 1. Overview on networks supporting quantum key distribution, April 2020.

[6] ITU-T Y.3801. Functional requirements for quantum key distribution networks, April 2020.

[7] ITU-T Y.3803. Quantum key distribution networks – Key management, Dec 2020.

[8] S. Fluhrer, P. Kampanakis, D. McGrew, and V. Smyslov, “Mixing Preshared Keys in the Internet Key Exchange Protocol Version 2 (IKEv2) for Post-quantum Security,” *IETF*, Jun. 01, 2020. https://datatracker.ietf.org/doc/html/rfc8784

[9] A. Kott, M. S. Golan, B. D. Trump, and I. Linkov, “Cyber resilience: By design or by intervention?”, arXiv.org, https://arxiv.org/abs/2201.11152

[10] M. Mosca, “Cybersecurity in an Era with Quantum Computers: Will We Be Ready?”, *IEEE Security & Privacy*, vol. 16, no. 5, pp. 38–41, Sep. 2018, doi: https://doi.org/10.1109/msp.2018.3761723.

[11] NATO Communications and Information Agency Technology Strategy Towards 2030, NU, Brussels, 2024

[12] S. R. Soare, Digitalisation of Defence in NATO and the EU: Making European Defence Fit for the Digital Age, Aug 2023, The International Institute for Strategic Studies

[13] Barker, E., et al. Considerations for Achieving Crypto Agility. NIST CSWP 39. https://doi.org/10.6028/NIST.CSWP.39.ipdhttps://doi.org/10.6028/NIST.CSWP.39.ipd

[14] NIST Post Quantum Cryptography standardization project, https://csrc.nist.gov/projects/post-quantum-cryptographyhttps://csrc.nist.gov/projects/post-quantum-cryptography

[15] ITU-T Study Group 17 (SG17) – Security, https://www.itu.int/en/ITU-T/studygroups/2025-2028/17/Pages/default.aspx

[16] Industry Specification Group (ISG) on Quantum Key Distribution (QKD); https://www.etsi.org/committee/qkd

[17] BSI, “Implementation Attacks against QKD Systems v. 1.0,” Nov 2023. [Online]. Available: https://www.bsi.bund.de

On the road to satellite-based Quantum Information Networks: Hong-Ou-Mandel effect with pulsed coherent states

Mathieu Bertrand[1], Luca Paccard[1], Pedro Roldan Gomez[1], Laurent de Forges de Parny[1], Michel Sotom[1], Etienne Samain[2], Olivier Alibart[3], Matteo Schiavon[4], Sacha Gressani[5], Erik Kerstel[5], Thierry Lanz[6], Eleni Diamanti[4], Sébastien Tanzilli[3] & Mathias van den Bossche[1]

[1] Thales Alenia Space, 26 avenue Jean-François Champollion - BP 33787, 31037 Toulouse-Cedex 1, France
[2] SigmaWorks, 8, Allée Bellevue, 06460 Saint Vallier de Thiey, France
[3] Université Côte d'Azur, CNRS, Institut de Physique de Nice (INPHYNI), UMR 7010, Parc Valrose, 06108 Nice Cedex 2, France.
[4] Sorbonne Université, CNRS, LIP6, 4 Place Jussieu, F-75005 Paris, France.
[5] Centre Spatial Universitaire de Grenoble, 120 Rue de la Piscine, 38400 Saint-Martin-d'Hères, France.
[6] OGS Technologies, 7 Chemin des Pins, 06000 NICE
mathieu.bertrand@thalesalenispace.com

Abstract. The so-called Quantum Information Networks (QIN) promise to revolutionize the world with new applications based on the interconnection of quantum devices such as quantum computers, quantum sensors and physically secured cryptographic receivers. Such networks employ photons as a propagation mean of quantum information in order to create entanglement between the end-users' devices. Over long distances, satellites will become mandatory in the network as they offer a better optical losses scaling than fibres. In this paper, we focus on the physical principle at the heart of the two-photon Bell-State Measurement devices used in the QIN to swap entanglement between and inside the network nodes: the Hong-Ou-Mandel effect.

Keywords: Quantum Information Networks, Satellite, Bell-State Measurement, Hong-Ou-Mandel.

1 Satellite-based QIN architecture

1.1 Rationale

QIN architectures are a thriving topic in the quantum information community [1, 2]. Generally speaking, QIN are made up of nodes (often called Quantum Repeater Nodes, QRN) combining Entangled Photon Sources (EPS), Bell-State Measurement (BSM) devices and Quantum Memories (QM) [3, 4]. Long-distance connections on ground are achieved using optical fibres in the telecom bands, with typical losses of $0.2 dB/km$.

On the other hand, the link budget of satellite Free-Space Optical (FSO) connections scales as

$$\frac{P_{R_x}}{P_{T_x}} \propto \left(\frac{\pi D_{R_x} D_{T_x}}{4\lambda L}\right)^2, \quad (1)$$

with the subscript T_x (R_x) corresponding to the emitter (receiver), P the optical power, D the telescope diameter, λ the optical wavelength and L the distance between the satellite and the Optical Ground Station (OGS) [5]. For end-users' distances typically above 100km, direct satellite connections thus provide higher transmission rates than optical fibres. When designing the satellite segment of the QIN, two main constraints have to be taken into account.

The first one is linked to the limitations in Size, Weight and Power (SWaP) of the equipment a satellite payload can embark. This constraint rules out architectures where the BSM devices of the QIN are placed on-board the satellite. Indeed, these devices rely on Single Photon Detectors (SPD). Today, the most efficient type of SPD are Superconducting Nanowire SPD (SNSPD), which require cryogenic cooling in the 1K temperature range and are thus very challenging to embark. Moreover, efficient QM technologies are either based on cryostats for Rare-Earth doped ions, or on complex and high-SWaP apparatuses for cold atoms and trapped ions, ruling them out too for integration into the satellite payload [6].

The second constraint is due to the fact that uplink Free-Space Optical (FSO) connections are more prone to losses than downlinks, typically by a factor of 10 in the optical link power budget. The reason is that the turbulence induced by the atmosphere occurs in the first few tens of kilometres above the Earth surface, while the rest of the propagation channel is basically vacuum. In uplink, these turbulences induce a beam wandering, with a strong negative impact on the link budget. In downlink, the atmosphere disturbances only induce wave front deformations, which can be corrected by an Adaptive-Optics (AO) assembly.

Combining these two constraints leads to defining satellite-based QIN nodes with EPS on-board, and to placing BSM and QM on the ground segment. A simplified architecture of such a network is represented in Fig. 1.

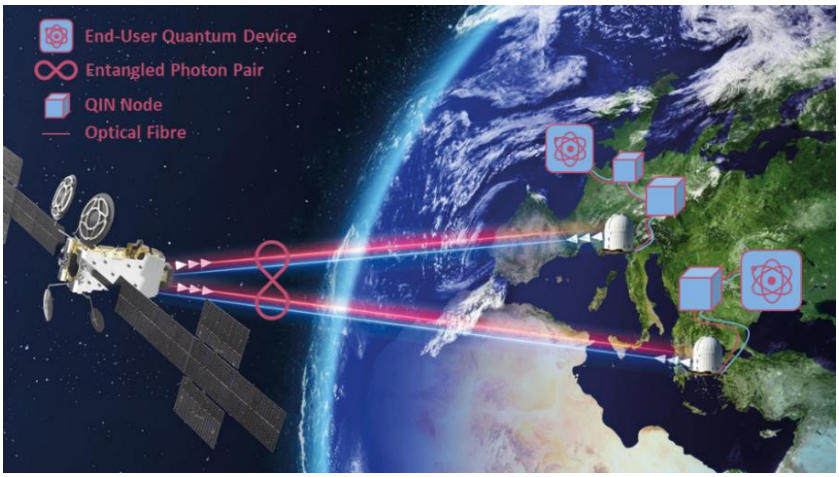


Fig. 1. Simplified architecture of a QIN including space and ground segments. Entanglement generated by EPS is propagated through satellite FSO and ground optical fibre links. QIN nodes include EPS, BSM and QM. Red lines correspond to qubits, while blue ones include all the auxiliary channels such as synchronization signals, polarisation references, etc.

1.2 Scope of the work

As part of this work, we develop an embryonic QIN space segment demonstrator. To end up with a fully fibered optical testbed at telecom wavelength, we greatly simplify the above architecture by making several assumptions, considered valid for a Proof-Of-Concept (POC) demonstration of such a satellite system. First, we remove the QM, as their maturity is still quite low and their cost is too high to be included in a mid-term project. This simplification leads to the schematic of the POC sketched in Fig. 2.

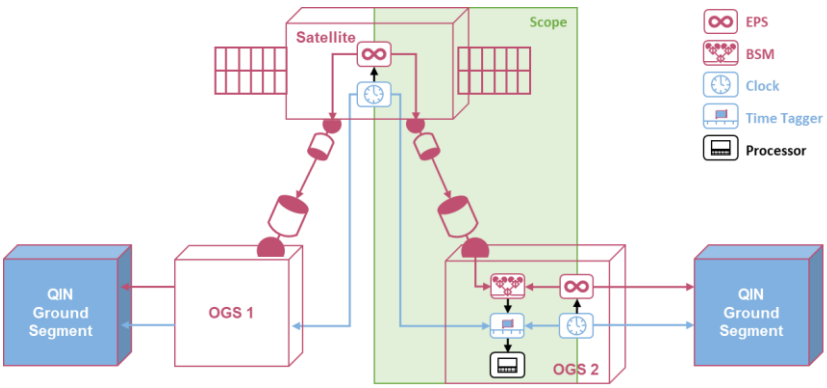


Fig. 2. POC demonstrator without QM, including a satellite containing an EPS and a clock and two OGS with dedicated quantum receivers.

In this picture, the BSM device is assumed to be a passive 2-photon polarisation BSM apparatus. The simplest way of implementing this element is to use a combination of Beam-Splitters (BS) and Polarising Beam-Splitters (PBS). The physical principle behind the projection onto the Bell-State basis is based on the so-called Hong-Ou-Mandel effect [7]. If all the degrees of freedom of the two incoming photons are identical, they will bunch on the BS if they have similar polarisations. Post-selecting the right correlation events at the BS output then allows to project the incoming photons on one or two of the four Bell states [8, 9]. The second simplification thus consists in reducing the BSM device to a simple BS.

Finally, the third and final simplification consists in considering only the two photons interacting on the BS. In our experiment, we then forget about the two other photons in Fig. 2, directed towards the OGS 1 and QIN ground segment. This last assumption allows us to work with attenuated laser sources instead of EPS to emulate the on-board and ground sources. The resulting HOM testbed is represented in Fig. 3.

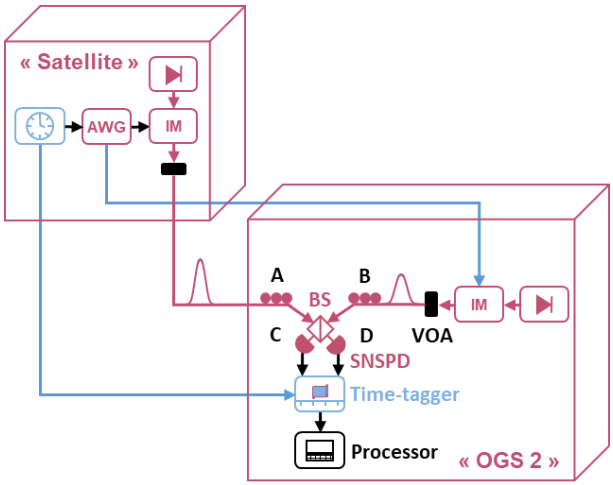


Fig. 3. Optical layout of the HOM testbed.
AWG: Arbitrary Wave Generator, IM: Intensity Modulator, VOA: Variable Optical Attenuator.

2 Hong-Ou-Mandel effect with pulsed coherent states

2.1 Experimental setup

First demonstrated in 1987 by the 3 physicists Chung Ki Hong, Zheyu Jeff Ou and Leonard Mandel, the HOM effect is a quantum interference between two photons impinging on each input port of a 4-way BS [7]. As the photons are bosons, if all their degrees-of-freedom are strictly identical, the two-photon wave function symmetrisation implies that they will go out in the same output port, even though this port is chosen randomly. Studying the correlation between the detections in the BS output ports with SPD thus reveals the existence of the effect, as depicted in Fig. 3.

In a satellite-based QIN, the interference between downlink and ground photons can be affected by 4 types of impairments: timing jitters due to synchronisation imperfections, frequency mismatch due to laser drifts and Doppler effect, polarisation alignment issues and link budget variations. In this work, we have studied the dependency of the so-called HOM dip, *i.e.* the reduction of the coincidences between the output ports, as a function of the arrival time delay and frequency offset between two attenuated optical pulses, as well as their mean photon number ratio.

To generate the pulses, we use two telecom C band lasers at 1545nm: one temperature/current controlled Distributed Feedback Laser (DFB) and a tuneable external cavity laser. The frequency offset between the lasers is preliminary matched and then let free-running. This offset Δ can be tuned by changing the optical frequency of the external cavity laser. The intensity of each laser is then carved by a Mach-Zehnder interferometer modulated by one output of a common Arbitrary Wave Generator (AWG). This setup allows to control the arrival time delay δ between the two coherent states. Moreover, the internal clock of the AWG serves as a time reference for the time-tagger. Fig. 4 shows the intensities of the optical signals obtained with this testbed for a repetition period of 30ns and a Full Width at Half Maximum (FWHM) of 2.5ns.

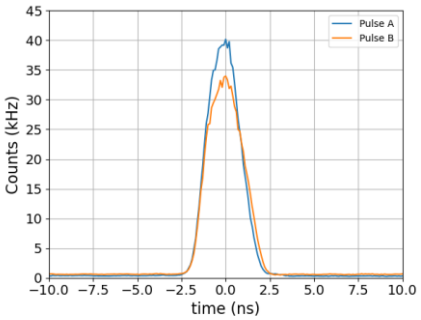


Fig. 4. Optical intensities of the two pulses measured with SNSPDs.

2.2 HOM effect theory

Taking the normalised square root of the optical pulses shown in Fig. 4 allows to define the wave function amplitudes $\epsilon_A(t)$ and $\epsilon_B(t)$ for the coherent states respectively incoming on ports A and B of the BS. The complete wave functions of these states are then expressed as

$$\xi_{A,B}(t) = \epsilon_{A,B}\left(t \mp \frac{\delta}{2}\right)e^{-i\left(\omega \mp \frac{\Delta}{2}\right)t}, \tag{2}$$

where ω stands for the central frequency of the lasers, δ for the time delay and Δ for the frequency offset between the two pulses [10]. The field operator corresponding to the annihilation of a photon in these ports writes

$$E^{+}_{A,B}(t) = \xi_{A,B}(t)a_{A,B}. \tag{3}$$

The BS transformation between the output C and D ports with respect to the input A and B ports gives

$$\begin{cases} E_C^+(t) = \frac{1}{\sqrt{2}}[-iE_A^+(t) + E_B^+(t)] \\ E_D^+(t) = \frac{1}{\sqrt{2}}[E_A^+(t) - iE_B^+(t)]. \end{cases} \quad (4)$$

In this experiment, we have worked with a low mean photon number $\mu_A = 0.02$. In this regime, only the events with a total of two photons, either incoming on the same input port or one for each port, are relevant for the coincidence calculation. The most important term for the HOM effect is the interference between a photon arriving at input port A and one at B, producing detections at the output port C at time τ_0 and D at $\tau_0 + \tau$. For an input state $|\psi_{in}\rangle = a_A^\dagger a_B^\dagger |00\rangle$, it is defined as

$$P_{CD}^{11}(\mu_A, \mu_B, \tau, \tau_0, \delta, \Delta) = \langle \psi_{in} | E_C^-(\tau_0) E_D^-(\tau_0 + \tau) E_D^+(\tau_0 + \tau) E_C^+(\tau_0) | \psi_{in} \rangle. \quad (5)$$

The different contributions P_{CD}^{kl} with k photons in C and l in D are then weighted by the probabilities P_{AB}^{mn} that m photons arrive at port A and n photons at port B, given by

$$P_{AB}^{mn}(\mu_A, \mu_B) = \mu_A^m \cdot \mu_B^n \cdot \frac{e^{-(\mu_A + \mu_B)}}{m!n!}. \quad (6)$$

The coincidence probability, called P_c, results from the weighted probabilities of events containing 2 photons in total and integrated over τ_0.

2.3 HOM dip measurements

The normalised coincidences on SNSPD[1] C and D as a function of the time delay between the optical pulses is represented in Fig. 5.

[1] Quantum efficiency ≥ 80%, dark count rate 10Hz, RMS jitter 30ps.

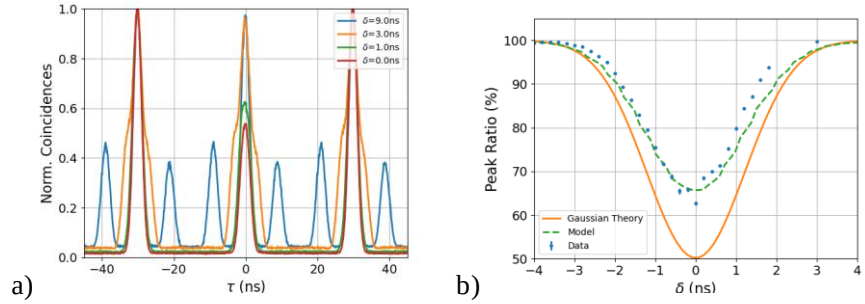


Fig. 5. a) Normalised coincidences as a function of the tagger delay τ for different optical pulses delays δ, with a mean photon number ratio $\frac{\mu_A}{\mu_B} = 1.1$. b) HOM dip defined as the ratio of the central peak over the lateral ones. The orange curve represents an analytical Gaussian approximation, and the green dotted curve is a model based on the experimental wave functions.

The results of Fig. 5.b show that our model, taking into account the actual wave functions and background noises of the experiment (limited by straight light in the laboratory), correctly reproduces the experimental results. The Gaussian theory is computed without any noise on the detectors and shows a well-known maximum HOM dip visibility of 50%. The dip as a function of the frequency difference between the two lasers is represented in Fig. 6.

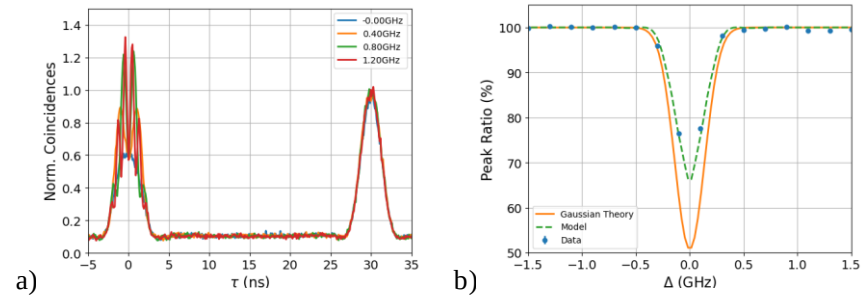


Fig. 6. a) Normalised coincidences as a function of the tagger delay τ for different optical pulses frequency shifts Δ for a mean photon number ratio $\frac{\mu_A}{\mu_B} = 0.8$. b) HOM dip in frequency.

Interestingly, as the time-tagger resolution ($200ps$) is smaller than the pulses width, we observe a frequency beat note inside the central peak of Fig. 6.a. The lateral peaks, produced by photons arriving 30 ns apart, do not interfere and only show the amplitude wave functions overlap. The resulting dip in Fig. 6.b is obtained by integrating the central and lateral peaks and computing their ratio. It is only sensitive to the pulses frequency overlap, Fourier limited to 400 MHz FWHM. The beat note frequency stability between the two lasers was typically around 100MHz over 1 second.

Finally, we have performed a HOM dip visibility analysis as a function of the photon ratio $R = \frac{\mu_A}{\mu_B}$. In the final system, EPS will have a different statistics than the Poissonian

photon number distribution of coherent states, but this study is still interesting for use cases such as teleportation between an EPS and a coherent state. The visibility, calculated with the correlation probabilities P_c, is defined as

$$V(\mu_A,\mu_B) = 1 - \frac{\int_{-\infty}^{+\infty} P_c(\mu_A,\mu_B,\tau,0,0)\,d\tau}{\int_{-\infty}^{+\infty} P_c(\mu_A,\mu_B,\tau,+\infty,+\infty)\,d\tau}. \quad (7)$$

The results for a set of experimental photon number ratios are shown in Fig. 7.

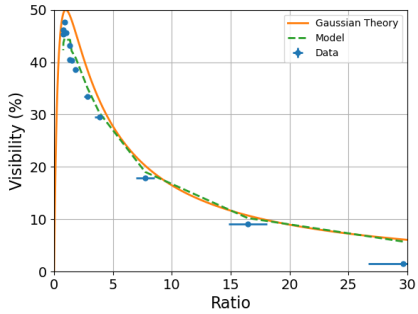


Fig. 7. HOM dip visibility as a function of the mean photon ratio between the two coherent states for $\delta = 0 \pm 0.1ns$ and $\Delta = 0 \pm 100MHz$. Data points (blue dots), theoretical curve obtained with Gaussian pulses (orange line) and model (green dotted curve) taking into account the experimental wave functions and noises.

The theoretical visibility obtained with Gaussian coherent states (represented by the orange line in Fig. 7), reaches the maximum of 50% as it is computed without any background noise. The green dotted line, corresponding to our model, is very close to the Gaussian curve shape as the actual pulses wave functions are almost Gaussian, but the visibility is reduced due to experimental noises. We still achieve a maximum of 47.5%. The slight discrepancy between the data points and the models are probably due to the fact that the photon number ratio estimation, based on the experimental wave functions integration, can be biased.

The experimental visibility of the HOM dip will be used to estimate the BSM success probability with EPS in the next phase of the project.

2.4 Conclusion

We have reported on a set of preliminary POC experiments that allows us to assess a simplified architecture for a future satellite-based QIN system. The results on the HOM effect between the equivalent of a satellite payload-emitted photon and a ground source photon reveal the dependency on the arrival time, the frequency offset and the mean photon number ratio of the two pulses. In the final system, we expect a higher visibility when considering entangled photon pairs instead of coherent pulses, as the statistics will be different.

In the next step of the project, this POC testbed will be used to assess more realistic synchronisation scenarios for a satellite system, where the on-board and ground sources have physically distinct timing references.

Acknowledgement

We thank the French space agency *Centre National d'Etude Spatiales* (CNES) for supporting this project.

References

[1] H. Kimble, «The quantum internet,» Nature, vol. 453, p. 1023–1030, 2008.

[2] S. Wehner, D. Elkouss et R. Hanson, «Quantum Internet: A vision for the road ahead,» Science, vol. 362, 2018.

[3] N. Sangouard et al, «Quantum repeaters based on atomic ensembles and linear optics,» Rev. Mod. Phys. , vol. 83, n° 33, 2011.

[4] K. Azuma et al, «Quantum repeaters: From quantum networks to the quantum internet,» Rev. Mod. Phys. , vol. 95, p. 045006, 2023.

[5] B. Klein et J. Degnan, «Optical Antenna Gain. 1: Transmitting Antennas,» Applied Optics, 1974.

[6] G. Gavin Brennen et al, «Focus on Quantum Memory,» New J. Phys. , vol. 17, p. 050201, 2015.

[7] C. Hong, Z. Ou et L. Mandel, «Measurement of Subpicosecond Time Intervals between Two Photons by Interference,» Phys. Rev. Lett., vol. 59, n° 2044, p. 10.1103/PhysRevLett.59.2044, 1987.

[8] S.-W. Lee et al, «Bell-state measurement and quantum teleportation using linear optics: two-photon pairs, entangled coherent states, and hybrid entanglement,» ArXiV, p. arXiv:1304.1214v1, 2013.

[9] C. Nölleke et al, «Efficient Teleportation Between Remote Single-Atom Quantum Memories,» Phys. Rev. Lett. , vol. 110, p. 140403, 2013.

[10] T. Ferreira da Silva et al, «Spectral characterization of weak coherent state sources based on two-photon interference,» J. Opt. Soc. Am. B , vol. 32, pp. 545-549 , 2015.

BenchQC - Scalable and modular benchmarking of industrial quantum computing applications

Florian Geissler[1], Eric Stopfer[2], Christian Ufrecht[2], Nico Meyer[2], Daniel D. Scherer[2], Friedrich Wagner[2], Johannes M. Oberreuter[3], Zao Chen[3], Alessandro Farace[3], Daria Gutina[4], Ulrich Schwenk[4], Kimberly Lange[4], Vanessa Junk[4], Thomas Husslein[4], Marvin Erdmann[5], Florian Kiwit[5], Benjamin Decker[5], Greshma Shaji[5], Etienne Granet[6], Henrik Dreyer[6], Theodora-Augustina Drăgan[1], and Jeanette Miriam Lorenz[1]

[1] Fraunhofer Institute for Cognitive Systems IKS, Munich, Germany (florian.geissler@iks.fraunhofer.de)
[2] Fraunhofer Institute for Integrated Circuits IIS, Nuremberg, Germany
[3] Reply SE, Machine Learning Reply GmbH, Munich, Germany
[4] OptWare GmbH, Regensburg, Germany
[5] BMW Group, Munich, Germany
[6] Quantinuum, Munich, Germany

Abstract. We introduce BenchQC, a quantum computing benchmarking project that focuses on application-centric benchmarking using diverse industry use cases. Building upon the open-source platform QUARK, we evaluate key metrics across the quantum software stack to identify trends towards quantum utility, and distinguish viable research directions. This initiative contributes to the broader effort of establishing reliable benchmarking standards that drive the transition from experimental demonstrations to practical quantum advantages.

1 Introduction

Quantum computing (QC) offers new avenues to tackle complex data analysis and optimization problems. In light of the ever-growing demand for computational resources, this puts the technology at the center of interest for both scientific researchers and industrial end users. Quantum computing promises to boost industrial applications in particular in the fields of optimization, simulation, and machine learning [7, 28]. A 2021 study estimates that, at tech maturity, quantum computing will unlock hundreds of use cases, with potential value creation of $100B and more in *each* of the above fields [15]. Being at the forefront of this emerging technology is therefore a competitive advantage that elicits strategic investments across various industrial domains (see, e.g., [7]).

With limited quantum hardware available in the current noisy intermediate-scale quantum (NISQ) era, quantum utility remains to be demonstrated in real-world applications [20]. To this day, the international quantum communities have each prioritized individual metrics, concepts, and applications for benchmarking (see Sec. 2). Nevertheless, the informative power remains limited for industrial

applications, particularly in emerging fields such as quantum machine learning (QML) [10,32]. For end users, it can be challenging to understand whether their problem category is well-suited to benefit from future quantum technology, what problem scales are realistic, or which hardware and algorithms are the best fit. As a lighthouse project of the Munich Quantum Valley (MQV), BenchQC therefore takes a different approach, and makes industrial applications the starting point of the benchmarking workflow. We discuss selected problems from different categories in Sec. 4, after reviewing related work in Sec. 2 and our methodology in Sec. 3.

2 Related work

Several international quantum computing benchmarking communities have outlined their strategic approaches. Selected examples include the French BACQ consortium [6], focusing on application-specific metrics like the Q-score and scalable aggregation using the Myriad-Q Tool, the Quantum Economic Development Consortium (QED-C) [1] in the USA, centering on hardware-centric metrics like quantum volume [22], or the Quantum Application Score (QuAS) proposed by the Dutch TNO ecosystem [23]. For a more detailed overview, we refer to Refs. [21,28].

Despite these initiatives, the systematic evaluation of quantum utility in industrial problems is still lacking. BenchQC addresses this gap with a structured approach uniquely characterized by i) use cases, which are derived from real industrial problems, ii) utilization of modular open-source frameworks like QUARK, iii) establishing interfaces for full-stack benchmarking, including various quantum hardware types, iv) a holistic range of metrics from hardware to application level, v) the inclusion of benchmarking for quantum machine learning applications, and vi) concepts to simulate problem complexity scaling.

3 Methodology

Problem decomposition and metrics: The overall workflow is illustrated in Fig. 1: Starting from a use case (a), we first derive a mathematical formulation (b). A sequence of problem sizes and complexities is formed (c). The problem is encoded as a hybrid algorithm by classical-quantum decomposition (d) in reference to the baseline of a classical solution (e). The hybrid problem is eventually executed on different quantum-computing hardware (f) or simulated on classical hardware (g). Relevant metrics quantify the performance of the various parts of this decomposed workflow to govern the overall benchmarking process (h).

Initiatives towards the standardization of quantum computing benchmarking metrics are ongoing [21]. In BenchQC, we collect application-agnostic metrics along the benchmarking pipeline, which reflect problem and solution architecture, see Fig. 2. Subsequently, we put those in context with the actual application performance to identify trends for potential quantum utility. Our metric categories are, with examples:

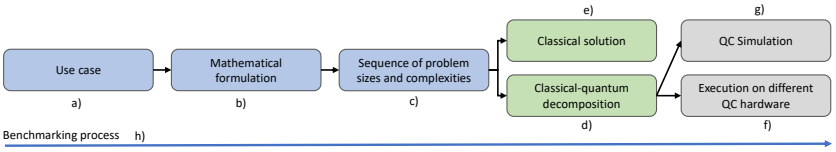


Fig. 1: Decomposition of the benchmarking workflow.

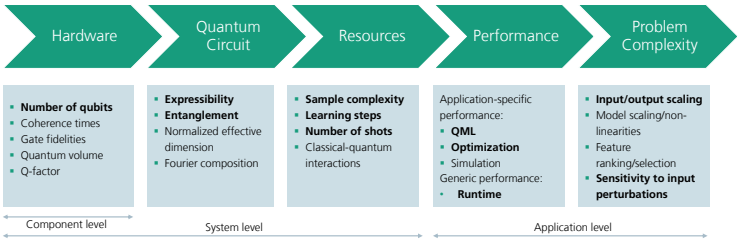


Fig. 2: Metrics across the benchmarking pipeline. Bold letters indicate that a form of this metric is currently implemented in QUARK.

- **Hardware:** Coherence times (qubit lifetime, dephasing time, crosstalk), gate fidelities, the quantum volume (QV), the Q-factor (quality factor) [22, 30].
- **Quantum circuit:** Expressibility (here one minus the Jensen-Shannon divergence) and entanglement (Meyer-Wallach measure) [33], normalized effective dimension [3], Fourier coefficients [27].
- **Resources:** Cost budget spent to derive a given solution, e.g., sampling complexity, number of learning steps.
- **Performance:** As a generic measure, we use the runtime or its quantum portion in a classical-quantum hybrid system. For application-specific measures, see the individual use cases in Sec. 4.
- **Problem complexity:** Scaling up the dimensionality of the system's input (e.g., resolution of an image) and output parameters, higher-order characteristics, number of input features, input preprocessing (e.g., adding noise).

The above categories can further be associated with component-level, system-level and application-level metrics, following the taxonomy of [21, 30], see Fig. 2. Hardware-related metrics may be associated with either component-level (e.g., gate fidelities) and system-level metrics (e.g., quantum volume). Our integration with the QUARK framework enables a systematic collection and analysis of these metrics across different system configurations, such as different use cases or problem sequences, see Fig. 1). Finally, any metrics aggregation approach (see e.g. Myriad-Q [6] or QuAS [23]) may be employed to arrive at a global utility score.

Tooling: The Quantum Computing Application Benchmarking (QUARK) framework [2] is an open-source platform with a focus on optimization and quantum machine learning applications [17]. Its architecture allows for a simple integration of custom use cases, as well as individual metrics and evaluation

pipelines. Users can run benchmarks locally, on quantum simulators, or on quantum computers using interfaces to hardware providers such as IBM, IonQ, or D-Wave. QUARK is organized in plugins (available as pip-packages) and provides standardized messaging interfaces for sequentially run modules. Benchmarking pipelines are constructed in an implementation-agnostic way, such that each module can be connected with another module given that the data type of their interfaces match. An example pipeline is shown in Fig. 3, in which a set cover problem is solved with a quantum annealer.

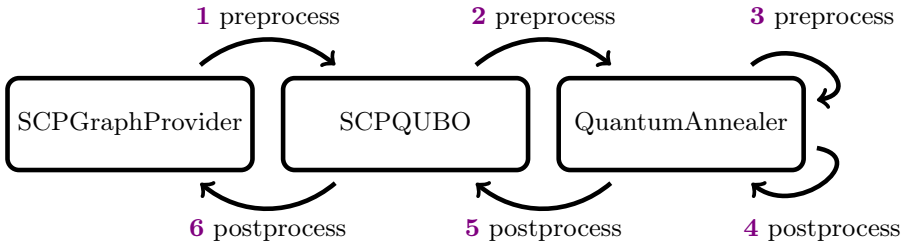


Fig. 3: The QUARK benchmarking pipeline is made up of modules that can pre- and postprocess data. In this case, the set cover problem is solved in two phases: By calling the preprocess functions of each module in order, a graph is created **(1)**, mapped to a QUBO formulation **(2)**, and solved on a quantum annealer **(3)**. Afterwards, calling the postprocess functions of each module in reverse order, the lowest energy sample is obtained **(4)**, re-interpreted as a list of nodes **(5)**, and compared with the graph **(6)**.

In BenchQC, we leverage QUARK as a platform to investigate industrial use cases. The following section gives a review of selected problems from each of the studied categories of optimization, machine learning, and simulation, along with preliminary insights from benchmarking. A full list is given in Tab. 1.

Table 1: Table of use cases in BenchQC.

Category	**Methods**	**Use case**
QML	Classification	Defect detection
	Generative models	Generative design for constructional elements
		Data augmentation for quality inspection
	QRL	6G Beam management
Optimization	Quantum Annealing (QA)	LiDAR sensor configuration
	QAOA and QA	Financial asset optimization
		Assembly line balancing
Simulation	Hamiltonian simulation	Dynamics of electrons in materials
		Low-temperature state preparation of materials
		Nuclear magnetic resonance simulation

4 Selected use cases

4.1 Optimization

Mixed-integer linear programs (MIP) represent a class of optimization problems with established solution approaches [11]. Non-linear problems pose additional computational difficulties [18]. For problems where classical solvers reach their limits, quantum methods can possibly offer solution paths with fewer computational steps than their classical counterparts [4, 35]. However, the current limitations of quantum hardware restrict their application to relatively small-scale problems. One fundamental metric for benchmarking multi-stage quantum algorithms like QAOA is the *runtime* including both training and sampling [22]. A suitable solution quality metric for the application-specific performance is the approximation ratio, which compares the best found solution relative to the optimal solution found by a classical solver. While it quantifies the gap to the optimal solution, it also remains reliable across different magnitudes of the underlying variables. When the optimal solution is unknown or hard to calculate, we may instead compare to a solution produced by a classical algorithm or even random sampling techniques (a popular metric for the latter is, for example, the *Q-Score* [31]).

The portfolio optimization problem is a fundamental challenge arising in finance. The goal is to select a set of assets that maximizes the return while minimizing the return variance (called *volatility*) [19, 29]. Due to the quadratic nature of the volatility, classical solvers scale poorly with an increasing number of assets. We first benchmark the scaling of classical solvers on different versions of the problem to identify its quantum potential. This multi-objective optimization problem can be tackled in various ways [19], like maximizing return with constrained volatility ("Maxret"), minimizing volatility with constrained return ("Minvola"), and maximizing a trade-off between return and volatility ("Multiobj"). Our analysis yields that the Minvola problem formulation requires the longest runtimes for classical optimization methods. For example, for 500 assets, the solving time of SCIP [8] to optimality is $\approx 16\times$ longer for Minvola than for Multiobj, and $\approx 62\times$ longer than for Maxret. This makes it our choice to investigate possible benefits of quantum-assisted solutions, which will be the subject of further investigations in BenchQC. In order to assess the potential of QA and QAOA, we compare the respective results to those obtained by not only exact classical optimizers but also classical heuristics like Simulated Annealing, Tabu Search and Steepest Descent. To maintain consistency across different hardware platforms, we impose a time limit of one minute for each method in order to fairly compare the solution methods on our Minvola test instances.

4.2 Machine learning

Hybrid classical-quantum neural networks (QNNs) are considered promising candidates for quantum computing [36]. We study two applications from the realms of reinforcement learning (RL) and image classification.

A challenging industrial RL problem is the *beam management* in 6G wireless communication, as next-generation antennas will produce directional beams to serve mobile users [12]. A RL agent learns to select optimal antenna indices and beam directions from a pre-defined codebook based on past interactions, without the need for the exact mobile trajectory [26]. Recent studies highlight potential advantages of quantum reinforcement learning (QRL) over classical RL [25], however, due to various stochastic influences, such conclusions have to be taken with great care. To achieve statistical significance, we hence simulate 100 independent training runs per environment instance [24], by far the largest number performed in this context to date. Using sample complexity [16] - which counts the number of interactions needed to achieve a performance threshold - we compare classical double Q-learning [34] and a quantum version of the algorithm, see left part of Fig. 4. The models of increasing complexity are: i) *small classical* neural network (2 hidden layers, 387 parameters), ii) *small quantum* circuit (4 layers on 14 qubits, i.e., 437 variational parameters), embedded in outer classical layers (with another 101 parameters), iii) *large classical* neural network (2 hidden layers, 4611 parameters). The results indicate that quantum models perform comparably to larger classical models, suggesting further improvement with increasing parameters in future research.

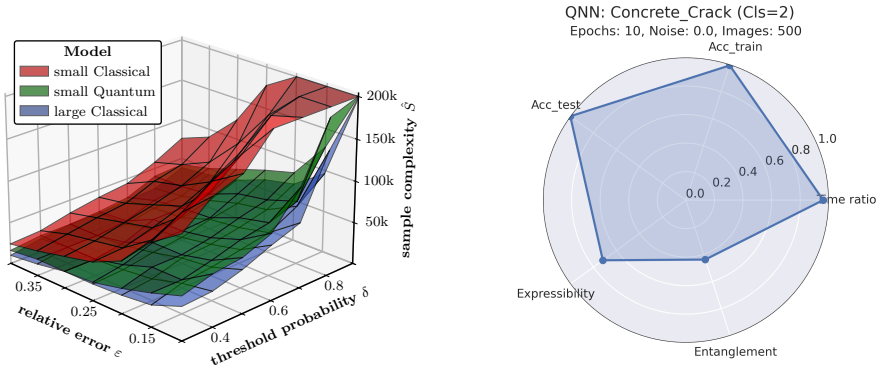


Fig. 4: Performance benchmarks for industrial QML use cases. (left) Empirical sample complexities (lower is better) of classical and quantum RL on a 6G beam management environment. $1-\epsilon$ is the chosen performance threshold, and δ is the confidence probability. Figure taken from [26]. (right) Defect detection: Selected key metrics for the performance of the classical-quantum hybrid QCNN trained on the road crack dataset.

As an example of image classification, we investigate how hybrid quantum-classical convolutional neural networks (QCNN) architectures can enhance a *defect detection* system, in particular, a classification model on a specialized dataset of road surface images with and without surface cracks [9]. The quantum part of the network is a parametrized quantum circuit with up to 8 qubits, that processes features compressed by preceding dense classical layers, which

allow to handle input data of realistic sizes. To evaluate the model performance we use the classification accuracy, while the ratio between the quantum and the overall execution time shows the compute efficiency. A comparison of the hybrid and classical models reveals several key insights, as depicted in the right part of Fig. 4: First, the quantum network achieves a high accuracy ($> 99.5\%$) on the train and test datasets, which is comparable to a purely classical implementation (not shown). Entanglement measurements (≈ 0.4) indicate a limited quantum entanglement in our circuit design, suggesting room for improvement in leveraging quantum correlations. Similarly, the expressibility score (≈ 0.7) indicates that our current circuit architecture may not fully explore the available Hilbert space. These quantum-specific metrics provide valuable guidance for circuit optimization, for example, to investigate the impact of typical quantum properties for final accuracy outcomes.

4.3 Simulation

The simulation of many-body quantum systems is one of the most obvious tasks that quantum computers can perform efficiently [14]. In quantum computing, simulation appears as a sub-routine in many applications, such as in quantum phase estimation to determine the ground state energy of molecules and reaction rates. Its direct industrial applications range from drug discovery to material discovery, enabling efficient simulations of crucial experiments such as nuclear magnetic resonance or neutron-scattering experiments.

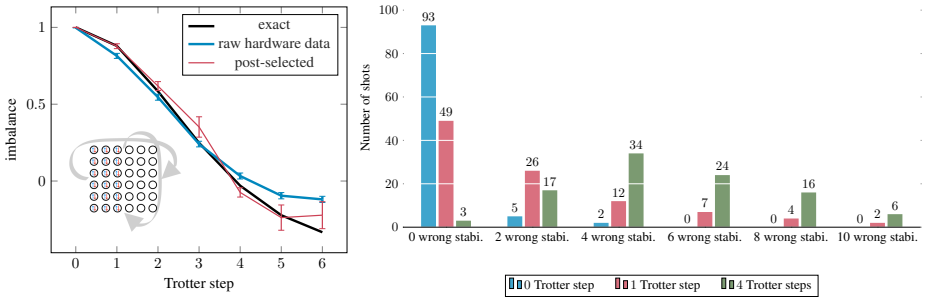


Fig. 5: Performance of Quantinuum's H2-2 ion-trap quantum computer on Bench-QC's simulation use case implemented in QUARK [2], in size 6×6, with Trotter step size $dt = 0.3$ and 100 shots per Trotter step. The circuits involve 54 qubits and around 440 two-qubit gates per Trotter step. (left) Imbalance as a function of number of Trotter steps, comparing exact value, raw hardware data, and data post-selected onto shots with 0 violated stabilizers. The raw data gets a score of 25 056 two-qubit gates, and the post-selected data a score of 59 616 two-qubit gates, which corresponds to the minimal number of gates to run on a perfect hardware to certify that the H2-2 data is incorrect. (right) Histogram of number of wrong stabilizers on hardware data, for different number of Trotter steps. A perfect noiseless hardware would have 100 shots with 0 wrong stabilizers for all Trotter steps.

As a representative industry-relevant use case we consider the simulation of the *dynamics of electrons in materials*. Our analysis starts from the Hubbard model, which is a widely studied model well known for example to describe high-temperature superconductivity in cuprates [5]. The classical computational cost to simulate the time evolution of this system is exponential in either system size for state-vector methods, or in simulation time for tensor networks methods. In order to define a benchmark protocol with an exact, classical reference for comparison, we switch off electron interactions. The evolution can then be classically simulated in polynomial time, while still involving very similar circuits to the Hubbard model, with non-Clifford gates and large entanglement. By time-evolving an initial state and measuring an observable, one can evaluate the accuracy of quantum computing hardware in a scalable and application-oriented way. To implement the fermionic statistics of the electrons, we use a fermionic encoding. This involves stabilizers that can be measured to provide further scalable information on the accuracy of the dynamics implemented on hardware. Varying the simulation time varies multiple features of the hardware panel of Fig. 2, such as the number of qubits and gate fidelity. We give in Ref. [13] a detailed description of the benchmarking protocol.

In Fig. 5, we show results obtained with a Quantinuum's H2-2 ion-trap quantum computing device. The benchmark protocol tests the dynamics of electrons on a 6×6 system, involving 54 qubits and up to $\sim$ 2500 two-qubit gates. It involves computing an imbalance of electron density between two halves of the system, shown in the left panel of Fig. 5, together with the exact, noiseless results. In the right panel, we show how many stabilizers are violated by hardware noise, directly constraining how well the electrons can be simulated. We see, for example, that after one Trotter step, only half of the shots implement the correct fermionic statistics for all the simulated electrons. The performance of the tested hardware can be evaluated with metrics of Fig. 2, e.g., runtime or number of shots to reach a given precision [13]. For example, to reach precision ϵ on all the points of the blue curve of the left panel of Fig. 5, corresponding to raw hardware data, we compute that on the machine tested one requires a runtime t_0/ϵ^2 with $t_0 = 1.1s$. This blue curve displays some difference with the exact black curve due to hardware noise. By post-selecting the shots onto zero violated stabilizers, one can remove a significant part of this noise. From the hardware results we estimate that approximately half of the shots have to be discarded per Trotter step to achieve this post-selection. This way we reach precision ϵ on all the points of the red curve of the left panel of Fig. 5, requiring a runtime t_1/ϵ^2 with $t_1 = 33.5s$. To define a metric that evaluates the precision of the output in a holistic way, we also defined a specific performance metric called the *distinguishability cost*, which helps to guide an optimal choice of quantum hardware. Given the output of a tested hardware, it quantifies the minimal number of gates to be run on an ideal perfect quantum computer to be able to detect an inaccuracy in the output [13]. The hardware data in Fig. 5 scores 25 056 two-qubit gates. This number increases with better hardware quality, as a noisy system with significant bias will require fewer shots to be distinguished from a noiseless quantum computer.

5 Conclusion

This report outlines the importance of an application-centric perspective for quantum computing benchmarking. Classical-quantum hybrid systems, as typical for NISQ, demand a holistic view on the full compute stack for meaningful benchmarking. Our concept is realized with the customizable QUARK framework. Three industry-ready use cases from the fields of optimization, machine learning and simulation are discussed. Preliminary results indicate that quantum solutions can reach competitive performance in such scenarios. Leveraging metrics across the solution pipeline, our work further identifies trends that guide the path towards possible quantum utility.

Acknowledgments

The project BenchQC is funded by grant BenchQC (DIK-2210-0011//DIK0425/02) of the Bavarian Ministry of Economic Affairs, Regional Development and Energy (StMWi). The experimental data reported in this work were produced by the Quantinuum H2-2 quantum computer, Powered by Honeywell, in August 2025.

References

1. The Quantum Consortium - Enabling the Quantum Ecosystem, https://quantumconsortium.org/
2. QUARK: A Framework for Quantum Computing Application Benchmarking (2025), https://github.com/QUARK-framework/QUARK
3. A. Abbas, D. Sutter, C. Zoufal, et al.: Nature Computational Science **1**(6) (2021)
4. A. Acuaviva, D. Aguirre, R. Peña, et al.: Benchmarking Quantum Computers: Towards a Standard Performance Evaluation Approach (2024), arXiv:2407.10941
5. D.P. Arovas, E. Berg, S.A. Kivelson, et al.: Annual Review of Condensed Matter Physics **13**(1), 239–274 (Mar 2022)
6. F. Barbaresco, L. Rioux, C. Labreuche, et al.: BACQ – Application-oriented Benchmarks for Quantum Computing (Mar 2024), http://arxiv.org/abs/2403.12205
7. A. Bayerstadler, G. Becquin, J. Binder, et al.: EPJ Quantum Technology **8** (2021)
8. K. Bestuzheva, M. Besançon, W.K. Chen, et al.: ACM Transactions on Mathematical Software **49**(2) (2023)
9. BMW: The road crack dataset (BMW Quantum challenge 2021) (2021), https://github.com/pasqal-io/bmwchallenge/tree/main
10. J. Bowles, S. Ahmed, M. Schuld: (2024), arXiv:2403.07059
11. M. Conforti, G. Cornuéjols, G. Zambelli: Integer Programming, Graduate Texts in Mathematics, vol. 271. Springer International Publishing, Cham (2014)
12. M.I. Corici, N. Franke, T. Heyn, et al.: A Fraunhofer 6G white paper: On the road to 6G: Drivers, challenges and enabling technologies. Tech. rep., Fraunhofer IIS (2021)
13. E. Granet, H. Dreyer: AppQSim: Application-oriented benchmarks for Hamiltonian simulation on a quantum computer (2025), https://arxiv.org/abs/2503.04298
14. R. Haghshenas, E. Chertkov, M. Mills, et al.: Digital quantum magnetism at the frontier of classical simulations (Mar 2025), http://arxiv.org/abs/2503.20870

15. Jean-Francois Bobier, Matt Langione, Edward Tao, et al.: What Happens When 'If' Turns to 'When' in Quantum Computing? (2021), https://www.bcg.com/publications/2021/building-quantum-advantage
16. M. Kearns, S. Singh: Finite-sample convergence rates for Q-learning and indirect algorithms. In: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II. p. 996. MIT Press, Cambridge, MA, USA (1999)
17. F.J. Kiwit, M.A. Wolf, M. Marso, et al.: KI - Künstliche Intelligenz **38**(4) (2024)
18. J. Lee, S. Leyffer: Mixed integer nonlinear programming, vol. 154. Springer Science & Business Media (2011)
19. Z.X. Loke, S.L. Goh, G. Kendall, et al.: IEEE Access **11**, 33100–33120 (2023)
20. J.M. Lorenz: Modern Physics Letters A **39**(21n22), 2430006 (Jul 2024)
21. J.M. Lorenz, T. Monz, J. Eisert, et al.: Systematic benchmarking of quantum computers: status and recommendations (2025), http://arxiv.org/abs/2503.04905
22. T. Lubinski, S. Johri, P. Varosy, et al.: Application-Oriented Performance Benchmarks for Quantum Computing (Jan 2023), http://arxiv.org/abs/2110.03137
23. K.J. Mesman, W.v.d. Schoot, M. Möller, et al.: QuAS: Quantum Application Score for benchmarking the utility of quantum computers (Jun 2024), http://arxiv.org/abs/2406.03905
24. N. Meyer, C. Ufrecht, M. Periyasamy, et al.: Qiskit-Torch-Module: Fast Prototyping of Quantum Neural Networks. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). vol. 1, p. 817 (2024). https://doi.org/10.1109/QCE60285.2024.00101
25. N. Meyer, C. Ufrecht, M. Periyasamy, et al.: A Survey on Quantum Reinforcement Learning. arXiv:2211.03464 (2022). https://doi.org/10.48550/arXiv.2211.03464
26. N. Meyer, C. Ufrecht, G. Yammine, et al.: Benchmarking Quantum Reinforcement Learning. In: 2025 Forty-Second International Conference on Machine Learning (ICML). PMLR (2025). https://doi.org/10.48550/arXiv.2501.15893
27. M. Monnet, N. Chaabani, T.A. Dragan, et al.: Understanding the effects of data encoding on quantum-classical convolutional neural networks (2024), http://arxiv.org/abs/2405.03027
28. C.A. Riofrío, J. Klepsch, J.R. Finžgar, et al.: Quantum Computing for Automotive Applications (Mar 2025), arXiv:2409.14183
29. W. Sakuler, J.M. Oberreuter, R. Aiolfi, et al.: A real world test of Portfolio Optimization with Quantum Annealing (2023), http://arxiv.org/abs/2303.12601
30. W.v.d. Schoot, R. Wezeman, P.T. Eendebak, et al.: Quantum Information Processing **22** (2023)
31. W.v.d. Schoot, R. Wezeman, N.M.P. Neumann, et al.: Extending the Q-score to an Application-level Quantum Metric Framework (2024), http://arxiv.org/abs/2302.00639
32. F.J. Schreiber, J. Eisert, J.J. Meyer: Physical Review Letters **131**(10) (2023)
33. S. Sim, P.D. Johnson, A. Aspuru-Guzik: Advanced Quantum Technologies **2** (2019)
34. H. Van Hasselt, A. Guez, D. Silver: Deep Reinforcement Learning with Double Q-Learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 30, p. 2094 (2016)
35. E. Wybo, M. Leib: Missing Puzzle Pieces in the Performance Landscape of the Quantum Approximate Optimization Algorithm (2024), http://arxiv.org/abs/2406.14618
36. R. Zhao, S. Wang: A review of Quantum Neural Networks: Methods, Models, Dilemma (Sep 2021), http://arxiv.org/abs/2109.01840

BACQ - Application-oriented Benchmarks for Quantum Computing

Frédéric Barbaresco[1], Félicien Schopfer[2], Emmanuelle Vergnaud[3], Laurent Rioux[1], Christophe Labreuche[1], Michel Nowak[1], Noé Olivier[1], Damien Nicolazic[4], Olivier Hess[4], Anne-Lise Guilmin[4], Robert Wang[4], Tanguy Sassolas[5], Stéphane Louise[5], Kyrylo Snizhko[5], Grégoire Misguich[5], Alexia Auffèves[6], Robert Whitney[6]

[1] THALES, FRANCE
[2] LNE, FRANCE
[3] TERATEC, FRANCE
[4] BULL SA, FRANCE
[5] CEA, FRANCE
[6] CNRS, FRANCE

frederic.barbaresco@thalesgroup.com; felicien.schopfer@lne.fr; emmanuelle.vergnaud@teratec.fr

Abstract. Delivering an application-oriented benchmark suite for objective multi-criteria evaluation of quantum computing performance, a key to industrial uses. With the support of the national program on measurements, standards, and evaluation of quantum technologies MetriQs-France, a part of the French national quantum strategy, the BACQ project is dedicated to application-oriented benchmarks for quantum computing. The consortium gathering THALES, BULL SA (previously EVIDEN), CEA, CNRS, TERATEC, and LNE aims at establishing performance evaluation criteria of reference, meaningful for industry users

Keywords: Quantum Computer, Quantum Algorithm, Quantum Emulator, Quantum Annealer, NISQ, FTQC, Benchmark, Multi-Criteria Decision.

1 Introduction

Quantum computing promises to revolutionize multiple technical fields and activity sectors, from optimization in logistics, to simulation for research in physics or chemistry, engineering or industry, passing through cryptography. Measuring the progress towards the quantum advantage and the realization of such promises, with objectivity and reliability, is of high interest for potential end-users and crucial for the future development of the domain, now subject of hype and high competition. The challenges, especially to achieve comparable measurements, comes from the diversity of the hardware platforms, their specificities in terms of physical characteristics and applications, their maturity that can still be low, and the potential rapid evolution of the technologies.

Dedicated to the whole value chain setting up from quantum hardware development to industrial use cases, BACQ is complementary to the benchmarking initiatives only

focusing on low-level hardware physical criteria. The envisioned benchmark suite will be based on the resolution of several classes of problems covering important application domains of quantum computing which are meaningful for industrial users (see Figure 1): simulation of Quantum Physics models, Optimization, Linear System Solving and Prime Factorization. Machine Learning could be included in Optimization application domain.

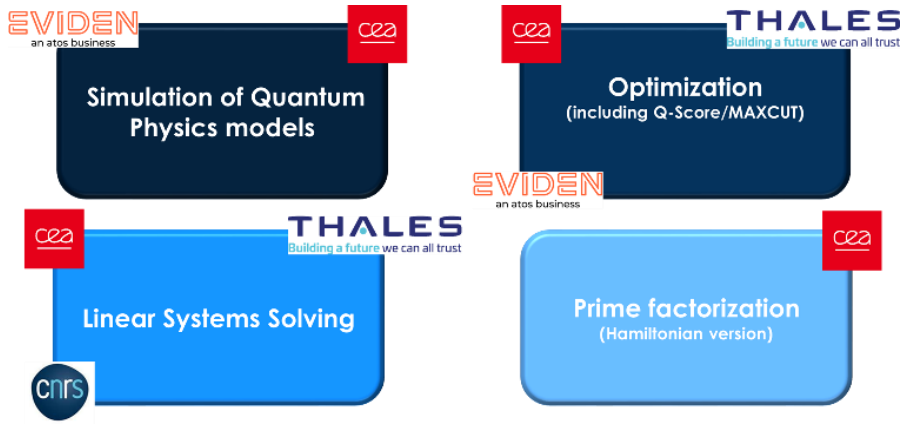


Fig. 1. Families of problems used for the QPUs benchmark

These problems are generic and could be relevant for different branches of industries and services (chemistry, aeronautics, electronics and energy for example). Criteria will be defined for the resolution of each problem, some being hardware-agnostic and others hardware-dependent (low level): computation time, latency, problem size, approximation rate, resolution probability, accuracy, fidelity… Importantly, the project also considers energetic criteria for the evaluation of the energetic performance of the machines.

The proposed methodology consists in the aggregation of low-level technical metrics and a multi-criteria analysis via the tool MYRIAD-Q in order to provide operational performance indicators of the different quantum computing solutions and point out the service qualities of interest to the end-users. The aggregation of the criteria and multi-criteria analysis allows fully explainable and transparent notations, comparisons between different quantum machines and with classical computers, as well as identification of the practical advantages of each quantum machine with respect to specific applications. The project will address both analog machines (quantum simulators and annealers) and gate-based machines, Noisy Intermediate Scale Quantum (NISQ) and Fault Tolerant Quantum Computing (FTQC). The followed practical approach is to have a suite of benchmarks, adaptive to some extent, appropriate to the capabilities of the available machines and able to demonstrate their respective advantages, including, in the longer term, exponential speedup of specific algorithms on FTQC machines.

As part of the project, a first action has already been launched regarding Q-Score, where Eviden developed on the MAXCUT optimization problem, to test and validate the benchmark on different types of quantum machines.

Sharing of the benchmark suite, as broadly as possible, is an important objective to establish common reference measurement methods and ensure no bias with inclusion of all the technologies. Consultations of the technology providers, as well as end-users is

key to develop an instrument that meets the needs. Once specified and developed, the benchmark suite will be available to the community for its use.

The BACQ project, with 3-year duration and almost 4 M€ budget, including 2.5 M€ funding by France 2030 via ANR within MetriQs-France. This project precisely aims at developing, exploiting and promoting a reliable, objective and long-lasting measurement instrument for measuring the performance of quantum computing.

2 Objectives

2.1 MetriQs-France Program

MetriQs-France is the French National Program on Measurement, Evaluation, and Standards for Quantum Technologies. MetriQs-France objective is to develop, exploit, and promote measurement capabilities of reference, validated and harmonized, for characterization and performance assessment of quantum technologies with reliability, impartiality and comparability: metrology, test & evaluation, international standardization are at the core of the program. MetriQs-France is coordinated by LNE, the French national metrology and testing laboratory, a public institution under the supervision of the French Ministry for Economy and Finance, in charge of industry.

2.2 BACQ project

While looking to the future, it is critical to measure the progress towards the realization of the quantum computing promises. Application-oriented benchmarks, enabling to evaluate the real performance of quantum computing from the user perspective appear to be useful within this perspective. The challenge comes from the diversity of the hardware platforms, their specificities in terms of physical characteristics and applications, their maturity and the potential rapid evolution of the technology. Developing an objective, long-lasting, widely shared measurement instrument, to serve as a common reference.

The evaluation of the practical quantum computing performance will be considered through benchmarks close to real applications, meaningful for industrial (as well as academic) end-users. The main objective is to measure the progress towards a practical quantum advantage. In that regard, the consortium plans to perform comparisons between the different quantum computing solutions as well as to compare them with current classical computers. This benchmarking initiative will eventually highlight the assets of each quantum computing solution given specific applications.

The benchmark suite resulting from this project will be maintained by LNE, an independent and trusted third party. Thanks to their interaction with the community of end-users, this tool will be operated in order to analyze the results obtained from the machines tested with the different benchmarks using the explainable aggregation tool. LNE will establish a list of performances, maintain it over time, and update the definition of the tests. Furthermore, adoption of this initiative by French, European and eventually international partners will be encouraged thanks to the development of communication tools

regarding the approach taken and visual representations of the aggregation of the results obtained by the machines from the different benchmarks. International dialogues and collaborations on the subject of quantum computers benchmarking will be promoted so that the approach, supported by MetriQs-France, is and remains an international reference. This BACQ project promotes the development of an international standardization in regard of the methods used to evaluate the specifications of quantum machines.

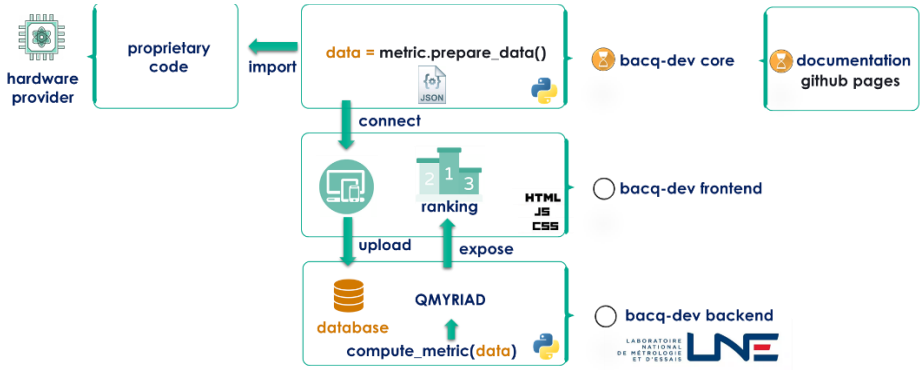


Fig. 2. BACQ Benchmark DevOp

In view of fulfilling these objectives, BACQ has initiated collaborations and dialogue with other international benchmarking initiatives: Fraunhofer IKS & FOKUS (BenchQC Project) in Germany, TNO & TU Delft (QPack Project) in The Netherlands, Qilimanjaro (CUCO Project) in Spain, QuIC (Use cases WG) in Europe, HamLib Project (Intel, LBNL et al.) in US, Unitary Fund (Metriq project) in US and QED-C (QC Benchmarking WG) in US.

3 Benchmark tests definition

As seen in the introduction, the set of benchmarks is divided into four categories: Optimization problems (including, but not limited to, EVIDEN's MaxCut QScore), linear system solving, quantum physics simulations and prime factorization. To this end, the set of problems constituting each benchmark family should be viewed as representative enough of the field (although pursuing exhaustivity would be out of the scope of the project at this time).

In order to design useful and relevant benchmark programs, it is necessary to adhere to a set of important principles.

As current quantum machines have a relatively high level of noise, it is of uttermost importance to compare the measurements we produce against a pure random machine. As these QPUs are good random generators, the usefulness of quantum computing that would be undistinguished from random results must be considered as null.

Every family of benchmark should provide at least two different applications per family whenever it makes sense (i.e. except for Prime Factorization which describe one and only one possible application, even though the implementation can still vary),

As the strong points of a given quantum processor can differ from another, the exact algorithm would better be fitted to match these strong points for a given application. Nonetheless, the design and specification process must be accurately and thoroughly followed and described so that it will be easier to create versions adapted to new hardware platforms. One simple illustration of that can be illustrated by the fact that quantum annealers (e.g. D-Wave QPUs) cannot apply the same algorithm as gate-based QPUs.

Last but not least, each design processes toward the implementation in software of each benchmark program must be thoroughly detailed so that it must be easy for a third party to re-implement the benchmark program from its documentation but also to design easily a new implementation for a new QPU.

The following sections describe each of the four benchmark families.

1) Optimization Benchmarks

Optimization is a field for which quantum computing is believed to be the most apt at for short to mid-term industrial applications. Nonetheless, it is also believed that BQP, the set of problems solvable in polynomial time on a quantum computer is different from NP, therefore one should not have unrealistic expectations from quantum computing approaches. However, it is a domain of application for quantum computing of the highest importance, because it may improve the time to solution, or the quality of approximate solution, or, even more importantly, the energy required to achieve a given level of quality for a solution.

The family of optimization problems and programs will be articulated around 3 to 4 sets of problems:

- MaxCut (as part of Eviden's QScore), as an example of NP-hard quadratic binary unconstrained model,
- Maximum Cardinality Matching, as an example of Polynomial constrained problem which is famously hard for annealing approaches,
- Higher Order Binary Optimization (HOBO) problems, which are not necessarily NP-hard, but will probe the case of non-quadratic problems which can be important in the general case,
- Optionally, some NP-hard constrained problems.

The Maximum Cardinality Matching family of problems is known to be only polynomial but they laid the base to demonstrate the worst-case convergence of Simulated Annealing. CEA-LIST was pioneer in utilizing the G_n series of problem to probe the capabilities of D-Wave Quantum Annealers. This work has since been extended by other research teams in the world. Work is currently underway to extend the study to the latest iterations of D-Wave QPUs and expose the best strategies toward obtaining the best results. These results should be easily extended to other Quantum Analog QPUs (e.g. Pasqal) and QAOA approaches to solve optimization problems.

HOBO problems are important because, while it is always possible to convert them to quadratic problems by introducing ancilla variables (a process called quadratization), doing so introduces many complications for the transformed problems. Therefore, QPUs

that allow for avoiding such transformations would take an increased edge against QPUs that would necessitate a quadratization of the problem before solving.

2) Linear Systems Benchmarks

Linear algebra in general provides a set of problem that are as important as optimization problems in term of applications, and is a field where ideal quantum computing has a known exponential speedup. Although it is clear that no short term exponential computing advantage can be expected, it remains an important field to investigate.

To that end, two sets of applications stand out:

- Linear system solving, which is the set for which HHL will be the most relevant algorithm for gate-based QPUs.
- Eigenvalues – eigenvectors of a given matrix, as it is a family of application in which some hybrid algorithms like VQE could be applied.

For the latest, VQE is already known as a family of algorithms which can be a potential short term application of quantum computing.

3) Quantum physics simulations Benchmarks

Quantum physics problems arise in various fields of science. For instance, in solid-state physics the understanding of new quantum phenomena and the design of new materials with potential applications often require to simulate some quantum problems with many electrons. Quantum physics with many electrons also play a central role in chemistry. They are also central problems in other domains of basic research, such as in particle or nuclear physics an even quantum information theory. Among the simple but yet nontrivial many-body Hamiltonians defined on lattices one can mention quantum spin-1/2 models (Heisenberg interactions, Ising model in transverse field, …) spinless fermion models, spin-1/2 fermion models (Hubbard model) or boson models (ex: Bose-Hubbard). Simulating these quantum many-body systems is notoriously difficult on a classical computer, for the same reason as a quantum computer is more powerful than a classical one, namely the Hilbert space dimension of a quantum system grows exponentially with the number of particles. However, with an analogue quantum simulator or with a digital (gate based) quantum computer, one would in principle be able to study many-body systems which are un-tractable on a classical machine. In that context one can mention two distinct tasks: 1) from a given Hamiltonian H, determine the expectation values of some observables (including H itself) in the ground-state of H. 2) Given a simple (product) initial state $|\psi_0\rangle$, determine the expectation values of some observables in the time-evolved state $|\psi_t\rangle = \exp(-iHt/\hbar)\,|\psi_0\rangle$. The ground-state problem (1) can for instance be tackled using VQE on a gate-based machine or via a (quasi) adiabatic evolution on an analogue simulator. The dynamical (also called quantum quench problem (2) can be studied via a discretization of the time evolution (Trotterization) on gate-based architectures, and somehow more directly with the native Hamiltonian dynamics on an analogue device (provided the interactions between the qubits can be tuned to match the Hamiltonian H of interest).

4) Prime Factorization Benchmarks

The discovery of Shor's algorithm for prime factorization is one of the results which generated immense expectations on the part of quantum computing. While it is clearly unrealistic to expect that this particular algorithm can be put to applications without having first developed fault-tolerant quantum computers, there exists other possible approaches. One is utilizing Hamiltonian based approaches. The basic principle is rather simple because it simply requires the minimization of the following cost function:

$$C(p_1, p_2) = (N - p_1 p_2)^2$$

Where p_1 and p_2 are 2 possible prime numbers that ought to be found and N is the integer which is the subject of the prime factorization.

Approaches that are not based on Shor's algorithm are not expected to provide any kind of computational advantage, it can however not be ruled out at the moment that it can provide some kind of advantage on the energy side.

4 Energy criterion by resource estimation for the resolution of simple algorithms on gate-based computers

For any given task, an energy efficiency can be parametrically defined as the metric chosen to quantify the performance of the task, divided by its energy cost. The efficiency is thus tightly dependent on the choice of metric, as well as the list of energy costs.

The objective of the work package is to propose and study the behavior of such energy efficiencies for some of the useful algorithms studied in this project. In principle, all machines can give rise to such figures of merit. Our approach will rely on the MNR methodology introduced, which consists in finding relations between Noise, Metric of performance and Resource cost, then exploit these relations to minimize the resource cost under the constraint of a given performance.

In the spirit of the BACQ project, we will care to remain as hardware-agnostic as possible to propose useful and flexible numerical tools to end-users and industrials. The focus shall thus be put on algorithmic resources, the connection with energy costs being handled via effective models. In close connection with the HQI project of the French quantum strategy, we will build a user-friendly interface accessible to any hardware provider can use to estimate the energy efficiency of the quantum processor at play.

5 Multi-Criteria Approach: MYRIAD-Q

The construction of the Multi-Criteria model is organized around three steps with MYRIAD-Q tool.

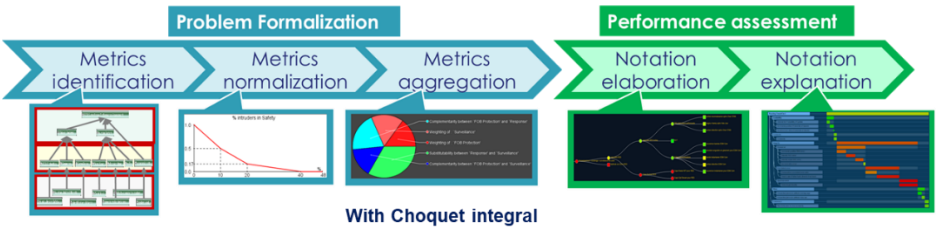


Fig. 3. : MYRIAD-Q Multicriteria Agregation Tool

1) 1st step: structuring phase

Given the plurality of QPUs, we first need to identify the subsets of QPUs that make sense to be compared. A MYRIAD-Q model is then constructed for each subset. For instance, it might not be relevant to compare Quantum Annealers with NISQ.

Once we have selected the scope of a MYRIAD-Q model, the multi-criteria approach focuses firstly on identifying operational metrics which have a real operational impact (fundamental objectives for experts) and not technical metrics which only have an indirect impact. Then the associated criteria are organized hierarchically.

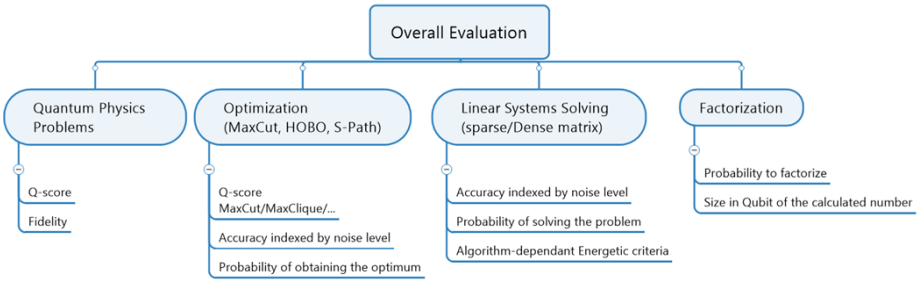


Fig. 4. : Sketch of a hierarchy of criteria

2) 2nd phase: normalization

It does not make sense to directly combine the metrics as they are given in different units. The aim of the marginal utility functions is to normalize the metrics and transform them in the same scale. This normalization corresponds to an *interval scale* in the terminology of measurement theory. An interval scale is given up to an affine transformation. One thus needs to define two reference levels to entirely fix it. For scale $S = [0, +\infty)$, we introduce the following two reference elements: B_i (it stands for Bad) is an element of metric X_i that is judged not satisfactory at all, and G_i (Good) is an element of metric X_i that is judged *satisficing*. This latter level is a central ingredient of the Bounded Rationality of Simon. It is a contraction of satisfying and sufficient. The bad element is the saturation level from below (no element has a worst utility); and there exists elements that are preferred to the satisficing level. In order to ensure commensurateness among criteria, we impose that all reference levels of the same nature across the criteria are assigned to the same utility.

3) 3rd phase: aggregation

As already mentioned, the aggregation function F is organized hierarchically. The first phase returns a tree described by a set M of nodes including a root $r \in M$ and by a function $P: M \to 2^M$ returning the set of parents $P(k)$ for every node $k \in M$. The leaves of the tree are the set N of criteria. An aggregation function $F_k: S^{P(k)} \to S$ is defined at each aggregation node $k \in M \setminus N$.

One needs to elicit each aggregation function F_k. The simplest aggregation function is the weighted sum $F_k\left(a_{P(k)}\right) = \sum_{l \in P(k)} w_l a_l$, where w_l is the weight of node l. This model cannot capture real-life decision strategies as it assumes that all its inputs are independent.

In order to overcome the limitations of this model, we propose to use the Choquet integral, which has the capacity to represent interactions among criteria. Its so-called 2-additive version is a good compromise between simplicity and expressivity of the model. It takes the form

$$F_k\left(a_{P(k)}\right) = \sum_{l \in P(k)} w_l a_l + \sum_{\{l,m\} \subseteq P(k)} \left(w_{l,m}^{\wedge} min(a_l, a_m) + w_{l,m}^{\vee} max(a_l, a_m) \right)$$

where coefficients $w_l, w_{l,m}^{\wedge}, w_{l,m}^{\vee}$ are non-negative coefficients that sum-up to one. Compared to the weighted sum, the min terms represent *complementarity* between the two criteria l, m (necessarily both scores have to be high to produce a high overall score; in other words, if one score is larger than the other one – say $a_l > a_m$ – then the better evaluation on criterion l is penalized by the worst score on criterion m). Likewise, the max terms represent redundancy between the two criteria l, m (it is sufficient that one score is high; in other words, if one score is larger than the other one – say $a_l > a_m$ – then the bad evaluation of criterion m is saved to a degree $w_{l,m}^{\vee}$ by the better evaluation on criterion l.

6 Conclusion

The benchmark tools and methodologies developed in the BACQ project will be used to help select the QPUs for the PROQCIMA program (notified DGA framework agreements to five companies, Alice&Bob, C12, Pasqal, Quandela and Quobly, with a view to identifying solutions enabling the development of these computers). The PROQCIMA program will last for at least 10 years. After four years, only the three most successful projects will continue the program to develop the best logical qubits capable of scaling up. After eight years, the competition will be limited to the two most successful technologies, which will continue the program to move from computer prototypes (target: 128 logical qubits) to industrial products usable by their first customers (target: 2,048 logical qubits).

References

1. BACQ: Application-oriented Benchmarks for Quantum Computing, arXiv:2403.12205v1 [quant-ph]; https://doi.org/10.48550/arXiv.2403.12205
2. Bench-QC: Application-driven benchmarking of quantum computers, Fraunhofer Institute IKS, Germany; https://www.iks.fraunhofer.de/en/projects/bench-qc-application-driven-benchmarking-of-quantum-computers.html
3. QPack: Benchmark for quantum computing. Delft University of Technology, Delft, The Nederlands; https://github.com/koenmesman/QPack?tab=readme-ov-file
4. CUCO project "Computacion Cuantica En Industrias Estrategicas", Spain; https://www.cuco.tech/en/project/
5. Metriq "Community-driven quantum benchmarks", Unitary Fund; https://metriq.info/
6. QED-C benchmark; https://github.com/SRI-International/QC-App-Oriented-Benchmarks
7. DARPA QBI (Quantum Benchmarking Initiative); https://www.darpa.mil/news-events/2024-07-16 ; https://www.darpa.mil/news-events/2024-08-15
8. HamLib: A library of Hamiltonians for benchmarking quantum algorithms and hardware,in 2023 IEEE International Conference on Quantum Computing and Engineering (QCE), 2023. arXiv:2306.13126v3 [quant-ph]; https://doi.org/10.48550/arXiv.2306.13126
9. HamPerf: A Hamiltonian-Oriented Approach to Quantum Benchmarking, CF '24 Companion: Proceedings of the 21st ACM International Conference on Computing Frontiers: Workshops and Special Sessions, Pages 133 – 138, https://doi.org/10.1145/3637543.3653431
10. E. Vergnaud, F. Schopfer, F. Barbaresco, 1st TQCI Seminar "Overview of upcoming application-oriented benchmarks for quantum computing in France and abroad", Thales Research & Technology, Palaiseau, France, 11th May 2023; https://teratec.eu/Seminaires/TQCI/2023/Seminaire_TQCI-230511.html
11. E. Vergnaud, F. Schopfer, F. Barbaresco, 2nd TQCI Seminar "Overview of upcoming application-oriented benchmarks for quantum computing in France and abroad", 4th & 5th June 2024, Reims, France; https://teratec.eu/activites_quantiques/TQCI_240604_programme.html
12. E. Vergnaud, F. Schopfer, F. Barbaresco, 3rd TQCI Seminar "benchmarks for quantum computers", 24th & 55th June 2025, Palaiseau, France; https://teratec.eu/Seminaires/TQCI/2025/Seminaire_TQCI-250625_UK.html
13. UK Quantum Computing Metrics Initiative, UK Quantum Metrology Institute of National Physical Laboratory: https://www.npl.co.uk/quantum-metrology-institute
14. Basque Quantum Benchmark, Benchmarking Quantum Computers: Towards a Standard Performance Evaluation Approach, https://arxiv.org/pdf/2407.10941v1
15. GIFAS, Rapport sur les technologies quantiques pour les applications Aéronautique-Spatial-Défense, Groupe de Travail Technologies Quantiques, 2024, Paris
16. Académie des technologies, État de l'art de l'ordinateur quantique tolérant aux fautes – Questions et défis, 2025, Paris ; https://www.academie-technologies.fr/wp-content/uploads/2025/06/202505_ordinateur_quantique.pdf
17. EQCBC (European Quantum Computing Benchmarking Committee), Systematic benchmarking of quantum computers: status and recommendations; ; https://doi.org/10.48550/arXiv.2503.04905
18. Tobias Rohe & al, Quantum Computer Benchmarking: An Explorative Systematic Literature Review, https://doi.org/10.48550/arXiv.2509.03078
19. Thorsten Koch & al, Quantum Optimization Benchmarking Library - The Intractable Decathlon, https://doi.org/10.48550/arXiv.2504.03832

Quant²AI – An End-to-End Quantum AI Benchmarking Framework for Both Researchers and Practitioners

Cristobal Corvalan Morbiducci[1], Pascal Halffmann[2[0000−0002−3462−4941]],, Andrew Barlow[1], Alexander Geng[2[0000−0002−9955−4809]], Lautaro Hickmann[1], Ali Moghiseh[2[0000−0001−9126−3495]], Sabine Müller[2], Christine Priplata[3], Hans-Martin Rieser[1], Colin Stahlke[3], and Michael Trebing[2]

[1] DLR Institute for AI Safety and Security, Wilhelm-Runge-Straße 10, 89081 Ulm, Germany
cristobal.corvalanmorbiducci@dlr.de
https://www.dlr.de/en/ki
[2] Fraunhofer Institute for Industrial Mathematics ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany https://www.itwm.fraunhofer.de/en.html
[3] Conet Solutions GmbH, Bundeskanzlerplatz 2, 53113 Bonn, Germany
https://www.conet.de/EN/conet

Abstract. Quantum AI and Quantum Machine Learning (QML) are among the most promising and dynamic research fields, with a vast variety of QML models. However, standardized benchmarking is lacking, and end users often struggle to determine whether quantum AI — and which specific approach — is suitable for their use cases. Addressing these challenges is essential to evaluate the current state of quantum AI and advance toward quantum utility.

We have developed *Quant²AI*, a holistic benchmarking framework for systematic comparisons of quantum AI pipelines using high performance clusters and both quantum simulators and hardware. Our end-to-end approach evaluates not just QML models but the whole pipeline, including e.g., preprocessing and hyperparameter variations. Its modular design enables easy integration of new components, such as alternative data preparation. We provide standardized and real-world datasets, quantum and classical AI reference pipelines, state-of-the-art evaluation metrics, and intuitive visualizations. Our framework offers benchmarking as a service for both researchers for testing their newly developed quantum AI components as well as end users for an intuitive way to identify promising quantum AI applications.

Keywords: Quantum AI · Benchmarking · Quantum Machine Learning · Use Cases.

1 Introduction

Quantum Machine Learning (QML) — harnessing the computational capabilities of quantum systems to improve machine learning (ML) models [5] — has

witnessed rapid advancements and substantial attention in recent years [3]. Theoretical propositions of quantum advantage suggest that QML could outperform classical models in training speed or quality [13], however, these prospects remain speculative: Critics point to the limited empirical evidence supporting substantial performance gains over classical baselines, particularly when idealized scenarios for quantum computers are considered [15]. As such, while the theoretical landscape of QML is rich with promise, its actual effectiveness in real-world applications remains largely uncharted. Rigorous benchmarking of QML pipelines [4] is thus essential to determine their practical utility and distinguish genuine quantum utility or even advantage from hype-driven expectations.

1.1 Previous Work

Benchmarking for classical machine learning methods [16] are numerous; an overview is beyond the scope of this paper, as also the major ML packages like Scikit-Learn, PyTorch and TensorFlow provide benchmarking suites. In the quantum realm, benchmarking has mostly been targeted at hardware with e.g., Supermarq [17]. As NISQ hardware becomes increasingly more viable for QML, it has increased the necessity for rigorous and reproducible benchmarking of machine learning methods on quantum devices with first approaches focusing on classification [2] or real-world use cases [10]. Recently, several frameworks have been developed: Besides the QML benchmarks provided by packages such as PennyLane, there exists the QUARK framework [6,11] focusing on application oriented benchmarks for general quantum methods. The MQT Bench framework [12] provides an easy-to-use library for benchmarking a wide variety of quantum tools and algorithms including QML methods like QNN. Automated quantum machine learning (AutoQML) [14] is a framework based on AutoML, which is a unifying interface for deployment of end-to-end QML pipelines. Most recently, BenchQC [7], a general framework to benchmark quantum methods in a modular fashion, also for optimization, has been published.

1.2 Our Contribution

We propose *Quant²AI*, a framework that encompasses the advantages of the aforementioned benchmarking frameworks for a rigorous, standardized, and reproducible benchmark of QML end-to-end pipelines. Our framework allows for an end-user friendly and fast execution on high performance clusters (hpc) using both simulators and real quantum devices. While we provide a set of quantum and classical reference pipelines comprising of standalone modules, both standard and real-world datasets, and evaluation metrics, users can not only just add their own components, but also fully customize the existing ones, including adding new quantum devices. Therefore, end-users can evaluate their use cases to determine any quantum utility, whereas researchers can seamlessly test their new components or whole pipelines.

Section 2 introduces the Quant²AI framework and its components, while in Section 3, we showcase the execution of a benchmark for a QML pipeline.

2 Benchmarking Framework

Robust benchmarking is a step to gain an understanding of the potential advantages, limitations, and applications of quantum machine learning algorithms relative to the classical case. In the following, we present the key ingredients of our benchmarking framework: We discuss the foundations and hpc integration in Section 2.1, followed by the presentation of modular end-to-end pipelines in Section 2.2. In the next section (Section 2.3), we introduce the integration of quantum hardware and conclude with our evaluation module and visualization tool in Section 2.4.

2.1 Foundation and HPC Execution

The framework follows the basic idea of allowing non-hpc experts to benchmark complete execution pipelines on cluster systems. Therefore, the actual interaction between user and hpc-backend is reduced to a minimum by different abstraction layers, all of them modular while being straightforward to exchange. The general workflow reflects this regime: The user provides a pipeline configuration, as well as general information about the cluster environment, e.g. the paths where the code and the data resides, to the top most abstraction layer: a python interface that automatically prepares the necessary job scheduling scripts. Here, we currently offer script generation for SLURM. However, tailoring job scheduling e.g. to Portable Batch System (PBS) is reduced to an exchange of this module.

As soon as the job is started, job dependent structures are prepared and all necessary information for a proper job execution are assembled: First, output folders for job-specific information are generated and necessary python environments are locally installed (in case of present local SSD hard drives on node-side). This environment ensures that on the framework side we provide and store the important information of the jobs so that reproducibility of results can be ensured. Once the environment is completely set, the pipeline starts execution. Based on a unique identifier, results as well as general log information about the job run itself combined with pipeline specific log information, are stored in these job-specific output folders. Generally, pipeline modules are executed sequentially, see Section 2.2. Once the pipeline is finished, relevant information is collected and stored in the parallel file system in the locations specified in the first step. A database incorporates and processes this information for the visualization of the collected metrics (Section 2.4). A schematic view of the framework with further details is given in Figure 1.

2.2 Pipelines and Modules

The implementation of pipelines is built around a modular design, enabling flexibility and re-usability across various machine learning and quantum machine learning tasks. The standard pipeline consists of five key modules, see Figure 1, but users are free to adjust the number and purpose of the modules. The framework already provides implementations:

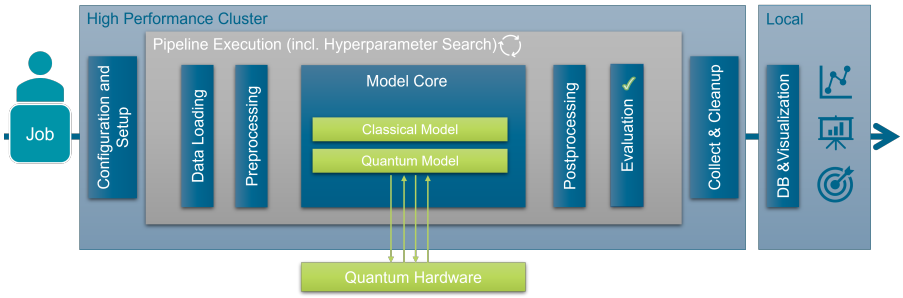


Fig. 1. Abstract visualization of the framework structure. The user submits a job consisting of the pipeline and configuration files (cluster-specific requirements and pipeline information such as hyperparameter values, loss function,...). A script is generated via SLURM that builds output folders and sets up the necessary python environments among others. This is a fresh python installation without any artifacts from previous runs. Once done, the pipeline is executed module by module. The most important module is the model core, where a classical, quantum, or hybrid model is executed. In the latter two cases, the quantum part can be executed on a simulator or on a real quantum device using a connection module. Information from the model core (and the quantum device connection module) is collected to compute suitable metrics. If hyperparameter search is invoked, multiple of these pipeline executions runs are carried out. After the pipeline execution, a cleanup process in invoked and the output is stored, ready to be processed by the visualization tool and its database on a local device.

- Data Loading: Prepares and partitions the dataset for training, validation, and testing. Example: Data loader to load and prepare the iris dataset.
- Preprocessing: Prepares the dataset for the model, e.g., using Principal Component Analysis (PCA).
- Model Core: Implements training of the model, also on a real quantum device, see Section 2.3. Example: Classical and quantum SVM, k-means.
- Postprocessing: Counterpart to the preprocessing (e.g., PCA), to reinterpret the results for the original data.
- Evaluation: Evaluates the performance of the pipeline, see Section 2.4.

Pipelines are constructed by chaining these modules together, with configuration and data passed between them via a centralized `JSON` file. This design enables logical interconnection between compatible modules while also enforcing consistency. Each module contains internal validation logic that checks the `JSON` input for structural and semantic correctness. This ensures that incompatible module combinations — such as applying an image processing pipeline to tabular data — are flagged and prevented early in the workflow making the integration of new modules seamless.

Built-in reference pipelines are provided for a range of use cases including anomaly detection, binary image classification and segmentation, clustering, multi-class classification, and regression – with two datasets for each use case available in the framework. These datasets range from famous toy examples like

iris or MNIST to real-world data and application such as identifying cracks on images of concrete or anomaly detection on warranty claims for purchased cars. Suitable data loaders preparing the data are provided. The reference pipelines serve as practical examples of how to structure and connect modules for various machine learning objectives. The user has full flexibility in building or customizing pipelines, for example adding or deleting modules, e.g., add a second preprocessing module or build a new data loader. This architecture ensures a high degree of modularity, robustness, and extensibility, suitable for a wide range of classical and QML applications.

2.3 Quantum Hardware Connection

The framework provides a quantum connection (qc_connect) module which handles communication between the running experiments and the quantum backend. Communication is mainly achieved through the widely used openQASM representation of quantum circuits. Therefore, the qc_connect module supports appropriate creation, interpretation, and storing of these representations for later usage.

The openQASM format allows for a "universal" wide incorporation of quantum computers of different hardware providers. In the same way, incorporation of non-local virtual simulators is also possible. Up to now, connectors for IQM, AQT, and IBM backends have been created. Verification steps such as sending access tokens are integrated in a secure fashion. Further devices can be added to the module with simplified integration steps.

The qc_connect module also handles the backend output, all information from the hardware execution. Different metrics such as the number of qubits and gate depth can also be extracted; further, the integration of new quantum metrics is possible and can then be forwarded to the evaluation module in the next section.

2.4 Evaluation and Visualization

Results of the model cores are processed in the evaluation modules. These modules collect all data from the model execution and compute metrics; besides general ML metrics, the framework already contains task-dependent metrics and evaluation modules for our five use cases. For the quantum pipelines, additional metrics are fetched from the quantum devices, both simulators and hardware, such as number of qubits or gate depth. Similar to the aforementioned modules, the evaluation is fully customizable, allowing the inclusion of newly defined metrics and the selection of a hyperparameter search objective. Indeed, a hyperparameter optimization using Optuna can be executed. Further, the aggregation of metrics across different cross validation folds for hyperparameter search is supported, which is crucial for stochastic stability. A prime example is the Inter-Quartile Mean [1] mitigating the impact of outliers to the final result.

In order to support transparent and extensible benchmarking of both classical and quantum-enhanced learning algorithms, we have developed a browser-based



Fig. 2. Dashboard with two linked cohorts. Each filter criteria panel defines the cohort for the results panel next to it. Left: accuracy distributions per model core (boxplots). Right: accuracy vs num_qbits, shown as an aggregate over selected models or, when grouped, as separate curves per model.

dashboard that couples an experiment repository with an interactive analytics layer. For each experiment, all metrics computed by the evaluation module (including hardware metrics stemming from the quantum device) and all metadata – like dataset, model core, or time-stamp – are ingested into a relational database that serves as a single source of truth. With this design, we eliminate hard-coded assumptions allowing the introduction of new metrics, data sets, or modules without refactoring front-end code. Visualization types such as histograms as well as plotted metrics, multi-scale comparisons, and aggregations are user-configurable. Predefined dashboards can be loaded for recurrent benchmarks while still retaining the full flexibility for ad-hoc analysis.

Figure 2 shows two experiment cohorts, arranged side-by-side by the user. Each cohort is defined by its filter criteria panel and rendered in the adjacent results panel. On the left, boxplots summarize accuracy per model core, to reveal variance and outliers. On the right, accuracy is plotted against the number of qubits. The curve aggregates over all model cores as selected in the right filter. The user can restrict the filter to one model core or group by model core to overlay model-specific curves. This pairing enables like-for-like comparisons of models and qubit scaling under identical, reproducible filter settings.

3 Illustrative Example

To demonstrate the implementation and benchmarking of a reference pipeline, we consider a hybrid quantum-classical image classification task — specifically, the real-world example of crack detection in concrete surfaces (HairWidthCracks [8]). The pipeline is based on quantum transfer learning, where a classical neural network is partially replaced by a parameterized quantum circuit [9]. This serves as a practical example of integrating QML within established deep learning work-

flows. The proposed pipeline is developed similar to Figure 1. Within the data loading module, the dataset is split into training, validation, and testing subsets. Moreover, a stratified K-Fold procedure and sanity checks are applied.

In the model core, the pipeline employs a pre-trained ResNet18 model as a feature extractor. The final fully connected (FC) layer is replaced with a hybrid structure consisting of a classical linear layer, a quantum circuit layer, and another linear layer. The quantum circuit operates on a number of qubits equal to the output dimension of the preceding linear layer. All training-related parameters are stored either for later use in the evaluation module or for monitoring the training process. The modular structure of the framework supports independent development of individual components. It should be noted, that ResNet18 is not the only possible choice for feature extractor and it could be easily replaced with another backbone (e.g., VGG16).

The evaluation module computes the final performance metrics. For this reference pipeline, these include accuracy, recall, precision, and F1-score. When a quantum circuit is included in the model core, additional quantum-specific metrics are also considered. Pipeline parameters are managed through a JSON configuration file, which defines adjustable hyperparameters such as learning rate, number of qubits, number of epochs, batch size, and Optuna optimization settings.

We have benchmarked both classical and quantum variants of the model on an HPC node using an Nvidia Hopper H100 GPU and dual Intel Xeon Gold 5420+ CPUs. The benchmark uses 5-fold cross-validation on the full HairWidthCracks dataset [8].

In addition to the benchmark framework information presented in Figure 2, the framework also records training-specific parameters, such as the progress shown in Figure 3. This figure illustrates the training dynamics across different model cores with varying parameters, using a 5-fold cross-validation on the complete HairWidthCracks dataset.

In our example, both the loss and accuracy curves in Figure 3 and the final evaluation metrics reported in Table 1 illustrate the performance of our pipelines: three neural networks with input and output dimensionality of 512 and 2 respectively. The hidden layer dimensions are set to none layer, 4 and 512. The last pipeline replaces its linear layer by a 4 qubit parametrized quantum circuit. Despite the higher computational overhead introduced by quantum circuit simulation and shot-based measurement, the quantum model achieves performance comparable to classical configurations across all key metrics, underscoring the potential of quantum machine learning (QML) in practical tasks.

4 Conclusion

This work has introduced a modular benchmarking framework designed to facilitate systematic comparisons between classical and quantum machine learning pipelines. By separating the workflow into well-defined modules, the framework enables reproducibility, transparent evaluation, and extensibility. Reproducibil-

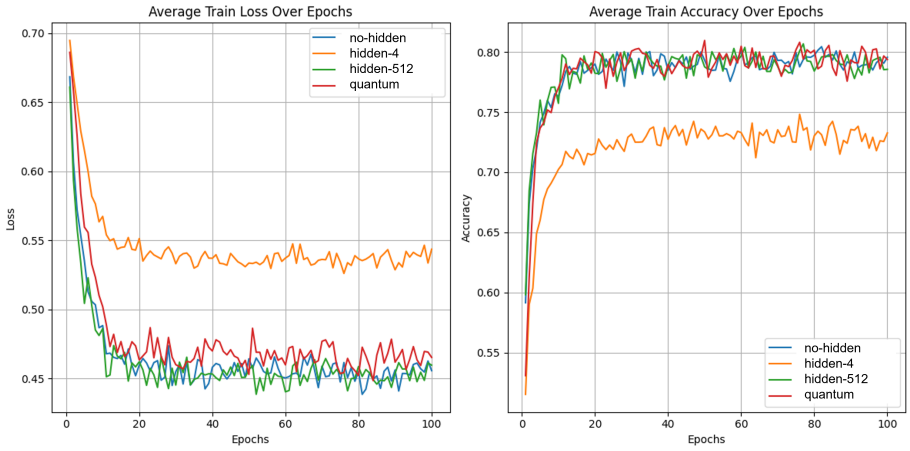


Fig. 3. Training loss and accuracy curves for classical and quantum pipelines over 100 epochs. We present four pipelines. Three classical fully connected linear layers with input and output dimensions of 512 and 2 respectively. The width of the hidden layer is set to none layer ("no-hidden"), 4 ("hidden-4") and 512 ("hidden-512"). The last pipeline replaces the hidden layer with a 4 qubit parameterized circuit ("quantum").

Table 1. Performance metrics on the unseen test dataset for the classic methods (as described in Figure 3) and the quantum version.

Metric	Classic			Quantum
	no-hidden	hidden-4	hidden-512	
Accuracy	0.8107	0.7640	0.8133	**0.8183**
Precision	**0.8473**	0.7755	0.8357	0.8426
Recall	0.7600	**0.7920**	0.7840	0.7833
F1 Score	0.8000	0.7763	0.8077	**0.8115**

ity is addressed by the job generation and environment creation, together with the well-defined available modules. Transparent evaluation is achieved by the synergy between our evaluation modules and its communication with the visualization dashboards, that allows to identify trends in different kind of graphs of aggregated metrics across experiments. Extensibility allows to integrate and test additional functionalities at different levels, such as new evaluation metrics that can be relevant to other use cases. If additional functions or evaluation metrics become relevant in the future, our benchmarking framework can readily accommodate them. Additional datasets, new pipeline steps, and in particular novel (Q)ML models can be integrated with minimal code modifications, allowing researchers to focus on experimentation rather than implementation overhead. Each module is accompanied by sample implementations that can either be used directly or serve as templates for further development. By combining modules in flexible ways, researchers can effectively construct new algorithms or pipelines

with minimal additional overhead. This adaptability makes the framework a practical tool for exploring a wide range of scenarios in both classical and quantum machine learning research. The framework comes with a large library of (real-world datasets), QML and ML models, and metrics. Finally, our software is able to integrate its workflow to communicate to different quantum hardware and simulations, highlighting a crucial aspect of the benchmarking paradigm: verification of the functionality of QML needs to consider and compare execution on the devices it is envisioned for. In this way, benchmarking will reveal its utility to do the leap from the research area to the application fields. The presented illustrative example demonstrate not only the validity of our framework but also the feasibility of integrating quantum components into established deep learning architectures and thus highlights the potential of such hybrid approaches for real-world applications. In conclusion, the proposed benchmarking framework is well suited for systematically comparing classical and quantum pipelines. Its modular design enables researchers to gain deeper insights into the impact of different parameter choices and architectural configurations. By recording both training dynamics and final evaluation results, the framework supports transparent and interpretable comparisons across methods.

Acknowledgments. This project was made possible by the DLR Quantum Computing Initiative and the Federal Ministry for Economic Affairs and Climate Action; https://qci.dlr.de/en/quant2ai/.

Disclosure of Interests. DLR, the contractor Fraunhofer ITWM, and the subcontractor Conet Solutions GmbH have a contractual relationship in the context of the project Quant²AI. This relationship is disclosed in the interest of transparency. This relationship is disclosed in the interest of transparency.

References

1. Agarwal, R., Schwarzer, M., Castro, P.S., Courville, A., Bellemare, M.G.: Deep reinforcement learning at the edge of the statistical precipice (2022), https://arxiv.org/abs/2108.13264
2. Aly, M., Fadaaq, S., Warga, O.A., Nasir, Q., Talib, M.A.: Experimental benchmarking of quantum machine learning classifiers. In: 2023 6th International Conference on Signal Processing and Information Security (ICSPIS). pp. 240–245. IEEE (Nov 2023). https://doi.org/10.1109/icspis60075.2023.10343811
3. Bauckhage, C., Bye, R., Iftikhar, A., Knopf, C., Mustafic, M., Piatkowski, N., Sifa, R., Stahl, R., Sultanow, E.: Quantum machine learning-state of the art and future directions. German Federal Office for Information Security p. 122 (2022)
4. Bowles, J., Ahmed, S., Schuld, M.: Better than classical? the subtle art of benchmarking quantum machine learning models. arXiv (Mar 2024). https://doi.org/10.48550/arXiv.2403.07059
5. Cerezo, M., Verdon, G., Huang, H.Y., Cincio, L., Coles, P.J.: Challenges and opportunities in quantum machine learning. Nature Computational Science **2**(9), 567–576 (Sep 2022). https://doi.org/10.1038/s43588-022-00311-3

6. Finzgar, J.R., Ross, P., Holscher, L., Klepsch, J., Luckow, A.: Quark: A framework for quantum computing application benchmarking. In: 2022 IEEE International Conference on Quantum Computing and Engineering (QCE). pp. 226–237. IEEE (Sep 2022). https://doi.org/10.1109/qce53715.2022.00042
7. Geissler, F., Stopfer, E., Ufrecht, C., Meyer, N., Scherer, D.D., Wagner, F., Oberreuter, J.M., Chen, Z., Farace, A., Gutina, D., Schwenk, U., Lange, K., Junk, V., Husslein, T., Erdmann, M., Kiwit, F., Decker, B., Shaji, G., Granet, E., Dreyer, H., Dragan, T.A., Lorenz, J.M.: Benchqc – scalable and modular benchmarking of industrial quantum computing applications. arXiv (Apr 2025). https://doi.org/10.48550/ARXIV.2504.11204
8. Geng, A., Moghiseh, A.: HairWidthCracks dataset. Fordatis - Forschungsdaten-Repositorium der Fraunhofer-Gesellschaft (2024), https://fordatis.fraunhofer.de/handle/fordatis/429
9. Geng, A., Moghiseh, A., Redenbach, C., Schladitz, K.: Hybrid quantum transfer learning for crack image classification on NISQ hardware. In: The National University of Science And Technology International Conference For Engineering Sciences. p. 130001 (2025). https://doi.org/10.1063/5.0246500, https://pubs.aip.org/aip/acp/article-lookup/doi/10.1063/5.0246500
10. Gujju, Y., Matsuo, A., Raymond, R.: Quantum machine learning on near-term quantum devices: Current state of supervised and unsupervised techniques for real-world applications. arXiv (Jul 2023). https://doi.org/10.48550/ARXIV.2307.00908
11. Kiwit, F.J., Marso, M., Ross, P., Riofrío, C.A., Klepsch, J., Luckow, A.: Application-oriented benchmarking of quantum generative learning using quark. In: 2023 IEEE International Conference on Quantum Computing and Engineering (QCE). pp. 475–484. IEEE (sep 2023). https://doi.org/10.1109/qce57702.2023.00061
12. Quetschlich, N., Burgholzer, L., Wille, R.: Mqt bench: Benchmarking software and design automation tools for quantum computing. Quantum 7, 1062 (2023) **7**, 1062 (Jul 2022). https://doi.org/10.22331/q-2023-07-20-1062
13. Ristè, D., da Silva, M.P., Ryan, C.A., Cross, A.W., Córcoles, A.D., Smolin, J.A., Gambetta, J.M., Chow, J.M., Johnson, B.R.: Demonstration of quantum advantage in machine learning. npj Quantum Information **3**(1) (Apr 2017). https://doi.org/10.1038/s41534-017-0017-3
14. Roth, M., Kreplin, D.A., Basilewitsch, D., Bravo, J.F., Klau, D., Marinov, M., Pranjic, D., Stuehler, H., Willmann, M., Zöller, M.A.: Autoqml: A framework for automated quantum machine learning (2025), https://arxiv.org/abs/2502.21025
15. Schuld, M., Killoran, N.: Is quantum advantage the right goal for quantum machine learning? PRX Quantum **3**(3), 030101 (Jul 2022). https://doi.org/10.1103/prxquantum.3.030101
16. Thiyagalingam, J., Shankar, M., Fox, G., Hey, T.: Scientific machine learning benchmarks. Nature Reviews Physics **4**(6), 413–420 (Apr 2022). https://doi.org/10.1038/s42254-022-00441-7
17. Tomesh, T., Gokhale, P., Omole, V., Ravi, G.S., Smith, K.N., Viszlai, J., Wu, X.C., Hardavellas, N., Martonosi, M.R., Chong, F.T.: Supermarq: A scalable quantum benchmark suite. arXiv (2022). https://doi.org/10.48550/ARXIV.2202.11045

Application of Multi-Criteria Decision Aiding to Quantum Benchmarking: Some Results on Scale Invariance

Christophe Labreuche

Thales cortAIx-Labs, 1 avenue Augustin Fresnel, 91767 Palaiseau, France
christophe.labreuche@thalesgroup.com

Abstract. Application-oriented benchmarking of quantum computing requires the evaluation of multiple, often conflicting, Key Performance Indicators (KPIs). To assess the overall quality of a single solution or to compare multiple solutions, these KPIs must be synthesized into a comprehensive metric. Multi-Criteria Decision Aiding (MCDA) provides a suitable methodological framework for this purpose. MCDA involves two main steps to produce an overall score from a vector of KPIs: first, normalizing the KPIs, and second, aggregating the normalized scores. It is important to note that normalization is not uniquely defined and may correspond to different measurement scales. We therefore examine the property of invariance, where switching from one admissible scale to another should not alter the comparative ranking of options. To facilitate aggregation, it is often assumed that some form of commensurability (or comparability) exists across the normalized KPI scores. However, this is a strong assumption that may not always hold in practice. As such, we investigate various aggregation functions that can accommodate different invariance assumptions, providing a more nuanced and robust approach to decision-making in quantum benchmarking.

Keywords: Quantum Computer Benchmarking · Multi-Criteria Decision Aiding

1 Introduction

Quantum computing benchmarking is a critical issue from the user's perspective, as it enables the assessment of performance and the analysis of tangible benefits compared to High-Performance Computing (HPC) or traditional computers [12]. In this context, we concentrate on application-level benchmarking, as opposed to component-level or system-level benchmarking. Application-level benchmarking represents the most comprehensive form of evaluation, subsuming the other levels, and aims to determine the capability of a quantum computer or simulator to solve specific problems of interest to users, such as optimization or fundamental physics problems.

Benchmarking requires the definition of Key Performance Indicators (KPIs) to evaluate different solutions. Application-level KPIs are directly interpretable and relevant to end-users. The Q-Score measures the largest problem size that can be solved with performance significantly above the noise level [13]. Originally defined for the Max-Cut optimization problem [13], it has since been extended to the Max-Clique problem [19]. The Quantum Application Score (QuAS) is defined as the volume under the curve of

a function fitting the Pareto front in the ⟨accuracy, runtime⟩ space [14]. TAQOS (Tight Analysis of Quantum Optimization Systems) introduces a KPI defined as the percentage of instances for which a solution is found with a cost function within a distance at most ε from the reference value within a time limit t, evaluated for different values of ε and t on Noisy Intermediate-Scale Quantum (NISQ) devices across various optimization problems [7]. The BACQ project aims to develop KPIs for a range of applications (such as optimization, linear system solving, prime factorization, and many-body problems), and also incorporates energy-related considerations [2].

When the user is interested in multiple KPIs, rather than a single one, it becomes necessary to combine these KPIs to enable meaningful comparisons between quantum computing solutions. Multi-Criteria Decision Aiding (MCDA) is a methodology that aims to elicit and represent the preferences of a stakeholder –typically a decision maker – regarding different options based on multiple, and often conflicting, decision criteria represented by KPIs [6].

QPack introduces an overall score that aggregates around a dozen KPIs [5] In this framework, the *pure* score denotes the raw value of the KPI, while the *mapped* score is a normalized value that facilitates the aggregation of different KPIs into a single comprehensive score. A similar methodology is applied in BACQ [2], where KPIs are first normalized and then aggregated. The distinction between BACQ and QPack is twofold: (1) BACQ proposes a MCDA-based elicitation approach that constructs normalization and aggregation schemes from decision maker input, whereas QPack relies on predefined formulas; (2) the aggregation models differ – BACQ employs a Choquet integral [4], while QPack uses a multi-linear model [10].

The aim of this paper is to analyze the applicability of MCDA to quantum benchmarking. The normalization of a KPI corresponds to defining a scale [11]. Scales are always specified up to some degree(s) of freedom, and comparisons among options made by the MCDA model should be invariant under valid transformations of these scales. Moreover, each KPI is normalized independently, and reference levels are introduced to establish relationships among the normalized KPIs. The number of reference levels dictates the degree of interrelation among the normalized KPIs. We derive the form of the MCDA model under varying assumptions about invariance on the reference levels (i.e., by imposing from 0 to 2 reference levels). We demonstrate that there is a tradeoff between the versatility of the model and the ease of defining normalized KPIs. Specifically, the more reference levels are imposed, the more challenging it becomes to normalize the KPIs, but a wider family of MCDA models becomes compatible.

After a brief overview of MCDA in Section 2, we discuss the notion of reference levels in Section 3. The main results related to invariance are presented in Section 4.

2 Reminder on MCDA

The aim of this section is to select a family of MCDA approaches suited for quantum benchmarking. The decision criteria are denoted by $N = \{1, \ldots, n\}$. Each criterion $i \in N$ is characterized by a KPI (also referred to as an *attribute*) that takes values in a set X_i. Thus, each alternative x is described by a vector $(x_1, \ldots, x_n)$, where $x_i \in X_i$ for all $i \in N$, and is therefore considered as an element of $X := X_1 \times \cdots \times X_n$. For

$x, y \in X$ and $i \in N$, we denote by (x_i, y_{-i}) the composite alternative in X that takes the value x_i for attribute i and y_j for all $j \in N \setminus \{i\}$.

The central problem in MCDA is to construct a model for comparing the elements of X. This is formalized through a binary relation $\succsim$ defined on X, representing the stakeholder's preferences. For $x, y \in X$, we write $x \succsim y$ to mean that the stakeholder considers x at least as attractive as y, taking all criteria comprehensively into account.

Example 1. For illustration, consider $n = 3$ KPIs corresponding to different optimization problems: (1) Q-Score for the Max-Cut problem; (2) Q-Score for the Max-Clique problem; (3) Maximum problem size for which the exact solution of the Max Cardinality Matching (MCM) problem is obtained. Suppose we have two quantum computers, $x = (50, 30, 40)$ and $y = (30, 20, 50)$, evaluated according to these three KPIs. The user may express a preference for the first option, that is, $x \succsim y$. ■

2.1 Overview of the main MCDA Approaches

A taxonomy of the main MCDA methods can be found in Figure 1.

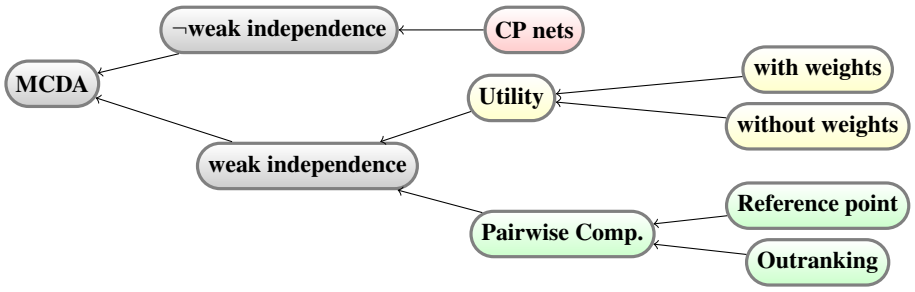


Fig. 1. Overview of the main MCDA methods

The first branching in Figure 1 arises from a property known as *Weak Independence* [11, Sect. 7.2.1]. A binary relation $\succsim$ is said to be *weakly independent* if, for every $i \in N$, and for all $x_i, y_i \in X_i$, $z_{-i}, z'_{-i} \in X_{-i}$, the following equivalence holds:

$$(x_i, z_{-i}) \succsim (y_i, z_{-i}) \iff (x_i, z'_{-i}) \succsim (y_i, z'_{-i}) \,. \tag{1}$$

This condition states that if value x_i is preferred to y_i for attribute X_i when all other attributes have certain fixed values z_{-i}, then the same preference between x_i and y_i must hold regardless of the values assigned to the other attributes, i.e., for any z'_{-i}. The MCDA method located at the top of Fig. 1 (in red) do not satisfy weak independence. The most prominent example is the CP-net (Conditional Preference network) [3].

Weak independence is satisfied for all KPIs relevant to quantum benchmarking. In Ex. 1, suppose we prefer a quantum computer with a Q-Score of $x_1 = 50$ over another with $y_1 = 30$, *ceteris paribus* (i.e., holding all other criteria constant). Given Eq. (1), we

can define the binary relation $\succsim_i$ on X_i by stating that $x_i \succsim_i y_i$ if $(x_i, z_{-i}) \succsim (y_i, z_{-i})$ for all $z_{-i} \in X_{-i}$. In Ex. 1, it follows that $50 \succsim_1 30$ for the first KPI.

The main classes of methods that satisfy weak independence are either based on pairwise comparisons between options – see the bottom of Fig. 1 (in green) – or on utility-based approaches (highlighted in yellow in Fig. 1). In the latter case, the preference relation is represented by a utility function $u : X \to \mathbb{R}$ (that is, $x \succsim y$ if and only if $u(x) \geq u(y)$). One of the most widely used models for representing the order relation $\succsim$ is the *transitive decomposable* model [11], in which

$$u(x) = F(u_1(x_1), \ldots, u_n(x_n)), \tag{2}$$

where $F : \mathbb{R} \times \cdots \times \mathbb{R} \to \mathbb{R}$ is an aggregation function, and $u_i : X_i \to \mathbb{R}$ ($i = 1, \ldots, n$) are referred to as *marginal utility functions*. The value function u_i quantifies the preference relation $\succsim_i$ of the stakeholder with respect to the elements of X_i: for $x_i, y_i \in X_i$, $x_i \succsim_i y_i \iff u_i(x_i) \geq u_i(y_i)$. In [5], the *mapped* scores correspond to the scores produced by the marginal utility functions.

2.2 Choice of a Type of MCDA Model

For quantum benchmarking, weak independence is satisfied. Furthermore, utility methods are generally more appropriate than pairwise comparison methods, as they permit the evaluation of a quantum computer independently of others. According to Fig. 1, utility methods can be classified according to whether they incorporate weights or not:

- Models with weights: The aggregation function F includes weights. This framework distinguishes between marginal utility functions – which serve to normalize the KPIs – and weights, which reflect the importance of individual criteria or subsets of criteria. Typical aggregation models include the weighted sum $F(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_i$, the Choquet integral [4], and the multi-linear extension model [10].
- Models without weights: Here, no explicit weights are used. The marginal utility functions u_i perform both normalization and the impact of each criterion. Common models include the additive utility model $u(x) = \sum_{i=1}^{n} u_i(x_i)$ (i.e., $F(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i$) and the *Generalized Additive Independence* model, which introduces additional terms $u_S(x_S)$ in the sum for selected subsets $S \subseteq N$ of criteria.

The principal advantage of separating weights from marginal utility functions is enhanced interpretability. This distinction makes it possible to specify, for example, that the value of a metric on a given criterion is *well-satisfied* (or, conversely, *ill-satisfied*), or that another criterion is *very important* (or *not important at all*). Such qualitative interpretation is not feasible with models that do not use weights.

The elicitation of models without weights typically rests on techniques such as UTA [9]. This process is challenging, as it involves the comparison of numerous vectors of values corresponding to quantum computing metrics. In contrast, elicitation of models based on weights usually separates the identification of marginal utility functions from the determination of weights, as in MACBETH [1] or AHP [17]. We believe this approach is better suited and less demanding for quantum computer benchmarking.

To differentiate between the importance weights of criteria and the level of satisfaction with individual metrics, it is necessary to introduce an additional assumption regarding the existence of reference levels. This is the object of the next section.

3 Existence of Reference Levels

In Eq. (2), the marginal utility functions and the aggregation function are not uniquely determined. The marginal utility function u_i defines a *scale* for measuring the preference relation $\succsim_i$ on X_i [11, 16]. Multiple *admissible* scales u_i can be constructed from a given relation $\succsim_i$. The scale type is characterized by the set Φ_i of admissible transformations $\phi_i \in \Phi_i$, mapping one admissible scale u_i to another, $u_i' = \phi_i \circ u_i$. In an *ordinal* scale, admissible transformations are all strictly increasing functions, allowing for only basic operations on the scale. A more expressive form is the *interval* scale, where Φ_i consists of all positive linear transformations,

$$\phi_i(a_i) = \alpha_i a_i + \beta_i \quad \text{with } \alpha_i > 0,\ \beta_i \in \mathbb{R}. \tag{3}$$

An interval scale is commonly constructed using a quaternary relation $\succsim_i^\star$ defined on $X_i \times X_i$ that assesses differences in satisfaction. For $x_i, y_i, z_i, t_i \in X_i$ with $x_i \succsim_i y_i$ and $z_i \succsim_i t_i$, the relation $(x_i, y_i) \succsim_i^\star (z_i, t_i)$ means that the difference in satisfaction between x_i and y_i is at least as great as that between z_i and t_i. This quaternary relation directly corresponds to a difference in utility u_i [11, Chap. 4, Theorem 2]:

$$(x_i, y_i) \succsim_i^\star (z_i, t_i) \quad \Leftrightarrow \quad u_i(x_i) - u_i(y_i) \geq u_i(z_i) - u_i(t_i). \tag{4}$$

Both $\succsim_i$ and $\succsim_i^\star$ are meaningful to stakeholders, enabling u_i to be elicited from these relations, for example using the MACBETH method. Since an interval scale has two degrees of freedom (α_i and β_i in Eq. (3)), the marginal utility function can be fully determined by specifying its value at two *reference elements* in X_i. These two reference levels are denoted $\mathfrak{B}_i, \mathfrak{G}_i \in X_i$ and their associated utility values are $\mathfrak{b}_i, \mathfrak{g}_i \in \mathbb{R}$, with $\mathfrak{g}_i > \mathfrak{b}_i$. Typically, $\mathfrak{B}_i$ (the *bad* element) represents a completely unsatisfactory outcome, set to the utility value $\mathfrak{b}_i$:

$$u_i(\mathfrak{B}_i) = \mathfrak{b}_i. \tag{5}$$

This serves as a saturation point, meaning for all $x_i \precsim_i \mathfrak{B}_i$, $u_i(x_i) = \mathfrak{b}_i$. In quantum benchmarking, $\mathfrak{B}_i$ might correspond to the performance delivered by a randomized algorithm. The upper reference level $\mathfrak{G}_i$ (the *good* element) is defined as the highest satisfactory level attainable given current technological capabilities. Its utility is set to $\mathfrak{g}_i \in \mathbb{R}$:

$$u_i(\mathfrak{G}_i) = \mathfrak{g}_i. \tag{6}$$

4 Various Qualitative Invariances to ease the Commensurateness Assumption of the Scales

In Sect. 3, we considered the marginal utility scales independently. The concept of commensurateness arises when comparing values from different marginal utility functions. Two scales u_i and u_k are said to be *commensurate* if $u_i(x_i) = u_k(x_k)$ implies that x_i and x_k represent the *same level of satisfaction*. Conversely, if $u_i(x_i) = u_k(x_k)$ does not convey any relationship of satisfaction, then the scales are not commensurate.

A convenient way to construct commensurate marginal utility functions is to assume that all reference levels $\mathfrak{B}_1, \ldots, \mathfrak{B}_n$ and $\mathfrak{G}_1, \ldots, \mathfrak{G}_n$ are commensurate (i.e., have the same meaning) and are therefore assigned identical utilities:

$$u_1(\mathfrak{B}_1) = \cdots = u_n(\mathfrak{B}_n) \quad \text{and} \quad u_1(\mathfrak{G}_1) = \cdots = u_n(\mathfrak{G}_n). \tag{7}$$

Example 2 (Ex. 1 cont.). All three KPIs are completely unsatisfactory for value 0: $\mathfrak{B}_1 = \mathfrak{B}_2 = \mathfrak{B}_3 = 0$. The upper reference level is assigned to the best value achievable with current technology. We thus set $\mathfrak{G}_1 = \mathfrak{G}_2 = 1000$ (achieved by the NEC computer with a simulated annealing [18]), and $\mathfrak{G}_3 = 512$ (attained by the NEC computer). ■

When the scales are commensurate (as per Eq. (7)), we have $\mathfrak{b}_1 = \cdots = \mathfrak{b}_n$ and $\mathfrak{g}_1 = \cdots = \mathfrak{g}_n$. Figure 2-Left illustrates two commensurate scales, i and k. The blue arrows represent the unit interval between the Bad and Good levels, which are identical in length due to commensurateness. In Figure 2-Right, different linear transformations ϕ_ℓ are applied to the scales; here, $\mathfrak{b}'_\ell = \phi_\ell(\mathfrak{b}_\ell)$ and $\mathfrak{g}'_\ell = \phi_\ell(\mathfrak{g}_\ell)$. As a result, the unit intervals between the Bad and Good levels now differ between the two KPIs.

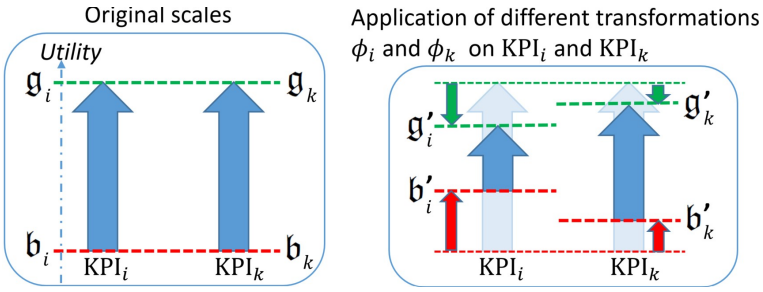


Fig. 2. Effect of affine transformations on the unit distance between the Bad and Good levels for two KPIs.

Let Φ denote the set of admissible tuples $(\phi_1, \ldots, \phi_n)$ of transformations. The property of scale invariance requires that the ranking of any set of options remains unchanged whether we use the original aggregation $F(u_1(x_1), \ldots, u_n(x_n))$ or any normalized version obtained by applying admissible transformations, i.e., $F(\phi_1 \circ u_1(x_1), \ldots, \phi_n \circ u_n(x_n))$ for $(\phi_1, \ldots, \phi_n) \in \Phi$. In other words, for $x, y \in X$, if x is at least as good as y under the aggregation F with marginal utility functions $u_1, \ldots, u_n$ (i.e., $F(u_1(x_1), \ldots, u_n(x_n)) \geq F(u_1(y_1), \ldots, u_n(y_n))$), this ordering must also hold for any admissible transformation of the utilities. Letting $a_i = u_i(x_i)$ and $a'_i = u_i(y_i)$, we express the invariance property for the aggregation function F as follows.

Definition 1. *F is* scale invariant under *Φ if, for all $(\phi_1, \ldots, \phi_n) \in \Phi$ and all $a, a' \in \mathbb{R}^n$,*

$$F(a_1, \ldots, a_n) \geq F(a'_1, \ldots, a'_n) \iff F(\phi_1(a_1), \ldots, \phi_n(a_n)) \geq F(\phi_1(a'_1), \ldots, \phi_n(a'_n)). \tag{8}$$

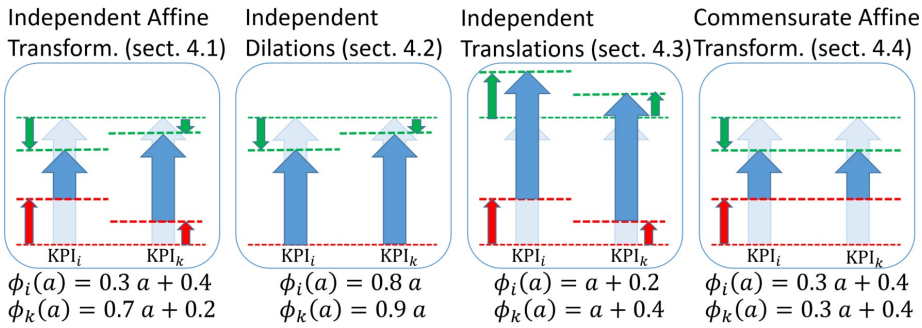


Fig. 3. Examples of different transformation classes Φ. We illustrate with explicit expressions for ϕ_i and ϕ_k, assuming $\mathfrak{b}_1 = \cdots = \mathfrak{b}_n = 0$ and $\mathfrak{g}_1 = \cdots = \mathfrak{g}_n = 1$.

In some situations, requiring the full commensurateness property of Eq. (7) may be overly restrictive. We now consider weaker assumptions. Let $\mathbb{R}^n_{\mathfrak{b}+} = \{a \in \mathbb{R}^n : \forall i \in N,\ a_i \geq \mathfrak{b}_i\}$.

4.1 Invariance w.r.t. Independent Affine Transformations on the KPIs

We first examine the extreme case in which the condition of Eq. (7) is simply omitted, meaning that there is no assumption of commensurateness among the reference levels. In this case, the invariance property must allow each marginal utility function $u_1, \ldots, u_n$ to be transformed independently (see Fig. 3):

$$\Phi_{\text{indep}} = \left\{ (\varphi_{\alpha_1,\beta_1}, \ldots, \varphi_{\alpha_n,\beta_n}) \mid (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n_+,\ (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n \right\}$$

where $\varphi_{s,t}(a) = s\ a + t$. In other words, the aggregation function must be able to combine scales that are not commensurate. The following result demonstrates that there is no aggregation function which satisfies scale invariance under Φ_{indep}, together with several mild conditions, such as continuity of F. Proofs of all results are omitted due to space constraints.

Theorem 1. *There does not exist a function F that is non-trivial (not all alternatives are indifferent), continuous, monotone, and scale invariant under Φ_{indep}.*

It is thus necessary to impose some level of commensurateness among the marginal utility functions. Various commensurateness conditions are explored in the next sections.

4.2 Invariance w.r.t. Independent Dilations on the Different Metrics

Here, we assume that all bad levels are commensurate and share the same interpretation, whereas the good levels may differ. This means that we can transform each utility function u_k into another u'_k, provided the utility on the bad level remains unchanged:

$$\forall k \in N \qquad u_k(\mathfrak{B}_k) = u'_k(\mathfrak{B}_k) = \mathfrak{b}_k. \tag{9}$$

Example 3 (Ex. 2 cont.). Requiring commensurate bottom levels is not restrictive in quantum benchmarking. For example, in most metrics that measure the largest problem size solvable, the bottom reference level is 0: $\mathfrak{B}_1 = \mathfrak{B}_2 = \mathfrak{B}_3 = 0$. ■

According to Eq. (9), since u_k and u'_k are related by an affine transformation, we have:

$$\forall k \in N \qquad u'_k(x_k) = \alpha_k \left(u_k(x_k) - \mathfrak{b}_k\right) + \mathfrak{b}_k, \tag{10}$$

for some $\alpha_k > 0$. The set of admissible transformations is (see Fig. 3):

$$\Phi_{\text{dilat}} = \left\{ (\varphi_{\alpha_1,\, \mathfrak{b}_1(1-\alpha_1)}, \ldots, \varphi_{\alpha_n,\, \mathfrak{b}_n(1-\alpha_n)}) \;\mid\; (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}_+^n \right\}.$$

The following result states that, under scale invariance with respect to Φ_{dilat}, there is only one compatible family of aggregation functions: the weighted geometric mean, up to a strictly increasing transformation.

Theorem 2. *If F is non-trivial (not all alternatives are indifferent), continuous, monotone, and scale invariant under Φ_{dilat}, then there exist $w_1, \ldots, w_n \geq 0$ with $w_1 + \cdots + w_n = 1$ and a strictly non-decreasing function $g : \mathbb{R} \to \mathbb{R}$ such that, for all $a \in \mathbb{R}^n_{\mathfrak{b}+}$,*

$$F(a) = g\left(\prod_{i \in N} (a_i - \mathfrak{b}_i)^{w_i} \right).$$

This outcome is intuitive: aggregating heterogeneous variables normalized by their worst value using a geometric mean is meaningful. For instance, in Cooperative Game Theory, the Nash bargaining solution relies on the geometric mean [15], and invariance to input scale is a key axiom. Notably, the geometric mean is also employed as an aggregation function in some optimization benchmarking settings [8, Sect. V.C]. However, the geometric mean has a limitation: a poor score on any KPI (a value close to its bad reference) cannot be compensated by high scores on other KPIs. Thus, it rewards solutions that show consistently strong performance across all KPIs, rather than those that are exceptional in some areas and weak in others.

4.3 Invariance w.r.t. Independent Translations on the Different Metrics

The next type of invariance we consider assumes that the dilation factor is fixed, while the translation factor is variable. This means that the transformation from one admissible utility function to another takes the form $\phi_i(a_i) = a_i + \beta_i$, where the translation factors $\beta_i \in \mathbb{R}$ may differ across criteria. The set of admissible transformations is given by (see Fig. 3):

$$\Phi_{\text{translat}} = \left\{ (\varphi_{1,\beta_1}, \ldots, \varphi_{1,\beta_n}) \mid (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n \right\}.$$

Example 4 (Ex. 3 cont.). Consider measuring the temperature of a quantum computer, assuming this metric is relevant to the user. A unit interval of $100°\text{C}$ between the Bad and Good levels may be meaningful. However, assigning the Bad level to $0°\text{C}$ is arbitrary. Therefore, the Bad level could be shifted, while the unit interval is preserved. ■

The following result establishes that, under scale invariance with respect to Φ_{translat}, the only compatible family of aggregation functions is the weighted sum, up to a strictly increasing transformation.

Theorem 3. *If F is non-trivial (not all alternatives are indifferent), continuous, monotone, and scale invariant under Φ_{translat}, then there exist $w_1, \ldots, w_n \geq 0$ with $w_1 + \cdots + w_n = 1$, and a strictly increasing function h such that for all $a \in \mathbb{R}^n_{\mathfrak{b}+}$,*

$$F(a) = h\left(\sum_{i \in N} w_i a_i\right).$$

4.4 Invariance w.r.t. Commensurate Affine Transformations

In this final case, we assume that the scales remain commensurate after applying transformations to the marginal utility functions. Two admissible scale tuples $(u_1, \ldots, u_n)$ and $(u'_1, \ldots, u'_n)$ (satisfying Eq. (7)) are therefore related by the same positive linear transformation, as in Eq. (3), across all criteria, with $\alpha_1 = \cdots = \alpha_n =: \alpha$ and $\beta_1 = \cdots = \beta_n =: \beta$. Thus, the set of admissible transformations is (see Fig. 3):

$$\Phi_{\text{commens}} = \Big\{(\varphi_{\alpha,\beta}, \ldots, \varphi_{\alpha,\beta}) : \alpha \in \mathbb{R}_+, \beta \in \mathbb{R}\Big\}.$$

Unlike in previous cases, we do not obtain a uniqueness (representation) theorem, since many aggregation models satisfy scale invariance under Φ_{commens}. For example, the weighted sum, the Choquet integral, and the multilinear model all fulfill this property.

5 Conclusion

MCDA is an essential tool for aggregating various KPIs in quantum benchmarking. Among the different MCDA approaches, we advocate using utility-based methods that distinguish between normalization and aggregation. The utility function associated with each KPI is treated as an interval scale, and positive linear transformations can be applied to each utility without affecting the overall benchmarking result. Traditionally, the reference levels are assumed to be commensurate. While this is a strong assumption, it enables a wide range of aggregation functions. We have demonstrated that no aggregation function remains invariant under independent positive linear transformations of each utility function. Therefore, it is necessary to consider more restrictive forms of invariance. Aggregation functions that are invariant when only the bad reference levels are commensurate (but not the good levels) are homomorphic to a weighted geometric mean. If the distance between the two reference levels is fixed but their positions can be translated freely, the only invariant aggregation function is homomorphic to a weighted sum. With these findings, the decision maker can choose the most suitable MCDA model, balancing the model's expressiveness with the practical challenge of defining appropriate reference levels.

Acknowledgment

As part of the MetriQs-France program, this work is supported by France 2030 under the French National Research Agency grant number ANR-22-QMET-0002 (BACQ project).

References

1. Bana e Costa, C.A., De Corte, J., Vansnick, J.C.: MACBETH. International Journal of Information Technology and Decision Making **11**, 359–387 (2012)
2. Barbaresco, F., et al: BACQ – application-oriented benchmarks for quantum computing. In: arXiv:2403.12205 (2024)
3. Boutilier, C., Brafman, R., Domshlak, C., Hoos, H., Poole, D.: CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. Journal of Artificial Intelligence Research (JAIR) **21**, 135–191 (2004)
4. Choquet, G.: Theory of capacities. Annales de l'Institut Fourier **5**, 131–295 (1953)
5. Donkers, H., Mesman, K., Al-Ars, Z., Moller, M.: QPack scores: Quantitative performance metrics for application-oriented quantum computer benchmarking. In: CoRR abs/2205.12142 (2022)
6. Figueira, J., Greco, S., Ehrgott, M. (eds.): Multiple Criteria Decision Analysis: State of the Art Surveys. Second edition. Kluwer Acad. Publ. (2016)
7. Gilbert, V., Louise, S., Sirdey, R.: TAQOS: A Benchmark Protocol for Quantum Optimization Systems. In: International Conference on Computational Science (ICCS). DOI: 10.1007/978-3-031-36030-5_13 (2023)
8. Hansen, N., Auger, A., Brockhoff, D., Tusar, T.: Anytime performance assessment in blackbox optimization benchmarking. IEEE Trans. on Evolutionary Computation **26**, 1293–1305 (2022)
9. Jacquet-Lagrèze, E., Siskos, Y.: Assessing a set of additive utility functions for multicriteria decision making: The UTA method. European J. of Operational Research **10**, 151–164 (1982)
10. Keeney, R.L., Raiffa, H.: Decision with Multiple Objectives. Wiley, New York (1976)
11. Krantz, D., Luce, R., Suppes, P., Tversky, A.: Foundations of measurement, vol. 1: Additive and Polynomial Representations. Academic Press (1971)
12. Lorenz, J.M., et al: Systematic benchmarking of quantum computers: status and recommendations. In: arXiv:2503.04905v1 (2025)
13. Martiel, S., Ayral, T., Allouche, C.: Benchmarking quantum co-processors in an application-centric, hardware-agnostic and scalable way. IEEE Trans. on Quantum Engineering **2** (2021)
14. Mesman, K., van der Schoot, W., Moller, M., Neumann, N.: QuAS: Quantum application score for benchmarking the utility of quantum computers. In: arXiv:2406.03905 (2024)
15. Nash, J.: The bargaining problem. Econometrica **18**, 115–162 (1950)
16. Roberts, F.: Measurement Theory. Addison-Wesley (1979)
17. Saaty, T.L.: A scaling method for priorities in hierarchical structures. J. Math. Psychology **15**, 234–281 (1977)
18. van der Schoot, W., Leermakers, D., Wezeman, R., Neumann, N., Phillipson, F.: Evaluating the Q-Score of Quantum Annealers. In: IEEE International Conference on Quantum Software (QSW), pp. 9–16. Barcelona, Spain (2022)
19. van der Schoot, W., Wezeman, R., Neumann, N., Phillipson, F., Kooij, R.: Extending the Q-Score to an Application-Level quantum metric framework. In: IEEE International Conference on Quantum Computing and Engineering (QCE), pp. 941–951. Montreal, Canada (2024)

Benchmark of MHT algorithm on Quantum Computer

Paul Rousset-Rouard [1], Bing Hong Teh [2] and Jean-Marc Divanon[2]

[1] CentraleSupélec, University of Paris-Saclay, France
[2] Thales Research & Technology, Singapore
paul.rousset-rouard@student-cs.fr
{bing-hong.teh, jean-marc.divanon}@thalesgroup.com

Abstract. This paper explores the application of quantum computing, specifically quantum annealing using D-Wave systems, to address the Multiple Hypothesis Tracking (MHT) problem in tracking systems. This involves the critical task of assigning sensor detections to relevant targets amidst various challenges such as false alarms, noise, and target manoeuvres. The combinatorial nature of MHT presents a significant obstacle, with the number of hypotheses escalating exponentially over time, leading traditional solvers to struggle under substantial computational load. By benchmarking classical solvers against D-Wave's hybrid quantum annealing approach, the study has demonstrated remarkable improvements in computational time, achieving up to 15-fold speed-ups during high false alarm density scenarios. This work highlights the potential of quantum computing over classical methods in efficiently tackling complex combinatorial optimization tasks, thereby offering a promising solution for demanding real-time tracking challenges. The research further suggests avenues for future exploration with different quantum computing technologies, underscoring the evolving landscape of technology in optimizing tracking system efficiency.

Keywords: Multiple Hypothesis Tracking (MHT), Data Association Problem (DAP), Maximum Weighted Independent Set (MWIS), NP-hard, Quantum Computing, D-Wave, Quantum Advantage.

1 Introduction

1.1 The Data Association Problem (DAP)

In tracking systems, sensors generate numerous detections that must be accurately attributed to relevant targets. The Data Association Problem (DAP) is the process of assigning these detections (or measurements) to corresponding targets such that each measurement is associated with at most one target, and each target receives at most one measurement per scan. This problem must be addressed following each scan and subsequently for all collected scans of measurements. Several complicating factors introduce significant complexity to the DAP:

- Presence of spurious plots, often referred to as false alarms,

- Noise contaminated measurements,
- Target manoeuvres,
- Missed detections of targets,
- Real-time requirements.

1.2 Multiple Hypothesis Tracking (MHT)

One of the well-known solution methods for DAP is Multiple Hypothesis Tracking (MHT), which was introduced by Reid [1]. At the heart of MHT lies the hypothesis tree structure (see Fig. 1), representing every feasible assignment of plots. For each potential assignment, MHT calculates a likelihood of truth [1][3][4]. Indeed, the MHT algorithm consists in evaluating possible tracks at each time step and updates these beliefs thanks to new observations. It basically keeps in memory all detections that happened at timesteps {0, 1, . . ., k} to make decisions about the most likely track at time step k. This is what makes this tracking algorithm powerful, but also memory and calculation intensive. The most probable hypotheses (meaning whose likelihood of truth are highest) are selected as current solution, though it is not definitive. Less probable hypotheses are nonetheless preserved, as they might become relevant with the receipt of new plots in future scans. This is illustrated in Figure 1 below, extracted from [3].

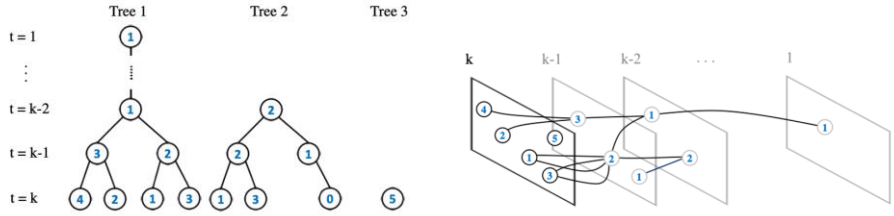


Fig. 1. Hypothesis tree structure built from radar measurements (*left*) and association at different timestep from 1 to k for each plot spatially (*right*). Extracted from Fig. 1 and Fig. 2 of [3].

A notable challenge in MHT is the combinatorial nature of this problem, where the number of associations escalates exponentially over time. This is further illustrated with example in section 3.2 below. Conventionally, this complexity is managed through various pruning techniques, eliminating less promising hypotheses and thus preventing combinatorial explosion. While this helps reduce the complexity of the DAP, some possible truths may be lost as a trade-off.

A brute force implementation of MHT endeavours to encompass all conceivable scenarios to accurately compute the most probable associations. In this exhaustive approach, every potential assignment of tracks and measurements is considered. For example, when dealing with 10 tracks and 100 measurements, assuming a detection probability of 1, the number of possible associations reaches approximately 17 trillion.

With the growing interest in quantum computing technology, the work presented in this paper provides a benchmark of the MHT algorithm implementation on D-Wave's quantum computer against state-of-the-art classical solvers.

1.3 Quantum Annealing (DWAVE)

Quantum annealing [2][9] is currently attracting significant interest for its effectiveness as a quantum algorithm in efficiently and quickly solving certain large-scale combinatorial optimization problems, particularly those with unknown structures, where quantum advantage has been indisputably proven [8]. It is commercially implemented almost exclusively by D-Wave.

D-Wave's native sets of problems are more specifically Quadratic Unconstrained Binary Optimization (QUBO) problems, like spin-glass systems, where the energy (in other words the Hamiltonian H) is the quantity to be minimized and can be written as:

$$\min_{\vec{x}\in\{0\,;\,1\}^n} H(\vec{x}) = \min_{\vec{x}\in\{0\,;\,1\}^n} \vec{x}^T Q\,\vec{x} = \min_{\vec{x}\in\{0\,;\,1\}^n} \sum_{i,j} Q_{ij}\, x_i x_j \tag{1}$$

where Q is real and symmetric. The fundamental principle quantum annealing is inspired from — adiabatic quantum computation — involves starting with an 'easy' Hamiltonian H_0 (one whose eigenvalues and corresponding system energies are known). The system then gradually evolves (typically over a few milliseconds) to the problem Hamiltonian H_P, whose eigenvalues are the target of optimization.

$$H(t) = A(t)H(0) + B(t)H_P \tag{2}$$

Where A (resp. B) is a decreasing (resp. increasing) function of t. Once the problem Hamiltonian is implemented, the system — under assumptions such as a slow enough evolution — naturally evolves to its lowest energy states, which correspond to the minimum of the optimization problem, according to the adiabatic theorem [10]. The qubits are then measured to determine the solution to the QUBO problem. Because D-Wave's quantum annealing chips are noisy and meant to be fast, they sample low-energy solutions, occasionally reaching the desired optimum. Such quantum annealing solvers are therefore more heuristics based.

Quantum annealing technology is quite mature, with D-Wave's quantum processing units (QPUs) featuring thousands of functional qubits — around five thousand. Additionally, D-Wave has developed a hybrid solver [14][15] that can tackle optimization problems ranging from 20,000 to 1 million variables by employing quantum annealing for subproblems in parallel with classical optimization techniques.

2 Methodology

2.1 MHT Framework

The framework for the implementation of Multiple Hypothesis Tracking (MHT) involves modelling a 2-D radar tracker with a linear Kalman filter and a constant velocity model to estimate target trajectories. Radar noise is considered normally distributed, and clutter appearance is uniformly distributed across the radar surface, with the number of clutter measurements per timestep following a Poisson distribution. Additionally, the radar's probability of detection is adjustable.

2.2 MHT Implementation

In the context of the MHT algorithm, gating, pruning, and other hypotheses reduction techniques are crucial to manage the computational complexity and optimize performance:

- **Gating** is a process that reduces the number of possible associations between measurements and tracks by setting a threshold based on the distance metric (e.g., Mahalanobis distance) from predicted positions to actual measurements using the Linear Kalman filter (taking measurement noise into account). Only measurements within this threshold are considered for association with existing track hypotheses.

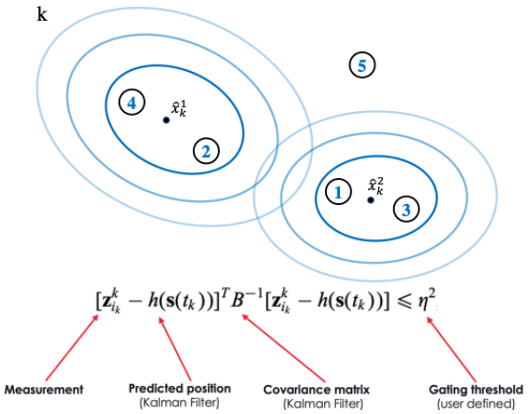


Fig. 2. Gating visualized *(top, extracted from [3])*: Measurements 2 and 4 (resp. 1 and 3) fall close to the prediction $\hat{x}_k^1$ (resp. $\hat{x}_k^2$) and are therefore considered as physically possible continuation of existing track hypotheses. 5 is not. Mathematical formulation *(bottom)* of the Mahalanobis distance between measurements and predictions.

- **Pruning** is a technique that reduces the problem size by eliminating hypotheses with low likelihood of truth from the hypotheses tree. This also involves cutting down branches in the hypothesis tree. Pruning [1] helps to keep the computational load reasonable by discarding hypotheses that have low probability or have become redundant, ensuring that the algorithm focuses resources on the scenarios that are most likely to be.
- Other techniques involve discarding track hypotheses with too many consecutive missed detections, or tracks that involve excessively high velocity. [3]

2.3 MWIS (Maximum Weighted Independent Set)

As it turns out, solving the DAP amounts to solving a MWIS problem. The latter can be described as an undirected graph problem, whose aim is to find the subset of non-conflicting hypotheses with the maximum total weight.

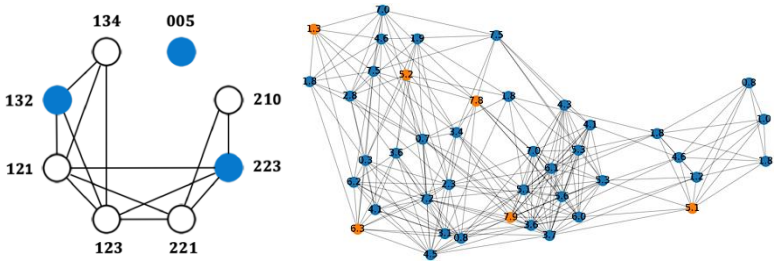


Fig. 3. A small undirected graph with edges (*left, extracted from [3]*): node 132 refers to the track hypothesis "Measurement 1 (M1) at timestep 1" + "M3 at timestep 2" + "M2 at timestep 3". Weights are not represented. The MWIS solution at timestep 3 are the blue nodes. An actual (yet simple) MHT-derived MWIS solution (*right*): on each node is written the score of the track hypothesis. Orange nodes are the MWIS solutions.

In this graph, each node represents a track hypothesis with a weight equal to its likelihood (or score). Edges connect nodes that conflict (i.e. that share a common measurement). The goal is to select a set of non-conflicting nodes (an independent set) with the highest total weight. Hence, solving the data association problem involves identifying this weighted independent set. The proper mathematical formulation implies dealing with binary values, where $x_i = 1$ means the node *i* is taken as a solution of the DAP at a given timestep. The exhaustive problem formulation is:

$$Given\ a\ graph\ (V,E):\ \text{find}\ \underset{\vec{x}\in\{0\,;\,1\}^V}{\text{argmax}} \sum_l w_l x_l \ \ s.t.\ x_i + x_j \leq 1, \forall (i,j) \in E \quad (3)$$

where w_l is the score (likelihood) associated to the node *l*.

As the graph scales up to thousands of nodes, finding this optimal set becomes increasingly complex, thus necessitating advanced approaches like quantum computing for efficient resolution. It is of interest to note that the MWIS problem is nothing but the Maximum Weight Clique problem (MWC) on the complementary graph.

2.4 Scenario for benchmarking

The goal of the study was to simulate complex enough scenarios that may arise in real settings and to solve it using the MHT algorithm with different solvers. The scenario in this paper involves a track moving along a straight line for 20 timesteps. Between frames 3 to 10, the expected number of false alarms increases by a factor of 100, introducing significant noise into the track's trajectory and adding complexity to the resolution of the MWIS problem.

Table 1: Input parameters for the scenario to be evaluated

Parameters	Description	Values taken
dth	Gating area for new detections	**30**
q	Kalman filter process variance on the acceleration ($m^2.s^{-3}$)	$\mathbf{10^{-5}}$
r	Estimate of measurement variance	**1**
n	Number of scan(s) before pruning	**2**
n_{miss}	Maximum consecutive missed detections	**4**
p_d	Probability of detection of the radar	**0.9**
$\lambda_{clutter}$	Expected number of false alarms per area unit per frame	$\mathbf{[10^{-5}, 10^{-4}]}$
$\lambda_{newtracks}$	Expected number of new tracks per area unit per frame	$\mathbf{[10^{-8}, 10^{-7}]}$
v_{max}	Maximum velocity of the detected objects (m/s)	**15**

The input parameters for the MHT algorithm are detailed in Table 1 above.
Various solvers were tested in this work and are summarized below:

1. Qualex-MS [11], a C-based heuristics solver with time complexity of $O(n^3)$.
2. NetworkX [5], an exact solver using Branch & Bound.
3. SCIP [7], a well-known heuristics solver that uses LP-relaxation.
4. Cliquer [12], an exact solver running on C using Branch & Bound.
5. TSM-WC [6], an exact solver of MWC implemented on C.
6. DWAVE's hybrid heuristics solver [14][15].

3 Results

Firstly, all tested solvers that were exact, other than TSM-WC and DWAVE, become significantly slower once the graphs reached 1000 nodes — typically for a false alarm density greater than $5\ 10^{-5}$. Qualex-MS and SCIP, the heuristics solvers, could not return good enough solution quality to the MHT problem and were discarded. For each false alarm parameter setting, classical solvers often exhibited a peak in computation time, around frame 8 or 9 (see Fig. 4).

In response, DWAVE's hybrid solver [14] was implemented to handle these computational peaks, while the classical solver continued to manage the simpler instances. A comparison was then made between the TSM-WC solver alone and the TSM-WC solver supplemented by DWAVE during intensive computations.

As DWAVE's hybrid solver is a black box algorithm, no detailed information on its specific way of solving is disclosed, hence the absence of more sophisticated explanation. The QPU access time is however available after each call, and quantum annealing makes up approximately ten percent of the solver's computing time. It should also be noted that DWAVE's hybrid solver has a minimum runtime of 3 seconds, explaining the lower bound observed on Fig. 5.

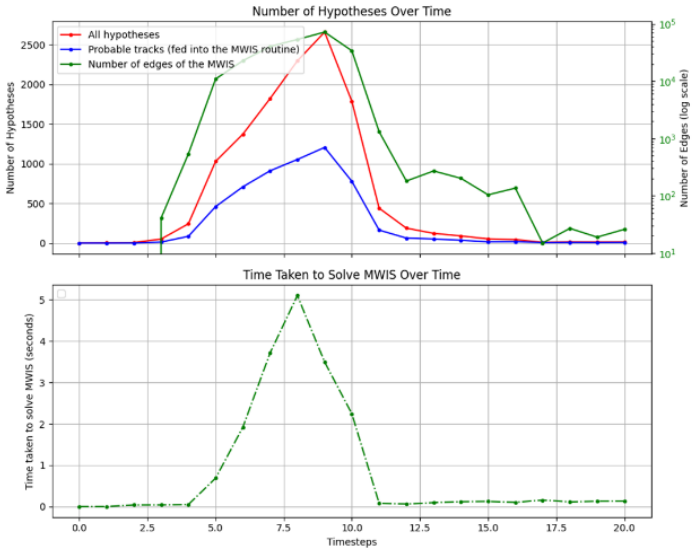


Fig. 4. Number of nodes (hypotheses) and edges (conflicting hypotheses) at each timestep for a given MHT routine (*top*): the probable tracks are those with positive score and are fed to the MWIS routine. (*bottom*): time taken to solve the MWIS routine for TSM-WC.

Table 2 Comparison on TSM-WC and DWAVE time to solution for the hardest instance.

	Time Taken (seconds)	
Solver / $\lambda_{clutter}$	**TSM-WC**	**DWAVE**
10^{-5}	0.1	3
$7.5 \cdot 10^{-5}$	1.48	3
$9 \cdot 10^{-5}$	6.8	3.6
10^{-4}	11.09	3.5
$1.1 \cdot 10^{-4}$	94	6

Table 2 above illustrates the time taken for TSM-WC and DWAVE for the hardest MWIS problem in each scenario (as depicted in Figure 4 above).

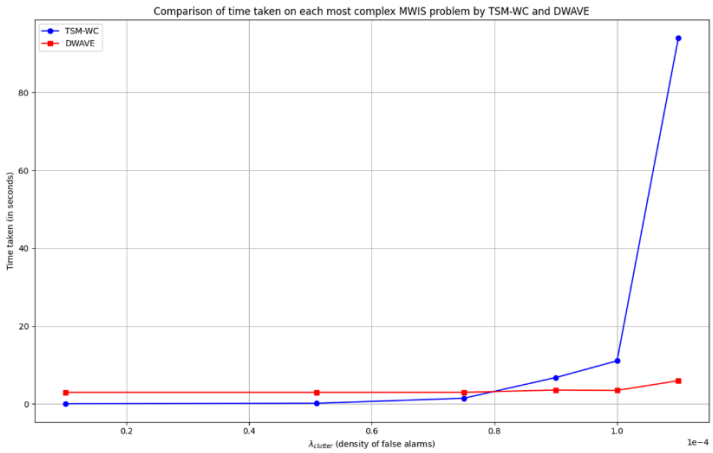


Fig. 5. Comparison on TSM-WC and DWAVE on those high clutter density calculations

Results are displayed in Figure 5. When comparing TSM-WC and DWAVE during particularly demanding calculations (peaks), the advantage of DWAVE's hybrid solver is clear.

As observed, quantum computing is beginning to offer substantial time advantages when the density of false alarms reaches 10^{-4}. False alarms densities higher than this value yielded MWIS problems too hard to be solved in reasonable times and were not explored. Nevertheless, a measured 15-fold speed-up in computation time was achieved, while maintaining similar results and solution quality. We emphasize this fact, as most classical heuristics solvers did not meet the required solution quality and were therefore not considered for the rest of the study. For this benchmark, if the track hypothesis corresponding to the ground truth (track going in a straight line) was not part of the MWIS solution at the end of the 20 frames — up to negligible missed detections and false alarms — then, the solution quality was deemed mediocre.

As the cost of QPU usage was non neglectable, no other runs were made, and no statistical values have been extracted from this study.

4 Conclusion

The integration of DWAVE with classical solvers shows promise in handling computationally intensive tasks more efficiently, marking a significant step forward in leveraging quantum computing for complex problems.

Current work is underway to evaluate MHT algorithms on C12's novel quantum computer using methods such as developed in [16][17][18]. C12 carbon nanotubes' spin qubits offer long coherence times [19], low error rates [20], and highly connected qubits [21], enabling both digital and analog modes with operation speeds comparable to superconducting qubits. The native use of spins also allows for a more direct implementation of Ising models.

All computations were performed on a MacBook Air (2022) equipped with an Apple M2 chip featuring 8 cores (4 performance cores and 4 efficiency cores), 8 GB of RAM, and running macOS Sonoma.

5 Acknowledgements

The authors wish to acknowledge DWAVE for the cloud access to its QPU.

References

1. D. Reid. An algorithm for tracking multiple targets. IEEE, 1979.
2. Sebenik et al. “Quantum annealing: A new method for minimizing multidimensional functions.” Chemical Physics Letters, 1994.
3. A. Ciptadi J. M. Rehg C. Kim, F. Li. Multiple hypothesis tracking revisited. IEEE, 2015.
4. P.P.A. Stormsa, F.C.R. Spieksma, An LP-based algorithm for the data association problem in multitarget tracking. IEEE, 2000
5. Hicks Illya V. Warren, Jeffrey S. Combinatorial branch-and-bound for the maximum weight independent set problem. Technical Report, Texas AM University, 2016.
6. Y. Liu F. Manya H. Jiang, C.-M. Li. A two-stage maxsat reasoning approach for the maximum weight clique problem. The Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
7. Tobias Achterberg, SCIP, Solving Constraint Integer Programs, Mathematical Programming Computation, Volume 1, Number 1, Pages 1–41, 2009
8. Vasil S. Denchev et al. What is the Computational Value of Finite-Range Tunneling? Phys. Rev. X 6, 031015, 2016.
9. Edward Farhi et al. “A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem” Science, 2001
10. Born, M. and Fock, V. A. "Beweis des Adiabatensatzes". Zeitschrift für Physik A. **51** (3–4): 165-180 (1928).
11. Stanislav Busygin, A new trust region technique for the maximum weight clique problem, 2006.
12. Sampo Niskanen and Patric R. J. Östergård, "Cliquer User's Guide, Version 1.0," Communications Laboratory, Helsinki University of Technology, Espoo, Finland, Tech. Rep. T48, 2003.
13. Jack Raymond et al. “Hybrid Quantum Annealing for Larger-than-QPU Lattice-structured Problems” ACM Transactions on Quantum Computing 4, 3, Article 17 (2023).
14. D-Wave, “Hybrid solvers for quadratic optimization,” 2022.
15. Osaba, E., & Miranda-Rodriguez, P. (2024). D-Wave’s nonlinear-program hybrid solver: Description and performance analysis.
16. Cazals, Pierre, et al. "Quantum Optimization on Rydberg Atom Arrays with Arbitrary Connectivity: Gadgets Limitations and a Heuristic Approach." *arXiv preprint arXiv:2508.06130* (2025).
17. Garrigues, Mathieu, et al. "A Scalable Heuristic for Molecular Docking on Neutral-Atom Quantum Processors." *arXiv preprint arXiv:2508.18147* (2025).
18. Tomesh, Teague, et al. "Divide and conquer for combinatorial optimization and distributed quantum computation." *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 1. IEEE, 2023.
19. Neukelmance, B., et al. "Microsecond-lived quantum states in a carbon-based circuit driven by cavity photons." *Nature Communications* 16.1 (2025): 5636.
20. Wu, Yi-Hsien, et al. "Simultaneous High-Fidelity Single-Qubit Gates in a Spin Qubit Array." *arXiv preprint arXiv:2507.11918*(2025).

21. Mammola, Andrea, M. M. Desjardins, and Q. Schaeverbeke. "Optimal connectivity from idle-qubit residual-coupling crosstalk in a cavity-mediated entangling gate." *Physical Review A* 111.4 (2025): 042621.

G-score, the Maximum Cardinality Matching G_n series as a benchmark for quantum computers

Stéphane Louise[1][0000−0003−4604−6453]

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
stephane.louise@cea.fr
https://list.cea.fr/

Abstract. The Maximum Cardinality Matching (MCM) Gn series represents a well-established set of problems designed to evaluate the capabilities of optimization heuristics. These problems have been instrumental in probing the optimization potential of various DWave quantum computers, offering insights into their performance and limitations. Complementary to the NP-hard yet unconstrained MaxCut problem used in the Q-Score benchmark, the MCM Gn series provides a constrained series of problem that further challenge quantum optimization techniques.
In this paper, we introduce the G-score benchmark specifically tailored to assess the capabilities of quantum optimization heuristics such as Quantum Annealing (QA) and the Quantum Approximate Optimization Algorithm (QAOA) or any other heuristic that can be relevant on quantum computers. This score aims to offer a comprehensive metric for evaluating the effectiveness of these methods in solving complex optimization problems.

Keywords: optmization · quantum computing · maximum cardinality matching · benchmark

1 Introduction and context

Optimization problems are ubiquitous across various scientific and industrial domains, ranging from logistics and supply chain management and financial modeling. The ability to efficiently solve these problems can lead to significant advancements and economic benefits, making optimization one of the most critical applications in computational science.

Techniques like Quantum Annealing (QA)[3] and the Quantum Approximate Optimization Algorithm (QAOA)[2] have shown potential in solving optimization problems more efficiently than classical methods. However, the practical realization of this potential hinges on the ability to accurately benchmark and evaluate the performance of quantum computers. Benchmarking is crucial for understanding the capabilities and limitations of quantum hardware and algorithms, guiding their development, and ensuring their effective deployment in real-world applications.

One well-established set of problems used to probe the capabilities of optimization heuristics is the Maximum Cardinality Matching (MCM) G_n series.

These problems have been instrumental in assessing the performance of Simulated Annealing (SA), one of the major main-stream optimization heuristic. Recently, it was also used to characterize the family of DWave quantum computers, providing valuable insights into their optimization capabilities[12, 8, 13]. The MCM G_n series complements the NP-hard but unconstrained MaxCut problem utilized in ATOS's Q-Score benchmark[7], by offering a set of constrained problems which are widely spread in real-world applications.

In this paper, we introduce a benchmark score, the G-Score specifically designed to evaluate the capabilities of quantum optimization heuristics such as QA and QAOA on quantum computers. Nonetheless, it is worth noting that any new and relevant quantum heuristic can be tested by the protocol and can provide a value of G-score. This score aims to provide a comprehensive metric for assessing the effectiveness of these methods in solving complex optimization problems.

Applicative benchmarks can be a cornerstone of measuring the gap to quantum utility. Several such initiatives are already into inception e.g. in the U.S.A. with QED-C[6], in Germany with BenchQC[4], or France with BACQ[1]. We place our contribution here presented in the French initiative MetriQs/BACQ as part of the evaluation of Quantum Optimization heuristics, in complement of the aforementioned Q-Score.

2 The evaluation of quantum optimization heuristics

The set of problems included in the optimization benchmark of BACQ encompasses a wide variety of classes so that to test different aspects of the capabilities of optimization heuristics on Quantum Processing Units (QPUs).

Table 1. Classes of optimization problems addressed by the BACQ benchmarking protocol in its current and planed form

Types	Quadratic Unconstrained	Quadratic Constrained	Higher order
Polynomial	–	MCM G_n	HORN-Sat
NP or NP-hard	QScore MaxCut	–	Max-Sat

While the MCM G-series is not NP-hard, it is specifically designed to be a challenging optimization problem for heuristics. Thus it is important to probe this class of problems, especially because most of real-world optimization problems are indeed constrained.

Higher-order problems that appears in the last column of Table 2 pose additional challenges for quantum optimization heuristics, particularly for Quantum Annealing (QA), due to the quadratization step. This step can easily increase the number of variables by one or even two orders of magnitude, depending on the problem at hand. It is also important to note that constrained optimization

problems are already difficult for QA. Addressing higher-order problems will represent a step further the current work in assessing the capabilities of quantum heuristics, a will constitute future work.

3 Maximum cardinality matching on the $\mathbf{G_n}$ series

3.1 Definitions and problem statement

Let $G = (V, E)$ be an non directed graph. Each edge $e \in E$ can considered as a set of two vertices (i.e.: $e = \{v_0, v_1\}$ with $v_0, v_1 \in V$).

Definition 1 (neighbor). *A vertex $v \in V$ is a neighbor of a vertex $u \in V$ if and only if $\{u, v\} \in E$. An edge $e \in E$ is a neighbor of an edge $e' \in E$ if and only if $e \cap e' \neq \emptyset$ and $e \neq e'$. The set of all neighbors of a vertex or an edge x is denoted $\Gamma(x)$. For undirected graphs, for all $(x, x') \in V^2 \cup E^2$, $x' \in \Gamma(x) \Leftrightarrow x \in \Gamma(x')$.*

Definition 2 (line graph). *The line graph of G is the graph $LG(G) = (E, E')$ where E' is defined as follows:*

$$(e_1, e_2) \in E' \text{ if and only if there exists } v \in V \text{ such that } e_1 \cap e_2 = \{v\} \quad (1)$$

Definition 3 (matching). *A matching of G is a set of edges $M \subseteq E$ without common vertices, that is, for all $\{e, e'\} \in M$, $e \cap e' = \emptyset$.*

The maximum cardinality matching (MCM) problem is the following : Given an undirected graph $G = (V, E)$, find a matching M with the largest number of edges (i.e.: for all M' matching of G, $|M| \geq |M'|$).

3.2 $\mathbf{G_n}$ series

Definition 4 (Complete bipartite graph $K_{n,m}$). *Let be $n, m \in \mathbb{N}^*$, a complete bipartite graph $G = (V, E)$ is a graph that can be partitioned into two subsets of vertices $V_l, V_r \subseteq V$ such that:*

- *$V_l \sqcup V_r = V$, that is, $V_l \cup V_r = V$ and $V_l \cap V_r = \emptyset$.*
- *for all $u, v \in V$, $\{u, v\} \in E \Leftrightarrow u \in V_l$ and $v \in V_r$.*

Such a graph is denoted $K_{n,m}$ with $n = |V_l|$ and $m = |V_r|$.

Definition 5 (G_n graph). *Let be $n \in \mathbb{N}^*$, the graph $G_n = (V_n, E_n)$ is composed of n bipartite graphs $K_{n+1,n+1}$ linked together with $n + 1$ edges between them. Bipartite graphs on each end are linked with $n+1$ edges to $n+1$ 1-degree vertices (see an example on Figure 1 for $n = 3$). The total number of vertices is $|V_n| = 2(n + 1)^2$ and the total number of edges is $|E_n| = (n + 1)^3$.*

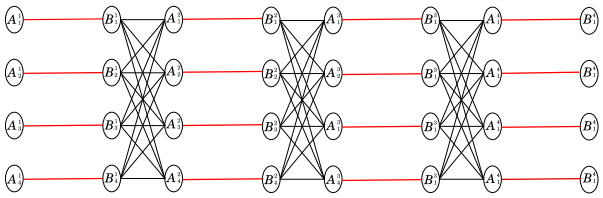


Fig. 1. G_3 graph with its maximum cardinality matching highlighted in red. The matching is done between A nodes and B nodes with a single match per node at most.

The solution of the maximum cardinality matching problem on each graph of that series is unique, composed of all edges between all the $K_{n+1,n+1}$ subgraphs (the red edges in Figure 1), for a total of $(n+1)^2$ edges. That series is interesting because G. Sasaki and B. Hajek proved that simulated annealing cannot find the solution of the maximum matching problem over the G_n graphs in polynomial time [9]. As simulated annealing can be considered as a classical equivalent (but without quantum properties) of the Quantum Annealing (QA)[3] and the Quantum Approximate Optimization Algorithm (QAOA)[2], two of the most commonly used quantum algorithms in the field of optimization, the same tool can be used for probing the optimization heuristics in quantum computing. Solving the MCM problem on a graph G is also equivalent to solve the MIS problem on its line graph $LG(G_n)$.

4 Measuring utility and setting the benchmark principles

Given the inherent randomness of QPU outcomes, any evaluation of quantum utility must discriminate between cases where a QPU achieves a solution through genuine computational effort versus sheer luck. This distinction is especially important for noisy quantum devices, which struggle to reliably execute complex computations or maintain the fidelity of the quantum states required for meaningful problem-solving. A robust benchmarking framework must therefore ensure that correct results reflect real algorithmic performance, not random sampling or noise-induced errors. This is critical to avoid overstating the capabilities of noisy quantum hardware.

4.1 Number of variables and number of qubits utilized

A key criterion in addressing this issue is the number $\mathcal{N}$ of binary variables (for computational tasks) or spin states (for physics simulations) relevant to the problem's outcome. Reducing this number $\mathcal{N}$ would be considered unfair, as it would artificially shrink the solution space and falsely enhance the perceived quantum utility measured by the benchmark.

It is, therefore, crucial to preserve this number $\mathcal{N}$ and ensure that the actual implementation on the QPU utilizes $\mathcal{N}$ as the basis for the benchmark application. Notably, for gate-based QPUs, the number of qubits required for the implementation of the problem is typically greater than $\mathcal{N}$, due to overheads such as error correction and auxiliary qubits. For analog QPUs (quantum simulators), the required number may also exceed $\mathcal{N}$, not due to the algorithm itself but because of embedding constraints imposed by the hardware.

In either case, it should be transparent in the benchmark outcome that at least $\mathcal{N}$ qubits are effectively used. Under such conditions, the likelihood of a random outcome from the QPU would be proportional to $\frac{1}{2^{\mathcal{N}}}$. To ensure that the observed results are not merely a consequence of random chance, it is essential to apply a rigorous statistical validation. Specifically, a 5-sigma (5σ) criterion—commonly used in physics to signify a result beyond reasonable doubt—should be employed if possible. Considering the cost of doing heavy computation and sampling on nowadays QPUs, in the short term, a 3-sigma criterion can be regarded as sufficient for a first evaluation.

As a general good practice, this number $\mathcal{N}$ will be pivotal in defining the scale at which a given test is performed and in case of a positive result, it is advised that it is a base of defining one element of the outcome of the benchmark.

4.2 Utility evaluation

In some scenarios, an approximate result may be seen as an acceptable outcome of the program—for example, obtaining an energy level sufficiently close to the ground state of a Hamiltonian or achieving a near-optimal solution to an optimization problem. However, in other cases, the solution is binary: it is either correct or incorrect. In these instances, the utility of an incorrect result is effectively zero, regardless of how many qubit values in the output were accurate. This distinction highlights the need for a case-by-case approach when evaluating each application of a given benchmark.

4.3 Measure tools for the evaluation

Hamming distance The Hamming distance is a metric that measures the number of bit positions at which two binary strings of equal length differ. In the context of evaluating the output of a quantum or classical program execution, it can serve as an effective measure of correctness and performance by quantifying how close the computed output is to the expected or optimal solution.

In the case of our specific need in Benchmarking, the Hamming distance is used for

- Correctness Assessment: For outputs where the exact solution is known (e.g., prime factorization or solving a linear system), the Hamming distance from the expected result indicates the effectiveness of the program and hardware. A low or zero Hamming distance reflects higher accuracy.

– Thresholding Success: A predefined Hamming distance threshold can be used to determine whether the execution is acceptable. For example, outputs with a Hamming distance below a certain value might still be considered useful approximations in certain applications.

Thus, the Hamming distance provides a practical, quantitative measure to evaluate and compare the effectiveness of programs and hardware implementations, especially in noisy or approximate computational environments.

Optimality of the solution Whenever an exact solution is not necessarily required—such as in combinatorial optimization, where a suboptimal but sufficiently good solution can be acceptable as long as the computational cost is low (i.e., a favorable quality-to-cost or quality-to-energy ratio)—a measure of distance to the optimal solution can be used for evaluation. While this is particularly challenging for NP-hard problems, the Q-Score Maxcut provides a framework to asymptotically assess the optimality of the best-found solution.

For problems involving constraints, which is often the case with real-world optimization tasks, it may be important to account for how well the proposed solutions satisfy those constraints. For instance, solutions that violate constraints could be penalized to reflect their reduced utility. Striking the right balance in weighting these factors is a challenging task, but the goal is to propose initial approaches and, through further development and testing, determine whether these choices are meaningful and effective.

4.4 Defining a metric for the MCM G_n series

In the case of the G_n series the problem is heavily combinatorial and allow only one optimal solution comprising $(n+1)^2$ over $\mathcal{N} = (n+1)^3$ variables for the problem. What we aim to obtain by the specified metric would be to have an evaluation of the number of variables that the QPU can handle and for which a relevant solution can be obtained. We can define the relevant thresholds and associated parameters:

– The value n_s for which the problem G_{n_s} can be solved exactly with a significant probability.
  • For G_1 which comprises 8 variables, the threshold should be $\frac{1}{32} \simeq 3.2\%$ over a minimum sampling of 1000 (the random case being $\frac{1}{256} \simeq 0.4\%$,
  • From G_2 as the number of variables is already significant, the threshold can be placed at $\frac{1}{1000}$ when a minimum of 10000 runs/samples are obtained. Otherwise, it should be at least $\frac{10}{\mathcal{S}}$ if $\mathcal{S}$ is the number of runs/samples, with a minimum number of sample of 1000 to be significant.
– For the first "failed" graph of the series $(n_s + 1)$, we can fine tune the evaluation by utilizing the two following metrics:
  • The relative Hamming distance to the optimal solution $r_H = \frac{d_H}{(n_s+1)^3}$ where d_H is the Hamming distance (see 4.3)

- An evaluation of the optimality taking into account the invalidation of constrains into the best found solution: $r_d = \left(0, \frac{\mathcal{L}-f}{(n_s+2)^2}\right)$ where $\mathcal{L}$ is the number of links in this best found solution and f being the number of failed constraints in it
- The choice of the "best" failed solution should be a reliably found solution (i.e. whose probability of occurrence should be higher than 1%) but that may take into account some light post-processing to account for glitches in the mapping of the variables (e.g. if we found inconsistencies in the values of the qubits corresponding to a single variable, it would be possible to do a "majority vote" to decide the value of this variable as the outcome of the QPU processing, and it may take into account degenerate solutions, i.e. solution that have the exact same optimality score)
- By these choices, we can define a ratio to optimality that we can choose to be

$$r_o = \frac{1}{2}(1 - r_H + r_d) \tag{2}$$

- Then the base metric in number of variables can be defined as:

$$S_{G_1} = (n_s + 1 + r_o)^3 \tag{3}$$

because $(n_s+1)^3$ is the number of edges in the G_{n_s} graph and $(n_s+2)^3$ is for the first failed G_{n_s+1} graph, so it interpolates smoothly between the two.

Illustration on DWave-2000Q: without pre-processing, according to the results presented in the PhD thesis of D. Vert [11], shows $n_s = 2$, and for G_3 the Hamming distance is 8, so $r_H = \frac{8}{64}$ and that there are 15 links with 3 broken constraints also on the best found G_3 so $r_H = \frac{8}{64} = \frac{1}{16}$ and $r_d = \frac{15-3}{16} = \frac{12}{16}$ therefore $r_o = \frac{1}{2}(1 - \frac{1}{16} + \frac{12}{16}) = \frac{29}{32}$ and the result of the DWave-2000Q for the G-Score is:

$$S_{G_1,dw2000q} = \left(2 + 1 + \frac{29}{32}\right)^3 = 59.6$$

The S_G score is open-ended and gives an approximation of the capability of the QPU utilizing the specified algorithm to operate simple cost functions with hard constraints.

One important caveat of S_{G1} score, is that it does not take into account the probability of finding the solution at n_s and this creates an obvious bias toward $(n_s+2)^3$ because of the non-linearity of the power function. This is noticeable on the experimental results shown in section 5. As a consequence, we decided to put more emphasis on the n_s value by taking into account the probability of finding the optimal solution for G_{n_s} by defining the G-score as follows:

$$S_{G_1} = (n_n + 1 + r_o \times p_{n_s})^3 \tag{4}$$

where p_{n_s} is the probability of finding the solution for G_{n_s} using the considered heuristic. As both r_o and p_{n_s} are in the interval $[0, 1]$ so is their product. This gives a much stronger bias toward n_s which is a more relevant way of characterizing the capabilities of the tested heuristic.

Example On the same evaluation of DW2000Q, the probability of finding the optimal solution for G_2 is $p_2 = 0.06$ (6%) so the S_G score for the DW-2000Q is:

$$S_{G,\mathrm{dw2000q}} = (2 + 1 + \frac{29}{32} \times 0.06)^3 = 28.5$$

which is a more accurate evaluation of the real solving capability of the DW2000Q without using pre-processing.

5 S_G score on the DWave QPU family

In this set of experimental validations, we used the same methods and principles (and a part of the open-source code) from the authors in [13]. We obtained the score over 3 generations of the annealers, thus measuring the improvements in the evolution of this family of QPUs.

Table 2. Results on several generations of DWave QPUs. In the table, n_s represent the highest problem solved with statistical significance, r_o is a ratio of optimality at the first failed rank $n_s + 1$ (see eq. 2), $p_n s$ is the probability of finding the solution at rank n_s and S_{G_1} and S_G are the values of the two (naive and improved) G-scores. It is worth noting that the experiments are done with enough sampling such that the scores are accurate at the number of significant digits written in the table. Scores from classical heuristics from [5]. For classical heuristics, the time to solution is enforced at 60s maximum.

QPU architecture	n_s	r_o	S_{G_1}	p_{n_s}	S_G
DWave-2000Q naive embedding	2	0.906	59.6	0.06	28.5
DWave-2000Q (Chimera)	2	0.938	61	0.849	54.7
DWave Advantage (Pegasus)	3	0.904	116	0.146	70.5
DWave Advantage-2-proto (Zephir)	3	0.936	122	0.315	79.2
Classical Heuristics tested					
Simulated Annealing	5	0.92	331	4%	220
NEC's Vector Annealing V.2	7	0.93	712	0.7%	513
NEC's Vector Annealing V.3	7	0.95	717	100	717

These results were sampled over sampling sets between 10^4 and 10^5 on several architectures when it was possible (i.e. at the moment, only on the Advantage QPUs). The difference between our results on the DWave-2000Q and the results derived in section 4.4 can be solely attributed to a pre-processing step for improving the minor-embedding of the problem on the QPU. This procedure is described and available as open-source code in [13] and was reused mostly verbatim in these new tests.

6 Potential critics

One another fair critic that arose of this series of problems is the reduction of the spectral gap rapidly when n grows. In [8] the authors evaluate that there is a risk of an exponential reduction of the spectral gap. While not rigorously demonstrated, the point is nonetheless valid. But it is also worth noting that such a statement is probably also true for any real-world optimization problem, as it is though that $BQP \neq NP$ and any heuristic would be poised by its ability to solve a NP-hard problem. Therefore, it is highly probable that for real-world problem, the spectral gap will also collapse exponentially as the number of variables increases.

Instead of viewing this risk as a limitation, our intuition leads us to consider it as a probing tool for assessing the utility of quantum heuristics in solving real-world problems. We expect that heuristics capable of accounting for the hard constraints of the G_n problem will perform much better than standard QA and QAOA. As the heuristic to be used is not specified, this benchmark will serve as a valuable basis also for comparing various quantum heuristics.

We noted in section 3.2 that solving the G_n series is equivalent to solving the Maximum Independent Set problem (MIS) on its line graph. MIS is also formally equivalent to the Maximum Clique (MaxClique) problem which is NP-hard in the general case. A well-known study in the literature defines a Q-Score MaxClique [10] by analogy to the Q-Score MaxCut from Eviden, which is already integrated into BACQ. One valid critique, therefore, is that the G-score might be redundant with the Q-Score MaxClique. However, experimental results indicate that the Q-Score MaxClique on DWave machines yields values nearly identical to those of the Q-Score MaxCut. Moreover, both scores on DWave machines appear to be primarily constrained by the hardware's capacity to embed a $(N, \frac{1}{2})$-Erdős-Rényi graph under either definition of the Q-Score. This limitation does not apply to the G-score, which provides a metric more closely aligned with the actual capabilities and utility limits of DWave hardware in solving optimization problems.

7 Future work

In this paper, we defined the G-Score benchmark that probe the capabilities of quantum optimization heuristic for constrained combinatorial problem, thus evaluating their real-world quantum utility. It will complement nicely with the Q-Score MaxCut aimed at evaluating these heuristics on unconstrained NP-hard problems.

The next steps will aim at generalizing the benchmark results on several other QPU families including potentially Neutral Atom based and Gate-based QPUs. The code and the instances as well as the code for computing the G-Score will be provided as open-source software, and will permit reproducibility tests. As noted previously, some future work will also be done toward a utility score based on the MaxSat family of problems.

This study was funded by the MetriQs/BACQ program under the France 2030/National Strategy for Quantum Technology initiative. MetriQs France 2030, ANR-22-QMET-0002. The authors have no competing interests and operate independently of any QPU provider.

References

1. Barbaresco, F., Rioux, L., Labreuche, C., Nowak, M., Olivier, N., Nicolazic, D., Hess, O., Guilmin, A.L., Wang, R., Sassolas, T., Louise, S., Snizhko, K., Misguich, G., Auffèves, A., Whitney, R., Vergnaud, E., Schopfer, F.: Bacq – application-oriented benchmarks for quantum computing (2024)
2. Farhi, E., Goldstone, J., Gutmann, S.: A Quantum Approximate Optimization Algorithm. arXiv preprint arXiv:1411.4028 (2014)
3. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M.: Quantum Computation by Adiabatic Evolution. arXiv preprint quant-ph/0001106 (2000)
4. Geissler, F., Stopfer, E., Ufrecht, C., Meyer, N., Scherer, D.D., Wagner, F., Oberreuter, J.M., Chen, Z., Farace, A., Gutina, D., Schwenk, U., Lange, K., Junk, V., Husslein, T., Erdmann, M., Kiwit, F., Decker, B., Shaji, G., Granet, E., Dreyer, H., Dragan, T.A., Lorenz, J.M.: Benchqc – scalable and modular benchmarking of industrial quantum computing applications (2025), https://arxiv.org/abs/2504.11204
5. Louise, S.: Vector annealing, a quantum-inspired technique: Benchmarking performance against quantum and simulated annealing within the bacq framework. In: IEEE International Conference on Quantum Computing and Engineering, QCE 2025, Alburqueque, NM, USA, Aug 31-Sept.5, 2025. IEEE (2025)
6. Lubinski, T., Goings, J.J., Mayer, K., Johri, S., Reddy, N., Mehta, A., Bhatia, N., Rappaport, S., Mills, D., Baldwin, C.H., Zhao, L., Barbosa, A., Maity, S., Mundada, P.S.: Quantum algorithm exploration using application-oriented performance benchmarks (2024), https://arxiv.org/abs/2402.08985
7. Martiel, S., Ayral, T., Allouche, C.: Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way. IEEE Transactions on Quantum Engineering **2**, 1–11 (2021). https://doi.org/10.1109/tqe.2021.3090207
8. McLeod, C.R., Sasdelli, M.: Benchmarking D-Wave Quantum Annealers: Spectral Gap Scaling of Maximum Cardinality Matching Problems. In: Groen, D., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) Computational Science – ICCS 2022. pp. 150–163. Springer International Publishing, Cham (2022)
9. Sasaki, G.H., Hajek, B.: The time complexity of maximum matching by simulated annealing. Journal of the ACM (JACM) **35**(2), 387–403 (1988)
10. van der Schoot, W., Wezeman, R., Neumann, N.M.P., Phillipson, F., Kooij, R.: Extending the q-score to an application-level quantum metric framework (2024), https://arxiv.org/abs/2302.00639
11. Vert, D.: Étude des performances des machines à recuit quantique pour la résolution de problèmes combinatoires. (Study of the performance of quantum annealing machines for solving combinatorial problems). Ph.D. thesis, University of Paris-Saclay, France (2021), https://tel.archives-ouvertes.fr/tel-03208838
12. Vert, D., Sirdey, R., Louise, S.: Benchmarking Quantum Annealing Against "Hard" Instances of the Bipartite Matching Problem. SN Computer Science **2**(2), 106 (Feb 2021). https://doi.org/10.1007/s42979-021-00483-1
13. Vert, D., Willsch, M., Yenilen, B., Sirdey, R., Louise, S., Michielsen, K.: Benchmarking quantum annealing with maximum cardinality matching problems. Frontiers Comput. Sci. **6** (2024). https://doi.org/10.3389/FCOMP.2024.1286057, https://doi.org/10.3389/fcomp.2024.1286057

Benchmarking Data Encoding Methods in Quantum Machine Learning

Orlane Zang[1,2], Grégoire Barrué[2], and Tony Quertier[2]

[1] Université de Picardie Jules Verne, Amiens, France
orlane.zang@u-picardie.fr
[2] Orange Innovation, Rennes, France
{orlane.zang,gregoire.barrue,tony.quertier}@orange.com

Abstract. Data encoding plays a fundamental and distinctive role in Quantum Machine Learning (QML). While classical approaches process data directly as vectors, QML may require transforming classical data into quantum states through encoding circuits, known as quantum feature maps or quantum embeddings. This step leverages the inherently high-dimensional and non-linear nature of Hilbert space, enabling more efficient data separation in complex feature spaces that may be inaccessible to classical methods. This encoding part significantly affects the performance of the QML model, so it is important to choose the right encoding method for the dataset to be encoded. However, this choice is generally arbitrary, since there is no "universal" rule for knowing which encoding to choose based on a specific set of data. There are currently a variety of encoding methods using different quantum logic gates. We studied the most commonly used types of encoding methods and benchmarked them using different datasets.

Keywords: Quantum Machine Learning · Data embedding · Optimization.

Introduction

Quantum Machine Learning (QML) is a research area that focuses on the development of Machine Learning (ML) algorithms that can be executed by a quantum computer. For solving any specific problem, QML algorithms can be designed as a quantum circuit with a sequence of different quantum gate operations. However, designing QML algorithms that perform well as or better than conventional ones is still a challenge, due to limited ressources of Noisy Intermediate Scale Quantum (NISQ) computers [1]. Quantum encoding involves the conversion of classical information into quantum states, enabling QML algorithms to operate efficiently. This process is fundamental, as the way in which data is encoded can significantly influence the results obtained by learning models [2]. Various encoding methods have been proposed in the literature [3], and in this work we make a comparative study of encoding methods that are recurrent in the literature. To this end, we have used a quantum neural network classification model, presented in Fig. 1, and trained it on real datasets, noting the results

obtained for each encoding. The rest of the paper is divided in four parts. The encoding methods used for our benchmarking are presented in Section 1, and Section 2 covers presentation of individual datasets used and preprocessing step. The results for each dataset are presented in Section 3, with interpretations. We end with a conclusion and outlook (Section 4).

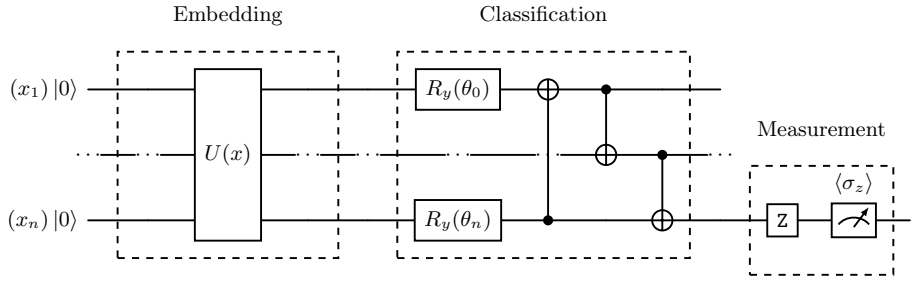


Fig. 1: A layer of the QML model used and the measurement part. Only the embedding part varies. The classification and measurement parts are the same for all the encodings tested.

1 Encoding methods

1.1 Angle encoding

The first type of encoding used is angle encoding [4]. For a classical dataset $D = (x, y)$, where $x \in \mathbb{R}^N$ is a N-dimensional input data (the features) and y the class label for the corresponding data, we can encode x into the quantum state $|\psi\rangle = \bigotimes_{j=1}^{N} |\psi_j\rangle$. Depending on the type of angle encoding used, the $|\psi_j\rangle$ state can take one of the following forms:

$$\begin{cases} |\psi_j\rangle = \mathtt{R_X}(x_j)\,|0\rangle = \cos\frac{x_j}{2}\,|0\rangle - i\sin\frac{x_j}{2}\,|1\rangle\,, & \textbf{Simple Angle}, \\ \\ |\psi_j\rangle = \mathtt{U}(x_j)\mathtt{H}\,|0\rangle = \cos x_j\,|0\rangle + \sin x_j\,|1\rangle\,, & \frac{\pi}{4}\textbf{-Angle encoding}, \end{cases} \tag{1}$$

where $\mathtt{R_X}$ is the Pauli rotation gate around the x-axis, $\mathtt{H}$ is the Hadamard gate, and $\mathtt{U}(x)$ is an unitary operation [5] with the following square matrix

$$\mathtt{U}(x) = \begin{pmatrix} \cos\left(\frac{\pi}{4} - x\right) & \sin\left(\frac{\pi}{4} - x\right) \\ \\ -\sin\left(\frac{\pi}{4} - x\right) & \cos\left(\frac{\pi}{4} - x\right) \end{pmatrix}. \tag{2}$$

Fig. 2 presents these two types of angle encoding.

Another extension of angle encoding used for our simulations is the so called ***Entangled Angle encoding*** [6], which is constructed by applying a *Hadamard*

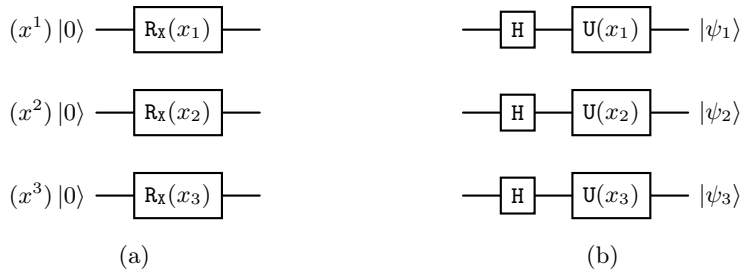


Fig. 2: *Simple Angle encoding (a) and* $\frac{\pi}{4}$*-Angle encoding (b).*

gate to each qubit, followed by the typical Simple Angle encoding (with $\mathtt{R_Y}$ gate), and a *Controlled-NOT* gate between each pair of adjacent qubits (Fig. 3). In all three cases of angle encoding, the number **n of qubits** required to encode **data of size N** is such that $\mathbf{n} = \mathbf{N}$.

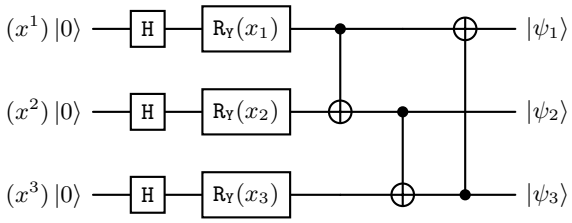


Fig. 3: *Entangled Angle encoding.*

1.2 Amplitude encoding

Amplitude encoding embeds classical features of the form $\sum_{i=1}^{N} \left|x_i\right|^2 = 1$, in an amplitude vector of a n-qubits quantum state $|\psi\rangle = \sum_{i=1}^{N} x_i\,|i\rangle$, where $N = 2^n$, $|i\rangle$ is the i-th computational basis state, and x_i is the i-th element of x. Among the methods used to carry out this encoding, we used *Algorithm 1* of the paper [7] called the *divide and conquer algorithm*. Indeed, it is a two-part algorithm, the first part computes the encoding rotation angles as follows: *starting from the input vector* x *of dimension* N*, a new vector* x_1 *of dimension* $\frac{N}{2}$ *is created such that*

$$x_1 = \left(\sqrt{\left|x[2k]^2 + x[2k+1]\right|^2}\right)_{k=0}^{\frac{N}{2}}. \tag{3}$$

As long as the dimension of the new vector x_1 *is greater than 1,* x *is assigned* x_1 *and the procedure is repeated. Then, for each new vector* x_1*, the rotation angles*

are calculated as follows:

$$\alpha = \begin{cases} 2\arcsin\left(\frac{x[2k+1]}{x_1[k]}\right), & \text{if } x_1[k] > 0 \\ 2\pi - 2\arcsin\left(\frac{x[2k+1]}{x_1[k]}\right), & \text{else.} \end{cases} \tag{4}$$

The second part of this algorithm generates the encoder circuit as presented in Fig. 4 for an input data of dimension 8.

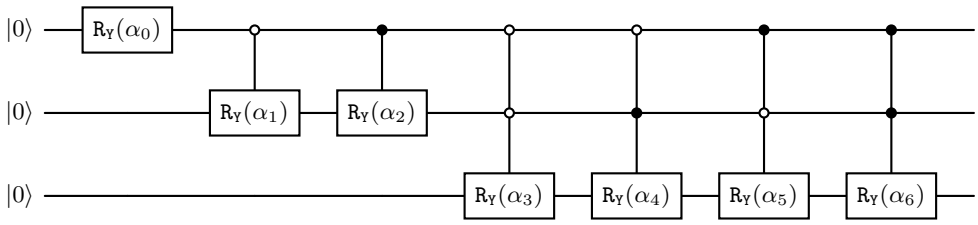


Fig. 4: *Architecture of our Amplitude encoding with* $N = 8$.

With n qubits, we can encode N data points, such that

$$\begin{cases} 2^{n-1} \leq N \leq 2^n, & \text{if } n = 1, \\ 2^{n-1} < N \leq 2^n, & \text{if } n > 1. \end{cases} \tag{5}$$

The cases $2^{n-1} < N < 2^n$ and $N = 2^n$ require the same number of qubits. However, the first case generate fewer rotation angles than the second. Hence, when $2^{n-1} < N < 2^n$, the remaining gates are set up with 0 as rotation angles.

1.3 Instantaneous Quantum Polynomial-time (IQP) encoding

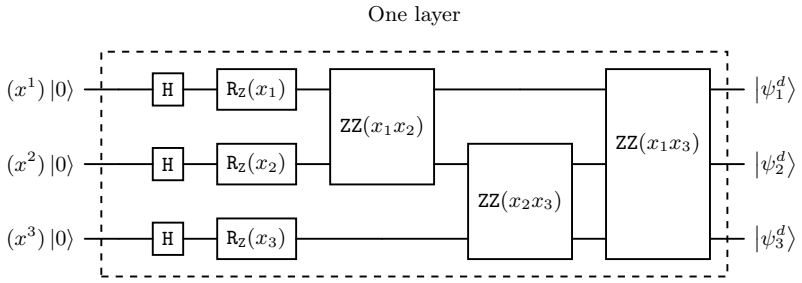


Fig. 5: *An IQP circuit on 3-qubits: the IQP embedding is composed of at least two layers of this IQP circuit.*

The instantaneous Quantum Polynomial-time (IQP) embedding [8] encodes the features into qubits using Hadamard gates $\mathtt{H}$, $\mathtt{R_Z}$ gates and the two-qubits $\mathtt{R_{ZZ}}$ for entanglement that encode higher-order data diagonal gates.

IQP circuit was originally introduced by [9] as a response to the difficulty of experimentally demonstrating a quantum advantage without having a complete universal computation. Indeed, in the standard circuit model, a universal set of quantum gates is used to generate the special unitary group $SU(2^n)$, which is non-abelian. Hence, the order in which the gates are applied is crucial and reflects the temporal complexity of the computation. IQP circuits are limited to Hamiltonians that commute with each other, thus generating an abelian subgroup of the unitary group. This commutative property means that all gates can be seen as applied simultaneously, hence the term *instantaneous*. Although mathematically much more limited than universal computation, these circuits generate probability distributions that are believed to be classically inapproximable. This makes it possible to design Alice and Bob type protocols, where Alice encodes a hidden structure in a problem and Bob must, using his IQP device, produce samples that respect a specific correlation that no classical computer can reproduce efficiently.

An IQP Embedding, as presented in [10] and [11], takes as its input the $|0\rangle^{\otimes n}$ state, and is composed of at least two layers of the IQP circuit configuration, as presented in Figure 5. The IQP embedding is capable of embedding N data points $\left(x_1, x_2, \cdots, x_j\right)$ into the state $|\psi\rangle = \mathtt{U_Z}\ \mathtt{H}^{\otimes n}\ \ \mathtt{U_Z}\ \mathtt{H}^{\otimes n}\ \ |0\rangle^{\otimes n}$ of $n = N$ qubits, with

$$\mathtt{U_Z} = \exp\left(\sum_{\substack{j=1\\k=1\\j\neq k}}^{n} -i\frac{x_j x_k}{2}\mathtt{Z}_j\mathtt{Z}_k + \sum_{j=1}^{n} -i\frac{x_j}{2}\mathtt{Z}_j\right).$$

2 Datasets and preprocessing

We used four benchmark binary datasets. The first is the popular dataset ***Wisconsin Diagnostic Breast Cancer (WDBC) dataset***, which comprises 569 total instances, including 357 benign cases and 212 malignant ones, and a total of 30 distinct attributes. For our simulations we used 455 samples for train and 114 for test (split ratio of 80 : 20). Next, two of the four datasets are binary datasets formed from the ***Modified National Institute of Standards and Technology (MNIST) dataset***, which consists of $60,000$ training images and $10,000$ testing images, black and white, of handwritten digits, normalized centered with 28 pixels on each side. The two datasets are **MNIST01** *(classes 0 and 1, 12665 training and 2115 test samples)* and **MNIST08** *(classes 0 and 8, 11774 training and 1954 test samples)*. The choice of these classes is motivated by the fact that while 0 and 1 have distinct characteristics that can make them easier to differentiate, the distinction between 0 and 8 presents a moderate challenge, which can be interesting for testing the robustness of the models without being too complex. The fourth dataset, called ***Malware dataset***, is a binary dataset consisting of malicious and benign Portable Executable (PE) files from datasets Bodmas [12] and PE Malware Machine Learning [13]. Bodmas

consists of 134,435 files. The 57,293 malicious files are fairly recent (collected between 2019 and 2020), and labeled by category (Trojan, Worm, Ransomware, $\cdots$). As for the PE Malware Machine Learning dataset, 114,737 files are malicious and come from sources such as VirusShare[3] and TheZoo[4]. The benign files (86,812) come from different versions of the Windows operating system (version $\geq$ Windows 7), for a total number of 200,549 files. We chose this dataset for our benchmarking because it corresponds to one of the key areas of our work: *cybersecurity*. Specifically, because of the fact that the dataset is big and not easy to classify, it is offers a valuable application for quantum classification tasks [14].

For the training, *4000 samples* are used for MNIST and Malware datasets. Regarding tests, we used *2000 samples* for Malware and MNIST01 datasets, while all *1954 test samples* of MNIST08 are used. We extracted from each dataset the 8 features with the highest impact and the lowest correlation with each other. The work involved in ranking the most impacting features of the Malware dataset is developed in *Benjamin Marais*'s thesis [15]. For WDBC dataset, we used the *Permutation feature importance* metric from *Scikit-learn*[5] to extract these features. Data from WDBC and Malware datasets were rescaled in $\left[0, \frac{\pi}{2}\right]$. The image data from MNIST was resized using the *Principal Component Analysis (PCA)* to suitable dimensions, and then also rescaled in $\left[0, \frac{\pi}{2}\right]$.

3 Results

For each of the 4 datasets, we carried out simulations using **Qiskit** simulators, on the 5 encodings for **N features** and **M layers**. First, we chose the number of features and layers as a power of 2. Then we looked at the behavior of Amplitude encoding in cases where N is not a power of 2. This is how we set $N \in \{4, 6, 8\}$, $M \in \{2, 4\}$. Test accuracies and F1-scores are presented in Tables 1, 2, 3 and 4 for all the four datasets. training has been done over 5 epochs, with a learning rate $\eta = 0.1$. In a longer version of this work available in pre-print [16], these train results are presented in the form of tables and figures, giving more precise training accuracies for each epoch.

On average, Amplitude encoding gives the best results on Malware and both MNIST datasets. On Malware dataset, it gives the best train and test accuracies for models where the number of features is a power of 2 (with F1-scores varying between 88% and 90%). On MNIST08, it performs best whatever the model, and on MNIST01, this better performance over the other models is only observed on 3 models: (4F, 4L), (6F, 4L) and (8F, 4F). Conversely, on WDBC dataset, Amplitude encoding gives the worst results on most models, with a max F1-score of 73.78%, and a min of 54.76%. As the WDBC dataset has fewer samples than Malware and MNIST datasets, these results could mean that Amplitude encoding is best suited to complex datasets with a large number of samples. Given that classification on MNIST08 is more difficult than on MNIST01, the

[3] VirusShare
[4] TheZoo
[5] Permutation feature importance

fact that Amplitude encoding is better than the others on MNIST08 would support this interpretation.

Table 1: Tests and F1-scores of Malware dataset.

	(4F, 4L)		**(4F, 2L)**		**(6F, 4L)**		**(6F, 2L)**		**(8F, 4L)**		**(8F, 2L)**	
Encoding	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S
Simple A.	0.87	0.88	0.815	0.830	0.847	0.856	0.811	0.828	0.848	0.857	0.819	0.835
$\frac{\pi}{4}$-Angle	0.882	0.872	0.80	0.821	0.9	**0.892**	0.833	0.830	0.888	**0.881**	0.830	0.828
Entangled A.	0.83	0.843	0.82	0.836	0.892	**0.891**	0.822	0.837	0.828	0.841	0.896	**0.891**
Amplitude	0.894	**0.891**	0.884	**0.884**	0.883	0.876	0.841	**0.851**	0.901	**0.9**	0.885	**0.886**
IQP	0.898	**0.891**	0.825	0.84	0.893	0.883	0.83	0.838	0.869	0.864	0.827	0.84

Table 2: Tests and F1-scores of WDBC dataset.

	(4F, 4L)		**(4F, 2L)**		**(6F, 4L)**		**(6F, 2L)**		**(8F, 4L)**		**(8F, 2L)**	
Encoding	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S
Simple A.	0.949	**0.923**	0.912	0.888	0.938	**0.913**	0.833	0.716	0.85	0.753	0.859	0.804
$\frac{\pi}{4}$-Angle	0.947	**0.92**	0.912	0.88	0.903	0.857	0.894	**0.846**	0.868	0.8	0.798	0.646
Entangled A.	0.938	0.909	0.937	**0.907**	0.938	**0.915**	0.868	0.788	0.929	**0.894**	0.929	**0.89**
Amplitude	0.614	0.62	0.684	0.632	0.666	0.547	0.701	0.653	0.763	0.737	0.719	0.673
IQP	0.938	0.917	0.833	0.707	0.912	0.868	0.877	0.815	0.859	0.771	0.807	0.656

Table 3: Tests and F1-scores of MNIST01 dataset.

	(4F, 4L)		**(4F, 2L)**		**(6F, 4L)**		**(6F, 2L)**		**(8F, 4L)**		**(8F, 2L)**	
Encoding	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S
Simple A.	0.971	0.972	0.785	0.824	0.987	0.987	0.616	0.736	0.979	0.98	0.802	0.844
$\frac{\pi}{4}$-Angle	0.97	0.973	0.984	**0.985**	0.991	**0.992**	0.989	**0.99**	0.972	0.974	0.99	**0.99**
Entangled A.	0.994	**0.994**	0.884	0.902	0.967	0.97	0.865	0.885	0.986	0.987	0.971	0.973
Amplitude	0.997	**0.997**	0.982	**0.983**	0.995	**0.995**	0.977	**0.979**	0.991	**0.992**	0.946	0.952
IQP	0.994	**0.994**	0.873	0.893	0.935	0.943	0.781	0.83	0.972	0.974	0.922	0.928

Unlike Amplitude encoding, Angle encodings are not as good on MNIST and Malware datasets. The lowest F1-score (82.11%) on Malware dataset comes from the $\frac{\pi}{4}$-Angle encoding, obtained with 4 features and 2 layers. Further on, we note that this encoding gives the lowest precision on all 2-layers models. On **MNIST08**, this encoding never gives the best results, whatever the model. Simple Angle encoding returns, on average, the lowest accuracies on the MNIST01 dataset. Case (6F, 2L) is particularly poor, with train accuracies ranging from $61\% - 67\%$, and a F1-score of 73.65%. These results are the worst of all MNIST datasets. Note that this encoding has the same behavior on both

Table 4: Tests and F1-scores MNIST08 dataset.

	(4F, 4L)		**(4F, 2L)**		**(6F, 4L)**		**(6F, 2L)**		**(8F, 4L)**		**(8F, 2L)**	
Encoding	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S	Test	F1-S
Simple A.	0.95	0.949	0.902	0.896	0.949	0.949	0.94	0.939	0.94	0.941	0.938	0.938
$\frac{\pi}{4}$-Angle	0.932	0.934	0.93	0.933	0.925	0.925	0.948	0.948	0.921	0.921	0.927	0.929
Entangled A.	0.925	0.923	0.919	0.917	0.967	**0.967**	0.852	0.868	0.958	0.957	0.905	0.909
Amplitude	0.935	0.931	0.941	**0.948**	0.957	**0.956**	0.94	0.936	0.964	0.963	0.943	0.941
IQP	0.942	**0.94**1	0.944	0.942	0.944	0.945	0.802	0.753	0.936	0.938	0.877	0.877

MNIST datasets. However, with this encoding we note that, on all the datasets, the training reaches a maximum value from the first epochs. So, the Simple Angle encoding do not need more epoch for the training. This last observation is also made with Entangled encoding on MNIST datasets. On average, of the three angle encodings, $\frac{\pi}{4}$-Angle and Entangled encoding returns the best precisions.

On WDBC, the best results are obtained with Angle encodings, where Simple Angle encoding returns the highest train and test accuracies ($> 94\%$), and the highest F1-score ($> 92\%$) with 4 features and 4 layers. One observed behaviour on the Malware dataset, that recurs with WDBC, is the fact that the best accuracies with Simple Angle encoding are achieved with just 4 features. Comparing Simple Angle and $\frac{\pi}{4}$-Angle encodings, it seems that, of the two angle encodings, it is better to use the Simple Angle in cases with very few data, such as the WDBC dataset. The $\frac{\pi}{4}$-Angle encoding, on the other hand, would be a better choice than the simple one for large amounts of data (Malware dataset). We remark that Entangled Angle encoding seems to behave on WDBC dataset the same way as Amplitude encoding behaves on MNIST and Malware datasets. That is, on several models, it is one of the best performers, unlike the two others Angle encodings, which perform very well on some models and less well on others.

On WDBC, IQP encoding seems to have the same behavior as on the Malware dataset. An important observation with IQP encoding is the fact that, with 8 features, the imbalance in the WDBC dataset (number of benign samples greater than the number of malignant samples) seems to be reflected, whatever the number of layers. Indeed, we have an F1-score of 77.14% with a train accuracy of up to 87% and a test accuracy of 85.96% on 4 layers. On 2 layers, the F1-score is lower (65.62%) with learning and test accuracies above 80%. Our intuition is that, with very few data and layers, the model trained with IQP encoding concentrates on learning the excess class in the dataset (benign class), neglecting the other class.

4 Conclusion

The aim of our work is to contribute to the optimization of QML algorithms in the current era of NISQ, by benchmarking 5 types of data encodings. Starting from a Quantum Neural Networks (QNN) model (Figure 1), we benchmarked

these encodings on 4 datasets. The results suggest that encodings such as $\frac{\pi}{4}$-Angle and IQP need few features to return good accuracies. To improve these accuracies, increasing the number of layers is a more effective solution than increasing the number of features. Simple Angle encoding does not require several epochs for training. However, of the three angle encodings, it is the one that performs the worst on the average of the models. On the other hand, the Amplitude encoding was better on several models on the Malware and MNIST datasets, while on WDBC, it returned the lowest performance. These performances tend to decrease as the number of features decreases.

In summary, for datasets with features ranked in order of importance, if the size of the dataset is large enough (thousands of data points), we can build an efficient model based on Amplitude encoding with very few features, and therefore very few qubits. However, to ensure good accuracy, it would be preferable to choose a number of features as a power of 2. In a case where the dataset does not have enough data (as in the case of the WDBC dataset), Angle encodings will be more suitable for a high-performance model, also using fewer qubits and few epochs. Entangled Angle and $\frac{\pi}{4}$-Angle encodings returned the best results compared with the Simple Angle encoding, while the latter requires little epoch to achieve its best accuracies. However, in the three cases, it is important to choose a large enough number of layers (at least 4 layers) to have good performances.

The primary goal of this work was the presentation of the results and their analysis. A possible direction for future work would be to go beyond this analysis by providing an interpretation related to the nature of each encoding. Moreover, the simulations were carried out only once, on simulators. It would also be interesting to repeat the same tests multiple times on the same simulators, and to consider performing them on real noisy quantum computers.

References

[1]J. Preskill, "Quantum Computing in the NISQ era and beyond", Quantum **2**, 79 (2018), http://dx.doi.org/10.22331/q-2018-08-06-79.

[2]M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models", Physical Review A **103**, 10.1103/physreva.103.032430 (2021), http://dx.doi.org/10.1103/PhysRevA.103.032430.

[3]D. Ranga, A. Rana, S. Prajapat, P. Kumar, K. Kumar, and A. V. Vasilakos, "Quantum Machine Learning: Exploring the Role of Data Encoding Techniques, Challenges, and Future Directions", Mathematics **12**, 10.3390/math12213318 (2024), https://www.mdpi.com/2227-7390/12/21/3318.

[4]M. Schuld and N. Killoran, "Quantum Machine Learning in Feature Hilbert Spaces", Phys. Rev. Lett. **122**, 040504 (2019), https://link.aps.org/doi/10.1103/PhysRevLett.122.040504.

[5]A. Chalumuri, R. Kune, and B. S. Manoj, "A hybrid classical-quantum approach for multi-class classification", Quantum Information Processing **20**, 119, 119 (2021).

[6]G. D. Luca, A. Vlasic, M. Vitz, and A. Pham, *Empirical Power of Quantum Encoding Methods for Binary Classification*, 2024, https://arxiv.org/abs/2408.13109.
[7]I. Araujo, D. Park, F. Petruccione, and A. da Silva, "A divide-and-conquer algorithm for quantum state preparation", 10.48550/arXiv.2008.01511 (2020).
[8]V. Havlicek, A. D. Corcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces", Nature **567**, 209–212 (2019), http://dx.doi.org/10.1038/s41586-019-0980-2.
[9]D. Shepherd and M. J. Bremner, "Temporally unstructured quantum computation", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **465**, 1413–1439 (2009), http://dx.doi.org/10.1098/rspa.2008.0443.
[10]T. Koike-Akino, P. Wang, and Y. Wang, *Autoqml: Automated Quantum Machine Learning for Wi-Fi Integrated Sensing and Communications*, 2022, https://arxiv.org/abs/2205.09115.
[11]Y.-Y. Hong, C. Arce, and T.-W. Huang, "A Robust Hybrid Classical and Quantum Model for Short-Term Wind Speed Forecasting", IEEE Access **PP**, 1–1 (2023).
[12]L. Yang, A. Ciptadi, I. Laziuk, A. Ahmadzadeh, and G. Wang, "BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware", in 2021 ieee security and privacy workshops (spw) (2021), pp. 78–84.
[13]M. Lester, *PE Malware Machine Learning Dataset*, https://practicalsecurityanalytics.com/pe-malware-machine-learning-dataset/.
[14]P. Bermejo, P. Braccia, M. S. Rudolph, Z. Holmes, L. Cincio, and M. Cerezo, *Quantum convolutional neural networks are (effectively) classically simulable*, 2024, https://arxiv.org/abs/2408.12739.
[15]B. Marais, "Améliorations des outils de détection de malwares par analyse statique grâce à des mécanismes d'intelligence artificielle", Theses (Normandie Université, Dec. 2023), https://theses.hal.science/tel-04416984.
[16]O. Zang, G. Barrué, and T. Quertier, *Benchmarking data encoding methods in quantum machine learning*, 2025, https://arxiv.org/abs/2505.14295.

Quantum Emulator Benchmarking Made Easy

Anna Leonteva[0000−0001−9521−0630], Maxime Outteryck, and Guido Masella[0000−0001−7108−2996]

QPerfect, 23 Rue du Loess, 67000 Strasbourg, France

Abstract. Classical quantum emulators are key tools for algorithm design, error correction, and noise mitigation. They rely on a variety of simulation methods such as statevector, tensor networks, matrix product states, decision diagrams and others. Their performance depends not only on simulation method and circuit properties, but also on emulator-specific parameters, pre-processing and implementation. Systematic benchmarking of emulators is essential for advancing simulation techniques and selecting the most suitable emulator for a given task. To address this need, we present an open-source, user-centric, problem-agnostic framework with a unified API for twelve leading emulators, delivering detailed performance insights for quantum computing research and development.

Keywords: Benchmarking · Quantum Emulator · Optimization.

1 Introduction

Classical simulation is a core technology for quantum computing research and development, driving progress on both hardware and software innovation. On the hardware side, emulators allow researchers to pinpoint performance bottlenecks, optimize quantum architectures, and validate error-correction protocols before physical implementation [1–3]. In software, emulators provide an environment for the design, testing, and refinement of quantum algorithms [4, 5].

Since classical emulation methods are continuously evolving, they demand systematic evaluation not as a mere leaderboard competition, but as a scientific methodology that bridges quantum computing and its potential applications by revealing each emulator's strengths and limitations. Moreover, robust benchmarking can systematically map which classes of quantum circuits today's emulators can tractably simulate and which remain beyond their reach, providing a clear roadmap for future research in quantum algorithms [6, 7].

However, benchmarking quantum emulators is a challenging task, because their performance depends on many factors. A key factor is the core simulation method. For instance, statevector methods typically faces exponential demands in computational time and memory with respect to the number of simulated qubits, restricting their effective use to systems of a few tens of qubits [7]. Whereas, Matrix Product States (MPS) methods can achieve polynomial scaling in time and memory, allowing for simulating hundreds or even thousands of qubits for circuits with limited entanglement. Secondly, emulator-specific parameters (e.g., bond dimension and accuracy cutoffs in MPS) can strongly influence

performance, depending on the circuit [6]. Furthermore, the programming language used to implement an emulator can significantly affect its performance. For instance, Python implementations are typically slower than those in C++, while Julia often achieves competitive run times thanks to just-in-time (JIT) compilation and efficient memory management. Finally, circuits themselves can affect results: e.g., the same emulator may scale poorly on one circuit, but achieve up to thousands of qubits on another [6, 7]. All of these factors introduce biases and lead to unreliable benchmarking results [8]. This implies that quantum emulator performance should not be characterized as a singular "empirical" property, but rather as a set of metrics (including run time, fidelity, and scalability) that are dependent on all mentioned factors.

Several open-source projects have addressed quantum emulator benchmarking from different angles. `QUARK` [9] is a modular framework automating application-level benchmarking, from problem definition to solver execution. However, its tight coupling to domain-specific encodings and limited backend support restricts its adaptability for broader, scalable comparisons. Jamadagni et al. [7] presented a toolchain for statevector simulators, profiling single- vs double-precision and single- vs multi-threaded regimes across leading libraries. `BenchPress` [10] measures circuit construction and transpilation performance across seven SDKs, but omits simulator execution metrics. Bowles et al. [11] introduced a PennyLane-based benchmark, reporting only machine learning quality metrics. While each tool excels in its niche — application workflows, simulator profiling, compiler performance, or quantum machine learning evaluation — none, to our knowledge, offers unified access to multiple quantum emulators with execution benchmarking across diverse circuits. Our framework addresses this gap.

The goal of this paper is to present an open-source framework, FENIQS[1], for automated benchmarking of quantum emulators across simulation methods, circuit types, and user-defined scenarios. It aims to streamline the benchmarking process, reducing time and manual effort involved.

2 Framework

FENIQS is an open-source `Python` framework designed to streamline the quantum emulator benchmarking process. Its architecture is user-centric, problem-agnostic, and language-independent, featuring a modular and extensible design, where all supported emulators are implemented as plugins. FENIQS benchmarking process is shown in Figure 1 and contains several principal blocks.

Workflow: Users initiate this workflow by supplying quantum circuits, in OpenQASM 2.0 format, to a designated input directory, which is monitored by a FENIQS daemon (Linux service). Each new circuit is processed by systematically invoking the configured emulator plugins via the `feniqs_lib` core library. Throughout the execution of each simulation, the framework rigorously enforces user-defined timeouts. Upon completion of a simulation run, all pertinent performance metrics, such as run time and fidelity (discussed in Section 3), along with

[1] `https://github.com/qperfect-io/feniqs_lite`

associated metadata, are captured and stored. FENIQS provides flexibility in data archiving, offering options for storage in a PostgreSQL database, facilitating robust and queryable long-term storage, or as individual JSON files. These JSON files can then feed into the framework's post-processing tools to generate a variety of insightful visualizations, including comparative plots of run time, fidelity, and scalability across different emulators and configurations.

*Core Library:*FENIQS currently supports twelve actively maintained quantum emulator frameworks, covering a wide range of simulation methods and hardware targets (summarized in Table 1). Internally, the core library `feniqs_lib` uses plugin and environment managers to load each emulator in an isolated virtual environment, avoiding conflicts and preserving reproducibility. All plugins conform to a unified interface for circuit execution, enabling fair comparisons across different emulators. This modular architecture also allows users to easily switch between simulation methods, versions or hardware platforms.

Configuration and Preprocessing: Before benchmarking, users can modify three distinct configuration files, in YAML format, to define FENIQS setup: `venv_deps.yaml` defines the environment, selecting emulator plugins and ensuring backend isolation; `config_backend.yaml` specifies emulator types and parameters; `optimizator.yaml` (optional) sets parameter ranges for performance tuning. To address the strong impact of emulator-specific parameters (such as bond dimension for MPS, accuracy thresholds, or circuit pre-optimizations) on both run time and fidelity, FENIQS includes a noise-aware hyperparameter optimizer. It supports CMA-ES for single-objective optimization, minimizing run time while satisfying a fixed fidelity constraint, ideal when high accuracy is required. For broader trade-off exploration, NSGA-II and MOEA-D enable multi-objective optimization, generating a Pareto front that balances fidelity and run time. All candidate configurations are evaluated under run time noise, with optimized solution exportable to the backend YAML file for direct use in benchmarking.

Post-processing: FENIQS includes a set of Python scripts for parsing JSON result files and generating comparative plots. These tools facilitate the analysis of benchmark metrics such as run time, fidelity, and scalability across different emulators.

3 Metrics

FENIQS evaluates performance based on two direct core metrics: *run time* and *mirror circuit fidelity*

The *run time* metric in FENIQS refers to the total wall-clock duration required to execute a quantum circuit. This includes all necessary operational steps, including circuit parsing from OpenQASM 2.0 format, quantum state initialization, circuit execution, and sampling of measurement outcomes. To account for system warm-up effects and inherent shot-to-shot variability in execution times, each benchmark circuit is executed multiple times, as specified by the users, and FENIQS reports the minimum run time (in seconds) achieved.

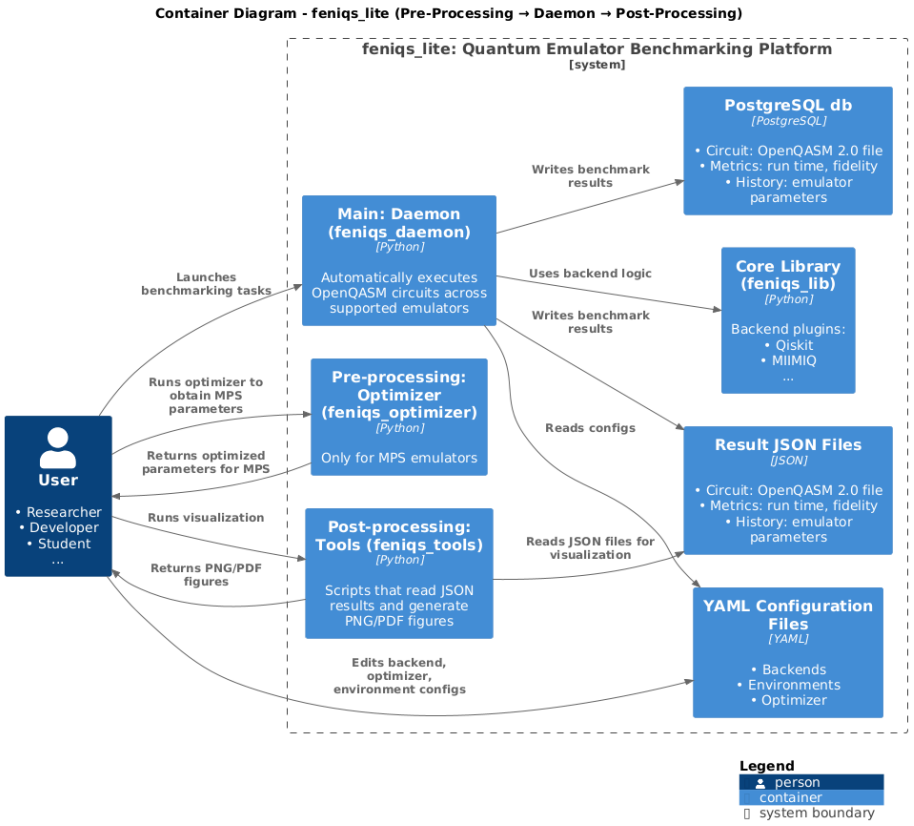


Fig. 1. Architectural diagram of the FENIQS benchmarking platform, illustrating key components, data stores, and user interactions. The large dashed rectangle delineates the system boundary of FENIQS. Boxes within this bondary represent primary software containers and data stores, while boxes positioned externally indicates persons interacting with the system. Arrows indicate the direction of control signals or data flows.

For assessing the accuracy of emulators, particularly those not based on statevector methods where direct fidelity calculation with a reference state can be intractable, FENIQS implements the *mirror-circuit fidelity*. This method evaluates an emulator's accuracy without relying on its internal fidelity estimates. It involves executing a circuit followed immediately by its inverse (the mirror circuit), and then measuring the probability of the system returning to its initial computational basis state. To prevent emulators from trivially optimizing the combined circuit to an identity operation, FENIQS supports the insertion of barrier operations after the forward circuit (where the emulator supports them).

Beyond presented primary metrics, FENIQS introduces indirect metric – *scalability*, which quantifies an emulator's ability to efficiently handle increas-

Table 1. Supported frameworks within FENIQS, detailing their primary simulation methods, API languages, open-source status, and CPU/GPU hardware targets.

Framework	Method	API Language	Open Source	CPU	GPU
Cirq [12]	StateVector	Python	✓	✓	✓
Qiskit [13]	StateVector	Python	✓	✓	✓
	MPS	Python	✓	✓	✓
MIMIQ	StateVector	Julia		✓	
	MPS	Julia		✓	
Pennylane [14]	StateVector	Python	✓	✓	✓
Qulacs [15]	StateVector	Python	✓	✓	✓
Yao [16]	StateVector	Julia	✓	✓	✓
ProjectQ [17]	StateVector	Python	✓	✓	
QMatchTea [18]	MPS	Python	✓	✓	✓
MQT-DDS [19]	Decision Diagrams	Python	✓	✓	
Pyqrack [20]	Factorized Ket	Python	✓	✓	
Qrisp [21]	Sparse Matrix	Python	✓	✓	
Quimb [22]	MPS	Python	✓	✓	
	NT	Python	✓	✓	

ing problem size without loss of accuracy or excessive run time. This metric is assessed by analyzing trends in the primary metrics – run time and fidelity – collected by FENIQS across multiple circuit instances of varying qubit numbers for a given algorithm or circuit class. Typically, this involves evaluating how the measured run time $T_{\mathcal{E}}(n)$ for the emulator $\mathcal{E}$ depends on the number of qubits n, subject to maintaining the mirror-circuit fidelity $\mathcal{F}_{\mathcal{E}} \geq 0.99$, and respecting an execution timeout of 300 seconds. Emulators exhibiting polynomial scaling of run time with n are generally considered more scalable for that class of circuits.

4 Results and Main Features

FENIQS robustness and versatility were demonstrated in a large-scale benchmarking work presented in [6], which assessed seven cutting-edge quantum emulators across diverse simulation techniques. This study included four MPS-based tools (Qiskit-MPS, Quimb-MPS, QMatchaTea-MPS, MIMIQ-MPS), as well as Quimb-TN, Pyqrack, and MQT-DDS. The benchmark suite featured 13 scalable circuits from the MQTBench library[2] , ranging from 4 to 1,024 qubits on CPU hardware. To present these results, Figure 2 reproduces a radar chart from [6], showing the maximum qubit counts achieved by each emulator across the 13 benchmark circuits. The radial axis uses a logarithmic scale, from 12 (innermost circle) to 1,024 qubits (outermost), providing a clear visual summary of scalability. Larger polygon areas indicate emulators capable of solving a greater fraction of circuits at higher qubit counts, highlighting their performance under utility-scale conditions.

Beyond its demonstrated use case, FENIQS offers several key features that make it a powerful tool for diverse benchmarking needs.

[2] `https://github.com/cda-tum/MQTBench`

Flexibility of the Benchmarking Process: FENIQS enables highly customizable benchmarking by allowing users to group emulators based on simulation method (e.g., statevector, tensor network), hardware platform (CPU/GPU), or criteria such as numerical precision or run time limits. YAML-based configurations let users compare backends, versions, and execution modes without writing code. For example, the left panel of Figure 3 shows run time scaling of the "Amplitude Estimation" circuit across seven emulators (mentioned above) with increasing qubit counts on CPU; the right panel compares GPU-implemented statevector simulators, Qiskit and PennyLane, executing the `hdyn` task (simulating XYZ-Heisenberg dynamics).

Extensibility: FENIQS supports the addition of new emulators and optimization strategies with minimal core changes. Benchmark suites can be extended by adding OpenQASM 2.0 circuits, and JSON-formatted results integrate easily with external tools for advanced post-processing and visualization, without altering existing workflows.

Efficiency and Accessibility: FENIQS streamlines benchmarking by automating setup, execution, and monitoring, reducing workflow time from hours to minutes. YAML configurations replace complex coding, making it accessible to users with less extensive programming and quantum emulators expertise.

Reproducibility and Clarity: YAML-based configurations and structured JSON results ensure reproducibility and simplify comparison across emulators.

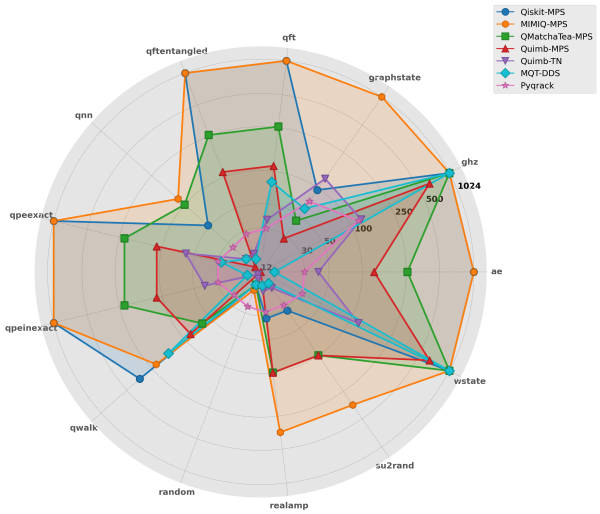


Fig. 2. FENIQS output visualizing comparative emulator scalability [6]. The radar chart displays the maximum achievable qubit counts for seven emulators across thirteen benchmark circuits. Each radial exis represents a specific benchmark algorithm, with concentric circles indicating the qubit scale from 12 to 1,024 (logarithmic). Larger polygon areas generally indicate superior overall scalability across the evaluated benchmark suite.

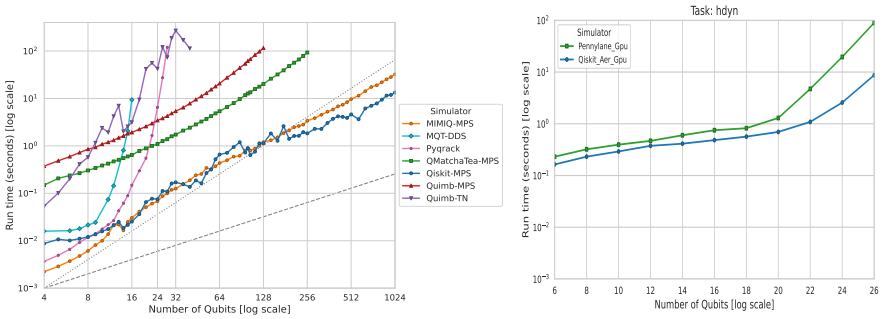


Fig. 3. Run time plots generated by FENIQS, comparing emulator performance as a function of qubit count for two distinct quantum tasks and emulator sets. Both panels present run time in seconds versus the number of qubits. Left panel illustrates run time (in log scales) scaling for the amplitude-estimation benchmark circuit [6] for seven different emulators, as detailed in the legend. Dashed and dotted lines provide visual references for linear and quadratic scaling behaviors, respectively. Right panel shows a performance comparison of two statevector GPU-based simulators (PennyLane and Qiskit) on the `hydn` benchmarks.

5 Conclusions

We introduced FENIQS, a user-centric, problem-agnostic, and plugin-based framework for benchmarking quantum emulators. FENIQS streamlines comparative emulator studies through a unified API supporting twelve distinct backends, a modular plugin system, automated hyperparameter tuning, and robust performance metrics. It automates key benchmarking steps, saving users time and reducing the need for specialized expertise, by supporting the full benchmarking life-cycle, from scenario setup to result visualization, and enabling custom experiments. The practical utility of FENIQS in generating reliable performance insights, such as detailed emulator scalability profiles and assessments of classical simulation boundaries — as demonstrated in a comprehensive utility-scale benchmarking study [6]. FENIQS thus serves as a valuable tool for the quantum computing community, aiding in emulator selection and fostering the advancement of simulation techniques. Future work includes extending support to quantum hardware to address execution variability and cost, and integrating machine-learning-based accelerators for parameter tuning. The full source code is openly available on GitHub[3].

[3] `https://github.com/qperfect-io/feniqs_lite`

Bibliography

[1] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[2] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505, 2023.

[3] Andrew W Cross, Lev S Bishop, Sarah Sheldon, Paul D Nation, and Jay M Gambetta. Validating quantum computers using randomized model circuits. *Physical Review A*, 100(3):032328, 2019.

[4] Alexis Morvan, B Villalonga, X Mi, S Mandra, A Bengtsson, PV Klimov, Z Chen, S Hong, C Erickson, IK Drozdov, et al. Phase transitions in random circuit sampling. *Nature*, 634(8033):328–333, 2024.

[5] Andrew D. King, Alberto Nocera, Marek M. Rams, Jacek Dziarmaga, Roeland Wiersema, et al. Beyond-classical computation in quantum simulation. *Science*, 388(6743):199–204, 2025.

[6] Anna Leonteva, Guido Masella, Maxime Outteryck, Asier Piñeiro Orioli, and Shannon Whitlock. Comparative benchmarking of utility-scale quantum emulators, 2025. URL `https://arxiv.org/abs/2504.14027`.

[7] Amit Jamadagni, Andreas M Läuchli, and Cornelius Hempel. Benchmarking quantum computer simulation software packages: state vector simulators. *preprint arXiv:2401.09076*, 2024.

[8] Thomas Lubinski, Joshua J Goings, Karl Mayer, Sonika Johri, Nithin Reddy, Aman Mehta, Niranjan Bhatia, Sonny Rappaport, Daniel Mills, Charles H Baldwin, et al. Quantum algorithm exploration using application-oriented performance benchmarks. *preprint arXiv:2402.08985*, 2024.

[9] Jernej Rudi Finžgar, Philipp Ross, Leonhard Hölscher, Johannes Klepsch, and Andre Luckow. Quark: A framework for quantum computing application benchmarking. In *2022 IEEE international conference on quantum computing and engineering (QCE)*, pages 226–237. IEEE, 2022.

[10] Paul D Nation, Abdullah Ash Saki, Sebastian Brandhofer, Luciano Bello, Shelly Garion, Matthew Treinish, and Ali Javadi-Abhari. Benchmarking the performance of quantum computing software for quantum circuit creation, manipulation and compilation. *Nature Computational Science*, pages 1–9, 2025.

[11] Joseph Bowles, Shahnawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models. *arXiv preprint arXiv:2403.07059*, 2024.

[12] Cirq Developers. Cirq (v1.5.0), 2025. URL `https://doi.org/10.5281/zenodo.15191735`.

[13] Matthew Treinish, Luciano Bello, Jake Lishman, Jay Gambetta, Diego M. Rodríguez, Manoel Marques, Julien Gacon, Paul Nation, Christopher J. Wood, ewinston, Juan Gomez, Richard Chen, Andrew Cross, Kevin Krsulich, Steve Wood, Ismael Faro Sertage, Naoki Kanazawa, Lauren Capelluto, Thomas Alexander, and qiskit bot. Qiskit/qiskit: Qiskit 2.0.0 (2.0.0), 2025. URL `https://doi.org/10.5281/zenodo.15116124`.

[14] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M Sohaib Alam, Guillermo Alonso-Linaje, B Akash-Narayanan, Ali Asadi, et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.

[15] Yasunari Suzuki, Yoshiaki Kawase, Yuya Masumura, Yuria Hiraga, Masahiro Nakadai, Jiabao Chen, Ken M Nakanishi, Kosuke Mitarai, Ryosuke Imai, Shiro Tamiya, et al. Qulacs: a fast and versatile quantum circuit simulator for research purpose. *Quantum*, 5:559, 2021.

[16] Xiu-Zhe Luo, Jin-Guo Liu, Pan Zhang, and Lei Wang. Yao. jl: Extensible, efficient framework for quantum algorithm design. *Quantum*, 4:341, 2020.

[17] Damian S Steiger, Thomas Häner, and Matthias Troyer. Projectq: an open source software framework for quantum computing. *Quantum*, 2:49, 2018.

[18] Simone Montangero. *Introduction to Tensor Network Methods*. Springer, 2018. `https://doi.org/10.1007/978-3-030-01409-4`.

[19] Robert Wille, Stefan Hillmich, and Lukas Burgholzer. Decision diagrams for quantum computing. In *Design automation of quantum computers*, pages 1–23. Springer, 2022.

[20] Daniel Strano, Benn Bollay, Aryan Blaauw, Nathan Shammah, William J. Zeng, and Andrea Mari. Exact and approximate simulation of large quantum circuits on a single GPU. *preprint arXiv:2304.14969*, 2023.

[21] Raphael Seidel, Sebastian Bock, René Zander, Matic Petrič, Niklas Steinmann, Nikolay Tcholtchev, and Manfred Hauswirth. Qrisp: A framework for compilable high-level programming of gate-based quantum computers, 2024. URL `https://arxiv.org/abs/2406.14792`.

[22] Johnnie Gray. quimb: a python library for quantum information and many-body calculations. *Journal of Open Source Software*, 3(29):819, 2018. `https://doi.org/10.21105/joss.00819`.

Disclosure of Interests. Guido Masella is a shareholder of QPerfect.

Evaluating Quantum Wire Cutting for QAOA: Performance Benchmarks in Ideal and Noisy Environments

R. M. J. (Michel) Meulen[1,2], Niels M. P. Neumann[1], and Jasper Verbree[1]

[1] Department of Applied Cryptograhpy & Quantum Algorithms, The Netherlands
Organisation for Applied Scientific Research, The Netherlands
{*first_name.last_name*}@tno.nl
[2] Fontys University of Applied Sciences, Eindhoven, The Netherlands

Abstract. Current quantum computers suffer from a limited number of qubits and high error rates, limiting practical applicability. Different techniques exist to mitigate these effects and run larger algorithms. In this work, we analyze one of these techniques called quantum circuit cutting. With circuit cutting, a quantum circuit is decomposed into smaller sub-circuits, each of which can be run on smaller quantum hardware. We compare the performance of quantum circuit cutting with different cutting strategies, and then apply circuit cutting to a QAOA algorithm. Using simulations, we first show that Randomized Clifford measurements outperform both Pauli and random unitary measurements. Second, we show that circuit cutting has trouble providing correct answers in noisy settings, especially as the number of circuits increases.

Keywords: Quantum Computing · Circuit Cutting · Wire Cutting · QAOA · MaxCut.

1 Introduction

Quantum computers have the potential to solve problems infeasible for any classical supercomputer [15,19]. Investments in quantum computers and the corresponding algorithms are continuously rising. However, current quantum computers, often called Noisy Intermediate-Scale Quantum (NISQ) devices, are still far away from their full potential. They have a limited number of qubits and high error rates [18]. Improving usability requires increasing the number of qubits and decreasing the error rates. Recently, techniques like error mitigation have been proposed to mitigate, rather than solve, these limitations [12,23].

In 2020, Peng et al. showed how to use an n-qubit quantum computer to run an $n + k$-qubit quantum circuit [17], introducing the subject of *circuit cutting*. Their method decomposes a quantum circuit in different blocks and combines the outcomes of the individual blocks to approximate expectation values produced by the original circuit. Follow-up work improved their method and provided explicit decompositions for certain quantum gates [1,14,16]. Typically, we distinguish

between two different circuit cutting techniques: wire cutting, where a qubit is measured in a certain basis and initialized again, and gate cutting, where a multi-qubit quantum gate is replaced by single-qubit gates and measurements with post-selection [1]. This work focuses on wire cutting.

First, we compare three variants of wire cutting, using exact and noisy simulations, and determine which of these variants performs best. Second, we apply the best wire cutting variant found to the quantum approximate optimization algorithm (QAOA) [5] applied to the MaxCut problem [25]. In this way, we learn the potential of quantum circuit cutting and its overhead for a practical problem showing how the exact theory translates to the noisy reality.

The remainder of this paper is organized as follows: Section 2 discusses in more detail wire cutting and the variants considered in this work. Section 3 presents the experimental setup: a benchmark comparison between Pauli-based and randomized measurement-based wire cutting; and the application of the randomized measurement-based wire cutting technique to QAOA in both ideal and noisy simulation environments. Section 4 gives the results for these experiments, and Section 5 discusses the results and concludes the work.

2 Wire cutting

Tang et al. introduced a new method for wire cutting based on decomposing quantum operations into linear combinations of Pauli matrices [22]. By measuring qubits in the X, Y, and Z bases and reinitializing them in the $|0\rangle$, $|1\rangle$, $|+\rangle$, or $|i\rangle$ state, a larger circuit can be approximated by smaller ones.

Once all measurement outcomes are obtained, the probability of a basis state being produced by the uncut circuit can be reconstructed. For instance, consider the state $|010\rangle$ where the second wire is cut. We define two auxiliary states $u_1 = |00\rangle$ and $u_2 = |01\rangle$ for the first and second qubit, and an output state $d = |10\rangle$ for the second and third qubit. The probability of measuring $|010\rangle$ is:

$$\Pr[|010\rangle] = \frac{1}{2}\sum_{i=1}^{4} U_i D_i, \tag{1}$$

where

$$U_1 = 2u_{1,Z}, \quad U_2 = 2u_{2,Z}, \quad U_3 = u_{1,X} - u_{2,X}, \quad U_4 = u_{1,Y} - u_{2,Y},$$
$$D_1 = d_{|0\rangle}, \quad D_2 = d_{|1\rangle}, \quad D_3 = 2d_{|+\rangle} - d_{|0\rangle} - d_{|1\rangle}, \quad D_4 = 2d_{|i\rangle} - d_{|0\rangle} - d_{|1\rangle},$$

and $u_{i,P}$ corresponds to the probability of measuring state u_i in Pauli basis P and $d_{|\phi\rangle}$ corresponds to the probability of measuring state d on input state $|\phi\rangle$. A schematic presentation of wire cutting is given in Figure 1. Note that cutting the middle qubit wire of a three-qubit circuit results in subcircuits of two qubits.

Lowe et al. introduced an alternative approach using two distinct quantum channels to reduce the sampling overhead of exact methods [14]. Their method is based on already existing methods to classically simulate the entanglement

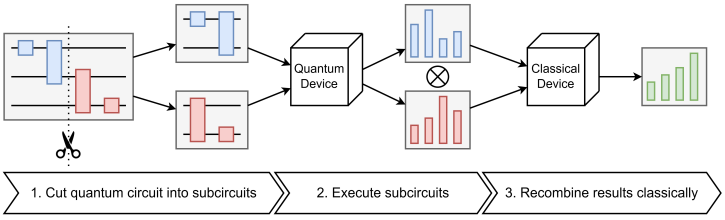


Fig. 1: Visual representation of wire cutting using Pauli measurements. Each subcircuit is run on a quantum device and the results are classically recombined. The tensor product symbol ($\otimes$) refers to the recombination of the different probability distributions as in Equation (1).

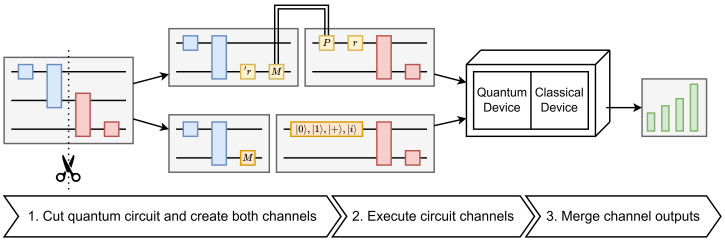


Fig. 2: Visual representation of randomized-measurement-based wire cutting. For every sample, the method randomly chooses between two quantum channels: The random Clifford channel (top) and the depolarization channel (bottom). The chosen channel is then executed on a hybrid-quantum-classical-device. The recombination of the probabilities follows Equation (2).

between matrix-product states [3,9]. One channel applies a random Clifford gate before the measurement and applies the conjugate of the Clifford gate as qubit reinitialization. The other channel acts as a depolarizing channel, where the qubit is measured and then reinitialized in a random basis state. Their method randomly selects between these two channels with probabilities $\frac{d+1}{2d+1}$ and $\frac{d}{2d+1}$, where $d = 2^k$ for k cut qubits. The probability corresponding to a state b is then given by

$$\Pr(b) = (d+1) \cdot \Pr_{Clifford}(b) - d \cdot \Pr_{depolarizing}(b), \tag{2}$$

where $\Pr_X(b)$ is the probability found for state b for channel X. Figure 2 gives an abstract overview of this randomized circuit cutting.

The Clifford gates used in the random measurement circuit cutting approach are examples of unitary 2-designs [14]. One might wonder whether other gates improve the results even further. A simple generalization of Pauli gates is rotational Pauli gates. Even though these gates no longer form unitary 2-designs, it is worth investigating whether they still help reduce the sampling overhead. We therefore also consider a modified random measurement circuit cutting ap-

proach where we apply random Pauli rotation gates before the measurement and initialize the qubits using its conjugated form.

3 Experimental setup

We now detail the experiments we have run and their implementation framework.

3.1 Comparing circuit cutting methods

In the first experiment we compared the exact wire cutting method based on Pauli measurements [22] with the randomized measurement approach [14] and a version where we apply arbitrary Pauli rotation gates. We consider a simple circuit that prepares a five-qubit GHZ state [8], which we chose due to its wide applicability in many applications such as distributed quantum computing. Additionally, the cascade-like nature of the quantum circuit makes it suitable for wire cutting. We added rotation gates before and after the CNOT gates to complicate the state and assure all measurement bases and initial states contribute.

3.2 Evaluating the performance for circuit cutting in QAOA

In the second experiment, we use the best performing method to test the performance for a QAOA algorithm applied to the MaxCut problem. Given a graph, the MaxCut problem is to partition the vertex set of the graph in two such that the number of edges with vertices in distinct partitions is maximized [6]. For the QAOA algorithm, we assigned every vertex in the graph a qubit in the circuit. The cost operator is implemented by applying parametrized ZZ-gates between two qubits if and only if they are neighbors in the original graph. The mixing operator is implemented using parametrized R_X-gates. We implemented the parametrized ZZ-gates using R_Z-gates conjugated by CNOT-gates (where the R_Z-gate is applied to the target qubits of the CNOT-gates).

Figure 3 shows the three layered graphs used in our experiments. Due to the layered structure of the graph, the cost circuit shows a cascading architecture. This means that few gates have to be cut in our experiment. In addition, determining the optimum for comparison is simple for these graphs, allowing easy verification of the obtained results.

The parameters of the QAOA algorithm are optimized classically using Simultaneous Perturbation Stochastic Approximation (SPSA) [20], known for its potentially faster convergence with fewer quantum circuit executions [21].

3.3 Code framework

We used a dual-platform approach for our simulation workflow. PennyLane [2] performs the ideal simulations, while Qiskit's Aer simulator [11] facilitates both ideal and noisy simulations. The workflow first constructs circuits in Pennylane and then converts them to the OPENQASM 2.0 format using built-in tools [4].

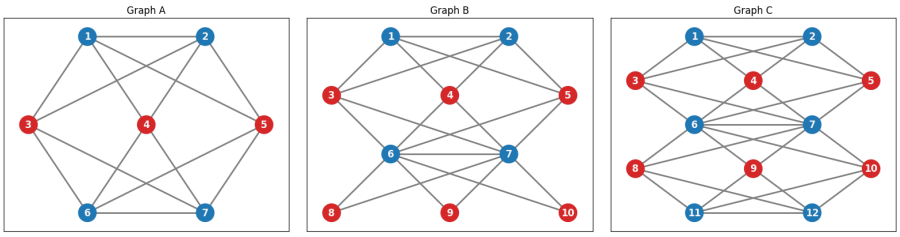


Fig. 3: Visual representation of the used layered graphs.

As the required mid-circuit measurements are not supported in this format, we developed a custom tool to upgrade the code to OPENQASM 3.0 [4]. The framework then sends this OPENQASM 3.0 code to construct a Qiskit quantum circuit object.

All experiments use a fixed shot budget evenly distributed over the sub-circuits. We measure the difference between two probability distributions $P = (p_i)_{i=1}^n$ and $Q = (q_i)_{i=1}^n$ with the Hellinger distance [7]:

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2}. \tag{3}$$

Results are averaged of 120 and 60 independent runs for noiseless uncut and cut experiments, respectively. Due to significant simulation cost in noisy versions, we averaged over 20 and 10 samples for the noisy experiments, respectively.

For the cut circuits, we assume that we have access to a quantum device of five qubits at most, requiring one, two, and three cuts for the graphs A, B and C of Figure 3. Our simulations use a noise model derived from the IBM Quantum Brisbane device [10].

Table 1 shows the number of qubits used for the different graphs in both the uncut and cut quantum circuits.

Table 1: Qubit requirements and the number of wirecuts for Graphs A, B, and C, demonstrating the reduction in qubits achieved through wirecutting.

Graph Name	**Wirecut Status**	**Qubits**	**Wirecut(s)**
Graph A	Uncut	7	0
	Wirecut	5	1
Graph B	Uncut	10	0
	Wirecut	5	2
Graph C	Uncut	12	0
	Wirecut	5	3

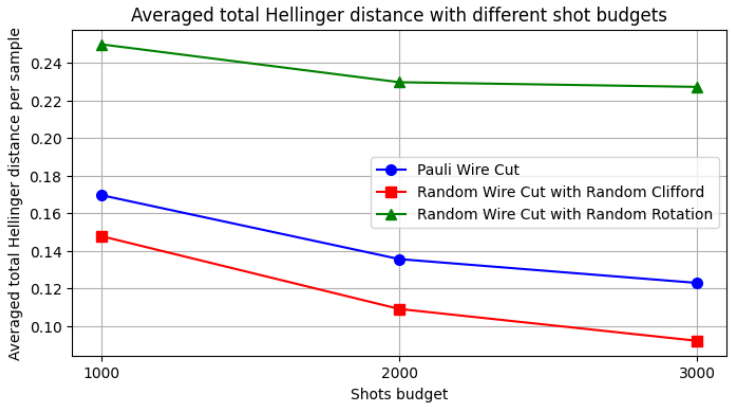


Fig. 4: Total Hellinger distance averaged over 60 independent runs for three wire cutting techniques (Pauli Wire Cut, Random Wire Cut with Random Clifford, and Random Wire Cut with Random Rotation) across varying shot budgets.

4 Results

This section presents the results of two distinct experiments: In Section 4.1, we investigate the efficiency and accuracy per shot of various wire-cutting techniques. Subsequently, in Section 4.2, we apply the randomized measurement wire-cutting technique [14] to QAOA for solving the MaxCut problem, evaluating its performance under both ideal and noisy simulation conditions.

4.1 Wire cutting techniques

We tested three circuit cutting methods: an exact method based on Pauli-measurements, a randomized method that randomly chooses between a random Clifford channel and a depolarizing channel, and a method that uses random Pauli rotation gates. Figure 4 shows the results for three different shot budgets. The randomized circuit cutting consistently outperforms the exact circuit cutting. Interestingly, applying random rotation gates before measurements and as state preparation (the third method) gives detrimental results.

All three methods show a decreased Hellinger distance with an increased shot budget. As more shots reduce the statistical variance, this is as expected. Lower Hellinger distance implies a better performance. The randomized approach using Clifford gates achieves the best performance with the same shot budget, implying a lower sampling overhead. From the same figure, we also see that, contrary to our hopes, the random Pauli rotations have a worse performance and higher sampling overhead even than standard Pauli cutting. We thus conclude that the randomized approach using Clifford gates works best and therefore use that method in our second experiment.

4.2 Wire cutting QAOA for MaxCut

This section uses randomized circuit cutting using Clifford gates to test how well a QAOA algorithm can solve MaxCut in both a cut and uncut version and in ideal and noisy settings. We varied the shot budget from 2000 to 4000.

Ideal Simulations Figure 5a shows how often the optimal cut is found in both an uncut and a cut run of the QAOA algorithm in an ideal simulation. We considered the three graphs shown in Figure 3.

Again, as shot budget increases, the optimal answer is found more often. As the complexity of the graphs increases, the success probability does however decrease. For the wire cut circuits, we see that especially for the larger graphs, suboptimal answers are found most of the time. This is quickly explained by the number of wire cuts applied (two for graph B and three for graph C), while fixing the shot budget. As a result, statistical variances have a greater effect on the overall performance. For improved performance, the shot budget should be significantly increased for the larger graphs: Every subcircuit requires sufficient samples to obtain the correct outcomes with high probability.

Noisy Simulations This section details the results of our noisy simulations for QAOA-MaxCut, comparing the performance of uncut and wirecut quantum circuits for the three graphs shown in Figure 3.

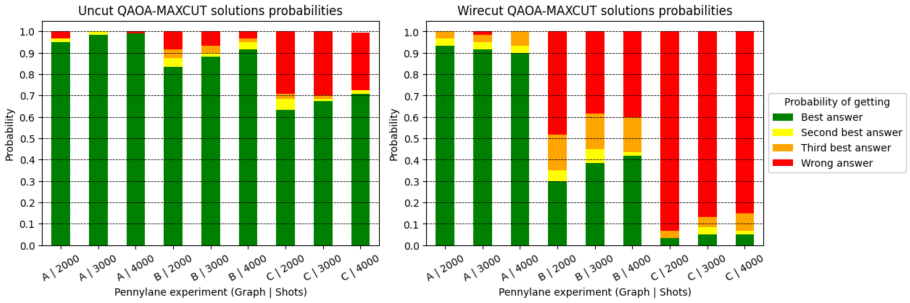
Figure 5b shows the results of the experiments with noisy simulations. We see that noisy simulations for the wirecut circuits are significantly worse than the uncut version. Contrary to what we expect, increasing the shot budget did not always result in higher accuracy. We believe this unexpected behavior stems from the fixed shot budget. With an increasing number of circuit cuts, the number of simulated subcircuits also increases. Each of these subcircuits gets assigned a smaller shot budget, making them more susceptible for statistical inaccuracies. The fact that for Graph C, the correct outcome was never found supports this feeling.

5 Conclusion

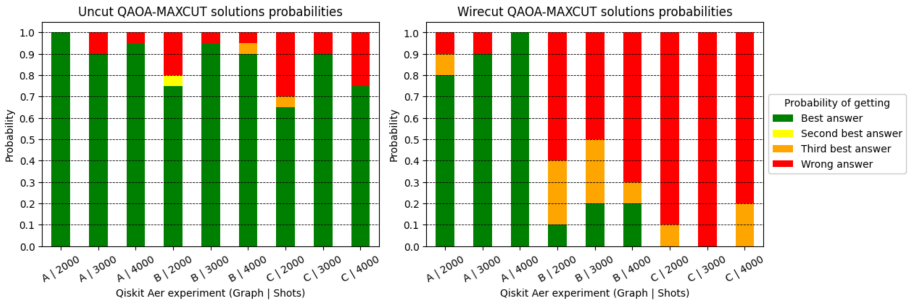
In this work we considered circuit cutting, a technique to break down a quantum circuit in smaller blocks. In all of our experiments, we fixed the shot budget, regardless of whether a circuit cut was used or not. In theory, an exponential number of shots is needed in the number of cuts performed. As a result, we expect statistical variance in our results, especially when multiple shots are being used.

We considered three different circuit cutting techniques and found that circuit cutting based on randomized Clifford gates achieves the best performance. We tested this using a circuit to prepare a GHZ state, where additionally random Pauli rotation gates are applied in between the CNOT gates.

In the second experiment we considered the effect of QAOA for solving a MaxCut problem. Here, we considered three different graphs, which required



(a) Results for the QAOA-MaxCut experiment with ideal simulation.



(b) Results for the QAOA-MaxCut experiment with noisy simulation. For the noise model of the IBM Brisbane quantum device was used.

Fig. 5: Numerical results on the performance of circuit cutting for QAOA. The left figures show results using uncut circuits, and the right figures show results using circuit cutting. Each bar shows the fraction of best, second best, third best, and wrong answer measured three graph instances and different shot budgets.

a different number of circuit cuts. Especially for the larger graphs, those that require two or even three circuit cuts, the performance lacks significantly compared to the uncut version. This effect is magnified in the noisy simulations. As mentioned before, we expect that this results in part from the need of an exponential number of shots, whereas a fixed number of shots was used. Overall, it can be concluded that applying wire cutting to QAOA reduces its accuracy significantly, though it enables execution on quantum hardware with fewer qubits. Furthermore, this loss in accuracy can be mitigated by increasing the shot budget for larger graphs with multiple wire cuts in ideal simulations but remains unknown in noisy circumstances.

One can imagine that the impact of individual terms in Equation (1) differs. Some terms have a large impact, while others have a smaller impact. In practice, it might therefore be advantageous to use a skewed shot budget, with more shots being allocated to large impact terms. This is an interesting topic for follow-up research.

For future research, we believe that larger sample sizes can help reduce statistical variances, this includes both larger shot budgets and averaging results over multiple independent runs. Especially the noisy simulations were computationally heavy, particularly due to the need for mid-circuit measurements. Next, it is vital to test the methods on actual hardware to see the performance in practice. Future research can also incorporate error-mitigation techniques, such as Pauli-twirling [24] or zero-noise extrapolation [23,13], to reduce the effect of noise.

Acknowledgments

This work is part of the project *Divide and Quantum* 'D&Q' NWA.1389.20.241 of the program 'NWA-ORC', which is partly funded by the Dutch Research Council (NWO).

References

1. Bechtold, M., Barzen, J., Beisel, M., Leymann, F., Weder, B.: Patterns for Quantum Circuit Cutting. In: Proceedings of the 30th Conference on Pattern Languages of Programs (PLoP '23). The Hillside Group (10 2023), in press
2. Bergholm, V., et al: Pennylane: Automatic differentiation of hybrid quantum-classical computations (2022), https://arxiv.org/abs/1811.04968
3. Chen, S., Yu, W., Zeng, P., Flammia, S.T.: Robust shadow estimation. PRX Quantum **2**, 030348 (Sep 2021). https://doi.org/10.1103/PRXQuantum.2.030348, https://link.aps.org/doi/10.1103/PRXQuantum.2.030348
4. Cross, A.W., Bishop, L.S., Smolin, J.A., Gambetta, J.M.: Open quantum assembly language (2017), https://arxiv.org/abs/1707.03429
5. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014), https://arxiv.org/abs/1411.4028
6. Filar, J.A., Haythorpe, M., Taylor, R.: Linearly-growing reductions of karp's 21 NP-complete problems. Numerical Algebra, Control and Optimization **8**(1), 1–16 (2018). https://doi.org/10.3934/naco.2018001
7. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International statistical review **70**(3), 419–435 (2002)
8. Greenberger, D.M., Horne, M.A., Zeilinger, A.: Going beyond bell's theorem. In: Bell's theorem, quantum theory and conceptions of the universe, pp. 69–72. Springer (1989)
9. Huang, H.Y., Kueng, R., Preskill, J.: Predicting many properties of a quantum system from very few measurements. Nat. Phys. **16**(10), 1050–1057 (Jun 2020). https://doi.org/10.1038/s41567-020-0932-7, http://dx.doi.org/10.1038/s41567-020-0932-7
10. IBM Quantum Computing: IBM Quantum (2025), https://quantum.ibm.com/
11. Javadi-Abhari, A., Treinish, M., Krsulich, K., Wood, C.J., Lishman, J., Gacon, J., Martiel, S., Nation, P.D., Bishop, L.S., Cross, A.W., Johnson, B.R., Gambetta, J.M.: Quantum computing with Qiskit (2024). https://doi.org/10.48550/arXiv.2405.08810

12. Kandala, A., Temme, K., Córcoles, A.D., Mezzacapo, A., Chow, J.M., Gambetta, J.M.: Error mitigation extends the computational reach of a noisy quantum processor. Nature **567**, 491–495 (2019). https://doi.org/10.1038/s41586-019-1040-7
13. Li, Y., Benjamin, S.C.: Efficient variational quantum simulator incorporating active error minimization. Phys. Rev. X **7**, 021050 (Jun 2017). https://doi.org/10.1103/PhysRevX.7.021050, https://link.aps.org/doi/10.1103/PhysRevX.7.021050
14. Lowe, A., Medvidović, M., Hayes, A., Oapos;Riordan, L.J., Bromley, T.R., Arrazola, J.M., Killoran, N.: Fast quantum circuit cutting with randomized measurements. Quantum **7**, 934 (Mar 2023). https://doi.org/10.22331/q-2023-03-02-934, http://dx.doi.org/10.22331/q-2023-03-02-934
15. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S.C., Yuan, X.: Quantum computational chemistry. Rev. Mod. Phys. **92**, 015003 (Mar 2020). https://doi.org/10.1103/RevModPhys.92.015003
16. Mitarai, K., Fujii, K.: Constructing a virtual two-qubit gate by sampling single-qubit operations. New Journal of Physics **23**(2), 023021 (feb 2021). https://doi.org/10.1088/1367-2630/abd7bc
17. Peng, T., Harrow, A.W., Ozols, M., Wu, X.: Simulating large quantum circuits on a small quantum computer. Phys. Rev. Lett. **125**, 150504 (Oct 2020). https://doi.org/10.1103/PhysRevLett.125.150504
18. Preskill, J.: Quantum computing in the NISQ era and beyond. Quantum **2**, 79 (2018)
19. Shor, P.W.: Scheme for reducing decoherence in quantum computer memory. Phys. Rev. A **52**, R2493–R2496 (10 1995). https://doi.org/10.1103/PhysRevA.52.R2493
20. Spall, J.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control **37**(3), 332–341 (1992). https://doi.org/10.1109/9.119632
21. Szava, A., Wierichs, D.: Optimization using SPSA. https://pennylane.ai/qml/demos/tutorial_spsa (03 2023), date Accessed: 2025-06-04
22. Tang, W., Tomesh, T., Suchara, M., Larson, J., Martonosi, M.: CutQC: using small quantum computers for large quantum circuit evaluations. In: Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems. p. 473–486. ASPLOS '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3445814.3446758
23. Temme, K., Bravyi, S., Gambetta, J.M.: Error mitigation for short-depth quantum circuits. Phys. Rev. Lett. **119**, 180509 (11 2017). https://doi.org/10.1103/PhysRevLett.119.180509
24. Wallman, J.J., Emerson, J.: Noise tailoring for scalable quantum computation via randomized compiling. Phys. Rev. A **94**, 052325 (11 2016). https://doi.org/10.1103/PhysRevA.94.052325, https://link.aps.org/doi/10.1103/PhysRevA.94.052325
25. Zhou, L., Wang, S.T., Choi, S., Pichler, H., Lukin, M.D.: Quantum Approximate Optimization Algorithm: Performance, mechanism, and implementation on near-term devices. Phys. Rev. X **10**, 021067 (Jun 2020). https://doi.org/10.1103/PhysRevX.10.021067

Quantum Minimal Learning Machine: A Fidelity-Based Approach to Error Mitigation

Clemens Lindner[1 [0009−0006−3734−5573]], Joonas Hämäläinen[1 [0000−0002−8466−9232]], and Matti Raasakka[2 [0000−0003−2636−0429]]

[1] University of Jyväskylä, Faculty of Information Technology, 40100 Jyväskylä, Finland
clindner@jyu.fi
[2] Aalto University, Department of Electronics and Nanoengineering, 02150 Espoo, Finland

Abstract. We introduce the concept of quantum minimal learning machine (QMLM), a supervised similarity-based learning algorithm. The algorithm is conceptually based on a classical machine learning model and adopted to work with quantum data. We will motivate the theory and run the model as an error mitigation method for various parameters.

Keywords: Quantum machine learning · Quantum computing · Machine learning · Error mitigation · Quantum information.

1 Introduction

Quantum machine learning [2] spans a wide range of possibilities: Leveraging fault tolerant quantum algorithms to accelerate core linear algebra routines [6], adopting quantum counterparts of classical kernel methods [11], and employing parametrized variational quantum algorithms (VQAs) optimized via a hybrid quantum-classical workflow [7]. When run on actual quantum computing hardware these models suffer from hardware noise, therefore methods of error mitigation are becoming another important avenue for the mingling of machine learning with quantum computing [9]. In this paper we will introduce a novel similarity based quantum learning model and explore its use for error mitigation. We will take inspiration from a classical model known as minimal learning machine (MLM) [4] where a linear mapping between distances in the input and output spaces is used to learn the labels of some given data-set. Several quantum k-nearest neighbor approaches have been implemented [13, 1], yet most deal with the calculation of distances between classical data instead of quantum data [8]. Since one of the hopes of quantum computers is that they will be able to simulate quantum systems and then provide us with synthetic data we aim to find a way of directly learning from that data. In section 2 we are first going to introduce the MLM model and then establish one possible way to use the same idea for quantum data. After we reformulated the model to work with quantum data we will use it in section 3 to demonstrate error mitigation of a given

noisy quantum state. The main aim in this specific use-case will be to assume we are given a noisy quantum state from a real quantum computer where we don't know the precise noise model and then learn its ideal noise-free version. In our implementation we will simulate this process with depolarizing noise and a dataset of variational quantum circuits using Qiskit and QiskitAer. We will analyze the models performance with different circuit parameter settings (number of qubits, rotational parameter ranges, Ansatz layers, noise levels) and provide the corresponding code.

2 Conceptual framework and preliminaries

2.1 MLM

Minimal learning machine (MLM) is a supervised learning method useful for learning the relationship between a set of input and output points. An exemplary use-case is multi label classification (MLC) [5] where the output points correspond to labels of the input points. In the following we will work with that example.

We are given a dataset $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_N, \mathbf{y}_N)\}$, where $\mathbf{x}_i \in \mathbb{R}^M$ corresponds to an input vector and $\mathbf{y}_i \in \{0,1\}^L$ to the multi-hot encoded output label vector for $i = 1, \ldots, N$. In $\mathbf{y}_i$, value 1 at index j represents that class-label j is associated with instance i. As an example, $\mathbf{y}_1 = (1, 1, 0, 0)$ could correspond to the vectorized picture data $\mathbf{x}_1$ including class-label 1 and 2 but not 3 and 4. For a MLC dataset, $\sum_{j=1}^{L} \mathbf{y}_i(j) > 1$ at least for one instance.

The approach goes as follows: We calculate the distances between the input space points and some set of reference points, this is done accordingly in the output space. We then store them in distance matrices $\mathbf{D_x}$ and $\mathbf{D_y}$ which are correlated via a parametrized matrix $\mathbf{B}$, i.e. $\mathbf{D_x B} = \mathbf{D_y}$. After having learned $\mathbf{B}$ with a set of training points we can use it to label new data points.

In MLM, the core subroutine is to solve the model's coefficient matrix $\mathbf{B}$ via ordinary least squares $\mathbf{B} = (\mathbf{D_x}^T \mathbf{D_x})^{-1} \mathbf{D_x}^T \mathbf{D_y}$ for feature space and output space distance matrices, $\mathbf{D_x} \in \mathbb{R}^{N \times K}$ and $\mathbf{D_y} \in \mathbb{R}^{N \times N}$, where $\mathbf{D_x}_{(i,j)} = \|\mathbf{x}_i - \mathbf{r}_j\|_2$, $\mathbf{r}_j \in \mathcal{R} = \{\mathbf{r}_k\}_{k=1}^{K} \subseteq \mathcal{X}$ and $\mathbf{D_y}_{(i,j)} = \|\mathbf{y}_i - \mathbf{y}_j\|_2$. The quality of the linear mapping $\mathbf{B}$ is affected by how feature space reference points $\mathcal{R}$ are selected. If $\mathbf{D_x}$ is a full rank matrix for $K = N$, we can use $\mathbf{B} = \mathbf{D_x}^{-1} \mathbf{D_y}$. In the prediction phase of MLM, for a given instance $\mathbf{x}$ distances are computed to $\mathcal{R}$ and then $\mathbf{B}$ is used to estimate output space distances to all training labels. An actual prediction for a given task is then extracted by post-processing of the distance estimates. The simplest approach for this is to identify the smallest predicted distance and its corresponding label vector $\hat{y} \in \mathcal{Y}$ [5]. Note that by selecting $\mathcal{R} = \mathcal{X}$ $(K = N)$, MLM is deterministic.

2.2 QMLM

In a 'quantized' version of MLM we can use the same model to work with quantum data. For illustrative purposes we keep with the multi label approach and introduce the space $\mathcal{H}_\mathcal{X}$, a Hilbert space in which our input data lives. More specifically the input data we wish to prescribe multiple labels to is a set of pure quantum states $|\Psi\rangle_i \in \mathcal{H}_\mathcal{X}$ with $i \in \{1, 2, \ldots, N\}$ or their corresponding density matrices $\rho_i = |\Psi_i\rangle\langle\Psi_i|$. As soon as we deal with quantum data we have to use a different distance measure to quantify data similarity. An intuitive way is to introduce the fidelity between two quantum input states as our distance measure, i.e.

$$F(|\Psi\rangle_i, |\Psi\rangle_j) = |\langle\Psi_i|\Psi_j\rangle|^2, \quad \text{where } |\Psi\rangle_i, |\Psi\rangle_j \in \mathcal{H}_\mathcal{X}. \tag{1}$$

We have $F(|\Psi\rangle_i, |\Psi\rangle_j) = F(|\Psi\rangle_j, |\Psi\rangle_i)$ and $0 \leq F(|\Psi\rangle_i, |\Psi\rangle_j) \leq 1$ where a value closer to one means the states are more similar. We choose $\mathcal{R} = \mathcal{X}$ and therefore define the matrix $\mathcal{D}_{\mathcal{H}_\mathcal{X}}$ to hold the fidelity values between all pairs of quantum states in the input space:

$$D_{\mathcal{H}_\mathcal{X}} = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1N} \\ F_{21} & F_{22} & \cdots & F_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N1} & F_{N2} & \cdots & F_{NN} \end{bmatrix}, \tag{2}$$

where $F_{ij} = F(|\Psi\rangle_i, |\Psi\rangle_j)$ and $i, j \in \{1, 2, \ldots, N\}$. In practice, these fidelity values can be determined by implementing a swap test.

For the multi label case we could naively assume an encoding of the labels through the computational basis states, i.e. $|\Phi_i\rangle = |01101\rangle \in \mathcal{H}_\mathcal{Y}$. This approach however would lead to fidelity values of either 0 or 1 and therefore our distance matrix in the label space wouldn't hold much information. We also have to find a way where a large Hamming distance between two label vectors, i.e. the number of bit positions where they differ $d_H(\mathbf{y}_i, \mathbf{y}_j)$, corresponds to a low fidelity between the two corresponding quantum states. This can be achieved by using two overlapping states in our representation. We encode the label vector $\mathbf{y}_i$ in the quantum state $|\Phi_i\rangle$ where $i \in \{1, 2, \ldots, N\}$. If we denote the k-th value in a bit string of length L corresponding to the label $\mathbf{y}_i$ as b_{ik} and it's corresponding quantum state as $|\phi_{ik}\rangle$ we can write

$$|\Phi_i\rangle = \bigotimes_{k=1}^{L} |\phi_{ik}\rangle = \bigotimes_{k=1}^{L} \begin{cases} |+\rangle, & \text{if } b_{ik} = 1 \\ |0\rangle, & \text{if } b_{ik} = 0, \end{cases} \tag{3}$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The bit string 01101 would hence be represented as $|\Phi_i\rangle = |0 + +0+\rangle$. Since $\langle 0|0\rangle = \langle +|+\rangle = 1$ and $\langle 0|+\rangle = \langle +|0\rangle = \frac{1}{\sqrt{2}}$ we have $\langle\Phi_i|\Phi_j\rangle = (\frac{1}{\sqrt{2}})^{d_H(\mathbf{y}_i, \mathbf{y}_j)}$ and therefore

$$F(|\Phi_i\rangle, |\Phi_j\rangle) = |\langle\Phi_i|\Phi_j\rangle|^2 = \left(\frac{1}{2}\right)^{d_H(\mathbf{y}_i,\mathbf{y}_j)}. \tag{4}$$

This way a large Hamming distance leads to a low fidelity value, as intended. We denote the distance matrix in the output space as $D_{\mathcal{H}_\mathcal{Y}}$ with entries $F'_{ij} = F(|\Phi_i\rangle, |\Phi_j\rangle)$ where $i, j \in \{1, 2, \ldots, N\}$. The rest of the algorithm can follow analogously to the classical MLM. We assume a linear mapping between the distance spaces, i.e. $D_{\mathcal{H}_\mathcal{X}}\mathcal{B} = D_{\mathcal{H}_\mathcal{Y}}$ and obtain $\mathcal{B}$ by taking the pseudoinverse of $D_{\mathcal{H}_\mathcal{X}}$. We then calculate the fidelities for a new test data state $|\Psi_t\rangle$ to all other input states, apply the mapping $\mathcal{B}$ and gain a vector encoding all mapped similarities in the label space. Since this vector encodes fidelities we now have to take the maximum value which corresponds to the highest similarity. The index of that value will give us the label corresponding to $|\Psi_t\rangle$, i.e. $|\Phi_t\rangle$, which after a quick remapping ($|+\rangle \mapsto 1, |0\rangle \mapsto 0$) will yield the labels bit string.

While this specific QMLM example could be used for multi label classification of quantum states it is primarily useful if we want to have the labels stored on a quantum computer and could still be improved upon by e.g. choosing a different label encoding. In the next section we are going to discuss a more near term use-case for QMLM, that being the use for error mitigation of noisy quantum states. Looking further ahead to fault-tolerant devices, QMLM may also prove valuable in any context where a systematic relationship can be established between the similarity structure of quantum states and that of a corresponding output space.

3 QMLM for error mitigation

In this section the aim of our QMLM model will be to learn an ideal quantum state when provided with its noisy version. The assumption is that we are given a set of noisy quantum states which could have been obtained from running quantum circuits that suffer from some hardware-specific noise. Our dataset consists of noisy states as inputs, i.e. $\{\rho_{i,\text{noisy}}\}$ and the corresponding ideal pure states $\{|\Psi_i\rangle\}$ as outputs. The fidelity in the input space would therefore take the general form

$$F(\rho_i, \rho_j) = \left(\text{Tr}\sqrt{\sqrt{\rho_i}\rho_j\sqrt{\rho_i}}\right)^2, \tag{5}$$

while the fidelity in the output space reduces to the form we have in equation 1. The assumption is now that since there always exists a completely positive trace preserving map $\mathcal{E}$, s.t. $\mathcal{E}(|\Psi_\text{ideal}\rangle\langle\Psi_\text{ideal}|) = \rho_\text{noisy}$, we can learn a useful relationship between the similarity matrices in the input and output space that could provide us with information about the specific hardware noise we encounter. In our implementation of the model we focused on depolarizing noise, i.e.

$$\mathcal{E}_\text{depol}(\rho_i) = (1-p)\rho_i + \frac{p}{d}I \tag{6}$$

where ρ_i is the initial quantum state, p is the depolarizing probability with $0 \leq p \leq 1$, d is the dimension of the Hilbert space (for an n-qubit system, $d = 2^n$) and I is the $d \times d$ identity matrix. Let ρ_1' and ρ_2' be the density matrices obtained after applying the depolarizing noise channel $\mathcal{E}_{\text{depol}}$ to the initial pure states ρ_1 and ρ_2 with depolarizing parameters λ_1 and λ_2, respectively:

$$\rho_i' = (1 - \lambda_i)\, \rho_i + \frac{\lambda_i}{d}\, I, \quad i = 1, 2\,. \tag{7}$$

Reordering for ρ_1 and ρ_2 and taking the fidelity formula for pure states $F(\rho_1, \rho_2) = \text{Tr}(\rho_1 \rho_2)$, we can relate the fidelity of our pure states with the trace over the resulting mixed ones:

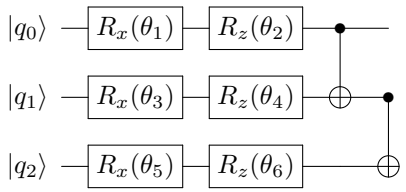
$$\text{Tr}(\rho_1' \rho_2') = \alpha F(\rho_1, \rho_2) + \frac{1 - \alpha}{d}, \tag{8}$$

where $\alpha = (1 - \lambda_1)(1 - \lambda_2)$. In theory, this would therefore motivate a mapping between fidelities for ideal states and the traces of their noisy counterparts which shows similarities to the approach found in [3], where a linear model is motivated for ideal and noisy expectation values. The benefit of our model (which corresponds to each ideal fidelity being a linear combination of noisy fidelities) is that it could model the underlying noise more generally. This could be useful since we are in reality never limited to just one specific noise model. In our setting, once the ideal version of a quantum state is obtained, the noise-free expectation value of an operator O can be evaluated via $\langle O \rangle_{\text{ideal}} = \text{Tr}(O \rho_{\text{ideal}})$.

Although our QMLM model has the benefit of broad applications, it does suffer from the same problem as quantum kernel methods, that being the problem of exponential concentration [12]. This is obvious since the fidelity matrix in equation 2 represents a Gram matrix whose off diagonal values go to zero if the dimension of the Hilbert space grows too large. In our implementation, we try to negate this effect by limiting the region of the accessible Hilbert space through small parameter changes in a specified variational quantum circuit. For real applications this could correspond to the knowledge that a set of given quantum data is limited to a certain region of Hilbert space or exhibits some inherent symmetry akin to approaches in geometrical quantum machine learning [10].

3.1 Qiskit Implementation

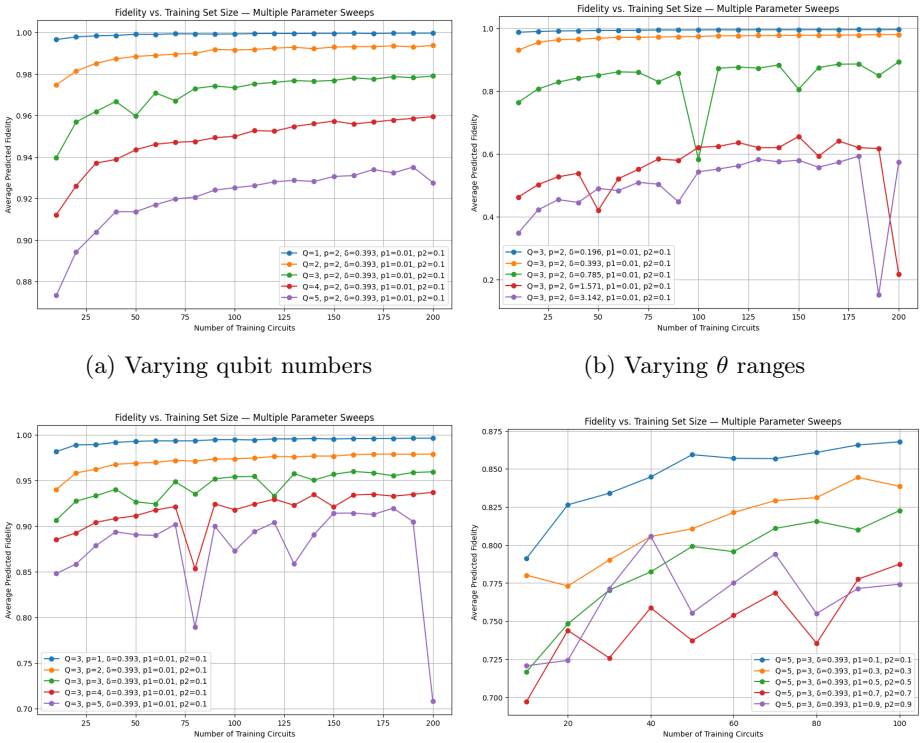
We implement the depolarizing noise model with QiskitAer and use the same noise parameters for each training data set. Our output space dataset consists of N variational quantum circuits with an Ansatz corresponding to the following form:



Because of the already mentioned vanishing fidelity problems, we will adjust the range of the randomly drawn θ from $\theta \in [-\pi, +\pi]$ to smaller intervals given by some δ, e.g. $\delta = \frac{\pi}{2}$ s.t. $\theta \in [-\delta, +\delta]$. The amount of repeated Ansatz layers can be specified by a parameter p, the depicted circuit here corresponds to $p = 1$. More specifically our output space dataset is then given by the statevectors we gain from such circuits using an ideal simulator. For each VQA circuit of the output space dataset we apply the depolarizing noise model which applies some error to each 1 and 2 qubit gate in the circuit depending on the chosen parameters p_1 and p_2. For each circuit we then gain the mixed density matrix as an end result of that noisy simulation and set the input space to be the collection of those.

After creating our dataset of ideal statevectors (the set $\mathcal{Y}$) and their noisy counterparts (the set $\mathcal{X}$) we calculate their corresponding fidelity matrices $D_{\mathcal{H}_\mathcal{Y}}$ and $D_{\mathcal{H}_\mathcal{X}}$. Through a standard pseudoinverse of $D_{\mathcal{H}_\mathcal{X}}$ we gain $\mathcal{B}$. Now a new data point is created by randomly drawing a new set of θ_i angles and applying the noise model to the corresponding circuit, meaning the new data point is some noisy state but by the nature of our process we have access to the ideal one. In reality such a new noisy state would be obtained from an actual quantum computation. After the QMLM procedure, we gain the quantum state of our set $\mathcal{Y}$ which exhibits the highest similarity to the ideal version of the new data point. This suggests that using more training samples improves performance, where performance is quantified by the fidelity between the predicted and the true ideal state. Because our data is generated randomly, we evaluate performance across many new test circuits and report it as the "average predicted fidelity".

In Fig. 1a we provide a result for that fidelity in comparison with the number of training dataset points (i.e. the number of VQA circuits) for $Q \in [1, 2, 3, 4, 5]$ qubits. For each visible point in all graphs we have averaged over 400 predictions. We limited the rotation range to $\theta \in [-\frac{\pi}{8}, +\frac{\pi}{8}]$ and have increased the noise to a level above known real hardware noise (approximately $p_1 = 0.001$ and $p_2 = 0.01$) to make its effect more prevalent. We can see that a higher dimensional Hilbert space leads to worse performance since we would need more points in order to approximately cover the whole space. Also while there is an initial increase in performance we can also see a leveling effect as soon as the reachable space is sufficiently covered by the training set data.
In Fig. 1b we compare different $\delta \in [\pi, \frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{16}]$ values and focus on three qubits. Despite the averaging we can see some statistical fluctuations which should be smoothed if we increase the number of predictions we average over



(a) Varying qubit numbers

(b) Varying θ ranges

(c) Varying Ansatz layers

(d) Varying noise levels

Fig. 1: Average predicted fidelity vs. dataset size under different experimental parameters.

while increasing model runtime. We observe that when the accessible Hilbert-space region grows, performance degrades, reflecting that both space dimension and VQA circuit expressibility impact results. Since more Ansatz layers also correspond to higher expressibility, the same effect should be a cause of worse performance as seen in Fig. 1c where we compare different Ansatz layer parameters p. The worse performance for higher p values can also in general be traced to the fact that any gate errors now have more chances to accumulate.

At last we will also compare different noise levels in Fig. 1d. Regarding the noise levels, our model runs into two singular cases: one where the noise is zero and $\mathcal{B}$ therefore becomes the identity matrix and the other one where the noise is maximal, leading to all input states being the maximally mixed one. In the first case the model basically acts as a lookup table, i.e. it checks if the new state is in the training data whereas the second case corresponds to $D_{\mathcal{H}_\mathcal{X}}$ consisting of only 1's in its entries which leads to the output similarity vector consisting of the same value in each entry hence providing us some trivial average performance

depending on $D_{\mathcal{H}_\mathcal{Y}}$. We plot the results for 5 qubits and note that the noise levels we chose are unrealistic, however increasing them still shows a general decrease in the models performance. This probably corresponds to the fact that the more noise we introduce the more similar the states become, leaving us with less information to learn from. We also observed in general larger fluctuations when increasing noise levels.

While we laid the groundwork for the QMLM model, further improvements can definitely still be done. Examples would be to change the post processing from a simple nearest neighbor approach to some weighted average over the closest points or to test if a different mapping between the distance matrices could lead to a better performance. Furthermore it would be interesting to run the model on actual quantum hardware and see if it sufficiently captures the more general noise models of real case scenarios. This however would lead to other problems, e.g. we can't just run a swap test on real hardware to measure the fidelity of noisy states since the swap test quantum gates themselves would be affected by the noise. We therefore either have to mitigate such effects or access some other methods like quantum state tomography.

4 Conclusion

In this paper, we presented an approach towards a 'quantized' MLM model. This was done by assuming the availability of a given set of quantum states and using the fidelity between them as a similarity measure. We specified use cases and ran the model for error mitigation. Our results show a good performance as long as we limit the accessible range of our created quantum states, more work is however still needed to test if the model is usable for higher dimensional data. QMLM could be a versatile approach to learning features of quantum states and a possible future avenue for learning immediately from the quantum data created on a fault-tolerant quantum computer.

5 Code availability

Our code was run in Jupyter notebooks using Qiskit and QiskitAer for the quantum simulations. The full implementation is provided in the following GitHub: https://github.com/clemensLin/QMLM.

Acknowledgments. We acknowledge the financial support of the Finnish Ministry of Education and Culture through the Quantum Doctoral Education Pilot Program (QDOC VN/3137/2024-OKM-4) and the Research Council of Finland through the Finnish Quantum Flagship project (JYU 359240, University of Jyväskylä).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Berti, A., Bernasconi, A., Del Corso, G.M., Guidotti, R.: The role of encodings and distance metrics for the quantum nearest neighbor. Quantum Machine Intelligence **6**(2), 62 (2024)
2. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**(7671), 195–202 (2017)
3. Czarnik, P., Arrasmith, A., Coles, P.J., Cincio, L.: Error mitigation with clifford quantum-circuit data. Quantum **5**, 592 (2021)
4. De Souza Junior, A.H., Corona, F., Barreto, G.A., Miche, Y., Lendasse, A.: Minimal learning machine: A novel supervised distance-based approach for regression and classification. Neurocomputing **164**, 34–44 (2015)
5. Hämäläinen, J., Hubermont, A., de Souza Júnior, A.H., Mattos, C.L., Gomes, J.P., Kärkkäinen, T.: Minimal learning machine for multi-label learning (2024), arXiv:2305.05518
6. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Physical review letters **103**(15), 150502 (2009)
7. Jerbi, S., Fiderer, L.J., Poulsen Nautrup, H., Kübler, J.M., Briegel, H.J., Dunjko, V.: Quantum machine learning beyond kernel methods. Nature Communications **14**(1), 1–8 (2023)
8. Kiani, B.T., De Palma, G., Marvian, M., Liu, Z.W., Lloyd, S.: Learning quantum data with the quantum earth mover's distance. Quantum Science and Technology **7**(4), 045002 (2022)
9. Liao, H., Wang, D.S., Sitdikov, I., Salcedo, C., Seif, A., Minev, Z.K.: Machine learning for practical quantum error mitigation. Nature Machine Intelligence pp. 1–9 (2024)
10. Ragone, M., Braccia, P., Nguyen, Q.T., Schatzki, L., Coles, P.J., Sauvage, F., Larocca, M., Cerezo, M.: Representation theory for geometric quantum machine learning. arXiv preprint arXiv:2210.07980 (2022)
11. Schuld, M.: Supervised quantum machine learning models are kernel methods (2021), https://arxiv.org/abs/2101.11020
12. Thanasilp, S., Wang, S., Cerezo, M., Holmes, Z.: Exponential concentration in quantum kernel methods. Nature communications **15**(1), 5200 (2024)
13. Zardini, E., Blanzieri, E., Pastorello, D.: A quantum k-nearest neighbors algorithm based on the euclidean distance estimation. Quantum Machine Intelligence **6**(1), 1–22 (2024)

No Scratch Quantum Computing by Reducing Qubit Overhead for Efficient Arithmetics

Omid Faizy[1,2,3,4], Norbert Wehn[3], Paul Lukowicz[4,5], and Maximilian Kiefer-Emmanouilidis[4,5]

[1] Laboratoire de Chimie de la Matière Condensée de Paris, UMR CNRS 7574, Sorbonne Université, 4, place Jussieu, 75252 Paris, France.
[2] Laboratoire de Mathématiques Pures et Appliquées Joseph Liouville, Université du Littoral Côte d'Opale, 50 rue Ferdinand Buisson, CS 80699, 62228 Calais, France.
[3] Microelectronic Systems Design (EMS), RPTU Kaiserslautern-Landau, Germany
[4] Department of Computer Science and Research Initiative QC-AI, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany.
[5] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany.

Abstract. Quantum arithmetic computation requires a substantial number of scratch qubits to stay reversible. These operations necessitate qubit and gate resources equivalent to those needed for the larger of the input or output registers due to state encoding. Quantum Hamiltonian Computing (QHC) introduces a novel approach by encoding input for logic operations within a single rotating quantum gate. This innovation reduces the required qubit register N to the size of the output states O, where $N = \log_2 O$. Leveraging QHC principles, we present reversible half-adder and full-adder circuits that compress the standard Toffoli + CNOT layout [Vedral et al., PRA, 54, 11, (1996)] from three-qubit and four-qubit formats for the Quantum half-adder circuit and five sequential Fredkin gates using five qubits [Moutinho et al., PRX Energy 2, 033002 (2023)] for full-adder circuit; into a two-qubit, 4×4 Hilbert space. This scheme, presented here, is optimized for classical logic evaluated on quantum hardware, which due to unitary evolution can bypass classical CMOS energy limitations to a certain degree.

While the present circuits target classical Boolean inputs (i.e., classically controlled operation) and prepare the outputs from the initialized state $|00\rangle$, we make this scope explicit: the blocks here are not intended as coherent adders acting on superpositions. Instead, they are resource-minimal *classical-to-quantum* primitives that realize full truth tables on two qubits with no scratch registers. Because the input parameters are binary, the resulting unitaries reduce to a finite set of two-qubit permutations that can be implemented exactly without T gates or approximation overhead. We further position these blocks as building units for *FPGA-like quantum-configurable logic*, e.g., programmable interferometer meshes or analog control blocks, where minimizing qubits and circuit depth is advantageous.

Keywords: Quantum Arithmetics · Quantum Hamiltonian Computing · Adder.

1 Introduction

The question of minimal and most efficient arithmetic operators is approaching the atomic limits. Below a few hundred atoms per transistor, leakage and energy dissipation at least $kT \ln 2$ per bit become prohibitive [4, 7, 17]. Reversible quantum logic can bypass these bounds to a certain degree [4, 7, 32]. Following this principle, early quantum adders [8, 32] implemented binary addition using Toffoli and CNOT gates, but at the expense of extra ancilla qubits for reversibility and state embedding. In this manuscript, we consider any qubit not needed to encode the output as an ancillary qubit. This distinguishes our approach from other ancilla-free methods which do not need any extra qubits when transpiled into universal gates, such as the three-qubit half adder, the four-qubit full adder. Here, the qubits necessary for input embedding are not treated as ancillas. To emphasize this distinction, we refer to scratch qubits or scratch registers instead, which identifies qubits needed for the embedding, as a form of ancilla.

For instance, standard reversible half-adder and full-adder circuits [21] require three and four qubits, respectively, in order to produce sum and carry outputs without erasing the inputs. This holds true even if the inputs are classical so that input and output are product states, and thus separable states. This "scratch" qubit overhead not only increases the Hilbert space (dimension 8 for three and dimension 16 for four qubits) but also adds numerous gates to uncompute intermediate results, complicating practical implementation. More recent designs reduce either the T-count or the ancilla count—sometimes to one or even zero [13, 28, 29], where additional ancillas refer to qubits necessary beyond the aforementioned input embedding for half-adders and four qubits for full-adders.

Quantum Hamiltonian Computing (QHC) encodes Boolean inputs in Hamiltonian matrix elements rather than basis populations, minimizing the state count for logic gates [1, 10, 11, 22]. Previous QHC studies implemented on single-molecule platforms realized 4-state half-adders and 8-state full-adders that required no qubits and relied on intrinsically irreversible dynamics; later work introduced multi-energy read-out schemes to compress the effective Hamiltonian even further [11, 22].

Theoretical studies show that classical data can be encoded as rotation angles in a QHC circuit by using a nano-graphene molecule doubly functionalized with nitro groups and bridged by graphene electrodes, whose nitro-group rotations serve as binary inputs that collectively implement a half-adder Boolean operation [31].

Adapting these ideas to quantum computing is to find a unitary $U(\overrightarrow{\theta}) = \exp(-\mathrm{i}H(\overrightarrow{\theta})t)$ that depending on the binary inputs formulates the wanted logical output of any given truth table for the given input $\overrightarrow{\theta}$. From this point of view we argue that we can formulate a quantum representation (in the form of a quantum gate) which yields a given output (following the corresponding truth table) with only $N = \log_2(O)$ qubits where O is the number of possible outputs.

Using the QHC paradigm, we present constructed reversible half- and full-adder circuits using only two qubits. This represents a significant reduction in

the required resources compared to conventional designs. Both our half-adder and full-adder operate within a 4×4 Hilbert-space (two qubits), rather than the 8×8 (three qubit), or 16x16 (four qubits) or larger spaces required by previous methods.

We have to emphasise that our design *does not* implement coherent addition on superpositions of inputs, nor is it meant to be a drop-in replacement for multi-bit quantum adders used inside phase-estimation or modular arithmetic. Rather, it is a resource-minimal *classical-to-quantum* primitive that can be useful in hybrid pipelines [3, 19], IO [6, 15], quantum routing [26], and programmable analog or photonic fabrics [27].

2 Circuit design

Throughout, we denote the two input bits by $a, b \in \{0, 1\}$ and the input carry by $c_{\text{in}} \in \{0, 1\}$. The outputs are read as SUM on the second qubit as least significant bits (LSB) and CARRY on the first qubit as most significant bit(MSB). Furthermore, we adopt units where the reduced Planck constant $\hbar = 1$.

Starting with the half-adder as depicted in Fig. 1 (a), the device employs two qubits prepared in $|\psi(t = 0)\rangle = |00\rangle$. A 4×4 unitary $U(a, b)$ confines dynamics to a protected sub-manifold and encodes the angles $(a, b) \in \{0, 1\}^2$ as logical inputs. The $U(a, b)$ filters and maps resulting amplitudes onto the final qubits, which are then measured in the computational basis for SUM (XOR) and CARRY (AND) outputs.

$$U(a,b) = \begin{pmatrix} A & B-F & B+F & 0 \\ B+F & A & B-F & 0 \\ B-F & B+F & A & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{1}$$

where

$$\begin{aligned} A &= \frac{1}{3}\left(2\cos\left(\frac{2}{3}\pi(a+b)\right)+1\right), \\ B &= \frac{1}{3}\left(1-\cos\left(\frac{2}{3}\pi(a+b)\right)\right), \\ F &= \frac{\sin\left(\frac{2}{3}\pi(a+b)\right)}{\sqrt{3}}. \end{aligned} \tag{2}$$

The corresponding half-adder Hamiltonian is

$$H_{\text{ha}}(a,b) = \frac{2\pi(a+b)}{3\sqrt{3}} h_{\text{ha}} = \frac{2\pi(a+b)}{3\sqrt{3}} \begin{pmatrix} 0 & -i & i & 0 \\ i & 0 & -i & 0 \\ -i & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3}$$

The time-evolution of this Hamiltonian $U(a, b) = \exp(-ih_{ha}t_{\text{eff}}(a, b))$ shows that the solution of the half-adder can be extracted on the timings given by $t_{\text{eff}}(a, b) = \frac{2\pi(a+b)}{3\sqrt{3}}$, see Fig. 2(b).

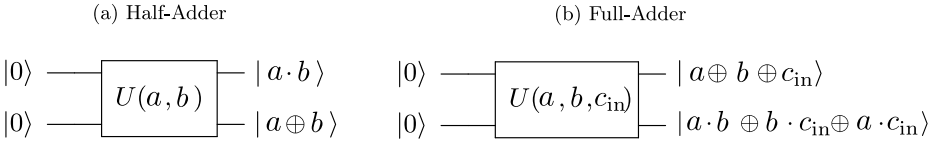


Fig. 1: (a) Two-Qubit QHC-Half-Adder circuit with classical binary inputs a and b implemented directly inside circuit. (b) Two-Qubit QHC-Full-Adder circuit with classical binary inputs a, b, and c_{in} implemented directly inside circuit.

Given the initial state $|00\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^T$ we can fulfill the half-adder truth table and response as:

$$\begin{aligned} U(0,0)\,|00\rangle &= |00\rangle \\ U(0,1)\,|00\rangle &= |01\rangle \\ U(1,0)\,|00\rangle &= |01\rangle \\ U(1,1)\,|00\rangle &= |10\rangle \end{aligned} \tag{4}$$

We can generalize now the protocol with full-adder as depicted in Fig. 1 (b). Again the device employs two qubits prepared in $|\psi(t=0)\rangle = |00\rangle$. A 4×4 unitary $U(a, b, c_{\text{in}})$ confines dynamics to a protected sub-manifold and encodes the angles $(a, b, c_{\text{in}}) \in \{0, 1\}^3$ as logical inputs. The $U(a, b, c_{\text{in}})$ filters and maps the resulting amplitudes to the final qubits, which are then measured in the computational basis for the SUM (XOR) and CARRY outputs.

The corresponding matrix has the following form:

$$U(a,b,c_{\text{in}}) = \tfrac{1}{4} \begin{pmatrix} l+2m+1 & -2n-p-iq+1 & l-2m+1 & 2n-p-iq+1 \\ 2n-p-iq+1 & l+2m+1 & -2n-p-iq+1 & l-2m+1 \\ l-2m+1 & 2n-p-iq+1 & l+2m+1 & -2n-p-iq+1 \\ -2n-p-iq+1 & l-2m+1 & 2n-p-iq+1 & l+2m+1 \end{pmatrix}, \tag{5}$$

where

$$\begin{aligned} l &= e^{i\pi(a+b+c_{\text{in}})}, \\ m &= \cos\left(\frac{1}{2}\pi(a+b+c_{\text{in}})\right), \\ n &= \sin\left(\frac{1}{2}\pi(a+b+c_{\text{in}})\right), \\ p &= \cos(\pi(a+b+c_{\text{in}})), \\ q &= \sin(\pi(a+b+c_{\text{in}})). \end{aligned} \tag{6}$$

See Sec. 5 Appendix for the corresponding steps to arrive at Eq. (5).

The corresponding full-adder Hamiltonian reads

$$H_{fa} = \frac{1}{4}\pi(a+b+c_{\text{in}})h_{fa} = \frac{1}{4}\pi(a+b+c_{\text{in}}) \begin{pmatrix} -1 & 1-i & -1 & 1+i \\ 1+i & -1 & 1-i & -1 \\ -1 & 1+i & -1 & 1-i \\ 1-i & -1 & 1+i & -1 \end{pmatrix}. \quad (7)$$

The time-evolution of this Hamiltonian $U(a,b) = \exp(-ih_{fa}t_{\text{eff}}(a,b,c_{\text{in}}))$ shows that the solution of the full-adder can be extracted on the timings given by $t_{\text{eff}}(a,b,c_{\text{in}}) = \frac{1}{4}\pi(a+b+c_{\text{in}})$, see Fig. 2(b).

Applying the full-adder gate we receive the corresponding truth table

$$\begin{aligned} U(0,0,0)\,|00\rangle &= |00\rangle \quad U(0,0,1)\,|00\rangle = |01\rangle \\ U(0,1,0)\,|00\rangle &= |01\rangle \quad U(0,1,1)\,|00\rangle = |10\rangle \\ U(1,0,0)\,|00\rangle &= |01\rangle \quad U(1,0,1)\,|00\rangle = |10\rangle \\ U(1,1,0)\,|00\rangle &= |10\rangle \quad U(1,1,1)\,|00\rangle = |11\rangle\,. \end{aligned} \quad (8)$$

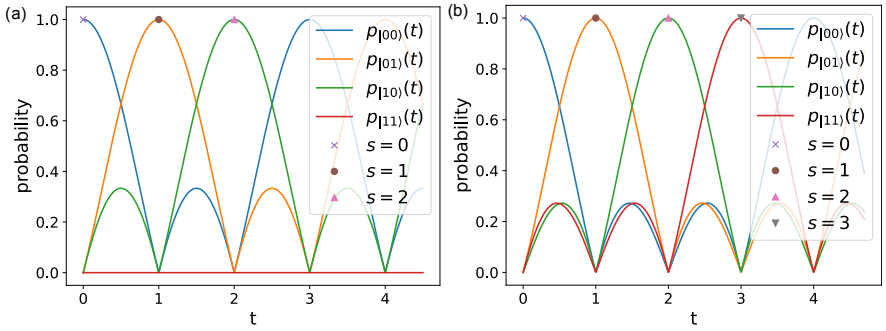


Fig. 2: Time simulation of probability density $p_{|\cdot\rangle} = |\langle\cdot|\exp(-iht_{\text{eff}})\,|00\rangle|^2$. (a) Simulation of half-adder Hamiltonian h_{ha} with $t_{\text{eff}} = \frac{2\pi t}{3\sqrt{3}}$. (b) Simulations of full-adder Hamiltonian h_{fa} and $t_{\text{eff}} = \frac{1}{4}\pi t$. Markers indicate the discrete steps provided by integer $s = a+b$ for the half-adder and $s = a+b+c_{\text{in}}$ for the full-adder.

Since the half-adder and full-adder unitaries correspond to reversible affine Boolean transformations on inputs (a,b) and (a,b,c_{in}), respectively, they can be implemented efficiently using Clifford operations. The optimal circuit design is $U = \big(\text{CNOT}(q_0 \to q_1);\ \text{X}(q_0)\big)^s$, which applies to both the half- and full-adder. It requires s CNOT and X gates, and must be reset after two iterations for the half-adder and after three iterations for the full-adder.

2.1 Hardware Realizations

Given the Hamiltonian formulations of h_{ha} and h_{fa}, the proposed gates can, in principle, be implemented on any hardware supporting universal quantum operations. Notably, h_{fa} has already been theoretically realized in a particle-conserving Rydberg system on a triangular grid [26], originally for quantum routing. This demonstrates how Rydberg platforms can directly realize complex-valued Hamiltonians across different lattice geometries [18, 24, 25].

Beyond Rydberg implementations, photonic and hybrid quantum–photonic platforms provide further motivation for Boolean logic. Besides aforementioned energy benefits of reversible logic [4,7,17] polarization-based encoding avoids wire delays and resistive losses, enabling massively parallel Boolean operations [6,15], while plasmonic optical memristive memory (OMM) allows local storage directly at the processing site, mitigating the von Neumann bottleneck [12]. Photonic hardware also offers substantial speed advantages: 10–100+ GHz electro-optic bandwidth and sub- ps all-optical switching, combined with wavelength and polarization multiplexing, yield high-throughput, low-latency control crucial for Measurement Based Quantum Computing (MBQC) and Quantum Error Correction (QEC) [27]. Since hybrid quantum algorithms demand heavy classical pre- and post-processing, implementing this functionality on the same photonic substrate eliminates costly conversion overhead and enables efficient quantum–classical co-processing [3, 19]. Experimental results already support this direction: plasmonic cavities have demonstrated stable Boolean operations via hot-electron dynamics over 1.6×10^5 iterations, confirming the physical feasibility of these QHC principles [9, 15].

3 Scalability, use cases, and limitations

Compared to textbook ripple-carry designs, these blocks eliminate scratch qubits by moving input embedding into classical controls. In settings where the inputs are classical, these unitaries correspond to Clifford operations, see end of Sec. 2, and can be implemented exactly without T gates or approximation overhead. Our two-qubit blocks are *classically controlled* state-preparation primitives: given classical inputs $(a, b, c_{\text{in}}) \in \{0, 1\}$, they prepare $|\text{CARRY}, \text{SUM}\rangle$ from $|00\rangle$. While these circuits efficiently implement truth tables—which might suggest applications in Quantum Random Access Memory (QRAM) [14] or Grover oracles [23]—they are **not** coherent adders. For coherent, superposition-valued inputs promoting (a, b, c_{in}) to quantum controls reintroduces ancillas and control overhead, recovering the usual trade-offs of coherent adders with the benefit that input and output registers are nicely separated.

Since this modification reintroduces ancilla qubits and control depth, thereby recovering the standard resource requirements of coherent quantum adders [8,32] the here presented circuits are best suited for applications where classical control suffices: hybrid classical-to-quantum interfaces [19] and configurable photonic/analog control systems where minimizing logical qubit count is paramount [5].

For multi-bit *classical* addition, a two-qubit accumulator can be rippled by applying the appropriate instance of U per bit slice with classical feed-forward of the carry, preserving the low logical footprint while keeping latency linear in the word size. By design, Eqs. (1)–(8) prepare outputs from $|00\rangle$ under classical control;

Our design is thus complementary to coherent ripple-carry or QFT-based adders rather than a replacement. The same construction also applies to other 2-bit truth-tables such as comparators or multiplexers, and in programmable photonics to configurable 2×2 blocks within larger meshes, while extensions to multipliers or modular arithmetic within a coherent algorithm likely require additional qubits and fall outside the present scope. Finally, we want to mention, that promoting (a, b, c_{in}) to non-integer values allows to construct superpositions again, see Fig. 2(a-b), furthermore, it allows a quantum unique path to differentiable logic [20].

4 Conclusion

In this study, we presented Quantum Hamiltonian Computing as a method for performing quantum arithmetic operations, which reduces qubit overhead by encoding classical inputs within a single quantum gate. Our two-qubit half- and full-adder circuits in a 4×4 Hilbert space require fewer resources than traditional textbook reversible design built from a Toffoli plus clean-up gates [2, 8, 30, 32]. Leveraging unitary evolution enables us to bypass classical CMOS energy limitations and achieve reversible logic with fewer qubits and gates, where the qubit count is only logarithmic in dependence on possible output states. Furthermore, the QHC unitary realizes the complete truth-table in one analogue pulse, as long as the gate can be implemented on the corresponding quantum hardware. While the current design targets classical logic without utilizing superposition of input and output states, it should be clear that this is possible in principle and can be as simple as converting the inputs from binary to real value inputs. Nevertheless, the found unitaries are not fully expressive but can only map a sub-manifold in the Hilbert space. Removing this limitation, would lead again to scratch qubits thus no benefit to previous implementations [32]. However, the sub-manifold can be optimized to evaluate to superposition states expected to become most prominent in the circuit without the need to introduce qubits.

We have shown that Quantum Hamiltonian Computing enables two-qubit implementations of the half- and full-adder truth tables (Eqs. (1)–(8)) with no scratch registers. Because the inputs are binary, the unitaries reduce to a finite set of two-qubit permutations that admit exact Clifford decompositions, eliminating T gates and approximation overhead. This makes the blocks attractive for hybrid classical-to-quantum interfaces and for FPGA-like quantum-configurable fabrics where the goal is to minimize logical qubits and depth.

This approach offers potential applications in quantum circuits, integrated quantum photonics, and other emerging technologies, especially as the opera-

tional speeds of integrated quantum photonic devices, for example, can exceed those of classical computers [16].

Acknowledgments

We would like to thank the Research Initiative 'Quantum Computing for Artificial Intelligence' (QC-AI) for their support.

5 Appendix: Full-adder unitary matrix $U(a, c_{\text{in}}, b)$ construction

Let the computational basis of two qubits be ordered as $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. We could fix the column vector (the *control* or *eigenvector*) as:

$$V \;=\; |00\rangle\,, \tag{A.1}$$

and three real parameters $a, b, c_{\text{in}} \in \mathbb{R}$. The task is to build a 4×4 unitary matrix $U(a, b, c_{\text{in}})$ such that, when each parameter is restricted to the Boolean set $\{0, 1\}$, the eight input triples are mapped as

(a,c_{in},b)	$U(a, c_{\text{in}}, b)\, V$
$(0, 0, 0)$	$\|00\rangle$
$(0, 0, 1)$	$\|01\rangle$
$(0, 1, 0)$	$\|01\rangle$
$(0, 1, 1)$	$\|10\rangle$
$(1, 0, 0)$	$\|01\rangle$
$(1, 0, 1)$	$\|10\rangle$
$(1, 1, 0)$	$\|10\rangle$
$(1, 1, 1)$	$\|11\rangle$

We define now a 4-cycle permutation matrix $\mathcal{R}$ as

$$\mathcal{R} \;=\; |01\rangle\langle 00| \;+\; |10\rangle\langle 01| \;+\; |11\rangle\langle 10| \;+\; |00\rangle\langle 11|\,. \tag{A.2}$$

Equation (A.2) permutes the basis cyclically $|00\rangle \to |01\rangle \to |10\rangle \to |11\rangle \to |00\rangle$; hence $\mathcal{R}^4 = I_4$. Written as a matrix,

$$\mathcal{R} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathcal{R}^\dagger \mathcal{R} = I_4.$$

Because $\mathcal{R}$ is unitary, the principal matrix logarithm

$$\mathcal{R} = e^{-iH}, \qquad H \;=\; i \log \mathcal{R}, \qquad H^\dagger = H, \tag{A.3}$$

is Hermitian. (The eigenvalues of $\mathcal{R}$ are the fourth roots of unity $\{1, i, -1, -i\}$.)

the three–parameter family could be Defined as

$$U(a, b, c_{\text{in}}) \;=\; \exp\!\big[-\,i\,(a + b + c_{\text{in}})\,H\big] \tag{A.4}$$

Some algebra leads to Eq. (5).

References

1. Ample, F., Faizy, O., Kawai, H., Joachim, C.: The design of a surface atomic scale logic gate with molecular latch inputs. In: Kolmer, M., Joachim, C. (eds.) On-Surface Atomic Wires and Logic Gates. pp. 139–155. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-51847-3_9
2. Amy, M., Maslov, D., Mosca, M., Roetteler, M.: A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **32**(6), 818–830 (2013). https://doi.org/10.1109/TCAD.2013.2244643
3. Arrazola, J. M., e.a., Bergholm, V., Brádler, K., et al.: Quantum circuits with many photons on a programmable nanophotonic chip. Nature **591**(7848), 54–60 (Mar 2021). https://doi.org/10.1038/s41586-021-03202-1
4. Bennett, C.H.: Logical reversibility of computation. IBM journal of Research and Development **17**(6), 525–532 (1973). https://doi.org/10.1147/rd.176.0525
5. Carolan, J., Harrold, C., Sparrow, C., et al.: Universal linear optics. Science **349**(6249), 711–716 (2015). https://doi.org/10.1126/science.aab3642
6. Caulfield, H.J., Dolev, S.: Why future supercomputing requires optics. Nature Photonics **4**(5), 261–263 (2010). https://doi.org/10.1038/nphoton.2010.94
7. Chiribella, G., Yang, Y., Renner, R.: Fundamental energy requirement of reversible quantum operations. Phys. Rev. X **11**, 021014 (Apr 2021). https://doi.org/10.1103/PhysRevX.11.021014
8. Cuccaro, S.A., Draper, T.G., Kutin, S.A., Moulton, D.P.: A new quantum ripple-carry addition circuit. arXiv quant-ph/0410184 (2004). https://doi.org/10.48550/arXiv.quant-ph/0410184
9. Dell'Ova, F., Brûlé, Y., Gros, N., et al.: Compact implementation of a 1-bit adder by coherent 2-beam excitation of a single plasmonic cavity. ACS Photonics **11**(2), 752–761 (2024). https://doi.org/10.1021/acsphotonics.3c01624
10. Dridi, G., Julien, R., Hliwa, M., Joachim, C.: The mathematics of a quantum hamiltonian computing half adder boolean logic gate. Nanotechnology **26**(34), 344003 (aug 2015). https://doi.org/10.1088/0957-4484/26/34/344003
11. Dridi, G., Namarvar, O.F., Joachim, C.: Qubits and quantum hamiltonian computing performances for operating a digital boolean 1/2-adder. Quantum Science and Technology **3**(2), 025005 (2018). https://doi.org/10.1088/2058-9565/aaa98b
12. Feldmann, J., Youngblood, N., Wright, C.D., Bhaskaran, H., Pernice, W.H.P.: All-optical spiking neurosynaptic networks with self-learning capabilities. Nature **569**(7755), 208–214 (2019). https://doi.org/10.1038/s41586-019-1157-8
13. Gidney, C.: Halving the cost of quantum addition. Quantum **2**, 74 (2018). https://doi.org/10.22331/q-2018-06-18-74
14. Giovannetti, V., Lloyd, S., Maccone, L.: Quantum random access memory. Phys. Rev. Lett. **100**, 160501 (Apr 2008). https://doi.org/10.1103/PhysRevLett.100.160501
15. Kumar, U., Cuche, A., Girard, C., et al.: Interconnect-free multibit arithmetic and logic unit in a single reconfigurable 3 μm^2 plasmonic cavity. ACS Nano **15**(8), 13351–13359 (2021). https://doi.org/10.1021/acsnano.1c03196
16. Labonté, L., Alibart, O., D'Auria, V., et al.: Integrated photonics for quantum communications and metrology. PRX Quantum **5**, 010101 (Feb 2024). https://doi.org/10.1103/PRXQuantum.5.010101
17. Landauer, R.: Irreversibility and heat generation in the computing process. IBM Journal of Research and Development **5**(3), 183–191 (1961). https://doi.org/10.1147/rd.53.0183

18. Lienhard, V., Scholl, P., Weber, S., et al.: Realization of a density-dependent peierls phase in a synthetic, spin-orbit coupled rydberg system. Phys. Rev. X **10**, 021031 (May 2020). https://doi.org/10.1103/PhysRevX.10.021031
19. McClean, J.R., Romero, J., Babbush, R., Aspuru-Guzik, A.: The theory of variational hybrid quantum-classical algorithms. New Journal of Physics **18**(2), 023023 (2016). https://doi.org/10.1088/1367-2630/18/2/023023
20. Miotti, P., Niklasson, E., Randazzo, E., Mordvintsev, A.: Differentiable logic cellular automata: From game of life to pattern generation (2025), https://arxiv.org/abs/2506.04912
21. Moutinho, J.a.P., Pezzutto, M., Pratapsi, S.S., et al.: Quantum dynamics for energetic advantage in a charge-based classical full adder. PRX Energy **2**, 033002 (Jul 2023). https://doi.org/10.1103/PRXEnergy.2.033002
22. Namarvar, O.F., Giraud, O., Georgeot, B., Joachim, C.: Quantum hamiltonian computing protocols for molecular electronics boolean logic gates. Quantum Science and Technology **4**(3), 035009 (2019). https://doi.org/10.1088/2058-9565/ab2412
23. Nielsen, M.A., Chuang, I.L.: Quantum computation and quantum information (2010). https://doi.org/10.1017/CBO9780511976667
24. Ohler, S., Kiefer-Emmanouilidis, M., Browaeys, A., Büchler, H.P., Fleischhauer, M.: Self-generated quantum gauge fields in arrays of rydberg atoms. New Journal of Physics **24**(2), 023017 (feb 2022). https://doi.org/10.1088/1367-2630/ac4a15
25. Ohler, S., Kiefer-Emmanouilidis, M., Fleischhauer, M.: Quantum spin liquids of rydberg excitations in a honeycomb lattice induced by density-dependent peierls phases. Phys. Rev. Res. **5**, 013157 (Feb 2023). https://doi.org/10.1103/PhysRevResearch.5.013157
26. Palaiodimopoulos, N.E., Ohler, S., Fleischhauer, M., Petrosyan, D.: Chiral quantum router with rydberg atoms. Phys. Rev. A **109**, 032622 (Mar 2024). https://doi.org/10.1103/PhysRevA.109.032622
27. Pan, B., Hu, J., Huang, Y., et al.: Demonstration of high-speed thin-film lithium-niobate-on-insulator optical modulators at the $2-\mu m$ wavelength. Opt. Express **29**(12), 17710–17717 (Jun 2021). https://doi.org/10.1364/OE.416908
28. Remaud, M.: Optimizing t and cnot gates in quantum ripple-carry adders and comparators. In: Proceedings of Recent Advances in Quantum Computing and Technology. p. 56–61. ReAQCT '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3665870.3665875
29. Remaud, M., Vandaele, V.: Ancilla-free quantum adder with sublinear depth. In: Glück, R., Kaarsgaard, R. (eds.) Reversible Computation. pp. 137–154. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-97063-4_11
30. Selinger, P.: Quantum circuits of t-depth one. Phys. Rev. A **87**, 042302 (Apr 2013). https://doi.org/10.1103/PhysRevA.87.042302
31. Srivastava, S., Kino, H., Joachim, C.: Quantum half-adder boolean logic gate with a nano-graphene molecule and graphene nano-electrodes. Chemical Physics Letters **667**, 301–306 (2017). https://doi.org/https://doi.org/10.1016/j.cplett.2016.11.009
32. Vedral, V., Barenco, A., Ekert, A.: Quantum networks for elementary arithmetic operations. Phys. Rev. A **54**, 147–153 (Jul 1996). https://doi.org/10.1103/PhysRevA.54.147

Optimizing Qubit Routing with Bridge Gates: Extending Quantum Circuit Efficiency Across Arbitrary Distances

Ward van der Schoot[1], Willem de Kok[1] and Frank Phillipson[1,2]

[1] TNO, The Hague/Groningen, The Netherlands
[2] Maastricht University, Maastricht, The Netherlands
f.phillipson@maastrichtuniversity.nl

Abstract. Qubit routing is a critical challenge in quantum computing, essential for implementing quantum circuits on hardware with limited connectivity. This paper introduces a novel perspective on qubit routing by exploring the use of bridge gates, which enable the execution of controlled NOT (CNOT) operations over non-adjacent qubits without rerouting the qubits. The study highlights the advantages of bridge gates compared to SWAP gates, particularly their ability to preserve qubit assignments and potentially optimize subsequent routing steps.
We propose an extension to the concept of bridge gates by generalizing them for arbitrary distances and provide constructions demonstrating their feasibility. By analysing their performance, we show that larger-distance bridge gates can significantly reduce the number of CNOT gates in certain circuits. Furthermore, a new qubit routing problem is defined, incorporating both SWAP and bridge gates, and we discuss how this impacts the complexity of the problem.

Keywords: Qubit Routing · Bridge Gates · Quantum Circuits · Quantum Gate Optimization

1 Introduction

Quantum computing is a revolutionary approach to computation that leverages the principles of quantum mechanics to process information in fundamentally new ways. Unlike classical computers, which rely on bits as their basic units of information, quantum computers use qubits, which can exist in multiple states simultaneously due to quantum superposition. This unique property, along with entanglement and interference, enables quantum computers to solve certain types of problems significantly faster than classical computers [1].

Its broad set of potential applications, including quantum chemistry and drug discovery [2, 3], machine learning [4, 5], optimization [6, 7], and simulation [8, 9], has fostered research efforts and corporate investment [10, 11]. Though some experiments have already hinted at the possibility of quantum supremacy [12, 13], substantial research efforts are needed before quantum computing can be employed in real-world settings.

Quantum algorithms are at the core of quantum computing, designed to harness these quantum properties to perform computations that would be infeasible for classical algorithms. Some well-known quantum algorithms include Shor's algorithm [14] for factoring large integers and Grover's algorithm [15] for database searching, both of which offer substantial speed-ups over their classical counterparts. To implement these algorithms on quantum hardware, they are typically expressed as quantum circuits, where qubits are manipulated through a series of quantum gates. Quantum circuits visually represent the sequence of operations that manipulate qubits to achieve the desired computational outcome, making them a crucial bridge between high-level quantum algorithms and the physical processes within a quantum computer.

When programming quantum algorithms, quantum circuits are often written in a generic way. To run a quantum circuit on specific hardware, the circuit must be translated to be compatible with the target hardware. This process is called compilation, and involves many different steps. There are different ways in which this process can be performed, but often it consists of at least the following steps, see for example [16, 17]:

1. Decomposition of quantum gates: decompose the gates in the quantum circuits into one-qubit gates and a single type of two-qubit gate, usually the CNOT or CZ gate. In this work, we will assume the CNOT gate, but everything below also holds for the CZ gate.
2. Initial qubit mapping: map the qubits of the quantum circuits to the qubits on the hardware.
3. Qubit routing: explained in more detail below.
4. Native gate decomposition: decompose the current quantum gates into gates that can be natively implemented on the quantum hardware

After this process, hardware-specific instructions are produced and the required quantum circuit can be run on the target hardware. The process of compilation can be seen as an optimisation process, in which the optimal translation to hardware-specific instruction is sought. Often, this optimisation process is not easy, and it can actually be shown to be as hard as an NP-hard problem. The hardness of this optimisation process mainly results from the step of qubit routing.

Qubit routing entails the process of changing the qubit order between different quantum gates within a circuit. After the step of qubit mapping, the circuit consists of arbitrary one-qubit and a single type of two-qubit gate, often the CNOT gate. While one-qubit gates (after further decomposition in step 4 above) can directly be applied to the corresponding qubit, this is not the case for two-qubit gates. Two-qubit gates can only be applied to hardware qubits that are directly adjacent within the target hardware architecture. However, as two-qubit gates between arbitrary qubits could be present in the circuit description, this yields a problem in performing these gates on hardware.

This problem is currently tackled by changing the mapping between the circuit and hardware qubits in between two-qubit gates, so that required hardware qubits are adjacent when performing the gate. This is done by inserting SWAP

gates between adjacent qubits. A SWAP gate swaps the mapping of two adjacent qubits and is built from 3 single CNOT gates, as depicted in Fig. 1. The process of remapping using SWAP gates is called qubit routing. If the quantum circuit is seen as a series of quantum gates, there is a step of qubit routing before every two-qubit gate within the circuit consisting of two non-adjacent qubits. SWAP gates are relatively expensive gates, in the sense that for current quantum hardware, the CNOT gate is the native gate with the highest fidelity error rate. For this reason, the main goal in qubit routing often is to minimise the number of SWAP gates. Formally, the following problem is considered: *Given a quantum circuit, a target qubit architecture and an initial mapping of the qubits, what is the minimal number of SWAP gates required to run the quantum circuit on the target hardware?* In literature, people worked on approaches using exact methods [18, 19], heuristics approaches [20–22], machine learning approaches such as reinforcement learning [23–25] and even quantum annealing approaches [26] to solve this problem. While these can give good or even optimal results for certain configurations, they do not offer optimal solutions for general cases (at least not in reasonable time).

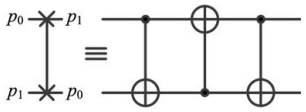


Fig. 1: The basic SWAP gate consisting of three CNOT gates.

However, the SWAP gate is not the only solution to enable the qubit routing. Also the so-called bridge gate can be used. In this paper, we investigate in what sense the qubit routing problem changes by using bridge gates and how it will influence the performance of the circuit. This work also proposes the notion of an arbitrary distance bridge gate, which has not been done before.

2 Bridge Gates

2.1 The Bridge Gate

A bridge gate is a gate that allows the operation of a CNOT gate between two qubits that have a common neighbouring qubit in the hardware lay-out. It is built out of four standard CNOT gates, as depicted in Fig. 2. These four CNOT gates together are exactly the operation of a CNOT from qubit 0 to qubit 2, without changing the qubit assignment. Both bridge gates and SWAP gates can be used to perform a CNOT between two qubits with a common neighbouring qubit. Bridge gates do this directly, while SWAP gates can do this by swapping the middle qubit with one of the two outer qubits, and performing a CNOT between the two required qubits, which are now adjacent. Both methods require the same

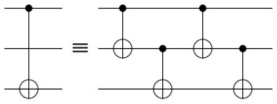


Fig. 2: A basic bridge gate consisting of four CNOT gates is equivalent to a distance 2 CNOT gate.

amount of four CNOT gates, and as the main cost of (currently) implementing quantum circuits comes from the number of CNOT gates to be implemented, the two gates are equally expensive. However, the difference lies in the qubit assignment after the operations: by using a SWAP gate, the qubit assignment is changed after the CNOT gate, while this is not the case when using a bridge gate. This means we have non-equivalent circuits.

As an example, consider a linear qubit assignment of three qubits numbered 0 to 2 from left to right, and a gate that needs to be executed between qubit 0 and 2. With a SWAP, this can be done by swapping 0 and 1 (costs three CNOT gates) and then performing a CNOT between 0 and 2 (total four CNOT gates), resulting in the qubit assignment 102. With a bridge gate, a CNOT can be directly executed between 0 and 2 (costs four CNOT gates), resulting in the qubit assignment 012. Using a bridge gate requires the same amount of CNOT gates, but the qubit assignment is unchanged after the full process, while this is not the case when using just SWAP gates.

As the qubit assignment influences the amount of SWAP (or bridge) gates in later qubit routing steps, both the SWAP and bridge gate options could be beneficial. In summary, the use of bridge gates in qubit routing could improve performance.

2.2 Bridge Gates over larger distances

In literature, such as [24], the only bridge gates that are being used, are used over distance two, i.e., with one neighbouring qubit in between the two qubits. However, in quantum algorithms, often gates need to be applied between qubits over larger distances. That is why this work introduces the notion of bridge gates over arbitrary distances. Both the construction of bridge gates over arbitrary distances, as well as their use in qubit routing, are introduced. Different constructions were considered, which are discussed below. It is important to note that this work does not claim this to be the most efficient way to perform bridge gates (in terms of number of CNOT gates). It presents the bridge gate as an alternative solution for re-routing that has advantages in some cases.

One way to construct bridge gates over larger distances, is to use the bridge gate over distance two as a building block. As an example, to construct a bridge gate over distance three, two distance-two CNOT gates and two simple CNOT gates are sufficient by following the original bridge gate building block, as shown in Fig. 3. Then, by replacing the distance-two CNOT gates by distance-two

bridge gates, the distance-three bridge gate can be constructed. This can also be done for larger distance bridge gates, as seen in Fig. 4. In the end, this gives a roughly quadratic number of CNOT gates in terms of the distance.

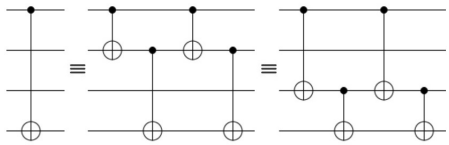


Fig. 3: Two ways of constructing a distance-three bridge gate by using a distance-two bridge gate architecture. The full bridge gate can be constructed by replacing the distance-two CNOT gates by distance-two bridge gates.

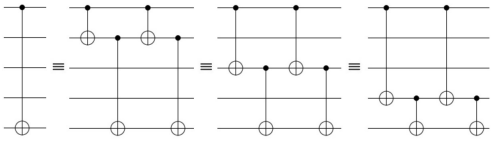


Fig. 4: Three ways of constructing a distance-four bridge gate by using a distance-two bridge gate architecture. The full bridge gate can be constructed by replacing the distance-two or -three CNOT gates by distance-two or -three bridge gates.

Next, another construction for bridge gates was constructed, as depicted in Fig. 5, specifically for a distance-five CNOT gates. This construction works for arbitrary distance CNOT gates. This yields a number of $2n + 2(n-1) = 4n-4$ CNOT gates for a distance n bridge gate.

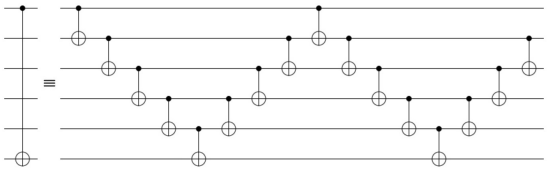


Fig. 5: A general construction for bridge gates. As an example, the construction is shown for distance-five, but the construction generalises to any distance CNOT gate.

Alternatively, a CNOT gate between two qubits at distance n can be applied by applying $n-1$ SWAP gates and one CNOT gate, as shown in figure 6 This

only requires $3(n-1)+1 = 3n-2$ CNOT gates. While this is less, it could still be beneficial to apply the distance n bridge gate, as this qubit assignment could be beneficial for later qubit routing steps. Notice in particular that it is worse to perform SWAP gates, perform the CNOT gate, and then perform the same SWAP gates to reach the same qubit ordering as before the gate, which would yield the same effect as the distance n bridge gate.

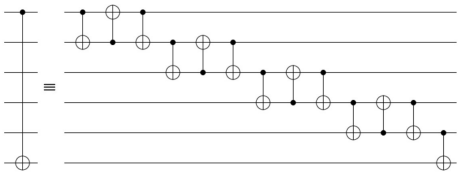


Fig. 6: A distance-five CNOT gate implemented by using only SWAP gates.

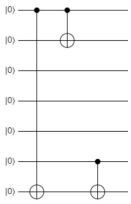


Fig. 7: A circuit showing the advantage of large distance bridge gates. Note that with large distance bridge gates, no SWAPs are required to perform this circuit.

To give an example as to why using a larger distance bridge gate could lead to less CNOT gates used, consider a linear qubit assignment of seven qubits numbered 1 to 7 from left to right. Using these qubits, a circuit is ran with three CNOT gates, the first one being between qubits 1 and 7, the second between qubits 1 and 2 and the third between qubits 6 and 7. The circuit is depicted in figure 7 Using large-distance bridge gates, this can be done with one distance-6 bridge gate (costs 20 CNOT gates) and then 2 CNOT gates, so 22 CNOT gates in total.

Consider now how many CNOT gates are required when using only the original distance-2 bridge gates and SWAP gates. First, the distance between qubits 1 and 7 needs to be reduced to two, which already costs at least 4 SWAPs, so $4 \times 3 = 12$ CNOT gates. After this, a bridge gate or another SWAP and a CNOT is required, bringing the total to 16 CNOT gates. After this, the qubit assignment is one of the following eleven forms: 1273456, 1723456, 2137456, 2173456, 2314756, 2317456, 2341576, 2341756, 2345167, 2345176, 2345617. Note that forms 1, 4, 9 and 10 require a single CNOT gate, and a distance-4 CNOT

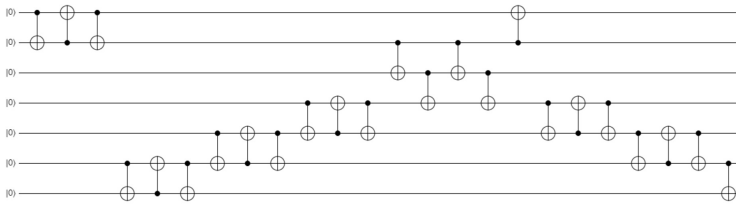


Fig. 8: The circuit in figure 7 implemented without using large distance bridge gates with a minimal number of 24 CNOT gates. First, qubits 1 and 7 are swapped to reach the qubits assignment 2137456, after which a distance-2 bridge gate is applied. Then, qubit 7 is swapped back to neighbour qubit 6 and to apply the final 2 CNOT gates.

gate. For the latter, three SWAPs and a single CNOT gate are optimal, requiring an additional 10 qubits. This yields a total of 27 CNOT gates. Equivalently, forms 3 and 7 require a single CNOT gate and a distance-3 CNOT gate, for which a total of 8 extra CNOT gates are required, yielding 24 CNOT gates in total. Form 5 requires two distance-2 CNOT gates, which also requires 8 CNOT gates, yielding again a total of 24 CNOT gates. Forms 6 and 8 require a distance-2 CNOT gate and a distance-3 CNOT gate, which needs 11 CNOT gates in total, yielding a total of 27 CNOT gates. Lastly, forms 2 and 11 require at least a distance-5 CNOT gate, which needs 4 SWAPs, so 12 CNOT gates, plus a single CNOT gate, bringing the total to 29 CNOT gates. All in all, it can be seen that by using only SWAP gates and distance-2 CNOT gates, at least 24 CNOT gates are required to perform the given circuit. An example of how this can be done with 24 CNOT gates, can be seen in figure 8. This is worse than what the larger distance bridge gate can achieve, which shows the advantage of using larger distance bridge gates.

2.3 New qubit routing problem

The addition of bridge gates introduces a different qubit routing problem, which is the same as before, but now allowing bridge gates as well. Note that this version of qubit routing has been proposed before, and is even used in some state-of-the-art compilers, such as the Quantinuum's compiler, as detailed in [27].

Problem (qubit routing with bridge gates): Consider a quantum circuit, a target qubit architecture and an initial mapping of the qubits. For every two-qubit gate, the gate can be implemented by using a combination of SWAP gates to change the qubit mapping and bridge gates over arbitrary distances. What is the minimal number of CNOT gates required to run the quantum circuit on the target hardware?

Note that for a single two-qubit gate, also the combination of SWAP gates and a bridge gate is allowed. Again, swapping only some of the distance between the

qubits could lead to a qubit mapping that is beneficial for later qubit routing steps. In particular, both the control and target qubit could be swapped some places before applying a bridge gate over a certain distance. This yields many different options for each single step of qubit routing.

3 Conclusion

This work considered the problem of qubit routing, and the use of bridge gates for this. These gates allow the implementation of CNOT gates over arbitrary distances, without using SWAP gates or changing the qubit mapping. This work is the first to construct bridge gates over larger distances. In addition, it is the first to define the problem of bridge gates over larger distances. Although no experiments with these bridge gates were performed in this work, it is easy to construct circuits in which bridge gates pose a considerable advantage over using just SWAP gates.

For further research, it would be particularly interesting to consider the extent of the impact of this larger distance bridge gate. Potential research questions are:

1. In which way can the qubit routing problem, including SWAP gates and bridge gates, be written as a mathematical formulation, to allow proving properties such as NP-hardness?
2. To which extent can bridge gates improve the performance of qubit routing, and which factors impact this performance increase? Potential factors are hardware architecture or circuit architecture.

References

1. M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.
2. G. Kumar, S. Yadav, A. Mukherjee, V. Hassija, and M. Guizani, "Recent advances in quantum computing for drug discovery and development," *IEEE Access*, 2024.
3. G. Mazzola, "Quantum computing for chemistry and physics applications from a monte carlo perspective," *The Journal of Chemical Physics*, vol. 160, no. 1, 2024.
4. D. Peral-García, J. Cruz-Benito, and F. J. García-Peñalvo, "Systematic literature review: Quantum machine learning and its applications," *Computer Science Review*, vol. 51, p. 100619, 2024.
5. F. Phillipson, "Quantum machine learning: Benefits and practical examples." in *QANSWER*, 2020, pp. 51–56.
6. A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, B. G. Elmegreen *et al.*, "Challenges and opportunities in quantum optimization," *Nature Reviews Physics*, pp. 1–18, 2024.
7. F. Phillipson, "Quantum computing in logistics and supply chain management-an overview," *arXiv preprint arXiv:2402.17520*, 2024.
8. S. Pal, M. Bhattacharya, S. Dash, S.-S. Lee, and C. Chakraborty, "Future potential of quantum computing and simulations in biological science," *Molecular Biotechnology*, vol. 66, no. 9, pp. 2201–2218, 2024.

9. B. Fauseweh, "Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges," *Nature Communications*, vol. 15, no. 1, p. 2123, 2024.
10. M.-L. How and S.-M. Cheah, "Business renaissance: Opportunities and challenges at the dawn of the quantum computing era," *Businesses*, vol. 3, no. 4, pp. 585–605, 2023.
11. H. Ko and S. Kwon, "Prominence of corporate science in quantum computing research," *Technological Forecasting and Social Change*, vol. 212, p. 123949, 2025.
12. G. Kalai, Y. Rinott, and T. Shoham, "Google's 2019 quantum supremacy claims: Data, documentation, and discussion," *Quantum Physics; Computational Complexity*, vol. 3, pp. 455–463, 2023.
13. R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya *et al.*, "Quantum error correction below the surface code threshold," *arXiv preprint arXiv:2408.13687*, 2024.
14. P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.
15. L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 1996, pp. 212–219.
16. B. Tan, "Layout synthesis for quantum computing," Ph.D. dissertation, University of California, Los Angeles, 2024.
17. A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, B. G. Elmegreen *et al.*, "Quantum optimization: Potential, challenges, and the path forward," *arXiv preprint arXiv:2312.02279*, 2023.
18. P. Zhu, X. Cheng, and Z. Guan, "An exact qubit allocation approach for nisq architectures," *Quantum Information Processing*, vol. 19, no. 11, p. 391, 2020.
19. J. Mulderij, K. I. Aardal, I. Chiscop, and F. Phillipson, "A polynomial size model with implicit swap gate counting for exact qubit reordering," in *International Conference on Computational Science*. Springer, 2023, pp. 72–89.
20. F. Wagner, A. Bärmann, F. Liers, and M. Weissenbäck, "Improving quantum computation by optimized qubit routing," *Journal of Optimization Theory and Applications*, vol. 197, no. 3, pp. 1161–1194, 2023.
21. P. Murali, J. M. Baker, A. Javadi-Abhari, F. T. Chong, and M. Martonosi, "Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers," in *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*, 2019, pp. 1015–1029.
22. M. Steinberg, M. Bandic, S. Szkudlarek, C. G. Almudever, A. Sarkar, and S. Feld, "Resource bounds for quantum circuit mapping via quantum circuit complexity," *arXiv preprint arXiv:2402.00478*, 2024.
23. L. Moro, M. G. Paris, M. Restelli, and E. Prati, "Quantum compiling by deep reinforcement learning," *Communications Physics*, vol. 4, no. 1, p. 178, 2021.
24. G. Pascoal, J. P. Fernandes, and R. Abreu, "Deep reinforcement learning strategies for noise-adaptive qubit routing," in *2024 IEEE International Conference on Quantum Software (QSW)*. IEEE, 2024, pp. 146–156.
25. M. G. Pozzi, S. J. Herbert, A. Sengupta, and R. D. Mullins, "Using reinforcement learning to perform qubit routing in quantum compilers," *ACM Transactions on Quantum Computing*, vol. 3, no. 2, pp. 1–25, 2022.
26. S. Müller and F. Phillipson, "Quantum annealing for nearest neighbour compliance problem," *Scientific Reports*, vol. 14, no. 1, p. 23340, 2024.

27. A. Cowtan, S. Dilkes, R. Duncan, A. Krajenbrink, W. Simmons, and S. Sivarajah, "On the Qubit Routing Problem," in *14th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2019)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), W. van Dam and L. Mančinska, Eds., vol. 135. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, pp. 5:1–5:32.

Resource Estimation for Matrix Inversion via QSVT with the Clifford+T Gate Set

Hiroaki Murakami[1][0009−0009−3974−238X], Kenzo Makino[1][0009−0000−6373−3947], Yasunori Lee[2][0000−0003−0209−3979], Keita Kanno[2][0000−0002−2426−9309], and Tomonori Fukuta[1][0009−0008−3579−6412]

[1] Advanced Technology R&D Center, Mitsubishi Electric Corp, 8-1-1 Tsukaguchi Honmachi, Amagasaki-shi, Hyogo 661-8661, Japan
[2] QunaSys Inc., 1-13-7 Hakusan, Bunkyo-ku, Tokyo 113-0001, Japan

Abstract. Solving linear systems of equations is a fundamental problem that serves as a crucial foundation in numerous fields. Initiated by the pioneering Harrow-Hassidim-Lloyd algorithm, the asymptotic performance of quantum linear system solvers has steadily improved over time. However, there remains a lack of quantitative evaluation of quantum resources based on explicit gate implementations. In this study, we particularly focus on the matrix inversion method via quantum singular value transformation and explicitly construct a quantum circuit to invert specific matrices using the standard universal gate set, Clifford+T gates. We examine the detailed methods for block-encoding, QSP angle finding, and gate decompositions, and numerically evaluate the overall resources required for the quantum circuit.

Keywords: Matrix inversion · Quantum Singular Value Transformation · Clifford+T gate set.

1 Introduction

Solving the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$ is one of the problems where quantum computers are expected to have an advantage over classical computers. The time and space complexity of existing classical algorithms, even for matrices with relatively good properties, is at best $\mathcal{O}(\mathrm{poly}(N))$ with respect to the matrix size $N = \max(N_{\mathrm{row}}, N_{\mathrm{col}})$ [14, 15]. This limits the size of problems that can be addressed by classical algorithms. Owing to the progress in recent quantum algorithm research, initiated by the pioneering work by Harrow et al. [6], an $\mathcal{O}(\mathrm{polylog}(N))$ complexity has been achieved for certain classes of matrices. However, there is still a lack of quantitative evaluation of computational cost based on explicit gate implementations. To bridge this gap, we constructed a quantum circuit for a specific problem using the standard universal gate set, Clifford+T gates, and numerically evaluated its accuracy and computational cost.

2 Backgrounds

Given a signal $x \in [-1, 1]$, there is a known method called Quantum Signal Processing (QSP) that achieves polynomial transformations of x by alternately applying a signal operator $W(x)$ that encodes the signal and a parametric signal processing operator $S(\phi)$;

$$W(x) \coloneqq \begin{pmatrix} x & i\sqrt{1-x^2} \\ i\sqrt{1-x^2} & x \end{pmatrix} = e^{i \arccos(x) X}, \tag{1}$$

$$S(\phi) \coloneqq \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{pmatrix} = e^{i\phi Z}. \tag{2}$$

For any polynomials $P(x), Q(x) \in \mathbb{C}[x]$ that satisfy certain conditions (see [5] for detail), there exists a sequence of angles $\boldsymbol{\Phi} = (\phi_0, \phi_1, \ldots, \phi_d) \in \mathbb{R}^{d+1}$ such that the following equation holds for $x \in [-1, 1]$ [8, 5].

$$S(\phi_0) \prod_{k=1}^{d} W(x) S(\phi_k) = \begin{pmatrix} P(x) & iQ(x)\sqrt{1-x^2} \\ iQ^*(x)\sqrt{1-x^2} & P^*(x) \end{pmatrix} \tag{3}$$

QSP can be extended to multi-qubit system by the method called Quantum Singular Value Transformation (QSVT) [5, 9]. With QSVT, when a matrix A is given in the form of a block-encoding, we can obtain a block-encoding of the matrix whose singular values $\{\sigma_i\}$ are transformed by the real part of the polynomial P.

In particular, if an appropriate polynomial P that approximates the function $1/x$ is given, it is possible to compute an approximate inverse of A by

$$P^{(\mathrm{SV})}(A)^\dagger \simeq W_\sigma \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_N} \end{pmatrix} V_\sigma^\dagger = A^{-1}. \tag{4}$$

Here, W_σ and V_σ are two unitary matrices whose columns are left and right singular vectors of A. Now we can obtain the solution to the linear system of equations as follows:

$$\begin{pmatrix} P^{(\mathrm{SV})}(A) \cdot & \\ \cdot & \cdot \end{pmatrix}^\dagger \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} \simeq \begin{pmatrix} A^{-1}\boldsymbol{b} \\ \cdot \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} \\ \cdot \end{pmatrix}. \tag{5}$$

In the next section, we briefly summarize the methods we adopt for each step of the matrix inversion using QSVT.

3 Methods

3.1 Matrix

In this study, we consider Toeplitz matrices, whose elements along each diagonal are constant. Due to its regular pattern, block-encoding circuit can be efficiently

constructed using rotation gates and adders [13]. Specifically, let N_{data} be the number of non-zero elements in the matrix, then the block-encoding of a Toeplitz matrix can be realized using N_{data} rotation gates and $4\log_2 N$ Toffoli gates. If N_{data} is $\mathcal{O}(\log(N))$, the total number of gates, including other Clifford gates, is $\mathcal{O}(\mathrm{polylog}(N))$.

In this study, we consider a simple matrix

$$A = \begin{pmatrix} 2 & 1 & & \\ & 2 & \ddots & \\ & & \ddots & 1 \\ & & & 2 \end{pmatrix}, \tag{6}$$

and explicitly implement the corresponding block-encoding circuit using the method in [13]. As an example, the block-encoding circuit for $(N = 4)$ is shown in Fig. 1. Here, we use the metod by Möttönen et al. [10] for data loading, and adopt two ripple carry adders [3] to compile the additions appearing in eq. (55) of [13].

We note that the condition number of this specific matrix asymptotically approaches 3 as the matrix size grows.

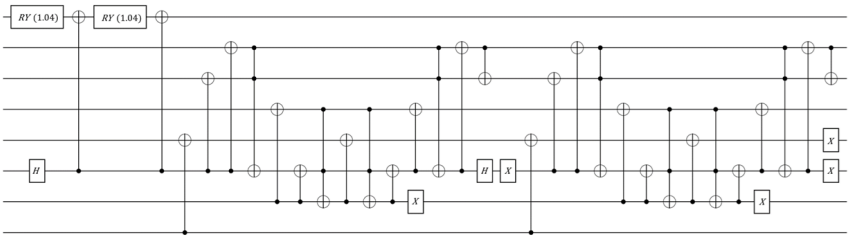


Fig. 1: A quantum circuit for block-encoding of Toeplitz matrix $(N = 4)$.

3.2 Polynomial approximation and angle finding

Several methods have been proposed to obtain a polynomial that approximates the function $1/x$ over a closed interval $D_{\kappa_{\mathrm{QSVT}}} = [1/\kappa_{\mathrm{QSVT}}, 1]$, given the degree d of the polynomial. We utilize the explicit form of the optimal polynomial P and the complementary polynomial Q, proposed by [2]. Since the condition number of the matrix can be considered almost constant as mentioned in Section 3.1, we fix $\kappa_{\mathrm{QSVT}} = 4$ for the following experiments. We then adopt the carving algorithm [5] to compute the corresponding QSP angle sequence $\boldsymbol{\Phi}$.

3.3 Decomposition of non-Clifford gates

To explicitly evaluate the computational cost, we decompose the non-Clifford gates included in the constructed QSVT circuit into the Clifford+T gate set. In

Table 1, we summarize the non-Clifford gates appearing in our implementation. The rotation gates are approximately decomposed into a sequence of H, S and T gates using the Ross-Selinger algorithm [11]. The multi-controlled gates are decomposed using doubly-controlled $\pm iX$ gates, following [12]. Toffoli gates in ripple carry adders [3] used for block-encoding are decomposed into Clifford+T gates via the method in [1].

By adding ancillary qubits, further reductions in T-count can be achieved. For example, the T-count of multi-controlled gates and ripple carry adders can approximately be halved by using "temporary logical-AND" structure (relative phase Toffoli gate) [4]. For data loading, there is a known trade-off between the number of ancillas and T-count [7]. While this paper does not delve into the specifics, these are important considerations for practical implementation.

Table 1: Non-Clifford gates.

Non-Clifford gates	# of appearance	T-count after decomp.	# of ancilla
R_Z for QSP	d	$\sim 3\log_2 \frac{1}{\epsilon_{\text{gate}}}$	0
R_Y for data loading	$N_{\text{data}}d$	$\sim 3\log_2 \frac{1}{\epsilon_{\text{gate}}}$	0
Multi-controlled gate	d	$8(\log_2 N_{\text{data}} + 1)$	$\log_2 N_{\text{data}}$
Adder in block-encoding	$2d$	$14\log_2 N$	0

4 Results

We numerically evaluated the implemented quantum circuit in terms of accuracy and computational cost.

4.1 Accuracy

First, we investigated the dependence of the output accuracy (evaluated by the operator norm $\left\|A^{-1} - P^{(\text{SV})}(A)^\dagger\right\|$) on the degree d of the approximate polynomial and the matrix size $N = 2^n$. As shown in Fig. 2, we confirmed that the output accuracy improves exponentially with respect to d, and is independent of N.

Next, we fixed the matrix size at $N = 8$ and decomposed the non-Clifford gates. Fig. 3 shows the dependency of the output accuracy on the decomposition error ϵ_{gate} of the rotation gates for different d. In the left region of the figure, where ϵ_{gate} is large, the decomposition error of the rotation gates is dominant. On the other hand, in the right region where ϵ_{gate} is sufficiently small, the error due to polynomial approximation (indicated by the dotted lines in the figure) becomes dominant, and the output accuracy of the quantum circuit saturates.

The error due to the decomposition of the rotation gates is considered to be proportional to the product of ϵ_{gate} and the number of rotation gates $(N_{\text{data}}+1)d$. Therefore, when each rotation gate is decomposed with an accuracy of ϵ_{gate},

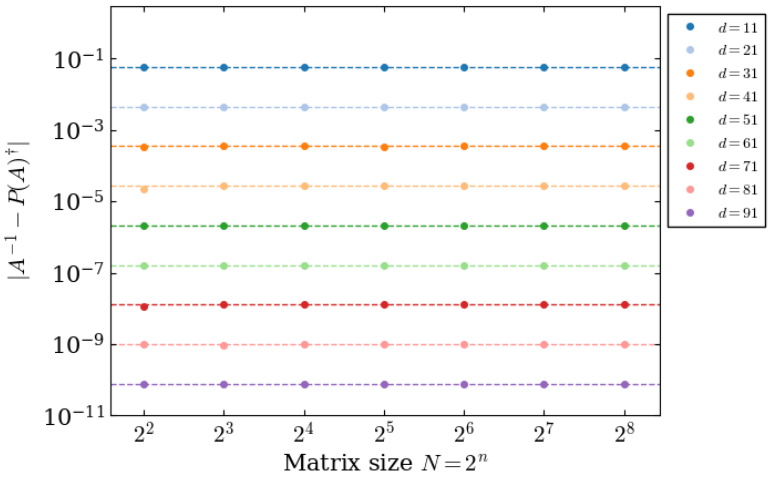


Fig. 2: Dependence of the error on the size of the matrix and the degree of the polynomial.

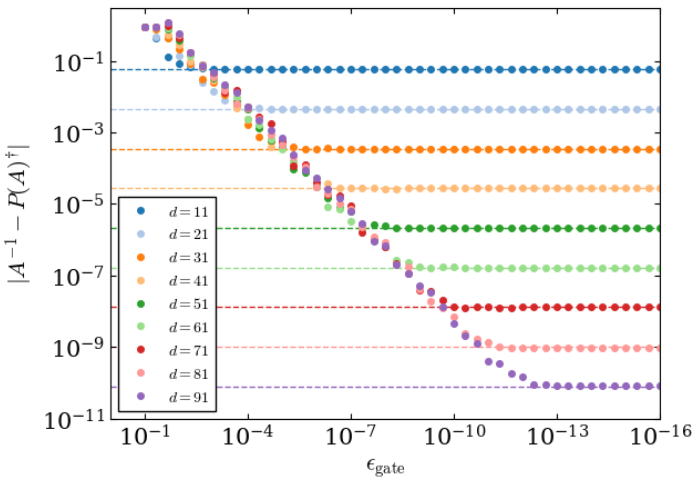


Fig. 3: Dependence of the error on the degree of the polynomial and the precision for decomposition of rotation gates.

and a final output accuracy of ϵ is required, the degree d of the approximate polynomial is determined by the minimum singular value of the scaled matrix A and some proportional constant. This, in turn, allows us to determine the computational cost discussed in the next section.

4.2 Computational cost

We evaluated the computational cost in regards of the number of qubits and the T-count. It was found that the number of logical qubits n_{qubits} is explicitly given as

$$n_{\mathrm{qubits}} = 2\log_2 N + \log_2 N_{\mathrm{data}} + 6, \tag{7}$$

which does not depend on the decomposition error ϵ_{gate} of the rotation gates or the degree d of the approximation polynomial.

Next, we examined the behavior of the T-count arises from the decomposition of non-Clifford gates. As can be seen from Fig. 4, it is confirmed that the T-count with respect to d and ϵ_{gate} matches the expectations outlined in Table 1 (indicated by the dotted lines in the figure).

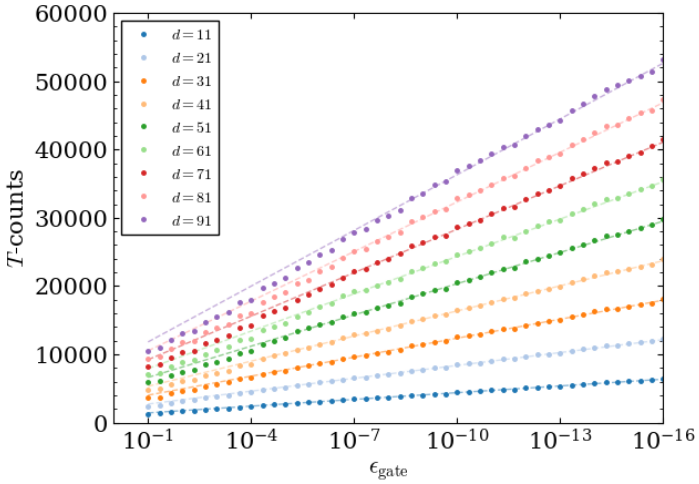


Fig. 4: Dependence of the T-count on the degree of the polynomial and the precision for decomposition of rotation gates.

5 Summary

We explicitly constructed a quantum circuit to compute the inverse of Toeplitz matrices via QSVT with the Clifford+T gate set, and evaluated its accuracy and computational cost. As a result, we confirmed:

(i) The output accuracy of the inverse matrix improves exponentially with respect to the degree d of the apporximate polynomial.
(ii) The output accuracy does not depend on the matrix size N, since the condition number of the matrix is bounded by a constant in our specific choice of matrix.
(iii) The T-count depends on d and the decomposition error of the rotation gates ϵ_{gate} as shown in Table 1, and it depends logarithmically on N through the adder in the block-encoding circuit.

In order to construct an end-to-end quantum linear system solver, the preparation of the input quantum states and the readout of the output states must also be implemented. It is known that both, if implemented naively, may require exponential cost, which would undermine quantum advantage. Addressing this issue remains a challenge for future work.

References

1. Amy, M., Maslov, D., Mosca, M., Roetteler, M.: A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **32**(6), 818–830 (2013). https://doi.org/10.1109/TCAD.2013.2244643
2. Berntson, B.K., Sünderhauf, C.: Two exact quantum signal processing results. In: 2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). pp. 625–626 (2024). https://doi.org/10.1109/ISVLSI61997.2024.00118
3. Cuccaro, S.A., Draper, T.G., Kutin, S.A., Moulton, D.P.: A new quantum ripple-carry addition circuit (2004). https://doi.org/10.48550/arXiv.quant-ph/0410184
4. Gidney, C.: Halving the cost of quantum addition. Quantum **2**, 74 (Jun 2018). https://doi.org/10.22331/q-2018-06-18-74
5. Gilyén, A., Su, Y., Low, G.H., Wiebe, N.: Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. pp. 193–204. STOC 2019, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3313276.3316366
6. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Phys. Rev. Lett. **103**, 150502 (Oct 2009). https://doi.org/10.1103/PhysRevLett.103.150502
7. Low, G.H., Kliuchnikov, V., Schaeffer, L.: Trading T gates for dirty qubits in state preparation and unitary synthesis. Quantum **8**, 1375 (Jun 2024). https://doi.org/10.22331/q-2024-06-17-1375
8. Low, G.H., Yoder, T.J., Chuang, I.L.: Methodology of resonant equiangular composite quantum gates. Phys. Rev. X **6**, 041067 (Dec 2016). https://doi.org/10.1103/PhysRevX.6.041067
9. Martyn, J.M., Rossi, Z.M., Tan, A.K., Chuang, I.L.: Grand unification of quantum algorithms. PRX Quantum **2**, 040203 (Dec 2021). https://doi.org/10.1103/PRXQuantum.2.040203
10. Möttönen, M., Vartiainen, J.J., Bergholm, V., Salomaa, M.M.: Quantum circuits for general multiqubit gates. Physical Review Letters **93**(13) (sep 2004). https://doi.org/10.1103/physrevlett.93.130502

11. Ross, N.J., Selinger, P.: Optimal ancilla-free Clifford+T approximation of z-rotations. Quantum Information and Computation **16**(11&12), 0901–0953 (Sep 2016). https://doi.org/10.26421/QIC16.11-12-1
12. Selinger, P.: Quantum circuits of t-depth one. Phys. Rev. A **87**, 042302 (Apr 2013). https://doi.org/10.1103/PhysRevA.87.042302
13. Sünderhauf, C., Campbell, E., Camps, J.: Block-encoding structured matrices for data input in quantum computing. Quantum **8**, 1226 (Jan 2024). https://doi.org/10.22331/q-2024-01-11-1226
14. Trefethen, L.N., Bau, D.: Numerical linear algebra. Society for Industrial and Applied Mathematics, Philadelphia (1997). https://doi.org/10.1137/1.9780898719574
15. Vorst, H.A.v.d.: Iterative Krylov methods for large linear systems. No. 13 in Cambridge monographs on applied and computational mathematics, Cambridge University Press, Cambridge; New York (2003). https://doi.org/10.1017/CBO9780511615115

Q-PORT: Quantum Portfolio Optimization with Resource-Efficient Encoding and Scalability Analysis

Alberto Marchisio[1,2], Muhammad Umair Hafeez[1,2], Nouhaila Innan[1,2], Muhammad Kashif[1,2], and Muhammad Shafique[1,2]

[1] eBRAIN Lab, Division of Engineering, New York University Abu Dhabi (NYUAD), Abu Dhabi, UAE
[2] Center for Quantum and Topological Systems (CQTS), NYUAD Research Institute, NYUAD, Abu Dhabi, UAE
alberto.marchisio@nyu.edu, muh2008@nyu.edu, nouhaila.innan@nyu.edu, muhammadkashif@nyu.edu, muhammad.shafique@nyu.edu

Abstract. Quantum Approximate Optimization Algorithm (QAOA) and Variational Quantum Eigensolver (VQE) have emerged as promising approaches for solving portfolio optimization tasks. However, the practical scalability of these methods remains a challenge due to the inherent noise and limitations of Noisy Intermediate-Scale Quantum (NISQ) devices. In this paper, we present *Q-PORT* (*Quantum Portfolio Optimization with Resource-efficient Encoding and Scalability Analysis*), a systematic study on the trade-offs between quantum circuit depth, stock encoding strategies, and scalability in quantum portfolio optimization.
We investigate the impact of multi-qubit representations per stock and multi-stock encodings per qubit while varying circuit repetitions and Ansatz types. Our experimental results indicate that increasing qubits per stock offers negligible precision gains compared to classical Mean-Variance Optimization (MVO), while encoding multiple stocks per qubit significantly improves efficiency with minimal precision loss. These findings provide a new pathway toward resource-efficient and scalable quantum portfolio optimization, paving the way for near-term financial applications in quantum computing.

Keywords: Quantum Computing · Portfolio Optimization · Quantum Finance.

1 Introduction

Portfolio optimization remains a cornerstone of quantitative finance, seeking to balance asset allocation to maximize returns while controlling risk [1]. Classical approaches, such as Mean-Variance Optimization (MVO), model this trade-off using expected returns and covariance matrices, with numerous extensions addressing practical constraints [2–4].

Nonetheless, the scalability limitations and computational overhead of classical methods have propelled interest in quantum computing as a next-generation

alternative [5]. Hybrid quantum-classical algorithms like the Quantum Approximate Optimization Algorithm (QAOA) [6] and the Variational Quantum Eigensolver (VQE) [7] have been successfully adapted to portfolio optimization [8–12]. Complementary advances, including quantum annealing for multi-objective portfolios [13], hardware-efficient ansatz design [14], and quasi-binary encoding techniques [15], have recently pushed the state of the art in this domain. However, the majority of these implementations assign a single qubit per stock, creating scalability barriers under the constraints of Noisy Intermediate-Scale Quantum (NISQ) hardware.

This work raises two fundamental research questions: (1) Can multi-qubit representations per stock improve solution fidelity, and do such gains justify the additional resource overhead? (2) Can multi-stock encodings per qubit meaningfully reduce resource requirements while preserving solution quality?

To address these questions, we propose *Q-PORT* (*Quantum Portfolio Optimization with Resource-efficient Encoding and Scalability Analysis*), a methodology to systematically investigate the impact of stock-to-qubit mappings, circuit repetitions, and Ansatz selection on portfolio optimization performance using VQE- and QAOA-based algorithms.

Our contributions are threefold: (i) we develop a comprehensive methodology for evaluating quantum portfolio scalability; (ii) we demonstrate that multi-qubit representations per stock yield negligible precision gains, indicating limited practical benefit; (iii) we show that multi-stock encodings per qubit significantly improve efficiency with minimal precision loss, enabling more scalable quantum portfolio optimization even under NISQ-era limitations.

2 Background and Related Work

2.1 Classical Portfolio Optimization

Classical portfolio optimization, rooted in Markowitz's Modern Portfolio Theory (MPT), aims to allocate assets to maximize expected return for a given level of risk, typically measured by portfolio variance [1]. Extensions to MPT have incorporated various constraints and considerations, such as transaction costs and investor preferences. Brandt et al. proposed a parametric portfolio policy that models portfolio weights as functions of asset characteristics, enhancing scalability for high-dimensional asset spaces [3]. Comprehensive treatments of these methodologies are detailed in the work of Elton et al.'s *Modern Portfolio Theory and Investment Analysis* [4].

2.2 Quantum Optimization Algorithms

Quantum computing introduces novel approaches in financial applications [16–28]. The most commonly used algorithms are the VQE [7] and the QAOA [6]. These hybrid quantum-classical algorithms utilize parameterized quantum circuits optimized via classical routines, effectively navigating complex solution

spaces inherent in Quadratic Unconstrained Binary Optimization (QUBO) formulations.

Recent studies have explored the application of these algorithms to portfolio optimization. Kerenidis et al. [29] developed a quantum algorithm for constrained portfolio optimization, demonstrating potential polynomial speedups over classical counterparts. Buonaiuto et al. [9] conducted experiments using VQE on real quantum devices, analyzing the impact of various hyperparameters on performance. Zaman et al. [11] introduced PO-QA, a framework for portfolio optimization using quantum algorithms, providing insights into circuit configurations and their effects on optimization outcomes.

Despite these advancements, challenges remain regarding scalability and precision, particularly under the constraints of NISQ devices. *Our work addresses these challenges by investigating multi-qubit representations per asset and multi-asset encodings per qubit, aiming to enhance scalability while maintaining solution quality.*

3 Q-PORT Methodology

Our proposed Q-PORT methodology systematically investigates the scalability of quantum portfolio optimization by analyzing various encoding strategies and quantum circuit configurations. The overall methodology, shown in Figure 1, consists of the following steps:

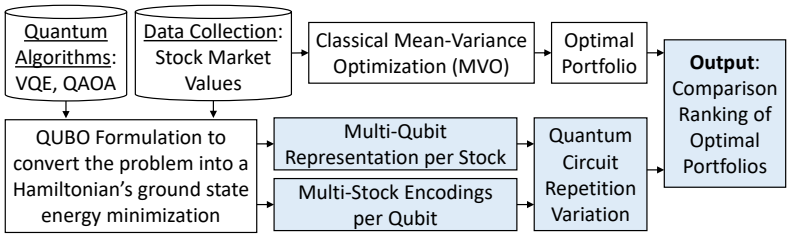


Fig. 1. Overview of our Q-PORT methodology.

– **Data Collection and Classical Benchmarking:** We begin by collecting historical stock market data, including expected returns and covariance matrices for a selected set of assets. As a classical reference, we employ the classical MVO method [1] to compute the optimal asset allocations under the classical framework. This serves as the baseline for evaluating quantum solutions.
– **QUBO Formulation:** The portfolio optimization problem is reformulated into a Quadratic Unconstrained Binary Optimization (QUBO) model, enabling its translation into the ground state energy minimization of a quantum Hamiltonian. This formulation captures both the expected returns and risk

terms, as well as budget and allocation constraints, in a structure amenable to quantum algorithms.

- **Quantum Encoding Strategies:** To explore scalability, we introduce two encoding strategies:
  - **Multi-Qubit Representation per Stock:** In this approach, each stock is mapped to a group of qubits, introducing redundancy that expands the solution space and enables potentially finer-grained portfolio allocations, albeit at higher resource cost. After measurement, individual qubit outcomes are aggregated using majority voting, where the most frequent state across the qubits determines the final allocation for the stock.
  - **Multi-Stock Encoding per Qubit:** In this scheme, multiple stocks are jointly encoded into a single qubit using compressed representations, significantly reducing the total number of qubits required and directly addressing the hardware limitations of NISQ devices. To identify optimal allocations within this compressed space, we generate a pool of candidate portfolios and evaluate each based on its expected return and associated risk. The resulting portfolio candidates are then clustered using the k-means algorithm to group solutions exhibiting similar performance characteristics.
- **Quantum Optimization and Circuit Repetition Variation:** We employ hybrid quantum-classical algorithms, specifically VQE [7] and QAOA [6], to solve the QUBO-formulated portfolio optimization problem. Within this stage, we systematically vary quantum circuit configurations, such as varying the number of circuit repetitions (i.e., layers).
- **Evaluation and Comparative Analysis:** The output portfolios obtained from different encoding strategies under varying quantum circuit depths are compared.

4 Evaluation of Q-PORT Methodology

4.1 Experiment setup

For our experiments, we collected historical stock price data from Yahoo Finance, selecting four distinct stock groups to evaluate scalability. The datasets cover the period from January 1, 2020, to January 1, 2024, and include:

- 3 stocks: Apple, Microsoft, and Google;
- 4 stocks: Apple, Microsoft, Google, and Amazon;
- 12 stocks: Apple, Microsoft, Google, Amazon, Tesla, Nvidia, Meta, Netflix, JP Morgan, Disney, Visa, and Berkshire Hathaway Inc.

We employed both VQE and QAOA algorithms, using layered ansatz circuits composed of parameterized single-qubit rotations and entangling gates arranged in a fixed connectivity pattern. Specifically, each ansatz begins with Hadamard gates to initialize qubits into superposition, followed by parameterized R_Z rotations for the cost Hamiltonian, two-qubit R_{ZZ} gates to induce entanglement reflective of the QUBO problem structure, and R_X gates implementing the mixer

Hamiltonian. By varying the number of layers (repetitions), from 1 to 5, we systematically control the circuit depth and the level of multipartite entanglement, thereby expanding the ansatz expressivity.

The risk tolerance parameter was fixed at 0.2 across all experiments, while the investment budget was defined as the total number of stocks minus one (e.g., budget of 3 for 4-stock scenarios). The classical optimizer COBYLA was employed with a maximum of 5000 iterations. All simulations were conducted using statevector-based simulators, avoiding sampling noise inherent to real quantum hardware. Experiments were executed on Google Colab instances equipped with Intel Xeon E5 CPUs.

4.2 Evaluation of Multi-Qubit Representations per Stock

Figures 2 and 3 present the ranking of the portfolios that correspond to the classical MVO-optimal solution within the output distributions of QAOA and VQE under the multi-qubit per stock encoding, across different numbers of circuit repetitions. In these plots, the value of *Ranking* on the y-axis denotes the position of the optimal solution (computed by the classical MVO) in the sorted list of quantum-generated portfolios. A lower ranking indicates that the quantum algorithm assigns a higher probability to the optimal portfolio, thus reflecting better alignment with the classical benchmark. Hence, a ranking of 1 implies that the solution found by the quantum algorithm matches the classically optimal solution.

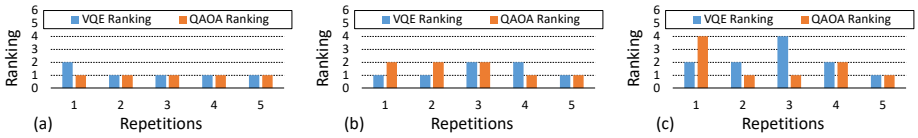


Fig. 2. Results of the multi-qubit representations per 3 stocks, as a function of circuit repetitions. (a) 1 qubit per stock, i.e., 3 qubits. (b) 2 qubits per stock, i.e., 6 qubits. (c) 3 qubits per stock, i.e., 9 qubits.

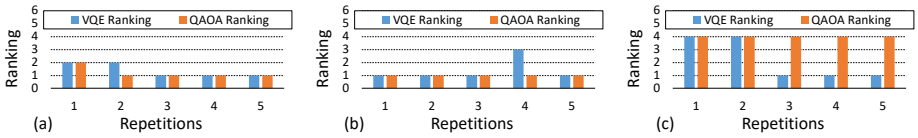


Fig. 3. Results of the multi-qubit representations per 4 stocks, as a function of circuit repetitions. (a) 1 qubit per stock, i.e., 4 qubits. (b) 2 qubits per stock, i.e., 8 qubits. (c) 3 qubits per stock, i.e., 12 qubits.

The results for 3-stock and 4-stock configurations demonstrate that increasing the number of repetitions generally improves the alignment between quantum and classical solutions, particularly for the QAOA. Notably, in the 3-stock case (Figure 2), the QAOA consistently ranks the optimal solution among the top positions even at low depth, whereas the VQE requires higher depth to converge to competitive rankings. For the 4-stock case (Figure 3), both QAOA and VQE exhibit increased variance, but higher repetitions still tend to yield lower (better) rankings overall.

Across subplots (a), (b), and (c) of Figures 2 and 3, we analyze the effect of increasing the number of qubits assigned per stock from 1 to 3. In subplot (a), the default configuration with one qubit per stock serves as a baseline. As shown, both QAOA and VQE are capable of identifying the classically optimal solution with relatively low rankings, particularly when the number of circuit repetitions is increased.

However, in subplot (b), where each stock is represented using two qubits, we observe a modest increase in variability and, in some cases, a decline in ranking performance. This trend becomes more pronounced in subplot (c), where three qubits per stock are used. Here, the quantum algorithms exhibit greater difficulty in consistently ranking the MVO-optimal solution near the top, particularly at lower circuit depths.

This behavior can be attributed to the expansion of the solution space due to redundant encoding. While redundancy through multi-qubit representation may increase robustness in noisy environments, it also introduces additional complexity in optimization, making it harder for the quantum algorithms to converge to the classically optimal solution. Overall, these results indicate that increasing the number of qubits per stock does not necessarily improve performance and may, in fact, degrade solution quality unless accompanied by deeper and more expressive quantum circuits.

4.3 Evaluation of Multi-Stock Encoding per Qubit

Figure 4 presents the ranking performance of the classically optimal MVO solution under different multi-stock per qubit configurations for a 12-stock portfolio. The three subplots (a), (b), and (c) correspond to encoding 1, 2, and 3 stocks per qubit, respectively. As before, lower rankings (closer to 1) indicate stronger alignment between quantum-generated results and the classical MVO benchmark.

In subplot (a), where each stock is assigned to a distinct qubit, we observe baseline performance for both QAOA and VQE across varying circuit repetitions. The QAOA generally ranks the optimal portfolio higher than the VQE, particularly at lower depths. Subplot (b) shows the configuration where two stocks are compressed into each qubit. Interestingly, despite a reduction in qubit resources, the QAOA continues to perform comparably well, often preserving the optimal solution within the top few ranks. The VQE, while showing slightly more variance, maintains a generally acceptable ranking trend with increased repetitions.

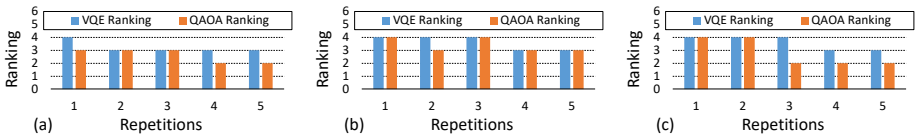


Fig. 4. Results of the multi-stock encoding for 12 stocks, as a function of circuit repetitions. (a) 1 stock per qubit, i.e., 12 qubits. (b) 2 stocks per qubit, i.e., 6 qubits. (c) 3 stocks per qubit, i.e., 4 qubits.

Subplot (c), which encodes three stocks per qubit, highlights the benefits and limitations of aggressive compression. While some degradation in ranking performance is expected due to the loss of individual stock granularity, the results demonstrate that the QAOA is still able to identify the optimal solution with competitive rankings, particularly at higher circuit depths. The VQE exhibits more fluctuation and tends to benefit from deeper circuits to stabilize performance.

Overall, these results suggest that multi-stock encoding is a viable strategy for scaling portfolio optimization to larger asset sets under limited quantum hardware resources. Although increasing the number of stocks per qubit introduces approximation errors, careful tuning of circuit depth can mitigate precision loss, particularly when using QAOA, which appears more resilient under compressed encodings.

5 Conclusion

In this work, we proposed *Q-PORT*, a systematic methodology to explore the scalability of quantum portfolio optimization under different encoding strategies and quantum circuit configurations. We introduced and analyzed two alternative encoding paradigms, the multi-qubit representation per stock and the multi-stock encoding per qubit. We then evaluated their effectiveness using VQE and QAOA across multiple portfolio sizes and circuit repetitions. Our findings demonstrate that increasing the number of qubits per stock yields negligible improvements in solution quality, while significantly increasing resource demands. In contrast, encoding multiple stocks per qubit enables substantial reductions in qubit usage with only minor losses in precision, especially when using QAOA with deeper circuits. These results highlight practical pathways for scaling quantum portfolio optimization under the constraints of NISQ hardware. As future work, we plan to extend our experimental evaluation to larger-scale portfolios and more complex financial datasets to further validate the generalizability and robustness of the proposed methodology.

Acknowledgment

This work was supported in part by the NYUAD Center for Quantum and Topological Systems (CQTS), funded by Tamkeen under the NYUAD Research Institute grant CG008.

References

1. H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
2. F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi, *Robust Portfolio Optimization and Management*. John Wiley & Sons, 2007.
3. M. W. Brandt, P. Santa-Clara, and R. Valkanov, "Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns," *The Review of Financial Studies*, vol. 22, no. 9, pp. 3411–3447, 2009.
4. E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, 9 ed., 2014.
5. J. Preskill, "Quantum computing in the nisq era and beyond," *Quantum*, vol. 2, p. 79, 2018.
6. E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," *arXiv preprint arXiv:1411.4028*, 2014.
7. A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a quantum processor," *Nature Communications*, vol. 5, p. 4213, 2014.
8. D. Venturelli and A. Kondratyev, "Quantum optimization of fully connected spin glasses," *Quantum Science and Technology*, vol. 4, no. 2, p. 025002, 2019.
9. G. Buonaiuto, F. Gargiulo, G. De Pietro, M. Esposito, and M. Pota, "Best practices for portfolio optimization by quantum computing, experimented on real quantum devices," *Scientific Reports*, vol. 13, no. 1, p. 19434, 2023.
10. T. Hodson, T. Green, L. Chakhmakhchyan, and I. Grout, "Portfolio optimization and variational quantum algorithms," *Quantum Information Processing*, vol. 21, no. 2, p. 76, 2022.
11. K. Zaman, A. Marchisio, M. Kashif, and M. Shafique, "Po-qa: A framework for portfolio optimization using quantum algorithms," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 1, pp. 1397–1403, IEEE, 2024.
12. N. Innan, A. Saleem, A. Marchisio, and M. Shafique, "Quantum portfolio optimization with expert analysis evaluation," *arXiv preprint arXiv:2507.20532*, 2025.
13. E. Aguilera, J. de Jong, F. Phillipson, S. Taamallah, and M. Vos, "Multi-objective portfolio optimization using a quantum annealer," *Mathematics*, vol. 12, no. 9, p. 1291, 2024.
14. S. N. Morapakula, S. Deshpande, R. Yata, R. Ubale, U. Wad, and K. Ikeda, "End-to-end portfolio optimization with quantum annealing," *arXiv preprint arXiv:2504.08843*, 2025.
15. B. Chen, H. Wu, H. Yuan, L. Wu, and X. Li, "Quasi-binary encoding based quantum alternating operator ansatz," *arXiv preprint arXiv:2304.06915*, 2023.
16. N. Innan, A. Marchisio, M. Bennai, and M. Shafique, "Qfnn-ffd: Quantum federated neural network for financial fraud detection," *arXiv preprint arXiv:2404.02595*, 2024.

17. S. Dutta, N. Innan, A. Marchisio, S. B. Yahia, and M. Shafique, "Qadqn: Quantum attention deep q-network for financial market prediction," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 2, pp. 341–346, IEEE, 2024.
18. N. Innan *et al.*, "Financial fraud detection using quantum graph neural networks," *Quantum Machine Intelligence*, vol. 6, no. 1, p. 7, 2024.
19. N. Innan, M. A.-Z. Khan, and M. Bennai, "Financial fraud detection: a comparative study of quantum machine learning models," *International Journal of Quantum Information*, vol. 22, no. 02, p. 2350044, 2024.
20. N. Innan, A. Marchisio, M. Bennai, and M. Shafique, "Lep-qnn: Loan eligibility prediction using quantum neural networks," *arXiv preprint arXiv:2412.03158*, 2024.
21. N. Innan, M. Kashif, A. Marchisio, M. Bennai, and M. Shafique, "Next-generation quantum neural networks: Enhancing efficiency, security, and privacy," in *2025 IEEE 31st International Symposium on On-Line Testing and Robust System Design (IOLTS)*, pp. 1–4, IEEE, 2025.
22. K. Zaman, A. Marchisio, M. A. Hanif, and M. Shafique, "A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead," *arXiv preprint arXiv:2310.10315*, 2023.
23. M. E. Alami, N. Innan, M. Shafique, and M. Bennai, "Comparative performance analysis of quantum machine learning architectures for credit card fraud detection," *arXiv preprint arXiv:2412.19441*, 2024.
24. P. K. Choudhary, N. Innan, M. Shafique, and R. Singh, "Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction," *arXiv preprint arXiv:2503.15403*, 2025.
25. A. Sawaika, S. Krishna, T. Tomar, D. P. Suggisetti, A. Lal, T. Shrivastav, N. Innan, and M. Shafique, "A privacy-preserving federated framework with hybrid quantum-enhanced learning for financial fraud detection," *arXiv preprint arXiv:2507.22908*, 2025.
26. P. Pathak, V. Oad, A. Prajapati, and N. Innan, "Resource allocation optimization in 5g networks using variational quantum regressor," in *2024 International Conference on Quantum Communications, Networking, and Computing (QCNC)*, pp. 101–105, IEEE, 2024.
27. N. Innan, A. Marchisio, M. Bennai, and M. Shafique, "Qfnn-ffd: Quantum federated neural network for financial fraud detection," in *2025 IEEE International Conference on Quantum Software (QSW)*, pp. 41–47, IEEE, 2025.
28. M. Kashif, S. Khalid, N. Innan, A. Marchisio, and M. Shafique, "Evaluating quantum amplitude estimation for pricing multi-asset basket options," *arXiv preprint arXiv:2509.09432*, 2025.
29. I. Kerenidis, A. Prakash, and D. Szilágyi, "Quantum algorithms for portfolio optimization," *arXiv preprint arXiv:1908.08040*, 2019.

Quantum Amplitude Estimation in Practice: A Case Study in Option Pricing

Nouhaila Innan[1,2], Muhammad Kashif[1,2], Alberto Marchisio[1,2], Muhammad Moonis Usman[1,2], and Muhammad Shafique[1,2]

[1] eBRAIN Lab, Division of Engineering, New York University Abu Dhabi (NYUAD), Abu Dhabi, UAE
[2] Center for Quantum and Topological Systems (CQTS), NYUAD Research Institute, NYUAD, Abu Dhabi, UAE
nouhaila.innan@nyu.edu, muhammadkashif@nyu.edu, alberto.marchisio@nyu.edu, mu2173@nyu.edu, muhammad.shafique@nyu.edu

Abstract. Accurately estimating expected payoffs is central to the pricing of European call options, especially when valuation depends on low-probability events in the distribution tail. This study evaluates the performance of Quantum Amplitude Estimation (QAE), using Iterative Amplitude Estimation (IAE) and Maximum Likelihood Amplitude Estimation (MLAE), in pricing European call options based on historical Apple market data. Despite the theoretical advantages of QAE, our experiments show that both quantum estimators return an expected payoff of zero, even in scenarios where classical methods, Black Scholes and Monte Carlo simulation yield significantly positive values. This outcome stems from the limited resolution imposed by limited uncertainty qubits, which inadequately encode small-amplitude, in-the-money price regions. While QAE circuits correctly identify the realized market outcome, they fail to capture the full expectation implied by the distribution. These results highlight the current limitations of QAE under realistic constraints and underscore the importance of enhanced encoding strategies for future quantum financial applications.

1 Introduction

Pricing European call options, contracts that give the holder the right to buy an asset at strike K on maturity T, is a cornerstone problem in quantitative finance. Classical methods such as the Black–Scholes (BS) model provide closed-form solutions under idealized assumptions [1], while Monte Carlo (MC) simulation offers a flexible simulation-based alternative at the cost of $O(1/\epsilon^2)$ sample complexity when targeting accuracy ϵ [2,3].

Recently, quantum computing-based techniques have been applied to financial applications [4–14]. Quantum Amplitude Estimation (QAE) promises an alternate route to expectation estimation, with asymptotic scaling $O(1/\epsilon)$ in the ideal, fault-tolerant limit [15]. Practical variants, Iterative Amplitude Estimation (IAE) and Maximum Likelihood Amplitude Estimation (MLAE) [16,17], adapt

QAE to near-term, noisy quantum devices by replacing deep phase-estimation subroutines with iterative Grover-based measurements and classical inference .

In this study, we do not assume that quantum methods will immediately outpace classical techniques [18]. Instead, our objective is to investigate whether QAE-based algorithms can faithfully replicate the structure and outcomes of classical option-pricing workflows. Specifically, we ask:

Can IAE and MLAE, when applied to a real-world European option pricing problem, produce expected-payoff estimates that align with BS and MC benchmarks, given current hardware constraints and state-preparation limitations?

To address this question, we construct a pipeline in Qiskit comprising:

1. **Distribution Encoding**: Discretize a log-normal model of asset returns into 2^n bins and prepare the corresponding quantum superposition.
2. **Payoff Circuit**: Implement the European call payoff $f(B) = \max(B-K, 0)$ as a controlled rotation on an ancilla qubit via a piecewise linear mapping.
3. **Amplitude Estimation**: Run IAE and MLAE to extract the amplitude corresponding to the expected payoff, each with carefully bounded Grover depths.
4. **Classical Benchmarks**: Compute analytic (BS) and numerical (MC) reference values using the same discretization grid for direct comparison.

By comparing quantum estimates against classical baselines on historical data, we identify the primary sources of deviation, such as finite-qubit discretization and limited iteration counts, and assess to what extent QAE algorithms can mirror classical pricing behavior. Our findings clarify the practical viability of quantum methods for financial expectation tasks and highlight concrete steps toward closing the gap between current quantum capabilities and established classical workflows.

2 Background

2.1 Black-Scholes Algorithm

The BS model is a foundational analytical framework for pricing European call options. Given an underlying asset price S, strike price K, risk-free interest rate r, time to maturity T, and volatility σ, the theoretical price C of the option is derived using the following expressions: $d_1 = \frac{\ln\left(\frac{S}{K}\right)+\left(r+\frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}, d_2 = d_1 - \sigma\sqrt{T}$, and $C = S\,\Phi(d_1) - K\,e^{-rT}\,\Phi(d_2)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. While efficient, this closed-form solution assumes constant volatility, frictionless markets, and log-normal price distributions, assumptions that may not hold in practice.

2.2 Monte Carlo Simulation

MC methods are widely used to simulate asset paths under the risk-neutral measure, providing a flexible numerical approach to option pricing [19]. The

terminal asset price under geometric Brownian motion is modeled as: $S(T) = S_0 \exp\left((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z\right), \quad Z \sim \mathcal{N}(0,1)$. For each simulated path, the payoff is computed as $\max(S(T) - K, 0)$, and the option value is estimated by discounting the average payoff: $V_{\text{MC}} = e^{-rT} \cdot \frac{1}{N}\sum_{i=1}^{N} \max(S_i(T) - K, 0)$.

MC simulation generalizes naturally to asset options by aggregating the weighted outcomes of multiple assets. Although accuracy improves with the number of samples, convergence rate is bounded by $\mathcal{O}(1/\epsilon^2)$, where ϵ is the desired error tolerance.

2.3 Quantum Amplitude Estimation

QAE is a quantum algorithm designed to estimate the amplitude a of a quantum state with quadratic speedup over classical sampling methods [20]. Let $\mathcal{A}$ be a unitary operator that prepares the state $\mathcal{A}|0\rangle = \sqrt{1-a}\,|\psi_0\rangle + \sqrt{a}\,|\psi_1\rangle$, where $a \in [0,1]$ is the target quantity encoding the expectation of a function (e.g., payoff). The standard QAE algorithm employs phase estimation on the Grover-like operator $Q = \mathcal{A}S_0\mathcal{A}^\dagger S_\chi$ to extract an estimate $\hat{a}$, using a number of queries that scales as $\mathcal{O}(1/\epsilon)$ in the noiseless setting. However, implementing this version requires deep quantum circuits with inverse operations and the quantum Fourier transform, which are impractical on current noisy devices.

Iterative Amplitude Estimation IAE avoids the need for quantum phase estimation and uses repeated Grover iterations combined with classical statistical inference. It performs a sequence of measurements after controlled applications of the Grover operator to refine the estimate of a iteratively (see Algorithm 1). The method preserves the asymptotic advantage while using shallower circuits, making it suitable for near-term hardware.

Algorithm 1 IAE

1: **Input:** Unitary $\mathcal{A}$, Grover operator $Q = \mathcal{A}S_0\mathcal{A}^\dagger S_\chi$, confidence level δ
2: Initialize estimate interval $[a_{\min}, a_{\max}] \leftarrow [0,1]$
3: **for** $j = 1$ to J **do**
4: Choose number of Grover iterations m_j
5: Apply $Q^{m_j}\mathcal{A}|0\rangle$, measure ancilla
6: Update confidence interval for a using binomial likelihood
7: **end for**
8: **Return:** Final estimate $\hat{a}$ as midpoint of final confidence interval

Maximum Likelihood Amplitude Estimation MLAE also avoids phase estimation but relies on sampling measurement outcomes for different Grover powers m and uses a classical maximum likelihood procedure to reconstruct

the amplitude [21]. For each m, the probability of measuring a success outcome is $\sin^2((2m+1)\theta)$, where $\theta = \arcsin(\sqrt{a})$. MLAE solves for the value $\hat{a}$ that maximizes the likelihood of the observed results (see Algorithm 2).

Algorithm 2 MLAE

1: **Input:** Unitary $\mathcal{A}$, list of powers $\{m_j\}$, number of shots N_j per m_j
2: **for** $j = 1$ to J **do**
3: Apply $Q^{m_j}\mathcal{A}|0\rangle$, measure ancilla qubit
4: Record number of successes k_j out of N_j trials
5: **end for**
6: Define log-likelihood function $\ell(\hat{a}) = \sum_j \log \Pr(k_j \mid m_j, \hat{a})$
7: **Return:** $\hat{a} = \arg\max_{a\in[0,1]} \ell(a)$

MLAE can be more robust than IAE under noise, especially when the likelihood function is sharply peaked and the number of samples is sufficient. Both IAE and MLAE serve as practical alternatives to QAE in noisy intermediate-scale quantum settings [22].

3 Methodology

This study explores the application of quantum amplitude estimation algorithms to the problem of pricing European call options. As shown in Fig. 1, the workflow comprises 4 main stages: (1) preparing the data, (2) modeling the uncertainty of the asset price using a log-normal distribution, (3) encoding the payoff function into a quantum circuit, and (4) estimating the expected payoff using two amplitude estimation algorithms, IAE and MLAE, and benchmarking the results against classical references.

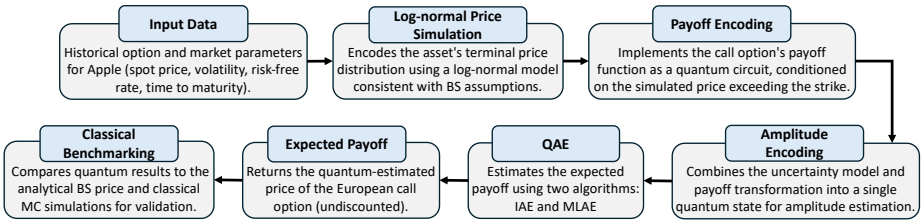


Fig. 1. Pipeline for quantum pricing of European call options. Starting from historical market data, the method proceeds through quantum encoding and amplitude estimation before comparing the quantum results with classical references (BS and MC).

3.1 Input Data and Preprocessing

We start with historical option and market data for Apple (AAPL) over the year 2024. The relevant features extracted include spot price S, strike price K, implied volatility σ, time to maturity T, and risk-free rate r. Only a subset of this data, filtered by data quality and maturity length, is used in quantum simulation. These features are then used to parameterize the log-normal model consistent with the BS model.

3.2 Log-Normal Price Simulation

We model the terminal asset price S_T using a log-normal distribution, consistent with the assumptions of the BS framework. The quantum state preparation uses the `LogNormalDistribution` component from Qiskit Finance[3] to encode the discretized probability distribution into a quantum circuit. The distribution parameters are derived from market inputs: initial spot price S, volatility σ, risk-free interest rate r, and time to maturity T. The support of the distribution is truncated to the interval $[S_{\min}, S_{\max}]$, spanning three standard deviations around the mean to ensure sufficient coverage. Let $\mu = \log(S) + \left(r - \frac{\sigma^2}{2}\right) T$, $\sigma_T = \sigma\sqrt{T}$, then the quantum state encodes the distribution as $|\psi\rangle = \sum_{i=0}^{2^n-1} \sqrt{p_i}\,|i\rangle$, where $p_i \approx \mathbb{P}(S_T \in \text{bin}_i)$.

3.3 Encoding the Payoff Function

The circuit is extended to represent the payoff function of a European call option, defined as $\max(S_T - K, 0)$. This is implemented as a piecewise linear transformation using Qiskit's `LinearAmplitudeFunction`. The operation applies controlled rotations on an objective qubit, conditioned on whether the simulated price exceeds the strike. A scaling factor c ensures the resulting payoff fits within the interval $[0, 1]$, satisfying amplitude estimation requirements. The quantum circuit consists of three key components:

- **State qubits**, which encode the log-normal distribution over asset prices. Each computational basis state $|i\rangle$ corresponds to a price bin s_i, with amplitude $\sqrt{p_i}$, forming the state $|\psi\rangle = \sum_i \sqrt{p_i}|i\rangle$.
- **An objective qubit**, initialized in $|0\rangle$, whose amplitude is modulated according to the payoff function. Using breakpoints and slopes, the payoff $\max(s_i - K, 0)$ is mapped linearly to amplitudes. The objective qubits are used to represent the payoff function with higher precision before it is mapped into amplitude space.
- **Ancilla qubits**, used to support the conditional logic required by the piecewise linear implementation. These ensure correct breakpoint handling and amplitude normalization.

Together, these elements form the composite state: $\mathcal{A}|0\rangle = \sqrt{1-a}\,|\psi_0\rangle + \sqrt{a}\,|\psi_1\rangle$, where $a \in [0, 1]$ is proportional to the expected payoff.

[3] https://qiskit-community.github.io/qiskit-finance/

3.4 Amplitude Estimation Algorithms

To extract the value of a, we apply two variants of amplitude estimation on the same quantum problem:

- IAE refines the estimate of a through successive rounds of Grover-like amplification and classical post-processing, avoiding the need for quantum Fourier transforms. It is initialized with a target precision ϵ and confidence level α, and proceeds with a fixed number of Grover operator applications per round. The final estimate is computed via a maximum likelihood strategy over all rounds.
- MLAE samples different Grover powers and formulates a classical optimization problem to infer the most likely amplitude $\hat{a}$. Unlike IAE, MLAE does not require pre-specifying a precision threshold and instead focuses purely on statistical likelihood maximization, making it potentially more resilient to sampling noise.

Both methods use identical quantum circuits for state preparation and payoff encoding, ensuring a fair comparison across the two estimation strategies.

3.5 Expected Payoff and Classical Benchmarking

The estimated amplitude a is rescaled to recover the expected payoff, expressed as $\mathbb{E}\left[\max(S_T - K, 0)\right] \approx a \cdot c$, where c is the payoff scaling factor. To validate the quantum estimates, we compute two classical references: The analytical price using the BS model, and the MC simulated price using standard sampling of log-normal paths. Both references are computed over the same discretized price space to ensure fair comparison. Additionally, delta estimates representing $\mathbb{P}(S_T > K)$ are derived from the quantum circuit by querying ancilla qubits associated with price thresholds.

4 Results and Analysis

4.1 Experimental Settings

As presented in Table 1, the experimental setup uses real-world financial data for Apple Inc. (AAPL), collected via the Yahoo Finance API [23], covering the period from January 2, 2022, to February 13, 2023. A 5% shift is applied to the mean log returns to simulate a bullish market scenario. The volatility is clipped to a minimum of 0.20 to ensure realistic risk estimates. The time to maturity is fixed at one year, and the strike price is set to 5% above the initial asset value. For quantum simulation using Qiskit [24], we allocate 6 uncertainty qubits to discretize the terminal asset price into 64 bins. The payoff is approximated using a piecewise linear function mapped onto 3 objective qubits, enabling finer-grained approximation of the payoff and more accurate encoding of small but significant values, with the help of 3 ancilla qubits. The amplitude estimation routines are configured as follows:

– **IAE**: Maximum of 3 Grover iterations, with precision $\epsilon = 0.01$ and confidence level $\alpha = 0.05$.
– **MLAE**: Evaluation schedule set to {0, 1, 3, 7}.

Classical benchmarks include a dot-product-based expected value from the discretized payoff distribution (BS) and a MC estimate using 100,000 samples.

Table 1. Simulation parameters for quantum and classical configurations.

Parameter	**Value / Description**
Assets	['AAPL']
Date Range	2022-01-02 to 2023-02-13
Return Shift	+5% to mean log returns
Min. Volatility	0.20 (capped)
Maturity Time T	1 year
Strike Price	5% above initial asset value
Uncertainty Qubits	6
Objective Qubits	3
Ancilla Qubits	3
IAE Parameters	$\epsilon = 0.01$, $\alpha = 0.05$, max 3 iterations
MLAE Schedule	{0, 1, 3, 7}
MC Samples	100,000
MC Passes	Single run per trial

4.2 Empirical Results and Analysis

Table 2 and Table 3 report the expected payoff and final asset price estimates for two distinct market scenarios, each calibrated from historical AAPL data with a strike set 5% above the initial asset value. In both cases, the realized payoff is zero (terminal price < strike), yet classical methods predict non-zero expected payoffs ($17.60 in 2024 and $17.06 in 2022), by integrating the tail of the price distribution. These values are obtained analytically via a dot product of the payoff and probability vectors and confirmed by MC (100,000 samples), which show deviations below 1%. In contrast, both quantum methods (IAE and MLAE) yield an expected payoff of 0.00 in both scenarios. This underestimation does not reflect a failure of the algorithms themselves, but rather the limitations imposed by coarse amplitude encoding: with only six qubits discretizing the price range into 64 bins, the contributions from low-probability, in-the-money regions are too small to be resolved. As a result, these significant but subtle amplitudes fall below the estimation threshold, causing QAE to return zero despite the positive expected payoff predicted by classical models.

This effect also appears in the estimated asset prices: classical approaches give averages near $198 (2023-2024) and $192 (2022-2023), whereas quantum outputs cluster around $185 and $179, respectively. The disparity underscores

the sensitivity of payoff estimation to distribution resolution, low-qubit implementations struggle to capture low-probability, high-payoff events that drive the true expectation.

Table 2. Expected payoff and asset price (2023-2024).

Method	**Expected Payoff (USD)**	**Asset Price (USD)**
Strike: 194.97 —	*Actual Payoff: 0.00*	
Classical (Exact)	17.5987	198.0063
MC	17.8259	198.1746
IAE	0.0000	185.3681
MLAE	0.0000	185.3681

Table 3. Expected payoff and asset price (2022-2023).

Method	**Expected Payoff (USD)**	**Asset Price (USD)**
Strike: 193.97 —	*Actual Payoff: 0.00*	
Classical (Exact)	17.0597	191.9414
MC	17.2799	192.1046
IAE	0.0000	179.6903
MLAE	0.0000	179.6749

4.3 Interpretation and Limitations

The disparity between quantum and classical estimates highlights the critical role of distribution resolution in QAE-based pricing. Six uncertainty qubits yield only 64 discrete bins, insufficient to capture the small but financially significant amplitudes above the strike. Consequently, low-probability events that dominate the expected payoff are effectively invisible to the amplitude estimation routines. Both IAE and MLAE correctly identify the most probable outcome (zero payoff), demonstrating stability and potential in binary classification or threshold-detection tasks. However, recovering the full expectation requires finer amplitude granularity, either via more qubits, advanced encoding schemes, or adaptive binning strategies. Our results also suggest that, even in a noiseless simulator with fully expressive circuits (e.g., IAE with 172,000 gates, MLAE with 12,600 gates), algorithmic limitations arise from the encoded probability landscape rather than circuit depth. Future work should therefore prioritize enhanced state preparation and payoff mapping, such as nonuniform bin spacing around strike regions or variational encoding, to better resolve tail events.

4.4 Discussion

The experimental results emphasize the current limitations of QAE for realistic financial estimation tasks, even when using expressive circuits. The coarse discretization from six qubits remains the dominant factor behind payoff underestimation, as bins near the strike fail to capture the relevant variations in asset prices. Though noise and hardware effects are excluded, transpilation reflects some realistic overheads, offering a practical perspective on achievable performance under ideal conditions. It is encouraging that both IAE and MLAE consistently predicted the realized market payoff, suggesting utility in binary event detection even under coarse resolution. These findings point toward promising avenues for improvement: increasing qubit counts, using nonuniform binning, or designing more targeted amplitude distributions. The presented pipeline offers a benchmark for testing such enhancements in future iterations of QAE.

4.5 Conclusion

This study evaluated the use of QAE algorithms for pricing European call options using historical asset data in a simulated quantum environment. While classical methods, dot-product integration (BS) and MC, accurately captured the expected payoff, quantum estimators returned a value of zero, underestimating the true financial expectation. The performance gap stems from resolution limitations and the inability to amplify low-amplitude but significant payoff states, not from insufficient circuit expressivity. This underscores the importance of encoding strategies and amplitude shaping in financial QAE applications.

Importantly, quantum circuits still produced stable and interpretable outputs aligned with the realized payoff, indicating near-term utility in classification or binary decision tasks. As quantum technology advances, the foundational pipeline developed here can guide progress toward accurate and scalable quantum financial analytics.

Acknowledgment

This work was supported in part by the NYUAD Center for Quantum and Topological Systems (CQTS), funded by Tamkeen under the NYUAD Research Institute grant CG008.

References

1. F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of political economy*, 1973.
2. P. Glasserman, *Monte Carlo methods in financial engineering*. Springer, 2004.
3. P. Sharma *et al.*, "Review of research on option pricing: A bibliometric analysis," *Qualitative Research in Financial Markets*, 2024.
4. N. Innan, A. Marchisio, M. Bennai, and M. Shafique, "Qfnn-ffd: Quantum federated neural network for financial fraud detection," in *2025 IEEE International Conference on Quantum Software (QSW)*, pp. 41–47, IEEE, 2025.

5. S. Dutta *et al.*, "Qadqn: Quantum attention deep q-network for financial market prediction," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 2, pp. 341–346, IEEE, 2024.
6. N. Innan, A. Marchisio, M. Bennai, and M. Shafique, "Lep-qnn: Loan eligibility prediction using quantum neural networks," *arXiv preprint arXiv:2412.03158*, 2024.
7. N. Innan, M. Kashif, A. Marchisio, M. Bennai, and M. Shafique, "Next-generation quantum neural networks: Enhancing efficiency, security, and privacy," in *2025 IEEE 31st International Symposium on On-Line Testing and Robust System Design (IOLTS)*, pp. 1–4, IEEE, 2025.
8. K. Zaman, A. Marchisio, M. A. Hanif, and M. Shafique, "A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead," *arXiv preprint arXiv:2310.10315*, 2023.
9. K. Zaman, A. Marchisio, M. Kashif, and M. Shafique, "Po-qa: A framework for portfolio optimization using quantum algorithms," in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 1, pp. 1397–1403, IEEE, 2024.
10. M. E. Alami, N. Innan, M. Shafique, and M. Bennai, "Comparative performance analysis of quantum machine learning architectures for credit card fraud detection," *arXiv preprint arXiv:2412.19441*, 2024.
11. P. K. Choudhary, N. Innan, M. Shafique, and R. Singh, "Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction," *arXiv preprint arXiv:2503.15403*, 2025.
12. A. Sawaika *et al.*, "A privacy-preserving federated framework with hybrid quantum-enhanced learning for financial fraud detection," *arXiv preprint arXiv:2507.22908*, 2025.
13. N. Innan, A. Saleem, A. Marchisio, and M. Shafique, "Quantum portfolio optimization with expert analysis evaluation," *arXiv preprint arXiv:2507.20532*, 2025.
14. M. Kashif, S. Khalid, N. Innan, A. Marchisio, and M. Shafique, "Evaluating quantum amplitude estimation for pricing multi-asset basket options," *arXiv preprint arXiv:2509.09432*, 2025.
15. G. Brassard *et al.*, "Quantum amplitude amplification and estimation," *Contemporary Mathematics*, 2002.
16. D. Grinko *et al.*, "Iterative quantum amplitude estimation," *npj Quantum Information*, 2021.
17. T. Tanaka *et al.*, "Amplitude estimation via maximum likelihood on noisy quantum computer," *Quantum Information Processing*, 2021.
18. N. Stamatopoulos *et al.*, "Option pricing using quantum computers," *Quantum*, 2020.
19. S. Woerner and D. J. Egger, "Quantum risk analysis," *npj Quantum Information*, 2019.
20. A. C. Vazquez and S. Woerner, "Efficient state preparation for quantum amplitude estimation," *arXiv preprint arXiv:2005.07711*, 2020.
21. Y. Suzuki, , *et al.*, "Amplitude estimation without phase estimation," *Quantum Information and Computation*, 2020.
22. M. Maronese *et al.*, "The quantum amplitude estimation algorithms on near-term devices: a practical guide," *Quantum Reports*, 2023.
23. R. Aroussi, "yfinance," 2023. https://pypi.org/project/yfinance/.
24. A. Javadi-Abhari *et al.*, "Quantum computing with qiskit," *arXiv preprint arXiv:2405.08810*, 2024.

Quantum Approaches for the Unit Commitment Problem - a literature survey

Frank Phillipson and Sven Muller

Maastricht University, Maastricht, The Netherlands
TNO, The Hague, The Netherlands
f.phillipson@maastrichtuniversity.nl

Abstract. This paper investigates the application of quantum computing to the Unit Commitment (UC) problem, a fundamental optimisation challenge in power system operations. From the literature, we can learn about various quantum approaches, including the Quantum Approximate optimisation Algorithm (QAOA), Quantum Annealing (QA), and hybrid quantum-classical methods. The review shows that in the literature it is believed that quantum computing can potentially solve the UC problem more efficiently than classical methods. QAOA shows promise in handling binary decision variables, while hybrid methods enhance computational efficiency and scalability. Quantum Annealing is effective for smaller UC instances, with larger problems requiring partitioning. Despite current hardware limitations, advancements in quantum algorithms and hybrid methods provide a strong foundation for future research. This study highlights the transformative potential of quantum computing in optimising power systems, emphasising the need for continued innovation in quantum hardware and error mitigation techniques.

Keywords: Quantum Computing · Energy Model · Unit Commitment · Optimisation

1 Introduction

The Unit Commitment (UC) problem is a critical optimisation task in electrical power systems, where the objective is to schedule the operation of generating units over a planning horizon to meet electricity demand at minimal cost while ensuring system reliability. This problem arises because large-scale energy storage is impractical, requiring real-time alignment of generation and consumption. UC is challenging due to the diversity of generation units, operational constraints such as minimum up/down times and ramp limits, and the need for advance scheduling to accommodate slow-starting units. Furthermore, the problem must account for network constraints, reliability requirements, and market rules, adding layers of complexity to the optimisation process.

Classical approaches to UC have evolved significantly. Since 2005, power system operators have transitioned from heuristic methods based on Lagrangian relaxation to Mixed-Integer Programming (MIP) formulations [13]. These models, solved using commercial solvers like CPLEX and Gurobi, have achieved

substantial cost savings—estimated at $5 billion annually in the United States. Despite these advances, solving real-world UC problems remains computationally demanding, often involving hundreds of generators, network constraints, and strict time limits for solution generation. Recent research focuses on improving MIP formulations to enhance computational efficiency, but challenges persist.

Quantum computing offers a promising alternative for addressing these computational bottlenecks. By leveraging superposition and entanglement, quantum algorithms could accelerate the exploration of vast combinatorial search spaces, potentially enabling faster and more scalable UC solutions. Although quantum computing for UC is in its early stages, it represents a promising step toward real-time optimisation for large-scale power systems.

This paper provides an overview of existing quantum approaches to the UC problem. Section 2 introduces key concepts in quantum computing. Section 3 outlines the UC problem and its mathematical formulation. Section 4 reviews existing work on quantum methods for UC. Finally, Section 5 summarizes findings and identifies directions for future research.

2 Quantum Computing

Quantum computing is a revolutionary approach to computation that leverages the principles of quantum mechanics to solve problems that are challenging or even intractable for classical computers. Unlike classical systems that use bits to represent information as either 0 or 1, quantum computers utilise quantum bits or qubits, which can exist in a superposition of states, representing both 0 and 1 simultaneously. This property, combined with quantum entanglement and interference, enables quantum computers to process a vast number of possibilities in parallel [26].

Two primary paradigms define the current landscape of quantum computing: gate-based quantum computing and quantum annealing (QA). Gate-based systems operate similarly to classical computers, employing quantum logic gates to manipulate qubits in complex sequences. These systems are programmable and are often visualised through circuit diagrams, enabling the implementation of sophisticated quantum algorithms.

In contrast, quantum annealers are specialised machines tailored for optimisation problems. They function by encoding a mathematical problem into the physical states of qubits and evolving the system to find a configuration that minimises the problem's energy. Quantum annealing has its roots in the work of Kadowaki and Nishimori [12] and is particularly effective for problems formulated as Quadratic Unconstrained Binary optimisation (QUBO) tasks.

The promise of quantum computing lies in its potential to revolutionise fields requiring extensive computation, including optimisation, cryptography, and material science. For optimisation challenges such as the UC Problem, quantum computing offers innovative methods that could transcend the limitations of classical algorithms, paving the way for breakthroughs in efficiency and scalability [11]. In the quantum computing approaches in literature, we will see a

number of techniques used, which will be introduced in the remainder of this section.

Quantum annealing is a specialised approach to quantum computation designed to solve optimisation problems by finding the minimum of a specific energy function. This is achieved by initialising qubits in a superposition of states and allowing them to evolve under the influence of a problem-specific Hamiltonian, which encodes the optimisation objective. Unlike classical simulated annealing, QA leverages quantum phenomena such as tunnelling to explore and converge to optimal or near-optimal solutions. It is particularly effective for problems expressed in QUBO form and is implemented in devices like those developed by D-Wave Systems.

The **Quantum Approximate optimisation Algorithm** (QAOA) [6] is a hybrid quantum-classical approach for solving combinatorial optimisation problems on gate-based quantum computers. Also this approach often uses the QUBO formulation of the problem as input. QAOA approximates the process of adiabatic quantum evolution by alternately applying two parametrised quantum operators to a system of qubits. By iteratively tuning the parameters through classical optimisation, QAOA seeks to find a quantum state that minimises the problem's objective function. Its adaptability, noise resistance, and relatively low qubit requirements make it a promising candidate for near-term quantum optimisation applications.

There are several approaches within **Quantum Machine Learning** (QML). Quantum Neural Networks (QNNs) [28] combine quantum computing principles with neural network architectures, leveraging qubits and quantum gates to represent and process high-dimensional data efficiently. They exploit quantum superposition and entanglement for richer representations and potential computational speedups, with applications in classification, pattern recognition, and generative modeling. Similarly, Quantum Reinforcement Learning (QRL) [4] integrates quantum algorithms with reinforcement learning to accelerate policy optimization and improve exploration. By encoding states and actions in quantum systems, QRL aims to achieve faster convergence and better performance in domains such as robotics, game theory, and decision-making under uncertainty.

Benders decomposition [25] is a mathematical optimisation technique that divides a large-scale problem into smaller, more manageable sub-problems. This approach is particularly useful for mixed-integer programming and problems involving complex constraints. The main problem is split into a master problem, which involves high-level decision variables, and sub-problems, which handle detailed constraints and computations. Iteratively solving these components allows for efficient handling of computationally intensive problems. The method is widely applied in fields like energy systems, logistics, and supply chain optimisation and can be used to create sub-problems that are better solvable by quantum annealing or QAOA.

3 Unit Commitment Problem

The UC Problem is a critical optimisation challenge in power system operations, determining the most efficient schedule for turning power generation units on and off while meeting electricity demand and system constraints. It aims to minimise operational costs while adhering to technical constraints such as generator ramp rates, minimum up and down times, and reserve requirements for system reliability. Given its combinatorial nature and large-scale complexity, the UC problem is computationally demanding, especially in systems with high integration of renewable energy sources. Advanced computational approaches, including quantum computing, are increasingly explored to address the limitations of classical methods in solving this problem efficiently.

Knueven et al. [13] give a general UC problem formulation:

$$\min \sum_{g \in G} \sum_{t \in T} c_g(t), \tag{1}$$

$$\text{s.t.} \sum_{g \in G} A_g(p_g, \overline{p_g}, u_g) + N(s) = L, \tag{2}$$

$$(p_g, \overline{p_g}, u_g, c_g) \in \Pi_g \quad \forall g \in G. \tag{3}$$

Here, c_g is the cost vector associated with $p_g, \overline{p_g}, u_g$, such that the objective function, system operation cost, is minimised. The vectors $p_g, \overline{p_g}, u_g$ represent the feasible generation schedule, maximum power available, and the on/off status for generator g, respectively. The matrix $A_g(p_g, \overline{p_g}, u_g)$ determines how the generator interacts with the system requirements, which are written in matrix form.

A specific example formulation can be found in [7]. The objective of the UC problem is to minimise the total generation cost:

$$P_{i,t} = \text{argmin} \sum_{i=1}^{G} \sum_{t=1}^{T} (a_i P_{i,t}^2 + b_i P_{i,t} + c_i)$$

where $P_{i,t}$ is the generation of unit i at time t, a_i, b_i and c_i are coefficients of unit i-th cost function, T is the number of operation periods, and G is the number of units. The minimisation is subject to the following constraints:

- System Demand Constraints: The total generated power should meet the demand D_t at time t:

$$\sum_{i=1}^{G} P_{i,t} = D_t$$

- Generation Capacity Constraints: Generation $P_{i,t}$ of online units are constrained between minimum P_i^{min} and maximum P_i^{max}:

$$x_{i,t} P_i^{min} \leq P_{i,t} \leq x_{i,t} P_i^{max}$$

  where $x_{i,t}$ is the on/off decision variable of unit i at time t.

– Ramp-Rate Constraints: Ramp-rate constraints ensure that the change of power generation levels between two consecutive time periods does not exceed ramp rates:

$$P_{i,t} - P_{i,t-1} \leq R_i x_{i,t-1} + V_i(1 - x_{i,t-1})$$

$$P_{i,t-1} - P_{i,t} \leq R_i x_{i,t} + V_i(1 - x_{i,t})$$

where R_i is the ramp rate of unit i and V_i is the startup/shut-down ramp rate.

4 Literature overview

In this section, we delve into the various quantum computing approaches that have been explored for addressing the UC problem.

QAOA

The literature on quantum approaches to the Unit Commitment (UC) problem presents a range of hybrid and quantum-enhanced frameworks, primarily leveraging the Quantum Approximate Optimization Algorithm (QAOA) and QUBO formulations. A core trend involves decomposing the UC problem into subproblems solvable via quantum and classical methods, often within distributed or iterative schemes such as ADMM.

Magar, Nikmehr et al. [17, 20, 21, 31] reformulate the UC problem as a QUBO for QAOA and develop a distributed ADMM framework for large-scale systems. Similarly, Fontana et al. [8] address single-period UC problems using hybrid solvers and demonstrate fast performance with D-Wave's platform. Halffmann et al. [9] focus on compact QUBO formulations optimized for current quantum hardware, minimizing qubit and connectivity requirements, and later expand their work [10] to robust optimisation under uncertainty using QAOA and quantum annealing (QA).

Several papers propose hybrid ADMM frameworks. Mahroo et al. [18] decompose UC into three sub-problems with QAOA handling the QUBO part, showing convergence and improved accuracy via parameter updates. Yang et al. [30] integrate Tabu Search with ADMM and QAOA (H-ADMM-TS) to escape local optima. Salgado et al. [27] refine this hybridization by applying QAOA to binary variables while using classical solvers for continuous decisions, achieving high accuracy and scalability, especially when warm-starting QAOA.

Feng et al. [7] introduce in their paper the Quantum-Surrogate Lagrangian Relaxation (QSLR) approach for large-scale, distributed UC, emphasizing scalability, data privacy, and robustness in noisy environments. Stein et al. [29] propose the QuSO and LinQuSO frameworks, exploiting quantum simulation-based speedups using QAOA and singular value transformation. These show promise for exponential gains but rely on future advancements in quantum hardware and error correction.

Koretsky et al. [14] propose a hybrid method where QAOA handles binary decision variables, and classical optimisers address continuous power output. Reformulating UC as a QUBO problem enables efficient handling of discrete variables by quantum algorithms while classical methods ensure continuous variable optimisation. Simulations on systems with up to 10 units show improved probability of finding near-optimal solutions with increased QAOA depth. Despite current quantum hardware limitations, the study emphasises hybrid approaches' scalability advantages for larger systems.

Across the literature, QUBO reformulations and QAOA remain central, often within hybrid schemes to offset current quantum hardware limitations. Techniques such as warm-starting, robust optimisation, and integration with metaheuristics (e.g., Tabu Search) enhance solution quality and convergence. While most studies acknowledge scalability and noise challenges, results consistently demonstrate quantum methods' potential to complement or outperform classical counterparts in UC problems, especially as quantum technology matures.

Quantum Annealing

Ajagekar et al. [1] demonstrates the implementation of the UC problem on D-Wave's quantum annealer by discretising continuous variables into equally spaced grid points and reformulating the problem as a QUBO model. Their results show that smaller instances could be efficiently solved on the quantum processor, achieving solutions close to the global optimum. However, larger instances required partitioning into sub-problems, which impacted solution quality. They emphasised the trade-off between solution precision and the likelihood of obtaining global optima, concluding that advancements in quantum hardware and algorithms are necessary for solving larger and more complex problems effectively.

Building on this, Quinton et al. [24] benchmarked D-Wave's hybrid quantum annealing solver against classical solvers such as CPLEX, Gurobi, and IPOPT across several optimisation problems, including the UC problem. They confirmed that while quantum annealers show promise for small-scale UC problem instances, larger problems necessitate hybrid approaches that combine classical and quantum methods. The study underscores the potential of quantum annealing for optimisation but calls for further technological advancements to fully leverage its benefits for industrial applications.

Braun et al. [2] extended the exploration of quantum computing for the UC problem by introducing a QUBO formulation capable of addressing uncertainties in renewable energy supply, power demand, and machine failures. This formulation integrates probabilistic modelling of future scenarios, enabling optimisation that considers both cost and variance minimisation. Results showed that smaller UC problem instances could be effectively solved using D-Wave's quantum annealer, while larger problems faced the same partitioning challenges observed in previous studies. The authors highlighted the need for continued innovation in quantum hardware and algorithms to manage the complexity of large-scale,

real-world UC problems effectively.

Müller [19] focuses on balancing intermittent generation and compares two methods for encoding inequality constraints in QUBO formulations: the traditional slack-variable approach and the unbalanced penalization technique. Computational experiments show that unbalanced penalization reduces the number of variables and outperforms the slack-variable method, but its uneven penalty on feasible solutions limits scalability.

Quantum Machine Learning

Liu et al. [16] introduce the Partially Connected Quantum Neural Network (PCQNN), a knowledge-based quantum-classical model designed to efficiently solve the UC problem. By reducing network connectivity based on domain knowledge, PCQNN improves precision and reduces circuit depth compared to traditional QNNs. Tested on 5- and 10-unit systems, it achieves near-optimal and exact solutions with fewer quantum layers, making it suitable for distributed quantum environments and scalable to larger problems using a divide-and-conquer strategy.

Kruse et al. [15] propose a Hamiltonian-based Quantum Reinforcement Learning (QRL) method, combining quantum computing with Neural Combinatorial Optimisation. By directly modeling the UC problem's Hamiltonian in a variational quantum circuit, the approach improves trainability and general applicability beyond graph-based problems, offering a versatile framework for complex combinatorial optimisation tasks.

Hybrid Approach based on Benders Decomposition

Chang et al. [3] propose a hybrid algorithm combining quantum computing and classical computing through Benders decomposition. Their method splits MIPs into binary programming problems, handled by a so-called noisy intermediate-scale quantum (NISQ) [23] processor, and linear programming problems, solved on a classical processor. Tested on the D-Wave 2000Q, the approach demonstrated promise for small-scale cases and power system-inspired problems but faced challenges with quantum annealing hardware limitations and the reformulation of inequality constraints into QUBOs.

Paterakis et al. [22] introduce a hybrid framework for MILPs in power system optimisation using a multi-cut Benders decomposition scheme. This method applies QC to generate multiple valid cuts, accelerating convergence for problems like Unit Commitment. While Quantum Annealers showed faster computation for small-scale problems, noise and scalability remain barriers for larger problems. The study underscores the need for error mitigation and improved scalability, offering a Python tutorial to aid further research.

Ellinas et al. [5] expand on the hybrid paradigm for power system optimisation, addressing Optimal Transmission Switching and Neural Network verification for DC Optimal Power Flow. Their results align with previous findings, demonstrating QC's potential to enhance computation times for small-scale problems but encountering noise and scalability constraints for larger instances.

The study reiterates the importance of developing error mitigation techniques and scaling strategies while providing community resources to advance the field.

Table 1. Overview of the literature. The papers are clusters into their main topic: QAOA, Simulated Annelling, Benders Decomposition and Quantum Machine Learning. For each paper is indicated which of the other paper it cites.

Topic	Article	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
QAOA	1 Koretsky 2021 [14]													x								
QAOA	2 Nikmehr 2022a [20]																					
QAOA	3 Feng 2022 [7]		x																			
QAOA	4 Nikmehr 2022b [21]		x																			
QAOA	5 Mahroo 2022 [18]													x								
QAOA	6 Stein 2023 [29]																					
QAOA	7 Yang 2023 [30]																					
QAOA	8 Salgado 2024 [27]	x	x	x		x	x							x								
QAOA	9 Fontana 2024 [8]	x				x								x								
QAOA	10 Magar 2024 [17]	x			x			x						x								
QAOA/SA	11 Halffmann 2022 [9]													x								
QAOA/SA	12 Halffmann 2024 [10]	x											x									
SA	13 Ajagekar 2020 [1]																					
SA	14 Braun 2023 [2]												x									
SA	15 Quinton 2024 [24]																					
SA	16 Müller 2025 [19]	x	x					x	x	x				x					x	x	x	
BDC/SA	17 Chang 2020 [3]																					
BDC/SA	18 Paterakis 2023 [22]	x			x	x								x								
BDC/SA	19 Ellinas 2024 [5]		x			x													x			
QML	20 Liu 2024 [16]	x				x		x												x		
QML	21 Kruse 2024 [15]	x																				

5 Conclusions

In this paper, we have explored the application of quantum computing to the Unit Commitment problem, a critical optimisation challenge in power system operations. Our review of the literature highlights several promising quantum approaches, including the Quantum Approximate Optimisation Algorithm, Quantum Annealing (QA), and hybrid quantum-classical methods.

The findings suggest that quantum computing holds significant potential for solving the UC problem more efficiently than classical methods in the future. QAOA, in particular, has shown promise in handling binary decision variables, while classical optimisers address continuous power output. Hybrid approaches that combine quantum and classical techniques have demonstrated improved computational efficiency and scalability, making them suitable for large-scale UC scenarios. Quantum Annealing has also been effective for smaller UC instances, although larger problems require partitioning into sub-problems. The integration of quantum algorithms with robust optimisation techniques has shown potential in managing uncertainties in power systems, such as renewable energy supply and power demand fluctuations.

Despite the current limitations of quantum hardware, such as noise and scalability issues, the advancements in quantum algorithms and hybrid methods

provide a strong foundation for future research. Continued innovation in quantum hardware and error mitigation techniques will be crucial in realising the full potential of quantum computing for the UC problem.

In conclusion, quantum computing offers a transformative approach to solving the UC problem, with the potential to significantly enhance the efficiency and scalability of power system optimisation. Future research should focus on further developing quantum algorithms, improving hardware capabilities, and exploring new hybrid methods to fully leverage the benefits of quantum computing in this domain.

References

1. Ajagekar, A., You, F.: Quantum computing for energy systems optimization: Challenges and opportunities. Energy **179**, 76–89 (2019)
2. Braun, M., Decker, T., Hegemann, N., Kerstan, S., Lorenz, F.: Towards optimization under uncertainty for fundamental models in energy markets using quantum computers. arXiv preprint arXiv:2301.01108 (2023)
3. Chang, C.Y., Jones, E., Yao, Y., Graf, P., Jain, R.: On hybrid quantum and classical computing algorithms for mixed-integer programming. arXiv preprint arXiv:2010.07852 (2020)
4. Dong, D., Chen, C., Li, H., Tarn, T.J.: Quantum reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **38**(5), 1207–1220 (2008)
5. Ellinas, P., Chevalier, S., Chatzivasileiadis, S.: A hybrid quantum–classical algorithm for mixed-integer optimization in power systems. Electric Power Systems Research **235**, 110835 (2024)
6. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014)
7. Feng, F., Zhang, P., Bragin, M.A., Zhou, Y.: Novel resolution of unit commitment problems through quantum surrogate lagrangian relaxation. IEEE Transactions on Power Systems **38**(3), 2460–2471 (2022)
8. Fontana, C.: Energy Management Systems and Potential Applications of Quantum Computing in the Energy Sector. Ph.D. thesis, University of Palermo (2024)
9. Halffmann, P., Holzer, P., Plociennik, K., Trebing, M.: A quantum computing approach for the unit commitment problem. In: International Conference on Operations Research. pp. 113–120. Springer (2022)
10. Halffmann, P., Trebing, M., Lenk, S.: Harnessing inferior solutions for superior outcomes: Obtaining robust solutions from quantum algorithms. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 1950–1953 (2024)
11. Horowitz, M., Grumbling, E.: Quantum computing: progress and prospects. National Academies Press (2019)
12. Kadowaki, T., Nishimori, H.: Quantum annealing in the transverse Ising model. Physical Review E **58**(5), 5355 (1998)
13. Knueven, B., Ostrowski, J., Watson, J.P.: On mixed-integer programming formulations for the unit commitment problem. INFORMS Journal on Computing **32**(4), 857–876 (2020)
14. Koretsky, S., Gokhale, P., Baker, J.M., Viszlai, J., Zheng, H., Gurung, N., Burg, R., Paaso, E.A., Khodaei, A., Eskandarpour, R., et al.: Adapting quantum approximation optimization algorithm (qaoa) for unit commitment. In: 2021 IEEE

International Conference on Quantum Computing and Engineering (QCE). pp. 181–187. IEEE (2021)
15. Kruse, G., Coelho, R., Rosskopf, A., Wille, R., Lorenz, J.M.: Hamiltonian-based quantum reinforcement learning for neural combinatorial optimization. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). vol. 1, pp. 1617–1627. IEEE (2024)
16. Liu, J., Zhou, X., Zhou, Z., Luo, L.: Exact quantum algorithm for unit commitment optimization based on partially connected quantum neural networks. arXiv preprint arXiv:2411.11369 (2024)
17. Magar, M.R., Pandit, D., Nguyen, D., Nguyen, N.: DC optimal power flow in unit commitment using quantum computing: An admm approach. In: 2024 56th North American Power Symposium (NAPS). pp. 1–6. IEEE (2024)
18. Mahroo, R., Kargarian, A.: Hybrid quantum-classical unit commitment. In: 2022 IEEE Texas Power and Energy Conference (TPEC). pp. 1–5. IEEE (2022)
19. Müller, S., Dukalski, M., Phillipson, F.: Quantum annealing for optimizing unit scheduling in renewable energy systems: Formulation and evaluation. IEEE Transactions on Power Systems (2025)
20. Nikmehr, N., Zhang, P., Bragin, M.A.: Quantum distributed unit commitment: An application in microgrids. IEEE transactions on power systems **37**(5), 3592–3603 (2022)
21. Nikmehr, N., Zhang, P., Bragin, M.A.: Quantum-enabled distributed unit commitment. In: 2022 IEEE Power & Energy Society General Meeting (PESGM). pp. 01–05. IEEE (2022)
22. Paterakis, N.G.: Hybrid quantum-classical multi-cut Benders approach with a power system application. Computers & Chemical Engineering **172**, 108161 (2023)
23. Preskill, J.: Quantum computing in the NISQ era and beyond. Quantum **2**, 79 (2018)
24. Quinton, F.A., Myhr, P.A.S., Barani, M., del Granado, P.C., Zhang, H.: Quantum annealing versus classical solvers: Applications, challenges and limitations for optimisation problems. arXiv preprint arXiv:2409.05542 (2024)
25. Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The Benders decomposition algorithm: A literature review. European Journal of Operational Research **259**(3), 801–817 (2017)
26. Rietsche, R., Dremel, C., Bosch, S., Steinacker, L., Meckel, M., Leimeister, J.M.: Quantum computing. Electronic Markets **32**(4), 2525–2536 (2022)
27. Salgado, B., Sequeira, A., Santos, L.P.: A hybrid classical-quantum approach to highly constrained unit commitment problems. arXiv preprint arXiv:2412.11312 (2024)
28. Schuld, M., Sinayskiy, I., Petruccione, F.: The quest for a quantum neural network. Quantum Information Processing **13**, 2567–2586 (2014)
29. Stein, J., Müller, L., Hölscher, L., Chnitidis, G., Jojo, J., Farea, A., Çelebi, M.S.: Exponential quantum speedup for simulation-based optimization applications. arXiv preprint arXiv:2305.08482 (2023)
30. Yang, M., Gao, F., Wu, G., Gong, B., Cao, H., Shuang, F.: A quantum-aided algorithm for unit commitment problems. In: 2023 42nd Chinese Control Conference (CCC). pp. 1733–1736. IEEE (2023)
31. Zhang, P., Nikmehr, N., Nikmehr, N., Zhang, P., Bragin, M.: Quantum-enabled distributed unit commitment. Tech. rep., Brookhaven National Lab.(BNL), Upton, NY (United States) (2022)

Toward Quantum Utility in Finance: A Robust Data-Driven Algorithm for Asset Clustering

Shivam Sharma[1][0009−0001−6580−1058], Supreeth Mysore Venkatesh[2,3][0000−0002−9824−7399], and Pushkin Kachroo[1][0000−0002−7701−8925]

[1] University of Nevada, Las Vegas, Nevada, USA
sharms15@unlv.nevada.edu, pushkin@unlv.edu
[2] University of Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany
supreeth.mysore@rptu.de
[3] German Research Center for Artificial Intelligence (DFKI), Saarbruecken, Germany
supreeth.mysore@dfki.de

Abstract. Clustering financial assets based on return correlations is a fundamental task in portfolio optimization and statistical arbitrage. However, classical clustering methods often fall short when dealing with signed correlation structures, typically requiring lossy transformations and heuristic assumptions such as a fixed number of clusters. In this work, we apply the Graph-based Coalition Structure Generation algorithm (GCS-Q) to directly cluster signed, weighted graphs without relying on such transformations. GCS-Q formulates each partitioning step as a QUBO problem, enabling it to leverage quantum annealing for efficient exploration of exponentially large solution spaces. We validate our approach on both synthetic and real-world financial data, benchmarking against state-of-the-art classical algorithms such as SPONGE and k-Medoids. Our experiments demonstrate that GCS-Q consistently achieves higher clustering quality, as measured by Adjusted Rand Index and structural balance penalties, while dynamically determining the number of clusters. These results highlight the practical utility of near-term quantum computing for graph-based unsupervised learning in financial applications.

Keywords: Quantum Utility · Signed Graphs · Graph Clustering · Asset Clustering · Quantum Annealing

1 Introduction and Background

Information about relationships among financial assets has long been central to strategies in portfolio optimization and algorithmic trading. Since companies and their stock prices are interdependent, exploiting these correlations, especially in markets that are not perfectly efficient, can enhance decision-making in portfolio construction. The foundational work of Markowitz [18] introduced the mean-variance optimization framework, establishing diversification as a means to maximize return while minimizing risk [23].

More recent strategies, such as statistical arbitrage [2] and pair trading [5], use correlation or cointegration metrics but have become less effective due to increased market adoption. Consequently, graph-theoretic methods have emerged as robust alternatives [13], where assets are represented as nodes and correlation matrices are treated as signed, weighted graphs. Clustering these graphs aims to group assets with high intra-cluster and low inter-cluster correlations [14,28], recent works focusing on identifying the best-performing clustering algorithms [4].

Traditional clustering approaches such as k-Means, k-Medoids [10,15], or hierarchical clustering [19] face limitations when applied to signed graphs. These algorithms require positive-definite distance metrics and often rely on transforming signed correlation matrices $\rho \in [-1, 1]$ into non-negative distances using monotonic mappings like [7,17,29]:

$$d_{ij}^{(\alpha)} = \sqrt{\alpha(1 - \rho_{ij})}, \quad \alpha > 0 \tag{1}$$

Although this transformation preserves ranking, it loses semantic fidelity, for example, $\rho_{ij} = 0$ is mapped to $\sqrt{\alpha}$, incorrectly implying a fixed non-zero dissimilarity. Thus, even optimally solving the transformed problem does not recover the optimal clustering under the original signed graph. Centroid-based methods further deviate from the true objective, which is to maximize intra-cluster correlations and minimize inter-cluster correlations, not to minimize distances to centroids. Moreover, these classical methods require manual tuning of hyperparameters such as the number of clusters k, initialization strategies, or threshold values for merging/dividing. These choices are not generalizable across datasets and must be revalidated for unseen data. Spectral clustering methods provide a more natural fit for graph-structured data by using eigenvalue decomposition of the similarity matrix [24]. SPONGE [6], a spectral method for signed graphs, has shown strong performance but is effective mainly when k is large or the graph is very sparse [4,11], conditions not typically met in financial data.

Quantum methods have also been explored, notably in portfolio optimization using gate-based variational circuits [3], though existing hardware limit scalability. One prior attempt at quantum-assisted clustering formulates the problem as a maximum clique and translates it into a large QUBO instance [9], introducing additional binary variables and exceeding the capabilities of current annealers.

In this work, we propose applying the Graph-based Coalition Structure Generation algorithm (GCS-Q) [31] to the task of asset clustering using return correlation data. GCS-Q operates directly on the signed graph, avoiding lossy transformations. It starts with all assets grouped together and iteratively partitions the graph via minimum cuts, maximizing intra-cluster weights at each step and dynamically determining k. This approach is well-suited to the signed graph setting where classical solvers offer only heuristic workarounds. The minimum-cut subproblem is formulated as a QUBO and solved on D-Wave's quantum annealer, leveraging its ability to explore exponentially large solution spaces efficiently. We validate GCS-Q on both synthetic and real-world financial datasets, demonstrating its superior clustering quality compared to classical state-of-the-art methods.

2 Methodology

We consider a set of financial assets $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$, whose historical returns over a rolling window are denoted by $r_i(t)$ for asset a_i at time t. The pairwise Pearson correlation coefficients [22] are computed as:

$$\rho_{ij} = \frac{\mathrm{Cov}(r_i, r_j)}{\sigma_{r_i}\sigma_{r_j}}, \quad \rho_{ij} \in [-1, 1]. \tag{2}$$

These correlations define a signed, weighted, undirected graph $G = (V, E, w)$, where each vertex $v_i \in V$ uniquely corresponds to asset a_i, and each edge $(v_i, v_j) \in E$ is weighted by $w_{ij} = \rho_{ij}$. In contrast to traditional clustering methods that transform correlations into non-negative distances [7], we retain the signed correlations, thereby preserving both positive (co-movement) and negative (anti-correlation) relationships.

We apply the GCS-Q algorithm, previously developed for coalition structure generation in graphs [31], to this asset clustering setting. The core contribution lies in adapting GCS-Q to leverage its ability to operate directly on the signed graphs without resorting to heuristic transformations.

Formally, GCS-Q aims to find a coalition structure $\Pi = \{C_1, C_2, \ldots, C_k\}$, a partition of V into disjoint subsets, that maximizes the intra-cluster edge weights:

$$\max_{\Pi} \sum_{C \in \Pi} \sum_{\substack{i,j \in C \\ i<j}} w_{ij}. \tag{3}$$

This objective promotes grouping of assets with strong positive correlations while implicitly disincentivizing inclusion of negatively correlated pairs within the same cluster.

The GCS-Q algorithm proceeds iteratively, starting with the full graph $G_0 = G$. At each iteration t, it computes a minimum cut $(S_t, \bar{S}_t)$ on the current subgraph G_{t-1}, using the actual edge weights w_{ij} which is strictly NP-Hard. The resulting partition optimizes the objective in Eq. (3). The process is then recursively applied to the subgraphs induced by S_t and $\bar{S}_t$, eliminating the cut edges at each step, and terminates dividing a subgraph further when there is no cut whose value is lower than the sum of all the edges in that subgraph. This stopping criterion inherently infers the number of clusters k as part of the optimization, obviating the need for manual tuning or initialization required by classical approaches.

Each minimum cut in GCS-Q is formulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem [8], well-suited for near-term quantum devices. These QUBOs can be efficiently solved using quantum annealers (e.g., D-Wave) or variational algorithms like the Quantum Approximate Optimization Algorithm (QAOA) [32], enabling hybrid quantum-classical workflows.

Theoretical work such as [1] has shown that optimally clustering signed graphs with edge weights in $[-1, +1]$ is NP-Hard. Our practical contribution lies in demonstrating that GCS-Q can robustly address this challenge in financial asset clustering, highlighting its potential for real-world deployment and

showcasing practical applicability of near-term quantum technologies, particularly quantum annealing.

3 Experiments

We conduct two sets of experiments. First, we benchmark the clustering quality of GCS-Q against classical baselines on synthetically generated signed graphs designed to mimic financial correlation structures. Second, we evaluate the methodology on real-world asset return data to demonstrate its practical utility.

All implementations are in `Python 3.12`. The QUBO subproblems in GCS-Q are solved using the D-Wave `Advantage_system5.4` annealer, which comprises 5614 physical qubits and 40,050 couplers arranged in the pegasus topology. The device is accessed remotely via the `dwave-ocean-sdk` APIs. As noted in [30], most runtime overhead arises from internet latency and queue wait times due to the lack of onsite quantum hardware. As such, the experiments emphasize solution quality to assess the practical relevance of the methodology using exisitng quantum annealers.

All remaining computations are executed on a standard 64-bit workstation equipped with a 12th Gen Intel Core i7-12800H CPU and 64 GB RAM.

3.1 Synthetic Data

We generate structurally balanced signed graphs, sampling intra-cluster edges from $[0.1, 1.0]$ and inter-cluster edges from $[-1.0, -0.1]$ to emulate realistic correlation patterns. Unlike sparse graphs in [6] that use discrete weights $\in \{-1, 0, +1\}$ and equal cluster sizes, our graphs span a continuous range and incorporate non-uniform cluster sizes via Dirichlet-based stochastic allocation. This introduces heterogeneity while ensuring all clusters remain non-empty. No noise is added, isolating algorithmic behavior under ideal structural conditions. We extrapolate the setup later in Sec. 3.2 to evaluate robustness under noise.

We benchmark GCS-Q against classical baselines commonly used for clustering assets via return correlation data. SPONGE and its symmetric variant $\text{SPONGE}_{\text{sym}}$ [6] are among the strongest graph-based techniques [4]. We also include k-Medoids (via the PAM algorithm) as a robust baseline well-suited to financial data [7].

Unlike GCS-Q, which determines the number of clusters k dynamically, SPONGE and PAM require k as input. This requirement often demands heuristics like silhouette analysis [25] or the elbow method [27], which are expensive and lack generalizability. Since SPONGE parallels spectral clustering, we estimate k using the eigengap heuristic [12], which selects k based on the largest gap between successive Laplacian eigenvalues.

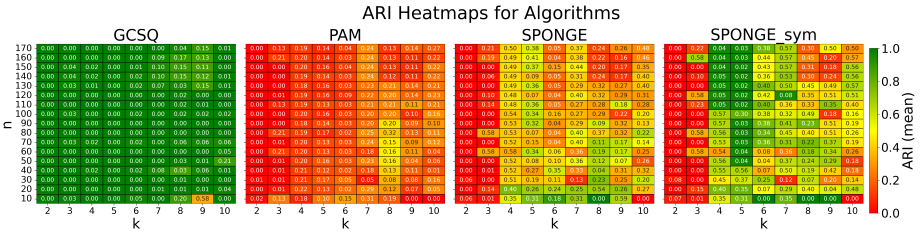


Fig. 1. ARI scores for synthetic data of varying number of nodes n and number of clusters k. Color represents mean and the (tiny) numerical annotations indicate variances across seeds.

We conduct experiments varying total nodes $\in \{10, 20, \ldots, 100\}$, with k ranging from 2 to 10 across 33 random seeds. Since ground-truth clusters are known, we assess accuracy using Adjusted Rand Index (ARI) [26].

Although GCS-Q is limited to roughly 175 nodes due to qubit constraints and hardware topology, Fig. 1 shows it consistently outperforms classical solvers. PAM performs worst, largely due to its reliance on transforming correlations into non-negative distances. While SPONGE and $\text{SPONGE}_{\text{sym}}$ perform well on graphs with weights in $\{-1, 0, 1\}$ and balanced cluster sizes, their performance degrades on continuous weights and stochastic sizes.

Moreover, SPONGE is designed for sparse graphs [4], which is misaligned with financial data where correlations close to zero are rare. The superior performance of GCS-Q is attributed to the expressive solution space explored by the quantum annealer at each step.

3.2 Yahoo Finance Data

We collect hourly closing prices of 50 assets from diverse sectors from `yfinance`, compute rolling log returns, and derive the Pearson correlation matrix, which is interpreted as a signed adjacency matrix. The same clustering solvers from Sec. 3.1 are applied. In the absence of ground truth, we assess clustering quality using the Penalty metric, which quantifies deviation from structural balance by penalizing negative intra-cluster correlations and positive inter-cluster correlations:

$$\text{Penalty}(\Pi) = \sum_{C \in \Pi} \sum_{\substack{i,j \in C \\ i<j \\ w_{ij}<0}} |w_{ij}| + \sum_{C_a \neq C_b} \sum_{\substack{i \in C_a, j \in C_b \\ i<j \\ w_{ij}>0}} w_{ij}$$

A lower penalty indicates better clustering, with zero representing a perfectly balanced structure. Such structurally sound clusters enable more effective asset diversification, aligning with the goal of minimizing portfolio variance in the Markowitz framework [18].

As Fig. 2 illustrates, GCS-Q consistently achieves the lowest penalty, indicating its superiority in uncovering market structures conducive to portfolio

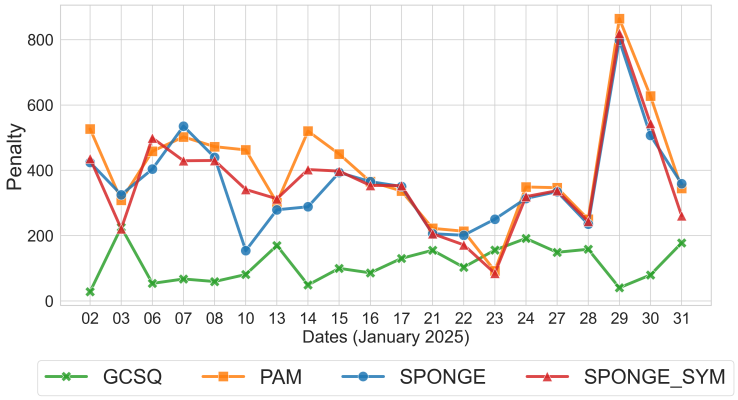


Fig. 2. Penalty comparison across clustering algorithms over business days in January 2025, based on correlations derived from hourly return intervals.

optimization and statistical arbitrage. Low penalty values indicate clusters that are internally cohesive, which supports mean-reversion strategies [4], and externally distinct, making them well-suited for capturing diverse risk-return profiles essential for effective diversification [23].

4 Discussion and Conclusion

This work is the first, to our knowledge, to demonstrate a quantum advantage in solution quality using real hardware for the practical task of asset clustering, validated on both synthetic and real financial data. GCS-Q's strength lies in its exponential solution space exploration at each iteration. When executed classically, its complexity is $\mathcal{O}(n2^n)$, where $\mathcal{O}(2^n)$ stems from solving the minimum cut and $\mathcal{O}(n)$ from dynamically determining $k \in [1, n]$. While commercial solvers like Gurobi and CPLEX exist for combinatorial optimization, they struggle to scale on dense graphs [33] and varying-sized clusters [16]. Although the adiabatic quantum process theoretically runs in constant time, apart from the hardware noise, current access via shared cloud infrastructure introduces latency and queue delays. In practice, clustering $n = 170$ assets with $k = 10$ took over 10 minutes on average, despite per-QUBO solve times being capped at 1 second.

GCS-Q is robust to real-world scenarios with heterogeneous cluster sizes and avoids the idealized constraints (e.g., uniform clusters, ternary edge weights) assumed in prior work [4,6]. It requires no manual tuning of k, operating in a fully data-driven manner. Unlike classical methods that convert signed correlations to Euclidean distances [7,17], GCS-Q works directly on the signed graph, preserving relational structure.

Future directions include applying the resulting clusters to downstream tasks like asset allocation and evaluating performance using metrics such as Sharpe Ratio and returns. From a quantum perspective, deploying GCS-Q on D-Wave's

hardware with Zephyr topology and better connectivity, exploring custom embeddings, and using gate-model quantum solvers with qubit-efficient strategies [21], as well as benchmarking against classical optimizers like Gurobi, are promising next steps. Moreover, similar iterative approaches are also used in optimizing machine learning models [20]. This work, we believe, marks a concrete advancement toward near-term quantum utility for signed graph clustering with an imperative application in finance.

Code Availability: All code necessary to reproduce the experiments and generate the plots are available at:
https://github.com/supreethmv/Quantum-Asset-Clustering

References

1. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning **56**(1), 89–113 (Jul 2004). https://doi.org/10.1023/B:MACH.0000033116.57574.95
2. Bertram, W.K.: Analytic solutions for optimal statistical arbitrage trading. Physica A: Statistical Mechanics and its Applications **389**(11), 2234–2243 (2010). https://doi.org/https://doi.org/10.1016/j.physa.2010.01.045
3. Buonaiuto, G., Gargiulo, F., De Pietro, G., Esposito, M., Pota, M.: Best practices for portfolio optimization by quantum computing, experimented on real quantum devices. Scientific Reports **13**(1), 19434 (Nov 2023). https://doi.org/10.1038/s41598-023-45392-w
4. Cartea, Á., Cucuringu, M., Jin, Q.: Correlation matrix clustering for statistical arbitrage portfolios. SSRN Electron. J. (2023)
5. CARTEA, A., JAIMUNGAL, S.: Algorithmic trading of co-integrated assets. International Journal of Theoretical and Applied Finance **19**(06), 1650038 (2016). https://doi.org/10.1142/S0219024916500382
6. Cucuringu, M., Davies, P., Glielmo, A., Tyagi, H.: Sponge: A generalized eigenproblem for clustering signed networks. In: Chaudhuri, K., Sugiyama, M. (eds.) Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 89, pp. 1088–1098. PMLR (2019), https://proceedings.mlr.press/v89/cucuringu19a.html
7. Duarte, F.G., de Castro., L.N.: Asset allocation based on a partitional clustering algorithm. In: Filho, C.J.A.B., Siqueira, H.V., Ferreira, D.D., Bertol, D.W., ao de Oliveira, R.C.L. (eds.) Anais do 15 Congresso Brasileiro de Inteligência Computacional. pp. 1–6. SBIC, Joinville, SC (2021)
8. Glover, F., Kochenberger, G., Du, Y.: Quantum Bridge Analytics I: a tutorial on formulating and using QUBO models. 4OR **17**(4), 335–371 (2019). https://doi.org/10.1007/s10288-019-00424-y
9. Kalra, A., Qureshi, F., Tisi, M.: Portfolio asset identification using graph algorithms on a quantum annealer. SSRN Electronic Journal (01 2018). https://doi.org/10.2139/ssrn.3333537
10. Kaufman, L., Rousseeuw, P.J.: Partitioning around medoids (program pam). In: Wiley Series in Probability and Statistics, pp. 68–125. John Wiley & Sons, Inc., Hoboken, NJ, USA (1990). https://doi.org/10.1002/9780470316801.ch2

11. Khelifa, N., Allier, J., Cucuringu, M.: Cluster-driven hierarchical representation of large asset universes for optimal portfolio construction. In: Proceedings of the 5th ACM International Conference on AI in Finance. p. 177–185. ICAIF '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3677052.3698676
12. Kong, W., Sun, C., Hu, S., Zhang, J.: Automatic spectral clustering and its application. In: ICICTA 2010. vol. 1, pp. 841–845 (2010). https://doi.org/10.1109/ICICTA.2010.164
13. Korniejczuk, A., Ślepaczuk, R.: Statistical arbitrage in multi-pair trading strategy based on graph clustering algorithms in US equities market. Working Papers 2024-09, Faculty of Economic Sciences, University of Warsaw (2024), https://ideas.repec.org/p/war/wpaper/2024-09.html
14. León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., Niño, J.: Clustering algorithms for risk-adjusted portfolio construction. Procedia Computer Science **108**, 1334–1343 (2017). https://doi.org/https://doi.org/10.1016/j.procs.2017.05.185, international Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland
15. Lloyd, S.: Least squares quantization in pcm. IEEE Transactions on Information Theory **28**(2), 129–137 (1982). https://doi.org/10.1109/TIT.1982.1056489
16. Macaluso, A., Venkatesh, S.M., Arenas, D., Klusch, M., Dengel, A.: Quantum-assisted correlation clustering. arXiv preprint arXiv:2509.03561 (2025)
17. Mantegna, R.N.: Hierarchical structure in financial markets. The European Physical Journal B - Condensed Matter and Complex Systems **11**(1), 193–197 (Sep 1999). https://doi.org/10.1007/s100510050929
18. Markowitz, H.: Portfolio Selection. Journal of Finance **7**(1), 77–91 (March 1952). https://doi.org/j.1540-6261.1952.tb01525.x
19. Murtagh, F., Contreras, P.: Algorithms for hierarchical clustering: an overview. WIREs Data Mining and Knowledge Discovery **2**(1), 86–97 (2012). https://doi.org/https://doi.org/10.1002/widm.53
20. Mysore Venkatesh, S., Macaluso, A., Arenas, D., Klusch, M., Dengel, A.: i-qls: Quantum-supported algorithm for least squares optimization in non-linear regression. In: Lees, M.H., Cai, W., Cheong, S.A., Su, Y., Abramson, D., Dongarra, J.J., Sloot, P.M.A. (eds.) Computational Science – ICCS 2025. pp. 19–34. Springer Nature Switzerland, Cham (2025)
21. Mysore Venkatesh, S., Macaluso, A., Nuske, M., Klusch, M., Dengel, A.: Qubit-efficient variational quantum algorithms for image segmentation. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). vol. 01, pp. 450–456 (2024). https://doi.org/10.1109/QCE60285.2024.00059
22. Pearson, K.: Vii. note on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London **58**, 240–242 (Jan 1895). https://doi.org/10.1098/rspl.1895.0041, publisher: Royal Society
23. Ponsich, A., Jaimes, A.L., Coello, C.A.C.: A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. IEEE Transactions on Evolutionary Computation **17**(3), 321–344 (2013). https://doi.org/10.1109/TEVC.2012.2196800
24. Pothen, A., Simon, H.D., Liou, K.P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM Journal on Matrix Analysis and Applications **11**(3), 430–452 (1990). https://doi.org/10.1137/0611030
25. Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics **20**, 53–65 (1987). https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7

26. Steinley, D.: Properties of the hubert-arable adjusted rand index. Psychological Methods **9**(3), 386–396 (2004). https://doi.org/10.1037/1082-989X.9.3.386
27. Thorndike, R.L.: Who belongs in the family? Psychometrika **18**(4), 267–276 (1953). https://doi.org/10.1007/BF02289263
28. Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N.: Cluster analysis for portfolio optimization. Journal of Economic Dynamics and Control **32**(1), 235–258 (2008). https://doi.org/https://doi.org/10.1016/j.jedc.2007.01.034
29. Traag, V.A., Bruggeman, J.: Community detection in networks with positive and negative links. Phys. Rev. E **80**, 036115 (Sep 2009). https://doi.org/10.1103/PhysRevE.80.036115
30. Venkatesh, S., Macaluso, A., Nuske, M., Klusch, M., Dengel, A.: Q-seg: Quantum annealing-based unsupervised image segmentation. IEEE Computer Graphics and Applications **44**(5), 27–39 (2024). https://doi.org/10.1109/MCG.2024.3455012
31. Venkatesh, S.M., Macaluso, A., Klusch, M.: Gcs-q: Quantum graph coalition structure generation. In: Computational Science – ICCS 2023. pp. 138–152. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-36030-5_11
32. Venkatesh, S.M., Macaluso, A., Klusch, M.: Quacs: Variational quantum algorithm for coalition structure generation in induced subgraph games. In: 20th ACM International Conference on Computing Frontiers. p. 197–200. Association for Computing Machinery (2023). https://doi.org/10.1145/3587135.3592192
33. Venkatesh, S.M., Macaluso, A., Nuske, M., Klusch, M., Dengel, A.: Quantum annealing-based algorithm for efficient coalition formation among leo satellites. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE) (2024). https://doi.org/10.1109/QCE60285.2024.10279

Hot-Starting Quantum Portfolio Optimization

Sebastian Schlütter[1], Tomislav Maras[2],
Alexander Dotterweich[2], and Nico Piatkowski[3](✉)

[1] Mainz University of Applied Sciences, 55128 Mainz, Germany
[2] PricewaterhouseCoopers GmbH, 80636 München, Germany
[3] LAMARR Institute, Fraunhofer IAIS, 53757 Sankt Augustin, Germany

Abstract. Combinatorial optimization with a smooth and convex objective function arises naturally in applications such as discrete mean-variance portfolio optimization, where assets must be traded in integer quantities. Although optimal solutions to the associated smooth problem can be computed efficiently, existing adiabatic quantum optimization methods cannot leverage this information. Moreover, while various warm-starting strategies have been proposed for gate-based quantum optimization, none of them explicitly integrate insights from the relaxed continuous solution into the QUBO formulation. In this work, a novel approach is introduced that restricts the search space to discrete solutions in the vicinity of the continuous optimum by constructing a compact Hilbert space, thereby reducing the number of required qubits. Experiments on software solvers and a D-Wave Advantage quantum annealer demonstrate that our method outperforms state-of-the-art techniques.

Keywords: Warm-Start · QUBO · Mean-Variance Analysis.

1 Introduction

The integration of prior knowledge into mathematical models is essential. Examples include: (I) informed machine learning [28], where knowledge about physical laws allows for useful models even if training data is scarce. And (II) numerical optimization, where the convergence rate of gradient descent can be shown to depend on the squared Euclidean distance between the initial solution $\boldsymbol{x}^{(0)}$ and the optimal solution $\boldsymbol{x}^*$ [20]. In the context of quantum computing, integrating prior knowledge about the solution is known as warm-start [2, 6]. An outstanding discussion of the topic can be found in [29].

In case of *Adiabatic Quantum Computation* (AQC) warm-start techniques act on the annealing schedule as the entry point since the initial state is built into the hardware and cannot be influenced—state preparation methods like [6] cannot be applied. Only a handful of practical methods is known, e.g. for training ML models to generate annealing schedules [5, 14] as well as techniques that rely on classical pre-processing [16, 29]. They can be seen as annealing equivalents

✉ Corresponding author: nico.piatkowski@iais.fraunhofer.de

to the techniques focusing on the parameter initialization for variational circuit, likewise pre-computing or generating annealing schedules instead of initial variational parameters.

At the core of our work lies the observation that some hard combinatorial optimization problems correspond to a surprisingly simple problem when allowing for continuous solutions. For example, *Quadratic Unconstrained Binary Optimization* (QUBO) with objective $f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}$ with $\boldsymbol{x} \in \{0,1\}^n$ is NP-hard [25]. However, when $\boldsymbol{x}$ comes from a simplex, the problem becomes a convex *Quadratic Program* (QP) that can be solved to optimality in polynomial time. Here, this observation is utilized by constructing a new QUBO that incorporates the location of the QP's optimal solution $\bar{\boldsymbol{x}}^* \in \mathbb{R}^n$. This approach allows us to restrict the discrete search space to solutions in the direct neighborhood of $\bar{\boldsymbol{x}}^* \in \mathbb{R}^n$, which results in a lower-dimensional Hilbert-space and thus in a reduced number of qubits when compared to the standard QUBO formulation. The situation is depicted in Fig. 1. The search space is endogenously determined to guarantee that it contains the optimum; in general it is not sufficient to round up or down the continuous solution.

Existing quantum approaches to portfolio optimization either model binary inclusion variables per asset [6, 30], or assume predefined investment bands of lower and upper bounds per asset [1, 23, 26]. To improve solution quality at limited hardware, papers have used insights to guide the quantum search, for instance, by identifying high-Sharpe-Ratio candidates in a pre-processing step [11] or using the results of a classical heuristic to seed the quantum annealer via reverse annealing [30]. Warm-starting methods using the solution from the continuous relaxation have been explored, including initial state preparation for simulated annealing [27] and the preparation of quantum states for algorithms on gate-based quantum computers [6]. To the best of our knowledge, our paper is the first one that integrates the solution of the continuous relaxation into the QUBO construction and thereby endogenously determines relevant investment bands.

We conduct experiments based on real-world S&P 500 stock market data on software solvers and a D-Wave Advantage quantum annealer. Our results show that our novel formulation can lead to a drastic reduction of qubits—effectively allowing for larger problem instances on actual quantum hardware.

2 Background

Quantum Computing. Let us quickly introduce the basic notion of what can be considered as quantum computation [22]. Today, practical quantum computing (QC) consists of two dominant paradigms: Adiabatic QC and gate-based QC. In both scenarios, a quantum state $|\psi\rangle$ for a system with n qubits is a 2^n dimensional complex and ℓ_2-normalized vector. The domain of $|\psi\rangle$ is known as Hilbert space. In the gate-based (universal) framework, a quantum computation is defined as a matrix multiplication $|\boldsymbol{\psi}_{\text{out}}\rangle = C\,|\boldsymbol{\psi}_{\text{in}}\rangle$, where the C is a $(2^n \times 2^n)$-dimensional unitary matrix (the circuit), given via a series of inner and outer products of low-

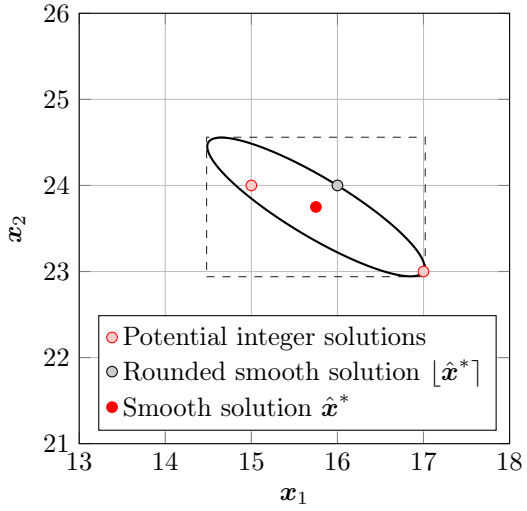


Fig. 1. Exemplary visualization of our proposed method. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ denote integer quantities of two assets. A naive approach would require roughly 5 qubits per dimension to allow for quantities up to $2^5 = 32$. Our approach constructs an ellipsoid around a smooth solution that can be obtained efficiently via a standard QP solver. Only integer points within the ellipsoid's dashed bounding-box are considered during optimization. Thus, the number of qubits required for each dimension is reduced to 2 and 3, respectively.

dimensional unitary matrices (the gates). The circuit is derived manually or via heuristic search algorithms [8]. In adiabatic quantum computation (AQC)—the framework that we consider in the paper at hand—the result of computation is defined to be the eigenvector $|\phi_{\min}\rangle$ that corresponds to the smallest eigenvalue of some $(2^n \times 2^n)$-dimensional Hermitian matrix $\boldsymbol{H}$ (also known as Hamiltonian). In practical AQC, $\boldsymbol{H}$ is further restricted to be a real diagonal matrix whose entries can be written as $\boldsymbol{H}_{i,i} = \boldsymbol{H}(\boldsymbol{Q})_{i,i} = (\boldsymbol{x}^{(i)})^\top \boldsymbol{Q} \boldsymbol{x}^{(i)}$ where $\boldsymbol{x}^{(i)} = \text{binary}(i, n) \in \{0,1\}^n$ is some (arbitrary but fixed) n-bit binary expansion of the unsigned integer i. Here, $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is the so-called QUBO matrix. By construction, computing $|\phi_{\min}\rangle$ is equivalent to solving

$$\min_{\boldsymbol{x} \in \{0,1\}^n} \quad \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \,. \tag{1}$$

Adiabatic quantum algorithms rely on this connection by encoding combinatorial problems into QUBO matrices.

In both paradigms, the output vector is 2^n-dimensional and can thus not be read-out efficiently for non-small n. Instead, the output of a practical quantum computation is a random integer i between 1 and 2^n, which is drawn from the probability mass function $\mathbb{P}(i) = |\langle i | \boldsymbol{\psi}_{\text{out}} \rangle|^2$. This sampling step is also known as collapsing the quantum state $|\boldsymbol{\psi}_{\text{out}}\rangle$ to a classical binary state $\text{binary}(i, n)$. AQC has been applied to numerous combinatorial optimization problems [15], ranging over satisfiability [10], routing problems [21] to machine learning [19].

The number of physically available qubits of an actual *Quantum Processing Unit* (QPU) is limited. As a result, a theoretically sound QUBO formulation might not be able to run on a real-world QPU as n exceeds the number of available qubits. Decomposition techniques can be applied in these cases, cutting the original QUBO into smaller pieces. It is, however, not guaranteed that recombining partial solutions will lead to a good (or even feasible) global solution.

Hence, another approach is to come up with an alternative QUBO formulation that requires a smaller number of qubits. One example is the integration of inequality constraints via the unbalanced penalty method [13], which requires less qubits than naive integration of the constraints. In the paper at hand, the QUBO dimension is reduced by reducing the search space to discrete solutions in the vicinity of the optimum of a relaxed convex QP.

Portfolio Theory. In the sense of the classical mean-variance approach introduced by [17], investors select non-negative portfolio weights $\boldsymbol{w}$ on a continuous scale, with each entry $\boldsymbol{w}_i$ defining the fraction of wealth invested in asset i. Let $\boldsymbol{\mu}$ denote the expected returns and $\boldsymbol{\Sigma}$ the covariance matrix of asset returns. The selection can be formalized by an optimization problem identifying the maximal expected portfolio return subject to a specified level of portfolio variance:

$$\max_{\boldsymbol{w}} \ \boldsymbol{\mu}^\top \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \leq c, \sum_{i=1}^{n} \boldsymbol{w}_i = 1, \boldsymbol{w}_i \geq 0 \text{ for all } i\,. \tag{2}$$

Problem (2) is the classical starting point to identify an optimal portfolio with integer amounts of asset purchases. Let $\boldsymbol{p} \in \mathbb{R}_+^n$ collect the price per unit of each asset, and let $B > 0$ denote the budget to be invested. Similar to [4], to identify the optimal units of purchased assets $\boldsymbol{x} \in \mathbb{Z}_+^n$, one considers the problem

$$\max_{\boldsymbol{x}} \ \tilde{\boldsymbol{\mu}}^\top \boldsymbol{x} \quad \text{subject to} \quad \boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{x} \leq c \cdot B^2, \sum_{i=1}^{n} \boldsymbol{p}_i \cdot \boldsymbol{x}_i = B, \boldsymbol{x}_i \in \mathbb{Z}_+ \text{ for all } i\,. \tag{3}$$

Notice how (2) gives rise to (3) by adding the integrality constraint $\boldsymbol{x}_i \in \mathbb{Z}_+$. Their solutions can be translated back and forth via the following equations[3]:

$$\boldsymbol{w} = \frac{\boldsymbol{x} \circ \boldsymbol{p}}{B}, \quad \tilde{\boldsymbol{\Sigma}} = \operatorname{diag}(\boldsymbol{p}) \cdot \boldsymbol{\Sigma} \cdot \operatorname{diag}(\boldsymbol{p}), \quad \tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} \circ \boldsymbol{p}, \tag{4}$$

In order to construct a convex problem over the simplex, the expected portfolio return and the portfolio variance are combined into a quadratic objective function via the risk aversion parameter $\gamma > 0$. Consequently, we arrive at

$$\max_{\boldsymbol{w}} \ \boldsymbol{\mu}^\top \boldsymbol{w} - \frac{\gamma}{2} \cdot \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} \quad \text{subject to} \quad \sum_{i=1}^{n} \boldsymbol{w}_i = 1, \boldsymbol{w}_i \geq 0 \text{ for all } i\,. \tag{5}$$

In (3), it would be restrictive to require that the budget is invested exactly, $\sum_{i=1}^{n} \boldsymbol{p}_i \cdot \boldsymbol{x}_i = B$. Instead, we assume that the difference between the invested amount and the budget B is invested at the risk-free interest rate r_f. In this case, the $\boldsymbol{w}_i$ represent the weights of each risky asset, while $(1 - \sum_{i=1}^{n} \boldsymbol{w}_i)$ represents the weight to be invested at the risk-free rate. Consequently, the constraint $\sum_{i=1}^{n} \boldsymbol{w}_i = 1$ can be removed and the expected return becomes

$$\boldsymbol{\mu}^\top \boldsymbol{w} + \left(1 - \mathbf{1}^\top \boldsymbol{w}\right) \cdot r_f = \left(\boldsymbol{\mu} - r_f \cdot \mathbf{1}\right)^\top \boldsymbol{w} + r_f \tag{6}$$

[3] The operator $\circ$ denotes element-wise multiplication.

with $\mathbf{1}$ denoting an n-dimensional vector of ones. Moreover, we omit the no-shortselling constraint $\boldsymbol{w}_i \geq 0$, as in practice transaction costs may imply that the investor slowly trades from an existing portfolio to the preferred one. Therefore, the optimal solution typically remains long-only, so the constraint is empirically non-binding (as seen in Sec. 4). The new problem can be formulated as

$$\max_{\boldsymbol{w}} \ \underbrace{(\boldsymbol{\mu} - r_f \cdot \mathbf{1})}_{\boldsymbol{\mu}_f}^{\top} \boldsymbol{w} - \frac{\gamma}{2} \cdot \boldsymbol{w}^{\top} \boldsymbol{\Sigma} \boldsymbol{w} \ . \tag{7}$$

Setting $\tilde{\gamma} = \gamma / B$ and invoking Eq. 4 for $\boldsymbol{\mu}_f$, we arrive at

$$\max_{\boldsymbol{x}} \ \tilde{\boldsymbol{\mu}}_f^{\top} \boldsymbol{x} - \frac{\tilde{\gamma}}{2} \cdot \boldsymbol{x}^{\top} \tilde{\boldsymbol{\Sigma}} \boldsymbol{x} \quad \text{subject to} \quad \boldsymbol{x}_i \in \mathbb{Z} \text{ for all } i \ . \tag{8}$$

The latter represents the NP-hard portfolio *Quadratic Integer Program* (QIP). However, since $\boldsymbol{\Sigma}$ is a covariance matrix which is typically positive definite, dropping the integrality constraint of (8) leads to a smooth convex quadratic program for which solutions can be readily obtained via standard solvers. For empirical situations with sample covariance matrices being ill-conditioned or singular, regularization methods are widely used to ensure positive definiteness; see, e.g., [12]. Furthermore, the structural form of the mean-variance framework in (8) can be derived, at least approximately, for homogeneous risk measures such as Value-at-Risk or Expected Shortfall, as well as for deterministic regulatory capital standards (see [24]). Even when (8) is extended to include additional linear or convex quadratic terms, such as transaction costs, a smooth convex quadratic program can be established (see Sec. 4).

In order to construct an equivalent QUBO, $\boldsymbol{x}$ must be binarized, e.g., by reserving a fixed amount of k bits for each asset and interpreting the bits as unsigned binary expansion of some integer. Clearly, the choice of k has an immediate impact on both, the accuracy to which we can represent the asset weights and the dimension of the resulting QUBO. We will re-visit this topic in our experiments.

3 Methodology

By adopting ideas from [3], we derive a novel method to reduce the number of qubits required for the portfolio optimization QUBO. The proposed method relies on an informed re-formulation of the portfolio QIP.

Theorem 1 (Informed Ellipsoid Formulation). *Let $\boldsymbol{x}^*$ be an optimal solution to (8) with $\tilde{\boldsymbol{\Sigma}}$ positive definite, $\hat{\boldsymbol{x}}^*$ the smooth maximizer of $f(\boldsymbol{x}) = \tilde{\boldsymbol{\mu}}_f^{\top} \boldsymbol{x} - \frac{\tilde{\gamma}}{2} \cdot \boldsymbol{x}^{\top} \tilde{\boldsymbol{\Sigma}} \boldsymbol{x}$, and $\hat{f}^* := f(\hat{\boldsymbol{x}}^*)$ the corresponding optimal function value. Assume that at least one entry of $\hat{\boldsymbol{x}}^*$ is not integer-valued. Let further*

$$\boldsymbol{E} := \left\{ \boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \hat{\boldsymbol{x}}^*)^{\top} \left(\frac{\tilde{\gamma}}{2C} \cdot \tilde{\boldsymbol{\Sigma}} \right) (\boldsymbol{x} - \hat{\boldsymbol{x}}^*) \leq 1 \right\} \tag{9}$$

with $C = \hat{f}^* - f(\lfloor \hat{\boldsymbol{x}}^* \rceil) > 0$. Then, $\boldsymbol{x}^* \in \boldsymbol{E}$. Thus, optimizing over the ellipsoid $\boldsymbol{E}$ is sufficient to compute $\boldsymbol{x}^*$. Here, $\lfloor \cdot \rceil$ denotes component-wise rounding to the nearest integer.

Proof. First, notice that

$$\begin{aligned}(\boldsymbol{x} - \hat{\boldsymbol{x}}^*)^\top \left(\frac{\tilde{\gamma}}{2}\tilde{\boldsymbol{\Sigma}}\right)(\boldsymbol{x} - \hat{\boldsymbol{x}}^*) \leq C \Leftrightarrow\ & (\boldsymbol{x} - \hat{\boldsymbol{x}}^*)^\top \left(\frac{\tilde{\gamma}}{2}\tilde{\boldsymbol{\Sigma}}\right)(\boldsymbol{x} - \hat{\boldsymbol{x}}^*) - \hat{f}^* \leq -f(\lfloor \hat{\boldsymbol{x}}^* \rceil) \\ \Leftrightarrow\ & (\hat{\boldsymbol{x}}^*)^\top \left(\tilde{\gamma}\tilde{\boldsymbol{\Sigma}}\right)\boldsymbol{x} - \frac{\tilde{\gamma}}{2}\boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}}\boldsymbol{x} \geq f(\lfloor \hat{\boldsymbol{x}}^* \rceil)\ .\end{aligned}$$

Moreover, since $\hat{\boldsymbol{x}}^*$ satisfies the first order condition of f, it follows that

$$\tilde{\boldsymbol{\mu}}_f - \tilde{\gamma}\tilde{\boldsymbol{\Sigma}}\hat{\boldsymbol{x}}^* = 0 \Leftrightarrow \hat{\boldsymbol{x}}^* = \frac{1}{\tilde{\gamma}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}_f\ .$$

Plugging this into the last inequality yields

$$\tilde{\boldsymbol{\mu}}_f \boldsymbol{x} - \frac{\tilde{\gamma}}{2}\boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}}\boldsymbol{x} = f(\boldsymbol{x}) \geq f(\lfloor \hat{\boldsymbol{x}}^* \rceil)\ .$$

We see that all integer points contained in $\boldsymbol{E}$ (if any) achieve a better objective function value than the rounded smooth solution $\lfloor \hat{\boldsymbol{x}}^* \rceil$. Since all points outside of $\boldsymbol{E}$ are strictly worse than $\lfloor \hat{\boldsymbol{x}}^* \rceil$, the best integer solution must be included in $\boldsymbol{E}$ as well. ■

Based on these insights, we compute the extent of $\boldsymbol{E}$ for each asset i in terms of upper and lower bounds. The lower and upper boundaries of the ellipsoid in all dimensions can then be computed via

$$L = \hat{\boldsymbol{x}}^* - \sqrt{\operatorname{diag}(\boldsymbol{M}^{-1})} \quad \text{and} \quad U = \hat{\boldsymbol{x}}^* + \sqrt{\operatorname{diag}(\boldsymbol{M}^{-1})}$$

with $\boldsymbol{M} = \frac{\tilde{\gamma}}{2C} \cdot \tilde{\boldsymbol{\Sigma}}$. Finally, instead of optimizing over $\boldsymbol{x} \in \mathbb{Z}^n$, we exploit the limits of the ellipsoid and optimize over

$$\boldsymbol{x}_i \in [L_i, U_i] \cap \mathbb{Z} \text{ for all } i \tag{10}$$

This corresponds to optimization over a bounding hypercube that tightly encloses the ellipsoid[4] (see Fig. 1). Since $[L_i, U_i]$ can be different for each i, the effective search space is adapted to the individual extent of each dimension. Thus, this procedure has the potential to reduce the search space, and hence the number of required qubits, by a large margin. We call this procedure *hot-start*.

[4] While this introduces sub-optimal solutions into the search space, it does not harm the soundness of our result. Nevertheless, the effectiveness of our approach can be enhanced through several methods; for example, by successively refining the search space based on the best integer solution found, and/or by including penalties in the objective function to more tightly bound the search space.

4 Experiments

Two experiments are conducted to demonstrate the viability of our approach. First, the number of qubits required by a baseline k-bits per asset construction is compared to the number of qubits required for the proposed approach for asset universes of increasing size. Second, we compare the solution quality obtained on a D-Wave Advantage Quantum Annealer and a randomized search heuristic, both utilizing the baseline and our novel construction.

We calibrate $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ using monthly S&P 500 equity returns from August 2003 to July 2023, sourced from Refinitiv. The vector $\boldsymbol{\mu}$ and the matrix $\boldsymbol{\Sigma}$ are estimated as the sample mean and sample covariance matrix of the stock returns over this period. The vector p contains stock prices as of July 2023; the risk-free interest rate is set to $r_f = 5.32\%$ based on the Federal Funds Effective Rate, as reported by the Federal Reserve Bank of St. Louis (FRED).

Our investment universe consists of the n stocks with the highest market capitalization as of July 2023. The investor has a budget of $B = 250{,}000\,\text{USD}$ and initially holds a portfolio with (approximately) equal weighting across all n assets in the investment universe. The initial portfolio is integer-valued, i.e., rounded as

$$\boldsymbol{x}_0 = \left\lfloor \frac{B}{n} \cdot \boldsymbol{p}^{-1} \right\rfloor \tag{11}$$

where $\boldsymbol{p}^{-1}$ denotes the vector of element-wise reciprocals of the price vector $\boldsymbol{p}$ and $\lfloor \cdot \rfloor$ denotes component-wise flooring. We consider problem (8), supplemented by transaction costs. As in [9], we assume that transaction costs are determined by the covariance matrix $\boldsymbol{\Sigma}$ and thus achieve

$$\max_{\boldsymbol{x}} \; \tilde{\boldsymbol{\mu}}_f^\top \boldsymbol{x} - \frac{\tilde{\gamma}}{2} \cdot \boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{x} - \tilde{\kappa} \cdot (\boldsymbol{x} - \boldsymbol{x}_0)^\top \tilde{\boldsymbol{\Sigma}} (\boldsymbol{x} - \boldsymbol{x}_0) \quad \text{subject to} \quad \boldsymbol{x}_i \in \mathbb{Z} \text{ for all } i \,. \tag{12}$$

We set $\gamma = 3$ and $\tilde{\kappa} = 50 \cdot \gamma / B$.[5]

For the $n = 4$ stocks with highest market capitalization, Table 1 shows how Hot-Start constructs a QUBO with six qubits. Starting from the initial portfolio $\boldsymbol{x}_0$, the smooth solution $\hat{\boldsymbol{x}}^*$ of (12) suggests reducing positions in Microsoft, NVIDIA, and Amazon, while increasing the position in Apple. Enclosing hypercube (10) around ellipsoid $\boldsymbol{E}$ from (9) implies that the optimal integer-valued solution of (12) must be either 194 or 195 for Microsoft, either 372 or 373 for Apple, and so on. Thus, (12) can be solved as a QUBO requiring $1+1+3+1 = 6$ qubits, where the number of qubits is determined by the binary encoding rule $\#\text{qubits} = \lceil \log_2(\#\text{integers}) \rceil$ for each stock.

Increasing n successively to 100, the maximum number of qubits required by Hot-Start for any single asset increases from 3 (for $n = 4$) to 10 (for $n = 100$).

[5] $\gamma = 3$ lies within the range of economically realistic and empirically estimated risk-aversion coefficients; for a meta-analysis see [7]. Compared to empirical estimates, our calibration of transaction costs is moderate; for example, [9, p. 2328] estimate κ for several commodities and find that, in the median, κ is a factor of 500 larger than γ.

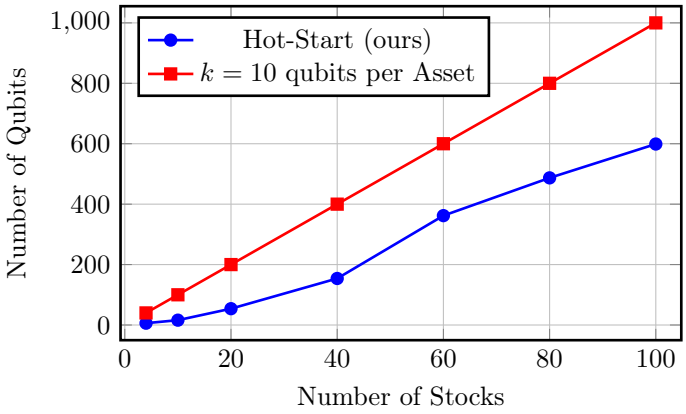


Fig. 2. Number of qubits required to encode the portfolio QUBO, as a function of the number of stocks. Comparison between our approach (red boxes) and a baseline (blue circles) where each stock quantity is encoded by 10 qubits.

To provide a conservative baseline for comparison, we calibrate a fixed k-bits per asset construction with $k = 10$. This baseline is as efficient as possible while still ensuring that the optimal solution is not overlooked.

Fig. 2 shows that in the first experiment, Hot-Start requires a substantially smaller number of qubits than the baseline. For the second experiment, we consider the D-Wave Advantage2 Prototype 2.6 QPU with 128 readouts (shots) per QUBO. Before sending the QUBO to the QPU, it must be mapped onto the Zephyr hardware topology. In case of the baseline, this step failed for all problems with more than 40 stocks. In case of Hot-Start, all problems with up to 80 stocks could be mapped successfully to the QPU. The energies of the found solutions coincide with those found by the D-Wave simulated annealing software solver. We refer to [18] for a comprehensive discussion on how to reduce integrated control errors and thus hardware noise on D-Wave annealers.

Table 1. Derivation of the number of qubits for problem (12) with $n = 4$ stocks.

Stock	$\boldsymbol{x}_0$	$\hat{\boldsymbol{x}}^*$	Integers in box		Decision variables	
			Smallest	Largest	# integers	# qubits
Microsoft	197	194.6	194	195	2	1
Apple	365	372.6	372	373	2	1
NVIDIA	1436	1434.2	1432	1436	5	3
Amazon	491	489.7	489	490	2	1
Total						6

5 Concluding Remarks

We presented Hot-Start, the first warm-start technique that lowers the qubit demand for quantum portfolio optimization. Experiments show favorable scaling: the conventional encoding lacks prior knowledge of the qubit count needed to represent a solution and therefore over-allocates, whereas Hot-Start infers that number automatically and uses qubits more efficiently. With Hot-Start we successfully embedded QUBOs for portfolios of up to 80 stocks on current QPUs, while the baseline approach already failed beyond 40 stocks. Because the approach relies only on the existence of a convex QP representation, we expect Hot-Start to transfer to a broad class of optimization problems well beyond finance, providing an effective pathway for harnessing near-term and future QPUs.

Acknowledgment

This research has been funded by the Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the LAMARR Institute for Machine Learning and Artificial Intelligence.

References

1. Aguilera, E., de Jong, J., Phillipson, F., Taamallah, S., Vos, M.: Multi-objective portfolio optimization using a quantum annealer. Mathematics **12**(9), 1291 (2024)
2. Beaulieu, D., Pham, A.: Max-cut clustering utilizing warm-start qaoa and ibm runtime (2021), `https://arxiv.org/abs/2108.13464`
3. Buchheim, C., Caprara, A., Lodi, A.: An effective branch-and-bound algorithm for convex quadratic integer programming. Math. Program. **135**(1-2), 369–395 (2012). https://doi.org/10.1007/S10107-011-0475-X
4. Castro, F., Gago, J., Hartillo, I., Puerto, J., Ucha, J.: An algebraic approach to integer portfolio problems. European Journal of Operational Research **210**(3), 647–659 (2011). https://doi.org/https://doi.org/10.1016/j.ejor.2010.11.007
5. Chen, Y.Q., Chen, Y., Lee, C.K., Zhang, S., Hsieh, C.Y.: Optimizing quantum annealing schedules with monte carlo tree search enhanced with neural networks (2022), `https://arxiv.org/abs/2004.02836`
6. Egger, D.J., Mareček, J., Woerner, S.: Warm-starting quantum optimization. Quantum **5**, 479 (2021). https://doi.org/10.22331/q-2021-06-17-479
7. Elminejad, A., Havranek, T., Irsova, Z.: Relative risk aversion: A meta-analysis. Journal of Economic Surveys (2022)
8. Franken, L., Georgiev, B., Mücke, S., Wolter, M., Heese, R., Bauckhage, C., Piatkowski, N.: Quantum circuit evolution on NISQ devices. In: IEEE CEC. pp. 1–8 (2022). https://doi.org/10.1109/CEC55065.2022.9870269
9. Gârleanu, N., Pedersen, L.H.: Dynamic trading with predictable returns and transaction costs. The Journal of Finance **68**(6), 2309–2340 (2013)
10. Kochenberger, G., Glover, F., Alidaee, B., Lewis, K.: Using the unconstrained quadratic program to model and solve Max 2-SAT problems. International Journal of Operational Research **1**(1-2), 89–100 (2005)
11. Lang, J., Zielinski, S., Feld, S.: Strategic portfolio optimization using simulated, digital, and quantum annealing. Applied Sciences **12**(23), 12288 (2022)

12. Ledoit, O., Wolf, M.: Robust performance hypothesis testing with the sharpe ratio. Journal of Empirical Finance **15**(5), 850–859 (2008)
13. Lee, L.K., Knaute, J., Gerhardt, F., Völker, P., Maras, T., Dotterweich, A., Piatkowski, N.: Formalizing, normalizing, and splitting the energy network redispatch for quantum annealing (2024), https://arxiv.org/abs/2409.09857
14. Lin, J., Zhang, Z., Zhang, J., Li, X.: Hard-instance learning for quantum adiabatic prime factorization. Physical Review A **105**(6) (2022). https://doi.org/10.1103/physreva.105.062455
15. Lucas, A.: Ising formulations of many NP problems. Frontiers in Physics **2** (2014). https://doi.org/10.3389/fphy.2014.00005
16. de Luis, A.P., Garcia-Saez, A., Estarellas, M.P.: Steered quantum annealing: improving time efficiency with partial information (2022), https://arxiv.org/abs/2206.07646
17. Markowitz, H.: Portfolio selection. The Journal of Finance **7**, 77–91 (1952)
18. Mücke, S., Gerlach, T., Piatkowski, N.: Optimum-preserving QUBO parameter compression. Quantum Mach. Intell. **7**(1), 1 (2025). https://doi.org/10.1007/S42484-024-00219-3
19. Mücke, S., Piatkowski, N., Morik, K.: Hardware acceleration of machine learning beyond linear algebra. In: Workshops of the ECML PKDD. vol. 1167, pp. 342–347. Springer (2019). https://doi.org/10.1007/978-3-030-43823-4_29
20. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization, Springer New York, NY, 1 edn. (2004). https://doi.org/10.1007/978-1-4419-8853-9
21. Neukart, F., Compostella, G., Seidel, C., von Dollen, D., Yarkoni, S., Parney, B.: Traffic Flow Optimization Using a Quantum Annealer. Frontiers in ICT **4** (2017)
22. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press (2016)
23. Palmer, S., Sahin, S., Hernandez, R., Mugel, S., Orus, R.: Quantum portfolio optimization with investment bands and target volatility. arXiv preprint arXiv:2106.06735 (2021)
24. Paulusch, J., Schlütter, S.: Sensitivity-implied tail-correlation matrices. Journal of Banking & Finance **134**, 106333 (2022)
25. Punnen, A.P. (ed.): The Quadratic Unconstrained Binary Optimization Problem: Theory, Algorithms, and Applications. Springer Cham, 1 edn. (2022). https://doi.org/10.1007/978-3-031-04520-2
26. Rosenberg, G., Haghnegahdar, P., Goddard, P., Carr, P., Wu, K., De Prado, M.L.: Solving the optimal trading trajectory problem using a quantum annealer. In: Proceedings of the 8th Workshop on High Performance Computational Finance. pp. 1–7 (2016)
27. Rubio-García, Á., García-Ripoll, J.J., Porras, D.: Portfolio optimization with discrete simulated annealing. arXiv preprint arXiv:2210.00807 (2022)
28. von Rüden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., Schuecker, J.: Informed machine learning - A taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans. Knowl. Data Eng. **35**(1), 614–633 (2023). https://doi.org/10.1109/TKDE.2021.3079836
29. Truger, F., Barzen, J., Bechtold, M., Beisel, M., Leymann, F., Mandl, A., Yussupov, V.: Warm-starting and quantum computing: A systematic mapping study. ACM Comput. Surv. **56**(9) (2024). https://doi.org/10.1145/3652510
30. Venturelli, D., Kondratyev, A.: Reverse quantum annealing approach to portfolio optimization problems. Quantum Machine Intelligence **1**(1), 17–30 (2019)

Constrained Quantum Optimization for Scheduling Aircraft Arrivals

Gaspard Cassassolles[1,2], Hian Lee Kwa[2], Teck Yoong Chai[2], and Yung Sze Gan[2]

[1] CentraleSupélec, Paris, France
gaspard.cassassolles@student-cs.fr
[2] CortAIx Labs Singapore, Thales Solutions Asia, Singapore
{hianlee.kwa, teckyoong.chai}@thalesgroup.com
yungsze.gan@asia.thalesgroup.com

Abstract. Air traffic scheduling near airports is a complex task, particularly during landings where aircraft must maintain strict temporal separations at waypoints. Disruptions such as in-flight delays can lead to trajectory conflicts, which are typically mitigated by adjusting arrival times, which effectively issues speed advisories to pilots. This paper introduces a quantum optimization approach for this problem by encoding it as a Quadratic Unconstrained Binary Optimization (QUBO) instance. We further investigate how two variational algorithms, Quantum Aproximate Optimization Algorithm (QAOA) and Variational Quantum Eigensolver (VQE), perform on this highly constrained problem compared to an unconstrained benchmark (MaxCut), highlighting the additional complexity introduced by constraints in quantum optimization.

Keywords: Air Traffic Management · Aircraft Scheduling · Combinatorial Optimization · Quantum Approximate Optimization Algorithm · Quantum Optimization · Variational Quantum Eigensolver

1 Introduction

When arriving at airports, flights typically follow predetermined routes, known as Standard Terminal Arrival Routes (STARs), which are defined by a series of waypoints. However, disruptions such as weather delays can cause deviations in the timing and spacing between aircraft. This may lead to a loss of separation standards, particularly in high-traffic airspace near airports. A common approach to maintain separation standards is to adjust an aircraft's estimated time of arrival at the STAR entry point by issuing a speed advisory, i.e., an instruction for the pilot to adjust their speed to meet a revised target time. This ensures that separation standards are adhered to without any deviations in the planned route, thereby reducing the amount of fuel consumed. This task is normally carried out by an Arrival Manager (AMAN) that considers multiple factors such as aircraft weight, weather, and runway availability.

The AMAN schedulers currently rely on classical optimization algorithms to find the optimal maneuvers for each aircraft in a given situation. Recently, there has also been multiple attempts of including the use of Reinforcement Learning models to enhance the performance of these tools [10]. However, these aforementioned classical computing tools may struggle to keep pace with the ever increasing volume of air traffic.

Quantum computing offers a promising approach for tackling the growing computational complexity of air traffic management, as it can consider large variable spaces and numerous possible solutions in parallel. This capability could improve the aircraft speed advisory process by reducing delays, minimizing conflict resolution maneuvers, and enhancing fuel efficiency. While Stollenwerk et al. [11] investigated quantum annealing for optimizing individual aircraft trajectories, our focus is on high-density traffic scenarios where arrivals follow fixed routes such as STARs.

In this paper, we explore the use of quantum computing in air traffic management, specifically the aircraft speed advisory problem, and formulate it as a combinatorial optimization task. Our main contribution is twofold: (1) we express the speed advisory problem using a QUBO formulation, making it suitable for quantum algorithms and (2) we provide a comparison of QAOA and VQE on this constrained problem against an unconstrained baseline (MaxCut) to illustrate the challenges of handling constraints in quantum optimization. We implement these formulations on hybrid quantum-classical solvers and evaluate their performance across different problem sizes and circuit depths. In doing so, we examine the trade-offs between solution quality and feasibility.

2 Problem Formulation

Aircraft are required to maintain minimum temporal and spatial separations throughout their flight to ensure safety. Conflicts occur when these separation standards are violated, which is more likely during approach due to the high air traffic density and limited runway resources. In this study, we focus on maintaining separation at the fixed waypoints defined by the STAR, under the assumption that doing so ensures conflict-free arrival sequencing. To achieve this, we apply controlled time shifts to each aircraft's arrival at the waypoint. In practice, this is implemented through speed advisories, which pilots execute to satisfy Time To Gain (TTG) or Time To Lose (TTL) requirements [7]. In our formulation, a speed advisory corresponds to selecting a discrete time shift for each aircraft, which becomes the decision variable in our combinatorial optimization model.

2.1 Combinatorial Optimization Formulation

While in practice, there may be multiple entry points along a STAR that results in the merging of routes at specific waypoints, for simplicity we only consider a single waypoint in this study. Let n be the number of aircraft, each with an estimated time of arrival (ETA) at the waypoint. We define a discrete set of

m possible time shifts for each aircraft, allowing both advances and delays. We define the binary decision variables as:

$$X = \{x_{ij} \mid i = 1, \ldots, n;\ j = 1, \ldots, m;\ x_{ij} \in \{0, 1\}\}. \tag{1}$$

Here, $x_{ij} = 1$ indicates that aircraft i is assigned time shift j; otherwise, $x_{ij} = 0$. To ensure that each aircraft is assigned exactly one time shift, we impose the following one-hot encoding constraint:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad \forall i \in \{1, \ldots, n\}. \tag{2}$$

We then define the actual time shift T_i applied to aircraft i as:

$$\Delta T_i = \Delta_d \sum_{j=1}^{m} \left(j - \frac{m+1}{2} \right) x_{ij}, \tag{3}$$

where Δ_d denotes the discretization step in minutes. This expression captures both advances and delays relative to the original ETA. Here, the maximum value possible for a time shift is the same for a delay or an advance, and is equal to $\Delta_d(\frac{m-1}{2})$ minutes.

A conflict between aircraft i and i' occurs if their temporal separation falls below the required minimum, T_{min}. This condition can be expressed as:

$$\Delta T_i - \Delta T_{i'} \in \left[-T_{\min} - (T_i - T_{i'}),\ T_{\min} - (T_i - T_{i'}) \right], \tag{4}$$

where T_i, T_j denote the ETA of aircraft i and j respectively and $T_{min} = 4$ minutes. We then seek to minimize the total time variation compared to the initial ETAs. This is expressed as:

$$f_{\text{total variation}} = \sum_{i=1}^{n} |\Delta T_i|. \tag{5}$$

2.2 Quantum Optimization

When solving QUBO problems, two main approaches emerge corresponding to the two primary types of quantum hardware. In their work finding optimal aircraft takeoff delays Stollenwerk et al. employed a quantum annealer developed by D-Wave [11]. Such quantum annealers encode the cost function into the energy of a Hamiltonian to be minimized. This method relies on the adiabatic principle, which states that a quantum system in an eigenstate undergoing slow enough changes will remain in that eigenstate [1]. By initializing the system in the ground state of a simple Hamiltonian and gradually transforming it into a *Cost Hamiltonian* that encodes the objective function, the system ideally transitions into the ground state of the cost Hamiltonian, thus identifying the optimal solution.

Another prominent method of solving quantum optimization problems is the Variational Quantum Eigensolver (VQE) that is run on gate-based quantum

computers. This is a hybrid quantum-classical algorithm that iteratively searches for the ground state of a Hamiltonian, i.e., the solution that yields the optimum value for a given cost function, using a parameterized quantum circuit. Using this framework, a parameterized circuit $U(\theta)$, also known as an *ansatz*, is used to generate trial wavefunctions $|\psi(\theta)\rangle = U(\theta)\,|0\rangle^{\otimes n}$, guided by a classical optimization algorithm that aims to solve the minimization problem:

$$\min_{\theta} \langle \psi(\theta)|H_C|\psi(\theta)\rangle = \min_{\theta} C(\theta), \tag{6}$$

where the left hand side expression is the expected value of the Cost Hamiltonian, representing the energy of the system. The solution to the QUBO problem is given by the wavefunction $|\psi(\theta^*)\rangle$, where $\theta^* = \arg min_{\theta} C(\theta)$ are the optimized parameters for the quantum circuit that yield the minimum cost. Here, $C(\theta)$ is the cost of the ansatz with the parameters θ. In practice, this cost function is estimated by running the quantum circuit multiple times, sampling the resulting bitstrings, and computing the average cost function value over the number of samples obtained.

QAOA is a specific case of VQE meant to replicate the principles of quantum annealing on gate-based quantum computers through creating an appropriate Ansatz from the cost Hamiltonian [3,5]. It can be shown that QAOA determines quantum state with guaranteed approximation ratio with respect to the ground state [3,4]. Furthermore, QAOA is the equivalent of adiabatic computation as the circuit depth, p, approaches ∞. This guarantees that QAOA will find the optimal solution of the combinatorial optimization problem on the condition that the ansatz is sufficiently deep and the optimal circuit parameters can be found. This ansatz is given by:

$$|\beta,\gamma\rangle = U_p(\beta,\gamma)|+\rangle^{\otimes N} = \left(\prod_{k=1}^{p} e^{-i\beta_k H_C}\, e^{-i\gamma_k H_M}\right)|+\rangle^{\otimes N}, \tag{7}$$

where β and γ are the variational parameters found by a COBYLA optimizer, p is the number of layers, or depth of the circuit, and H_M is the Mixer Hamiltonian, which is used to encourage exploration of the solution space by driving transitions between basis states.

2.3 QUBO Formulation

Applying VQEs to the formulated combinatorial optimization problem requires first recasting it as a QUBO. In this form, the objective is to find a binary vector $\mathbf{z} \in \{0,1\}^n$, that minimizes a quadratic cost function. This is given as:

$$\min_{\mathbf{z}\in\{0,1\}^n} \mathbf{z}^\top \mathbf{Q}\mathbf{z}, \tag{8}$$

where n is the number of variables and Q is an $n \times n$ real matrix that encodes the problem's objective functions and constraints. Since QUBO problems are inherently unconstrained, we incorporate one-hot encoding and enforce no-conflict

rules by appending weighted penalties to the cost function, which discourage any constraint violations. In this work, two penalty functions are used:

$$f_{\text{encoding}} = \lambda_{\text{encoding}} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_{ij} - 1 \right)^2, \tag{9}$$

which enforces one-hot encoding, and

$$f_{\text{conflict}} = \lambda_{\text{conflict}} \sum_{\substack{i,j,i',j' \\ \Delta T_i - \Delta T_{i'} \in [-T_{\min} - (T_i - T_{i'}),\ T_{\min} - (T_i - T_{i'})]}} x_{ij} x_{i'j'}, \tag{10}$$

which penalizes aircraft conflicts. In the two penalty functions $\lambda_{\text{encoding}}$ and $\lambda_{\text{conflict}}$ are sufficiently large penalty weights.

The overall cost function to be minimized can be written as

$$f = f_{\text{total variation}} + f_{\text{encoding}} + f_{\text{conflict}}. \tag{11}$$

2.4 Choice of Penalty Weights

Implementing constraints as penalty terms induces a trade-off between solution validity and objective optimality. While increasing penalty weights reduces invalid solutions, the use of excessive weights can dominate the final cost function and the cost distribution, thereby producing an ill-conditioned optimization landscape that impedes parameter tuning. Moreover in hardware implementations, large penalty coefficients map to stronger coupling terms in the Hamiltonian, which may be challenging to realize. As such, we aim to find the smallest pair of lambdas such that the ground-state bitstring satisfies all constraints. For this, we test different values of a 2D-grid, using an exact solver. In Fig. 1, the box-like boundary between the valid and invalid zones suggests that the validity of the two penalties is independent. For all four scheduling problems, we set $\lambda_{\text{encoding}} = 9$ and $\lambda_{\text{conflict}} = 2$, as these are the lowest integer values for these penalty weights that ensure a valid solution with an exact solver for all scheduling problems.

3 Experiment Setup

To solve this QUBO problem, we use IBM's Qiskit simulator [8] to emulate noise-free qubits. Each execution of a quantum circuit uses 10,000 shots. For parameter optimization, COBYLA is employed with a convergence tolerance of 1e−5. We define four problems of various sizes, as seen in Table 1, to be solved using both QAOA and VQE.

We also define four MaxCut problems of the same sizes, which are naturally unconstrained, to compare the solving of our constrained scheduling problem to a well-known unconstrained baseline. Including this comparison to the MaxCut

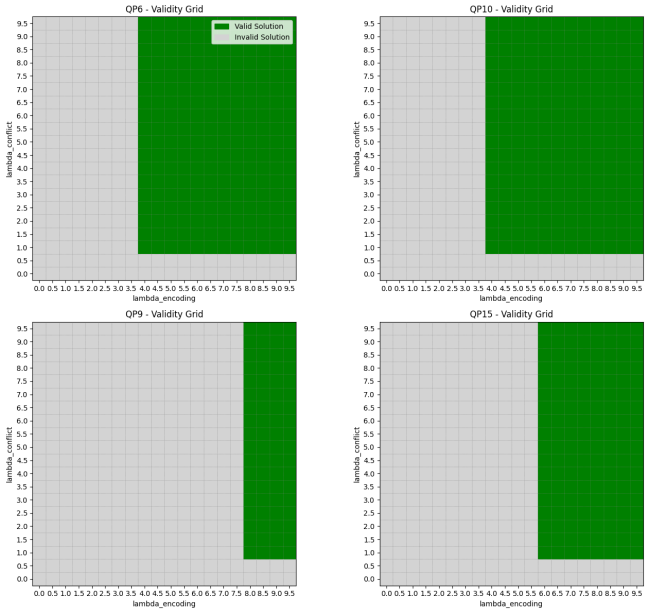


Fig. 1: Effect of penalty weights ($\lambda_{\text{encoding}}, \lambda_{\text{conflict}}$) on solution validity for four scheduling problems.

Name	n (No. of Planes)	m (Possible Time Shifts)	**Number of variables**
QP6	2	3	6
QP9	3	3	9
QP10	2	5	10
QP15	3	5	15

Table 1: Four aircraft speed advisory problem instances of different sizes.

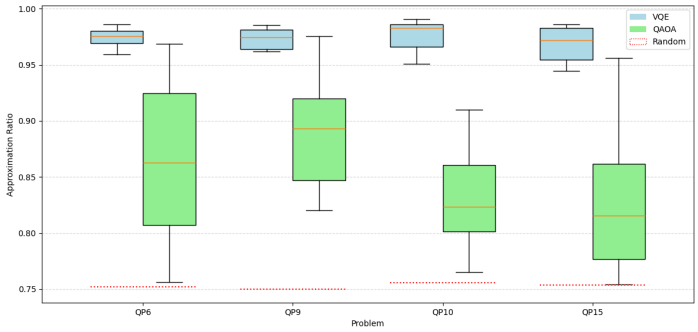
problem serves to demonstrate that unconstrained problems are inherently easier for variational quantum algorithms to solve compared to highly constrained ones like our scheduling problem, highlighting the additional complexity introduced by constraint handling in quantum optimization. Readers unfamiliar with the MaxCut problem are directed to [3] for an introduction.

To assess the performance of a given quantum solver, two metrics are used: (1) the probability of sampling the optimal solution, and (2) the normalized approximation ratio, defined as:

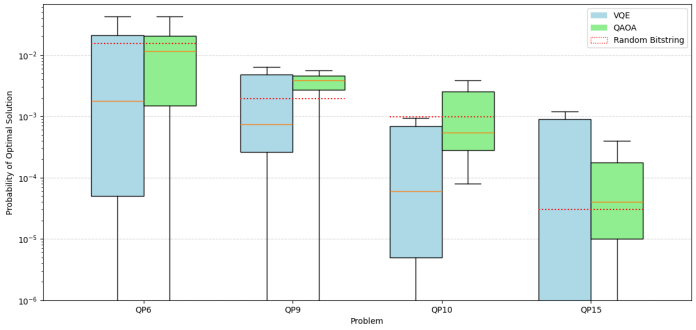
$$r = \frac{C(\theta) - C_{max}}{C(\theta^*) - C_{max}}. \tag{12}$$

As previously mentioned, we set the penalty coefficients as $\lambda_{\text{encoding}} = 9$ and $\lambda_{\text{conflict}} = 2$.

4 Results



(a) Approximation ratio



(b) Probability of sampling optimal solution

Fig. 2: Performance of using QAOA (green) and VQE (blue) when applied to the speed advisory problem.

When comparing the performance of both QAOA and a VQE (both with $p = 1$), Fig. 2a illustrates that both quantum algorithms give a better approximation ratio than a random choice; the expected cost of the result generated by these heuristics is better than that of a random bitstring. In addition, it can be observed that the VQE solver consistently outputs a better approximation ratio than the QAOA solver, with less variation in the performance. This is due to the higher expressivity of the VQE circuit compared to the QAOA circuit, which stems from the higher number of rotational angles used in the VQE.

However, the results shown in Fig. 2b indicate that the probability of selecting the optimal bitstring is essentially equal to the probability obtained by randomly choosing a bitstring. This shows that, although the optimization process tends to produce solutions with a lower expected cost than randomly chosen bitstrings, it does not significantly increase the likelihood of sampling the true optimum. This is because the classical optimizer minimizes the expected cost of the circuit, i.e., the average cost over all bitstrings sampled from the circuit, instead of directly maximizing the probability of the optimal solution.

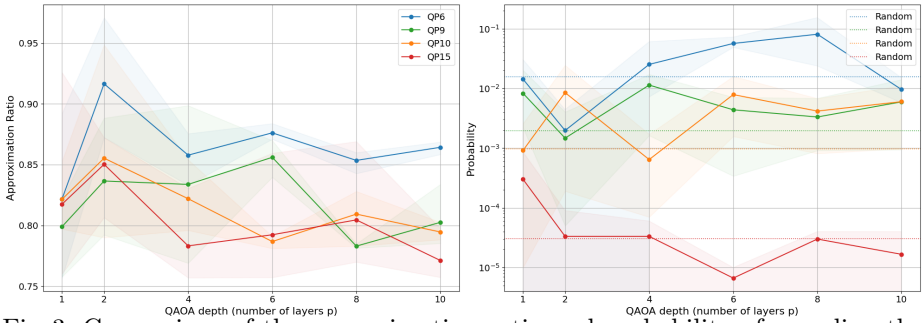


Fig. 3: Comparison of the approximation ratio and probability of sampling the optimum bitstring with increasing circuit depth.

Although deeper circuits are typically expected to yield higher-quality solutions [3], Fig. 3 shows no significant improvement in QAOA performance with increasing p. This lack of improvement may be due to the sparse and irregular energy landscape. Indeed, Fig. 4 shows a highly irregular and non-convex energy landscape where small shifts in the value of β and γ causes a drastic impact on the expected cost of the circuit. Increasing the circuit depth introduces more variational parameters, which can make it increasingly difficult for classical optimizers to find the global minimum and reliably reach the optimal bitstring.

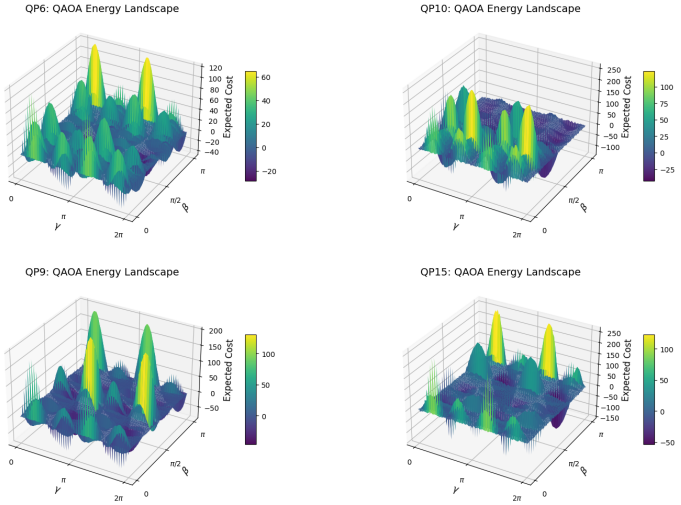


Fig. 4: Energy landscape among the parameter space for QAOA with p=1

Part of the challenge in sampling the optimum bitstring stems from the highly constrained problem formulation. As shown in Fig. 5, there are more possible values for the cost of an individual bitstring for the highly constrained speed advisory problem compared to the MaxCut problem, which is naturally unconstrained. This contrast validates our motivation for including MaxCut as a baseline. Unconstrained problems such as MaxCut present a more regular solution space with no invalid solutions, making them significantly easier for

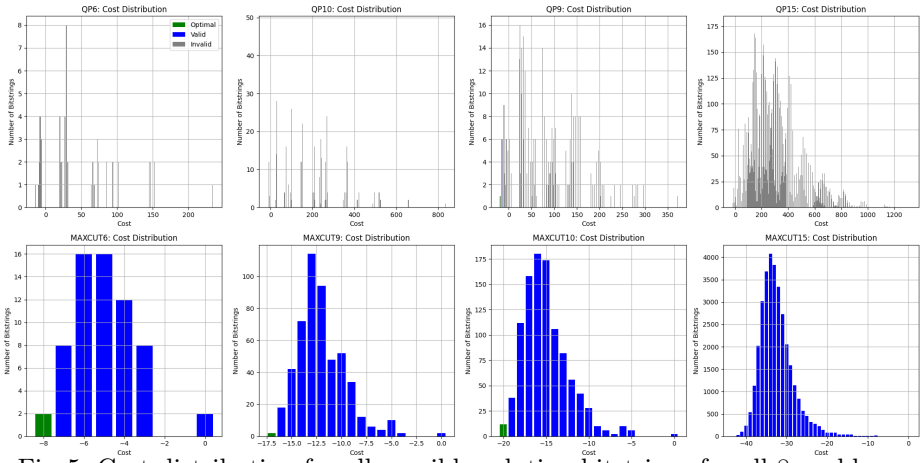


Fig. 5: Cost distribution for all possible solution bitstrings for all 8 problems

variational quantum algorithms to handle. In contrast, the strict constraints in the speed advisory problem create a sparse and irregular landscape. This reduces the proportion of valid solutions, thereby impeding the optimizer's ability to find low-cost configurations. This observation reinforces the inherent difficulty of constraint handling in quantum optimization and explains the performance gap between the two problem classes.

5 Conclusion and Discussion

In this paper, we presented a framework that translates the aircraft arrival scheduling problem into a combinatorial optimization problem, formulates it as a QUBO, and evaluates solutions obtained from both QAOA and VQE ansätze on a noise-free quantum emulator. Our findings show that while these algorithms can achieve good approximation ratios, the probability of sampling the exact optimal solution remains low due to the highly irregular cost landscape, which makes classical parameter optimization challenging.

Consistent with prior studies, VQE outperformed QAOA for small problem instances, owing to its higher circuit expressivity. While our results do not indicate a quantum advantage at this scale, they highlight the additional complexity introduced by constraints. These additional constraints makes the speed advisory problem significantly harder to solve with quantum optimization algorithms compared to unconstrained benchmarks such as MaxCut.

The present study is limited to single-waypoint scenarios and noiseless simulations. Future work will extend the model to multiple waypoints and route merging for greater realism and evaluate implementations on quantum hardware to capture noise and connectivity constraints. We also plan to develop the presented quantum methods in an integrated arrival and departure management (AMAN/DMAN) framework, where the optimizer coordinates both arrival and departure flights while considering runway capacity limits [9]. Even though classical optimization heuristics remain superior for small-scale cases, our long-term goal is to assess whether quantum approaches can offer scalability benefits as problem sizes and system complexity grow.

Future work could also explore advanced techniques to improve solution quality. For example, the Conditional Value at Risk (CVaR) cost function focuses optimization on the best-performing subset of sampled bitstrings, rather than the full expectation, which can increase the probability of sampling low-cost solutions [2]. Similarly, replacing the standard X-mixer in QAOA with an XY-mixer Hamiltonian, which is done using the Quantum Alternating Operator Ansatz, preserves Hamming weights of the qubits during evolution [6]. This ensures that the one-hot encoding constraints are respected and may reduce the number of invalid states sampled.

Acknowledgements The authors would like to thank Prof. Dimitris Angelakis, Mr. Gordon Ma, and Mr. Ioannis Leonidas from AngelQ for their valuable insights on VQEs, as well as Mr. Francois Balland and Mr. Andrei Purica from Thales Land and Air Systems for the discussions about ATM operations.

References

1. Albash, T., Lidar, D.A.: Adiabatic quantum computation. Reviews of Modern Physics **90**(1), 015002 (2018)
2. Barkoutsos, P.K., Nannicini, G., Robert, A., Tavernelli, I., Woerner, S.: Improving variational quantum optimization using CVaR. Quantum **4**, 256 (2020)
3. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014)
4. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm applied to a bounded occurrence constraint problem. arXiv preprint arXiv:1412.6062 (2014)
5. Farhi, E., Goldstone, J., Gutmann, S., Lapan, J., Lundgren, A., Preda, D.: A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. Science **292**(5516), 472–475 (2001)
6. Hadfield, S., Wang, Z., O'gorman, B., Rieffel, E.G., Venturelli, D., Biswas, R.: From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. Algorithms **12**(2), 34 (2019)
7. Huo, Y., Delahaye, D., Sbihi, M.: A dynamic control method for extended arrival management using enroute speed adjustment and route change strategy. Transportation Research Part C: Emerging Technologies **149**, 104064 (2023)
8. Javadi-Abhari, A., Treinish, M., Krsulich, K., Wood, C.J., Lishman, J., Gacon, J., Martiel, S., Nation, P.D., Bishop, L.S., Cross, A.W., Johnson, B.R., Gambetta, J.M.: Quantum computing with qiskit (2024)
9. Ma, J., Delahaye, D., Sbihi, M., Scala, P.: Integrated optimization of arrival, departure, and surface operations. In: International Conference for Research in Air Transportation 2018. ICRAT (2018)
10. Machrouh, J., Purica, A., Tevissen, J., Barbaresco, F.: Qualification/validation of AI-augmented ATM solutions for sustainable aviation. In: Towards Sustainable Aviation Summit 2025. 3AF (2025)
11. Stollenwerk, T., O'Gorman, B., Venturelli, D., Mandra, S., Rodionova, O., Ng, H., Sridhar, B., Rieffel, E.G., Biswas, R.: Quantum annealing applied to de-conflicting optimal trajectories for air traffic management. IEEE Transactions on Intelligent Transportation Systems **21**(1), 285–297 (2020)

QAOA in Quantum Datacenters: Parallelization, Simulation, and Orchestration

Amana Liaqat[1], Ahmed Darwish[2], Adrian Roman[1], and Stephen DiAdamo[2]

[1] Qoro Quantum, London, England
[2] Qoro Quantum, Munich, Germany
stephen@qoroquantum.de

Abstract. Scaling quantum computing requires networked systems, leveraging HPC for distributed simulation now and quantum networks in the future. Quantum datacenters will be the primary access point for users, but current approaches demand extensive manual decisions and hardware expertise. Tasks like algorithm partitioning, job batching, and resource allocation divert focus from quantum program development. We present a massively parallelized, automated QAOA workflow that integrates problem decomposition, batch job generation, and high-performance simulation. Our framework automates simulator selection, optimizes execution across distributed, heterogeneous resources, and provides a cloud-based infrastructure, enhancing usability and accelerating quantum program development. We find that QAOA partitioning does not significantly degrade optimization performance and often outperforms classical solvers. We introduce our software components—Divi, Maestro, and our cloud platform—demonstrating ease of use and superior performance over existing methods.

Keywords: Quantum computing, hybrid computing, QAOA, distributed quantum computing, parallel quantum computing, cloud quantum computing, quantum simulation, quantum datacenter.

1 Introduction

To scale quantum computing infrastructure to practical levels, both in the near and long term, networked systems are essential. In the near term, quantum computers can be integrated with classical High Performance Computing (HPC) infrastructure, forming quantum datacenters [1, 8]. These datacenters offer users unified access to classical and quantum resources through a single interface. In the long term, quantum systems are expected to evolve into fully networked clusters, where quantum computers share entanglement over quantum networks to enable distributed quantum computing.

In the current near-term scenario, quantum computers in datacenters are often not connected via quantum networks. Instead, they function as distributed systems in the traditional sense, relying on classical connections. While this setup offers performance gains [2, 3, 7], it also imposes limitations. Without quantum

entanglement between devices, qubit resources cannot be effectively shared, limiting scalability. Nevertheless, even classical networking can provide runtime improvements, though the specific benefits depend heavily on the application and how computational tasks are managed [5, 9].

Our proposed solution, aligned with efforts like [8, 11], integrates high-performance simulation and hardware execution under a unified interface. This approach allows users to focus on solving problems without needing to manage infrastructure complexities. While some solutions already exist, there remains a gap in the ecosystem when it comes to optimizing specific applications for massively parallel and distributed systems. Developers still face fundamental questions: How should quantum algorithms be parallelized for specific hardware? How can batch jobs be structured to avoid overwhelming resources? Which simulators are best suited for validating solutions? And, when transitioning from simulation to real hardware, how much of the workflow needs modification?

Our workflow incorporates the three essential components in the following way: (1) Problem decomposition, which partitions a constraint graph into manageable subgraphs; (2) Efficient subproblem generation; and (3) Automated batch job simulation, enabling the execution of diverse circuit sets at scale. We present a software library that automates these processes, reducing the user's workload and facilitating the efficient generation and execution of QAOA batch jobs through a unified interface.

On the simulation side, we develop a common interface that integrates multiple simulation engines. A key feature of this interface is its ability to act as a single point of entry for diverse simulators and simulation types. It automates the selection of the most suitable simulator for each circuit within a batch job, optimizing execution based on circuit characteristics such as complexity, size, and entanglement. Managing interoperability between different simulators is challenging due to variations in input formats, capabilities, and memory requirements. Our approach abstracts these differences, allowing users to run circuits on different backends without manual intervention, and maintaining performance of the simulator.

To ensure accessibility, we deploy these services through a cloud-based infrastructure. This system allows users to execute QAOA workflows without needing access to local high-performance computing resources. The cloud platform manages batch job submissions, allocates computational resources dynamically, and provides seamless scaling based on demand. This setup enables users to run large-scale QAOA experiments without worrying about the underlying infrastructure, ensuring efficient resource utilization and optimized performance.

The remainder of this paper is organized as follows: Section 2 details our automated parallelization workflow for QAOA and introduces Divi, our software library for managing this process. Section 3 discusses Maestro, our interface for integrating multiple simulators, and presents benchmark results. Finally, Section 4 concludes the paper and discusses future research directions.

2 Efficient Execution Workflow for QAOA

To simplify the execution of large-scale QAOA problems on distributed systems, we develop a workflow and software package that automates the parallelization process. A key question that arises is whether problem partitioning affects solution quality. We first evaluate this on the MaxCut problem, detailing our approach to partitioning. Beyond partitioning, we introduce parallelization at the optimization step, integrating these components into our software library, Divi.

2.1 Constraint-Graph Partitioning for QAOA

An integration of graph partitioning techniques with a quantum resource-aware approach is essential. To address the challenges posed by large-scale optimization problems within the constraints of near-term quantum computing, a flexible methodology that balances both the problem's specific characteristics and the hardware's limitations is required. A balanced and adaptive strategy is crucial for optimizing performance and achieving practical advancements in the current quantum computing era.

In our methodology, we employ spectral partitioning to decompose a large graph into smaller subgraphs, enabling the application of QAOA to solve the MaxCut problem on quantum processors with practical limitations. This carries over to simulators that also have strict limitations in qubit counts. Spectral partitioning methods are advantageous for partitioning graphs as they balance subgraphs' sizes effectively while preserving the overall structure of the principal graph. In our implementation, rather than employing spectral bisection, we use k-means clustering to partition the graph using the Python library *scikit-learn*, specifically its SpectralClustering algorithm [10]. This enables us to partition based on the desired number of clusters, which is constrained by the qubit size that the subgraph must adhere to. This approach helps to mitigate the information loss associated with recursive bisection methods.

Once the graph is partitioned, each graph is optimized independently. The results of each sub-problem are then aggregated, resulting in one final result. The approach leverages graph partitioning to break down a complex optimization problem into smaller, independent subproblems, which are solved separately before recombining the results.

2.2 Partitioned Optimization Evaluation

To evaluate the performance of partitioned QAOA, we use the MaxCut problem on various graph structures of varying size. The MaxCut problem entails partitioning a graph into two sets to maximize the number of edges between them. The MaxCut problem has applications in logistics, network design, and machine learning [4]. Additionally, the MaxCut problem serves as a standard benchmark for evaluating optimization algorithms due to its NP-hard complexity and will be used as the primary performance metric in our results.

When solving MaxCut with QAOA, each graph node corresponds to a qubit, totaling n qubits for n nodes. Edges are represented by controlled-Z (CZ) gates, entangling the qubits. This setup encodes the problem into a quantum circuit, with each qubit's state ($|0\rangle$ or $|1\rangle$) representing the node's set assignment.

While assessing the performance of solvers, the cut-size serves as the primary metric for solution quality, where the higher the cut size, the better the result. Fig. 1 illustrates the comparative outcomes between a hybrid approach (graph partitioning followed by QAOA execution under qubit-count constraints) and the standalone Goemans-Williamson (GW) algorithm on circulant graphs. GW is an approximate solver that provides a guaranteed performance ratio of at least 0.878 for the MaxCut problem through semidefinite programming relaxation [6]. We use this solver as our classical benchmark due to its computational efficiency and well-established theoretical guarantees, making it a standard reference point for evaluating quantum approaches to combinatorial optimization problems.

The data demonstrates QAOA achieves parity with GW for small-scale graphs of circulant type ($n \lessapprox 500$), but diverges progressively as problem size increases, revealing an optimality gap. Crucially, QAOA implementations with larger subproblem dimensions (e.g., with qubit counts $n = 14$ vs. $n = 6$) exhibit enhanced approximation ratios, narrowing the GW performance gap. This empirical correlation underscores an inverse relationship between partition-induced subproblem reduction and solution fidelity—a trade-off inherent to divide-and-conquer solution strategies.

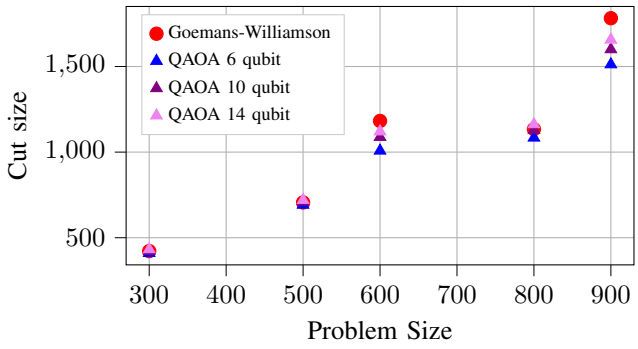


Fig. 1: Approximation quality comparison between partitioned QAOA and the Goemans-Williamson solver across different graph sizes. Larger subproblems yield higher fidelity.

2.3 Divi: Automated Parallelization of Quantum Algorithms

As quantum computing hardware advances, the complexity of quantum algorithms and their execution requirements increase accordingly. Many quantum

algorithms, particularly hybrid quantum-classical algorithms such as QAOA, VQAs, or quantum machine learning, are well structured to be parallelized and distributed across hybrid computing infrastructure. To leverage this, it requires an understanding of the underlying structure of the algorithms and a software infrastructure to perform the various tasks involved to generate and aggregate parallelized quantum programs.

To address these challenges, we have developed Divi, a Python library designed to streamline the execution of large-scale quantum algorithms by automating the parallelization and batching of quantum programs. Divi is designed to integrate seamlessly with cloud infrastructure, automating away the complexity of parallelizing, batching, and aggregating quantum programs. Divi brings the following key features:

(1) Automatic Parallelization of Quantum Programs. Divi automates the parallelization of hybrid quantum programs by decomposing complex problems into smaller, independently executable tasks, enabling efficient optimization while reducing user burden. Using a hierarchical structure, it supports both parallel and serial optimization methods, including Monte Carlo, genetic algorithms, and Nelder-Mead, and allows users to define multiple decision paths to generate additional jobs when beneficial. Divi manages iterative updates, computes expectation values across measurement bases, handles fallback cases, and aggregates results into a final output, automating job tracking and synchronization through a cloud interface to ensure completeness and reliability.

(2) Efficient Batch Job Generation. To serialize and execute quantum subprograms, Divi uses Pennylane to generate circuit abstractions due to its flexible modeling and QASM transpilation capabilities. However, Pennylane's QASM export (v0.40) mishandles expectation value computations by producing only one QASM file per Hamiltonian, rather than one per term with distinct measurement bases. Divi resolves this by implementing a custom export routine that correctly generates separate QASM programs per term. Additionally, Pennylane's repeated transpilation of fixed-structure circuits incurs heavy computational overhead, which is inefficient for variational algorithms. Divi avoids this by creating a symbolic QASM scaffold once per subprogram and binding parameters at runtime, dramatically reducing redundant compilation. This approach improves performance significantly, e.g., reducing generation time for 1,500 circuits by a factor of 13—highlighting a broader inefficiency in frameworks lacking symbolic caching.

3 Simulating Batches of Circuits

Quantum circuit simulation uses various methods, each offering trade-offs between accuracy, scalability, and computational efficiency. State-vector simulation is the most direct approach, representing the quantum state as a complex vector of dimension 2^n, where n is the number of qubits. This method allows exact simulation and full access to state amplitudes, making it ideal for debugging and analyzing quantum algorithms. However, its exponential memory requirements

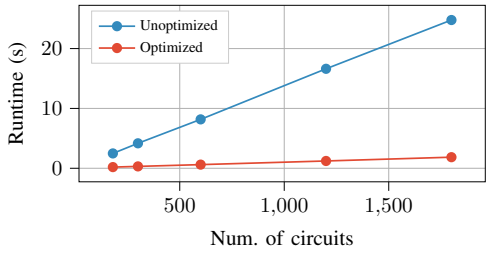


Fig. 2: Time it takes to generate a batch of circuits for QAOA for 5 nodes and varying the number of samples with MC optimization.

limit scalability—simulating more than 30 qubits can require hundreds of gigabytes of memory—making it impractical for large systems but effective for small- to medium-scale circuits.

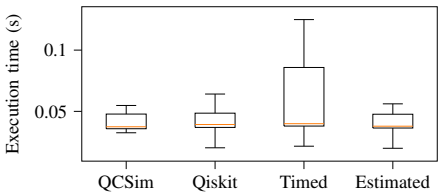
Choosing the right simulation method involves balancing precision, scalability, and computational resources. State-vector simulators are ideal for exact results in small circuits, while MPS and tensor networks extend capabilities to larger or structured systems. GPU-accelerated methods can enhance performance for large, parallelizable workloads but may struggle with smaller or memory-intensive tasks. Additionally, factors like hardware configuration—cores, memory, and architecture—impact performance, complicating simulator selection, especially in batch processing scenarios with diverse circuits.

Maestro, our platform, addresses these challenges by integrating multiple simulation paradigms into a unified interface. It allows users to leverage the most appropriate simulator for their specific needs while maintaining interoperability across quantum hardware and software ecosystems.

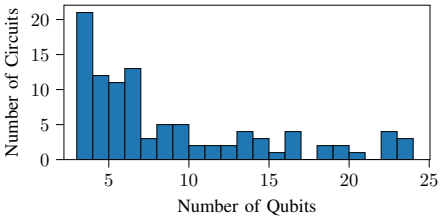
Maestro provides an automated simulator selection mechanism to optimize performance. When a circuit is to be run, at runtime, Maestro has two ways to decide which simulator to use. Firstly it can execute a single shot across multiple supported simulators in parallel, records execution times, and selects the right backend for the remaining shots, with an understanding of the multi-shot optimization performance for each. This method is simple and flexible in that it works regardless of software updates to the simulators and the underlying hardware. The downside is, running one shot on many simulators occupies computing resources and can delay the computation in some case.

The second method we use, as an alternative, uses simulation algorithm complexity with circuit information in addition to hardware features (determined at compile time) to make the decision. At compile time, the constants for the algorithm complexity are determined based on the hardware, and then used to select the right simulator for the input circuit. Regression methods are used to extrapolate runtime. This method is very efficient, as it is essentially a lookup task, but it can be challenging to determine the complexity of the simulation algorithm for each simulator, especially when they use parallelization and multi-

threading, and requires care. This is an ongoing research direction which we aim to explore more deeply.



(a) Execution times for a batch job of circuits.



(b) Distribution of circuits executed.

Fig. 3: Execution time distribution for a batch job of circuits. All circuits run for 5,000 shots.

4 Outlook and Conclusion

In this work, we present a comprehensive framework that transforms the execution of large-scale QAOA workflows by addressing key scalability challenges in quantum computing. By automating complex decision-making, problem partitioning, and batch job execution, we reduce the need for user intervention and streamline quantum workflows for greater scalability and usability. Our integrated system includes Divi, which parallelizes and batches QAOA through intelligent decomposition; Maestro, which dynamically selects optimal simulators via a unified interface; and our Cloud Platform, which orchestrates distributed execution across heterogeneous resources, including both simulators and real devices.

This framework bridges the gap between theoretical algorithm design and practical implementation, offering a critical step toward realizing useful quantum computing. We are actively extending its capabilities with realistic noise modeling and GPU-accelerated simulators to further improve performance. Looking ahead, we see distributed quantum computing—enabled by multi-vendor, multi-modal networks—as a promising path to scale quantum circuits beyond classical limits. Achieving this requires robust software infrastructure that abstracts complexity while coordinating diverse resources. Our work contributes to this goal by promoting standardization and automation, paving the way for quantum computing to deliver real-world impact.

References

1. Alexeev, Y., Amsler, M., Barroca, M.A., Bassini, S., Battelle, T., Camps, D., Casanova, D., Choi, Y.J., Chong, F.T., Chung, C., et al.: Quantum-centric super-

computing for materials science: A perspective on challenges and future directions. Future Generation Computer Systems **160**, 666–710 (2024)
2. Carrera Vazquez, A., Tornow, C., Ristè, D., Woerner, S., Takita, M., Egger, D.J.: Combining quantum processors with real-time classical communication. Nature pp. 1–5 (2024)
3. Chen, K.C., et al.: Noise-aware distributed quantum approximate optimization algorithm on near-term quantum hardware. arXiv preprint arXiv:2407.17325 (2024)
4. Dunning, I., Gupta, S., Silberholz, J.: What works best when? a systematic evaluation of heuristics for max-cut and qubo. INFORMS Journal on Computing **30**(3), 608–624 (2018)
5. Giortamis, E., Romão, F., Tornow, N., Lugovoy, D., Bhatotia, P.: Orchestrating quantum cloud environments with qonductor. arXiv preprint arXiv:2408.04312 (2024)
6. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. Journal of the ACM (JACM) **42**(6), 1115–1145 (1995)
7. Mineh, L., Montanaro, A.: Accelerating the variational quantum eigensolver using parallelism. Quantum Science and Technology **8**(3), 035012 (2023)
8. Mohseni, M., Scherer, A., Johnson, K.G., Wertheim, O., Otten, M., Aadit, N.A., Bresniker, K.M., Camsari, K.Y., Chapman, B., Chatterjee, S., et al.: How to build a quantum supercomputer: Scaling challenges and opportunities. arXiv preprint arXiv:2411.10406 (2024)
9. Moretti, R., Giachero, A., Radescu, V., Grossi, M.: Enhanced feature encoding and classification on distributed quantum hardware. arXiv preprint arXiv:2412.01664 (2024)
10. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
11. Schulz, M., Ruefenacht, M., Kranzlmüller, D., Schulz, L.B.: Accelerating hpc with quantum computing: It is a software challenge too. Computing in Science & Engineering **24**(4), 60–64 (2022)

Towards Large-Scale Satellite Acquisition Scheduling with Hybrid Quantum-Classical Optimization

Amer Delilbasic[1,2 [0000−0001−7845−5193]], Morris Riedel[1,2,3 [0000−0003−1810−9330]], Kristel Michielsen[3,4 [0000−0003−1444−4262]], and Gabriele Cavallaro[1,2,3 [0000−0002−3239−9904]]

[1] Forschungszentrum Jülich, 52428 Jülich, Germany
a.delilbasic@fz-juelich.de
[2] University of Iceland, 102 Reykjavík, Iceland
[3] AIDAS, 52425 Jülich, Germany
[4] RWTH Aachen University, 52056 Aachen, Germany

Abstract. The need for on-demand satellite acquisitions for Earth observation is rapidly increasing, along with the availability of public and commercial satellite constellations capable of fulfilling such requests. Scheduling acquisitions optimally remains a computationally challenging task, especially in complex scenarios with multiple satellites, constraints, and high number of requests. Today, heuristic algorithms are commonly used to find sub-optimal solutions that balance computational time and acquisition value, for example in terms of economic return. The potential role of quantum optimization in this domain is still an open question. While recent demonstrations have explored quantum approaches, they are limited to small problem instances due to current hardware constraints and the difficulty of simulating large quantum systems with classical hardware.

In this work, we present a hybrid quantum-classical optimization approach to address acquisition scheduling problems of any size. At each iteration, a classical heuristic generates a global candidate solution, from which a smaller local sub-problem is extracted and solved using a quantum optimization algorithm. We validate the method on large-scale instances, leveraging D-Wave Advantage to solve the sub-problems with quantum annealing. Initial results show the feasibility of the approach and highlight the capability of quantum annealing to guide the hybrid solver towards better overall acquisition schedules than simulated annealing in a comparable amount of time, in particular as the complexity increases.

Keywords: quantum optimization · hybrid quantum-classical computation · quantum annealing · scheduling

1 Introduction

The growing demand for timely and high-resolution Earth observation data has led to a rapid increase in both public and commercial imaging satellite constellations [4, 7]. These systems are expected to support diverse applications, ranging from environmental monitoring and disaster response to defense and agriculture [3]. However, fully exploiting these satellite resources requires intelligent and efficient scheduling of imaging tasks [19]. Acquisition scheduling involves assigning requests to specific satellites in a way that maximizes acquisition value while satisfying a wide set of operational constraints. The total value of an acquisition schedule can be related to economic return, for the case of commercial imaging systems, or to urgency in performing the acquisition, for the case of rapidly evolving extreme events such as fires or eruptions. The underlying constraints can involve satellite orbits, sensor agility, on-board resources, weather conditions, and ground station availability.

Due to the problem's combinatorial complexity, exact solutions are often infeasible for large scenarios. Heuristic and metaheuristic algorithms, such as greedy methods, genetic algorithms, or Simulated Annealing (SA), are commonly used to find good-enough solutions within reasonable time frames [2]. However, as the problems become larger in practical cases, the performance of existing methods degrades. In recent years, quantum optimization has been explored. Early studies have shown promising results on small satellite acquisition scheduling problems, e.g., using Quantum Annealing (QA) [16, 11] and Quantum Approximate Optimization Algorithm (QAOA) [14, 12]. However, scalability remains limited, due to algorithmic limitations, current quantum hardware constraints, and the difficulty of mapping large problems to existing quantum systems. For example, QAOA has only been tested in simulation, already showing degrading performance with more than 10 requests, and QA can only realistically solve problems with less than 100 requests.

In the optimization literature, hybrid quantum-classical methods have emerged as a practical strategy for tackling large, NP-hard problems that exceed current quantum hardware limits. A common pattern is to use a classical algorithm, such as metaheuristics, mathematical programming, or constraint-based search, to generate and maintain a feasible global solution, then identify smaller, well-defined sub-problems that can be mapped to a quantum solver. This approach has been applied in domains such as scheduling [18], routing [8], and portfolio selection [15]. Makarov et al. [12] also successfully solved satellite acquisition scheduling problems with up to 209 requests with the D-Wave `LeapBQMHybrid` solver. However, it cannot be determined whether and to which extent this specific solver benefits from quantum optimization, leaving the door open to research in the direction of hybrid optimization.

In this paper, we present a hybrid quantum-classical method designed for large-scale satellite acquisition scheduling problems. The method relies on a classical heuristic to generate a global candidate solution, from which local, constraint-aware sub-problems are extracted and solved using a quantum annealer. This hybridization allows us to exploit quantum hardware without being

limited by its size. We validate the approach on large problem instances with up to 20000 requests using the Advantage quantum annealer from D-Wave Systems [5]. The rest of the paper is structured as follows: Sect. 2 provides a problem formulation for satellite acquisition scheduling. Sect. 3 introduces the proposed approach. Sect. 4 provides results from experimental validation. Sect. 5 concludes the paper and summarizes possible future research directions.

2 Problem Formulation for Satellite Acquisition Scheduling

Satellite acquisition scheduling varies significantly across systems and applications. It fundamentally involves assigning acquisition requests to satellites in a way that maximizes the overall value of collected data while minimizing acquisition time. These problems are subject to a range of operational constraints, which may involve satellite configuration, sensor characteristics, on-board memory capacity, cloud coverage, illumination conditions, or ground station communication windows.

In this work, we focus on a simplified acquisition scheduling formulation, first introduced in the work of Quetschlich et al. [14]. The setup involves a single satellite tasked with fulfilling up to R acquisition requests $\{r_i : i = 1, ..., R\}$ during one orbital pass. Each acquisition i is associated with a binary decision variable $x_i \in 0, 1$, where $x_i = 1$ indicates that the acquisition is selected. Each request also has an associated value v_i, representing, for instance, economic importance or urgency, as highlighted in Sect. 1.

The primary constraint arises from the time required to rotate the satellite's sensor between targets located at different off-nadir angles. Let $T_{rotation}(i, j)$ denote the sensor rotation time between acquisitions i and j, and $T_{transition}(i, j)$ the available orbital transition time between the two. The optimization problem is:

$$\max_{x \in 0,1^R} \sum_{i=1}^{R} x_i v_i \quad \text{subject to: } T_{rotation}(i, j) \leq T_{transition}(i, j) \quad \text{if } x_i = x_j = 1 \tag{1}$$

A visual depiction of this acquisition setup is provided in Fig. 1.

The problem is reformulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem, the constraint being integrated into the energy (objective) function as a penalty term:

$$\max_{x \in \{0,1\}^R} \sum_{i=1}^{R} x_i v_i - p \cdot \sum_{i=1}^{R-1} \sum_{j=i}^{R} x_i x_j \cdot c(i, j) \tag{2}$$

Here, p is a penalty coefficient, and $c(i, j) = 1$ if $T_{rotation}(i, j) > T_{transition}(i, j)$, and 0 otherwise. This ensures compatibility with quantum optimization methods such as QA [10], as this formulation is fundamentally equivalent to the Ising

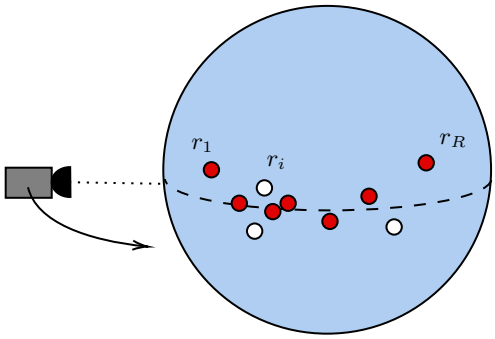


Fig. 1: Considered setting for satellite acquisition scheduling, with acquisition requests r_i represented as circles along the satellite trajectory projection on the Earth's surface. The objective consists in finding a subset of acquisition requests (here, a possible subset is indicated in red) that maximizes the acquired value and satisfies the rotation constraint. The image was first introduced in [6].

model. Note that we effectively convert Eq. 2 to a minimization problem by inverting the sign of the terms, as this is a stardard for optimization. Despite being a simplified formulation with a linear cost function, one satellite and one set of constraints, computational limitations arise when the number of requests R increases, as sensor rotation contraints increase at an even faster rate.

3 Hybrid Quantum-Classical Optimization for Satellite Acquisition Scheduling

As mentioned in the previous sections, finding optimal acquisition schedules for a high number of acquisition requests R is prohibitive. Heuristic methods based on quantum optimization cannot be directly applied, due to the aforementioned limitations in computational capabilities. Here, we introduce an iterative approach to integrate quantum optimization within an acquisition scheduling algorithm.

Full problem solution sampling. A conventional (classical) heuristic method can start tackling the large problem and sampling candidate solutions. The solution associated with the lowest energy, corresponding to a sub-optimal acquisition schedule, can be the initial solution for a subsequent local refinement.

Sub-problem generation. In order to locally improve on the initial solution, a suitable sub-problem is generated. In Fig. 2, the proposed sub-problem generation method is shown. The method selects a local subset of acquisition requests within a window, for which rotation constraints (see Eq. 1) are in place with respect to the acquisitions delimiting the window. First, two acquisitions r_i and r_j belonging to the initial solution are chosen. This ensures that the updated schedules for the selected acquisitions present no constraint violations with the initial

acquisition schedule outside the window. The window is iteratively extended until a S-dimensional sub-problem is reached. An independent sub-problem is thus generated.

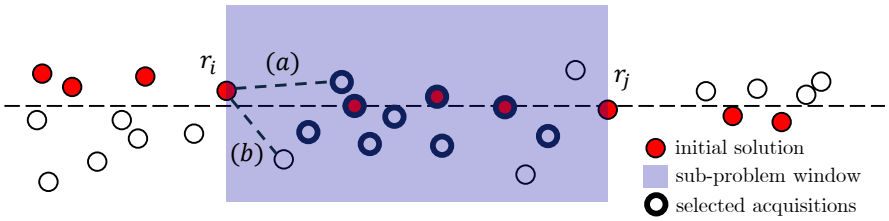


Fig. 2: Example of sub-problem generation. Circles represent acquisition requests distributed along the satellite orbit projection. Two acquisitions r_i and r_j belonging to the initial solution to the acquisition scheduling (in red) define a window (in blue). The selected acquisitions for the sub-problem generation (circled in blue) are all acquisitions within the window for which the rotation constraint is satisfied, i.e., $T_{rotation}(r_{i,j}, r) \leq T_{transition}(r_{i,j}, r)$ (a), excluding the incompatible ones (b).

Sub-problem solution sampling. The size S of the sub-problem can be adaptively selected so that a quantum optimization algorithm, such as QA, can sample solutions for it. The initial solution can also be exploited, e.g., using advanced schedules such as reverse quantum annealing [13, 6] or for warm-starting QAOA [14]. The best solution is used to update the schedule, which then becomes the initial solution of the next iteration of full and sub-problem solution sampling.

4 Experimental Validation

In this section, we show initial results for hybrid optimization on large acquisition scheduling problems, following the approach of Sect. 3. We generated two problems, with requests $R = 10000$ and $R = 20000$, each request having value $v_i = \{1, 2\}$, with overall penalty coefficient $p = \max(x_i + x_j) + 1 = 5$. In order to validate the iterative approach and understand the impact of hybrid optimization, we worked with two experimental setups:

1. *Simulated annealing - simulated annealing (SA-SA)*: SA is used for both full problem and sub-problem solution sampling. It is executed on 2 Intel Xeon Cascade Lake Platinum 8260M CPUs from the DEEP HPC system located at the Jülich Supercomputing Centre (JSC) [17].
2. *Simulated annealing - quantum annealing (SA-QA)*: in this case, forward QA is used for sub-problem solution sampling. QA is executed using the `Advantage_system_6.4` solver, accessed through D-Wave Leap [1].

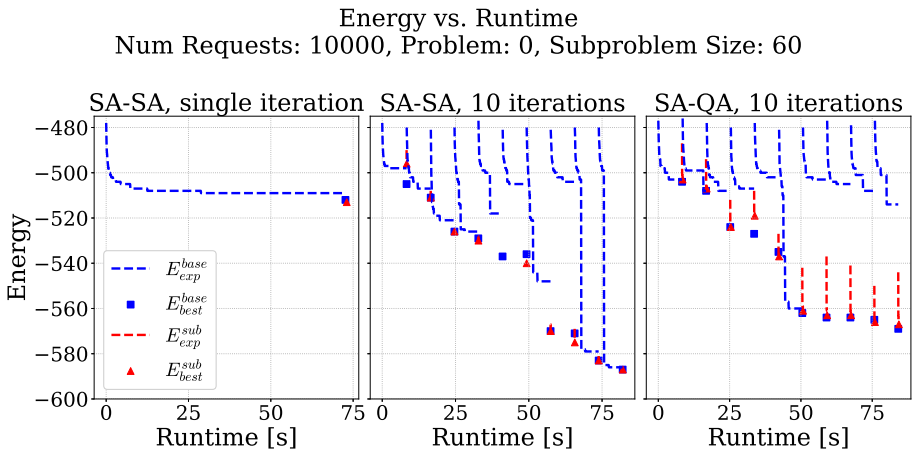
We generate subproblems of a fixed size of 60 requests, to ensure a feasible problem embedding into the Advantage system. We perform a fixed number of runs for each solver (100 runs at 1000 sweeps for SA, 1000 runs at $100\mu s$ annealing time for QA) and 10 iterations. The high number of runs ensures statistical significance of the obtained results, from which we collect suitable metrics for practical use. We compute a discrete cumulative distribution function (CDF) for the statistical distribution of the sample energies collected by all the runs. We can derive the expected energy for each iteration I as the energy value for which $CDF(\text{expected_energy}_I) = 1 - 0.5^I$. A continuous inverse CDF, needed for finding the expected energy, is obtained by interpolating the discrete CDF using PCHIP 1-D monotonic cubic interpolation [9] and inverting it. The results for the two setups are compared to a single SA-SA iteration with 1000 runs. The code for the method and the experimental setup is made available for reproducibility[5].

A plot for the expected and best solution energies vs. runtime is shown in Fig. 3. The best solution found in our runs for each subroutine was used as initial solution of the subsequent subroutine. This initial experiment confirms the feasibility of the iterative sub-problem generation approach, as the solution search generally improves at each iteration. QA is also able to run and sample feasible solutions for the generated sub-problems. The question is whether it is worth running QA over SA. For the $R = 10000$ problem instance, this is not the case, except for a better energy found by the SA-QA setup at iteration 6, after which no significant improvement is reached. In the $R = 20000$ problem instance, solutions updated by QA actually lead to better final acquisition schedules compared to the SA-SA setup. This could be related to the fact that a higher acquisition density leads to higher problem complexity, where QA can have an edge. The higher solution variability of QA, signaled by the varying expected sub-problem energy, also indicated a difference in the exploration of the solution space. Another indicator is the high difference between expected and best energies obtained in the SA-QA case, showing that QA shifts from the conventional statistical distribution of SA solutions.

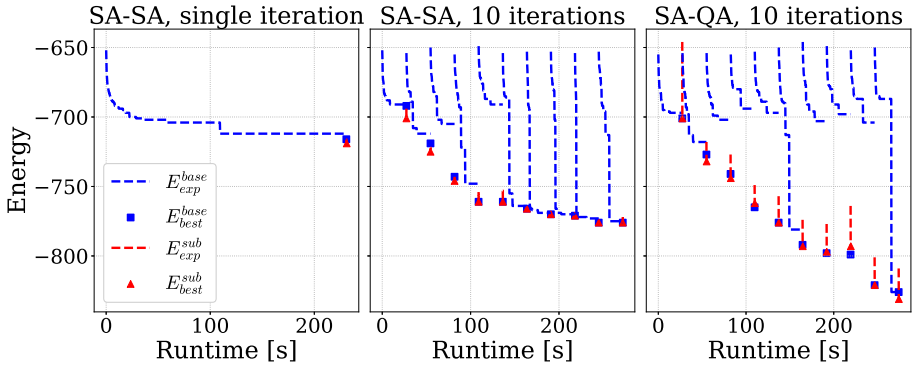
5 Conclusion

The presented hybrid approach to optimization enables experimental validation of quantum optimization subroutines in acquisition scheduling, with initial promising results for QA. A higher contribution of quantum optimization, finding better solutions and working on larger subproblems, is an expected outcome of future quantum machines. Practical use will also benefit from tight on-premise integration of quantum machines, which will reduce communication delays. Future work will focus on extending the validation to other optimization methods, beyond standard QA, including a deeper analysis of the benefits of quantum optimization. In addition, further scalability will also be studied, working with a higher number of requests and more complex acquisition scenarios.

[5] `https://gitlab.jsc.fz-juelich.de/sdlrs/qmp`



(a) $R = 10000$



(b) $R = 20000$

Fig. 3: Hybrid optimization results for two different acquisition scheduling problems, i.e., number of requests $R = 10000$ (a) and $R = 20000$ (b), and two algorithm combinations (SA-SA, SA-QA). Results are compared to one SA-SA iteration, executed for a comparable amount of time. Blue plots show the expected (E_{exp}^{base}) and best (E_{best}^{base}) energies recorded on the full problem of size 60. Red plots show the expected (E_{exp}^{sub}) and best (E_{best}^{sub}) energies recorded on the generated sub-problem of size 60.

Acknowledgments. The authors gratefully acknowledge support from the project JUNIQ that has received funding from the German Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the State of North Rhine-Westphalia. This work is part of the Quantum Computing for Earth Observation (QC4EO) initiative from the ESA Φ-lab. This work is co-financed by the EUROCC2 project funded by the European High-Performance Computing Joint Undertaking (JU) and EU/EEA states under grant agreement No 101101903. This work is partially supported by the Icelandic Centre for Research, project *"Hybrid Quantum-Classical Workflows for EO"*, RANNÍS grant number: 2511078-051 (`http://en.rannis.is/`).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. D-Wave Leap (May 2025), `https://cloud.dwavesys.com/leap/`
2. Agnèse, J.C., Bataille, N., Blumstein, D., Bensana, E., Verfaillie, G.: Exact and Approximate Methods for the Daily Management of an Earth Observation Satellite. Proceedings of SpaceOPS (01 1996)
3. Bovolo, F., Bruzzone, L., Fernández-Prieto, D., Paris, C., Solano-Correa, Y.T., Volden, E., Zanetti, M.: Big data from space for precision agriculture applications. IOP Conference Series: Earth and Environmental Science **509**(1), 012004 (jun 2020). https://doi.org/10.1088/1755-1315/509/1/012004, `https://dx.doi.org/10.1088/1755-1315/509/1/012004`
4. Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., Nieke, J.: The copernicus hyperspectral imaging mission for the environment (chime): Status and planning. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. pp. 5011–5014 (2022). https://doi.org/10.1109/IGARSS46834.2022.9883592
5. D-Wave Systems: https://www.dwavesys.com/, accessed: September 19, 2025
6. Delilbasic, A., Le Saux, B., Riedel, M., Michielsen, K., Cavallaro, G.: Reverse Quantum Annealing for Hybrid Quantum-Classical Satellite Mission Planning. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. pp. 432–436 (Jul 2024). https://doi.org/10.1109/IGARSS53475.2024.10640974, `https://ieeexplore.ieee.org/abstract/document/10640974`, iSSN: 2153-7003
7. Denby, B., Lucia, B.: Orbital edge computing: Nanosatellite constellations as a new class of computer system. In: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems. p. 939–954. ASPLOS '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3373376.3378473, `https://doi.org/10.1145/3373376.3378473`
8. Feld, S., Roch, C., Gabor, T., Seidel, C., Neukart, F., Galter, I., Mauerer, W., Linnhoff-Popien, C.: A Hybrid Solution Method for the Capacitated Vehicle Routing Problem Using a Quantum Annealer. Frontiers in ICT **6** (Jun 2019). https://doi.org/10.3389/fict.2019.00013, `https://www.frontiersin.org/journals/ict/articles/10.3389/fict.2019.00013/full`, publisher: Frontiers

9. Fritsch, F.N., Butland, J.: A Method for Constructing Local Monotone Piecewise Cubic Interpolants. SIAM Journal on Scientific and Statistical Computing **5**(2), 300–304 (1984). https://doi.org/10.1137/0905021, https://doi.org/10.1137/0905021
10. Kadowaki, T., Nishimori, H.: Quantum Annealing in the Transverse Ising Model. Physical Review E **58**(5), 5355–5363 (1998)
11. Makarov, A., Taddei, M.M., Osaba, E., Franceschetto, G., Villar-Rodríguez, E., Oregi, I.: Optimization of image acquisition for earth observation satellites via quantum computing. In: Quaresma, P., Camacho, D., Yin, H., Gonçalves, T., Julian, V., Tallón-Ballesteros, A.J. (eds.) Intelligent Data Engineering and Automated Learning – IDEAL 2023. pp. 3–14. Springer Nature Switzerland, Cham (2023)
12. Makarov, A., Pérez-Herradón, C., Franceschetto, G., Taddei, M.M., Osaba, E., del Barrio Cabello, P., Villar-Rodriguez, E., Oregi, I.: Quantum Optimization Methods for Satellite Mission Planning. IEEE Access **12**, 71808–71820 (2024). https://doi.org/10.1109/ACCESS.2024.3402990, https://ieeexplore.ieee.org/abstract/document/10534762
13. Pelofske, E., Hahn, G., Djidjev, H.N.: Advanced Anneal Paths for Improved Quantum Annealing. Proceedings - IEEE International Conference on Quantum Computing and Engineering, QCE 2020 pp. 256–266 (10 2020). https://doi.org/10.1109/QCE49297.2020.00040
14. Quetschlich, N., Koch, V., Burgholzer, L., Wille, R.: A Hybrid Classical Quantum Computing Approach to the Satellite Mission Planning Problem. arXiv preprint (Jul 2023), https://arxiv.org/abs/2308.00029v1
15. Sakuler, W., Oberreuter, J.M., Aiolfi, R., Asproni, L., Roman, B., Schiefer, J.: A real-world test of portfolio optimization with quantum annealing. Quantum Machine Intelligence **7**(1), 43 (Mar 2025). https://doi.org/10.1007/s42484-025-00268-2, https://doi.org/10.1007/s42484-025-00268-2
16. Stollenwerk, T., Michaud, V., Lobe, E., Picard, M., Basermann, A., Botter, T.: Agile Earth Observation Satellite Scheduling with a Quantum Annealer. IEEE Transactions on Aerospace and Electronic Systems **57**, 3520–3528 (10 2021). https://doi.org/10.1109/TAES.2021.3088490
17. Suarez, E., Kreuzer, A., Eicker, N., Lippert, T.: The DEEP-EST project. Porting Applications to a Modular Supercomputer-Experiences from the DEEP-EST Project; Schriften des Forschungszentrums Jülich IAS Series pp. 9–25 (2021), https://juser.fz-juelich.de/record/905812/files/IAS_Series_48-1.pdf
18. Tran, T., Do, M., Rieffel, E., Frank, J., Wang, Z., O'Gorman, B., Venturelli, D., Beck, J.: A hybrid quantum-classical approach to solving scheduling problems. In: Proceedings of the International Symposium on Combinatorial Search. vol. 7, pp. 98–106 (2016)
19. Zhang, B., Wu, Y., Zhao, B., Chanussot, J., Hong, D., Yao, J., Gao, L.: Progress and challenges in intelligent remote sensing satellite systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **15**, 1814–1822 (2022). https://doi.org/10.1109/JSTARS.2022.3148139

Quantum model for CVRPTW

Imran Meghazi[1,2], Éric Bourreau[2]

[1] La Poste, France
[2] LIRMM, France
{imran.meghazi, eric.bourreau}@lirmm.fr

Abstract. This paper proposes a quantum algorithm for the capacitated vehicle routing problem with time windows (CVRPTW) based on Grover Search framework. This problem is often faced by Postal services in the context of package delivery or other time-sensitive operations. We provide an implementation on gate based quantum computer of a model inspired by classical route first, cluster second technique. The quantum paradigm allows to overcome suboptimality inherent property of this decomposition. In the current NISQ (Noisy Intermediate-Scale Quantum) era, the most important limitation is the number of available qubits which makes time windows and capacity constraints hard to tackle. We introduce a qubit-efficient split-inspired modeling which adds only a linear number of decision qubits to standard quantum formulations for Traveling Salesman Problem (TSP).

Keywords: operations research, quantum computing, optimization.

Introduction

Operations research aims to tackle a wide range of combinatorial problems, among which the Vehicle Routing Problem (VRP) is a prominent example. The VRP generalizes the well known Travelling Salesman Problem (TSP). The TSP seeks the shortest cycle that visits every node in a graph exactly once. In a logistics context, it can be interpreted as finding the most efficient route for a single vehicle to deliver one item to each customer. The VRP extends this formulation by considering multiple vehicles operating simultaneously, each subject to capacity constraints, in order to serve all customers while minimizing the total travel cost. This problem arises daily in logistics companies and organizations, such as the French postal service *La Poste*. In practice, delivery services face additional constraints. In this paper, we focus on two of them : a capacity constraint, which reflects the limited space of each vehicle, and a time window constraint that allows customers to specify a preferred delivery period. Solving the VRP with such constraints, referred to as the Capacitated Vehicle Routing Problem with Time Windows (CVRPTW), is computationally difficult and is usually adressed using heuristics [7] and bio-inspired algorithms [1].

Quantum Computing is a promising paradigm that has already produced algorithms outperforming their classical counterpart. In some cases, such as Shor's

algorithm [11] it offers exponential speedup, in others, like Grover's algorithm [6], a quadratic improvement. Grover's algorithm provides an oracle-based general framework for addressing combinatorial problems. While quantum oracle-based models for the Traveling Salesman Problem and QUBO formulations for the CVRPTW have been proposed, to the best of our knowledge, no quantum oracle-based formulation has yet been introduced for the CVRPTW. In this work, we propose such a model using Grover Adaptive Search [5] on gate-based quantum computers.

1 The Capacitated VRP with Time Windows

1.1 Problem definition

A *Capacitated Vehicle Routing Problem with Time Windows* can be defined as follows: a fleet of vehicles must deliver packages to a set of n customers $V = \{v_1, v_2, \ldots, v_n\}$, where all packages are initially stored at a central depot d. Each customer v_i requests a specific quantity of goods denoted by q_i. In addition to capacity constraints, customers specify a preferred delivery time window during the day. In this work, we consider the case where each customer is assigned a single time window, meaning that customer v_i must be served within the interval $[a_i, b_i]$.

1.2 Methods

In classical operations research, the Vehicle Routing Problem is typically addressed using two main approaches. The most well-known is the *cluster-first, route-second* heuristic, which involves grouping customers into clusters and then solving a separate routing problem, such as a TSP, within each cluster. For instance, customers may be clustered based on their geographical proximity. It is then assumed that solving a TSP for each cluster yields a VRP solution that is reasonably close to optimal. However, this assumption breaks down in the presence of time window constraints, which can invalidate geographically efficient clusterings. Several heuristics based on this strategy can be found in the literature, such as in [7].

An alternative strategy, albeit less intuitive, is the *route-first, cluster-second* approach, originally introduced by Beasley [2]. In this method, a so-called *giant tour* is constructed that visits all customers, and this tour is subsequently divided into feasible *sub-tours* that satisfy the problem constraints. We specifically focus on the *split* procedure presented in [10], where the partitioning of the giant tour is typically performed according to vehicle capacity limits.

1.3 Split Procedure

In this section, we detail the *split* procedure, as our quantum approach is heavily inspired.

The *split* algorithm takes as input an auxiliary graph $H = (X, A, W)$ constructed from the solution obtained in the first routing phase, namely the *giant tour*. The set X contains the nodes, ordered from node 0 to n according to this tour. Let P be this tour. The set of arcs A contains an arc (i, j), with $i < j$, if a trip visiting every customers between P_{i+1} and P_j doesn't violate any constraints. The corresponding weight w_{ij} represents the total cost of that sub-tour, computed as $D_{d,i+1} + \sum_{k=i+1}^{j-1} D_{P_k,P_{k+1}} + D_{j,d}$ with D the distance matrix. Finding a minimum cost path from node 0 to n in H yields an optimal splitting of the giant tour, where every arc corresponds to a feasible sub-tour that respects the original problem's constraints.

As an illustrative example, let us consider a typical instance of the Capacitated Vehicle Routing Problem.

Distance matrix

	d	1	2	3	4	5	6
d	0	23	30	23	14	20	26
1	23	0	17	27	27	38	36
2	30	17	0	21	40	46	37
3	23	27	21	0	35	40	12
4	14	27	40	35	0	16	31
5	20	38	46	40	16	0	33
6	26	36	32	12	31	33	0

Nodes

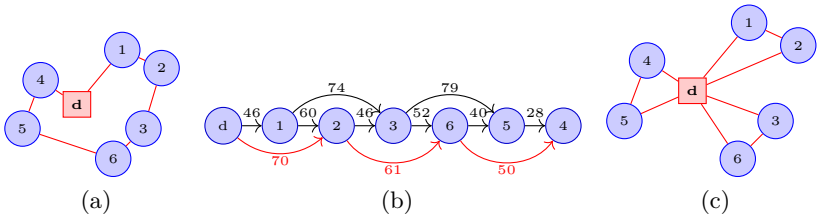
v_i	q_i
d	0
1	2
2	3
3	1
4	3
5	2
6	3

$C^{max} = 5$

The only constraints is that, for every subtour, the total must not exceed the vehicle's maximum capacity C^{max}. Stated differently, the cumulative load at k^{th} node, denoted c_{P_k}, must not exceed C^{max}.

$$\sum_{k=i+1}^{j} c_{P_k} \leq C^{max}, \ \forall (i, j) \in X, i < j \tag{1}$$

We first compute a giant tour, let P be the ordering of this tour, P_i is the ith node of the tour. Figure 1a show this giant tour and figure 1b shows the auxiliary graph.



(a) (b) (c)

Fig. 1. The steps to compute a solution. 1a the giant tour. 1b the auxiliary graph H. 1c the solution.

Finding an optimal split for this giant tour can be effectively done by computing the shortest path over figure 1b.

2 Quantum Model

In classical computing, a split-based formulation is inherently suboptimal, as the optimal VRP solution is not necessarily the optimal split of the optimal *giant tour*. Consequently, it requires evaluating every possible *giant tour* and all its splits, an exponential task. As a result, it is typically employed only as a highly efficient heuristic. In contrast, quantum superposition enables the parallel evaluation of all these possibilities, allowing the identification of the optimal *giant tour* for every possible splits.

2.1 Split Inspired model

Objective function Let $G = (V, E)$ be the graph instance. Let D denotes the distance matrix of this graph. Following the approach introduced in the previous section, we define separately the decision variables related to the ordering of customers and those related to the clustering (or splitting) of the route. Let P represent the *giant tour*, and let P_i denote its ith node.

Then we define the following variables :

$$y_i = \begin{cases} 1 & \text{if } P_i \text{ is the end of a sub-tour} \\ 0 & \text{otherwise} \end{cases}$$

The variables y_i encode the splitting decisions. Specifically, if $y_i = 1$, it indicates that the vehicle must return to the facility immediately after visiting node P_i. Consequently, we enforce $y_0 = 1$ and $y_n = 1$ to ensure that the tour starts and ends at the depot.

Thus the value to minimize is defined as follows :

$$\sum_{i=1}^{n-1} \left[D_{P_i,P_{i+1}} \left(1 - P_i\right) + \left(D_{P_i,f} + D_{f,P_{i+1}} \right) y_i \right] + D_{f,P_1} + D_{P_n,f} \tag{2}$$

As discussed in the previous section, the clustering process (or splitting, in this context) must be performed in accordance with the problem's constraints.

Uniqueness constraint Each customer must be visited exactly once, meaning that no customer may appear more than once in the tour.

$$P_i \neq P_j, \ \forall (i,j) \in V \tag{3}$$

Capacity constraint This constraint ensures that the total load of each sub-tour does not exceed the vehicle's maximum capacity. Given a fixed customer ordering, it can be formalized as shown in Equation 1. In practical terms, this means that at each node i, the cumulative load c_i must be less than or equal to the vehicle's capacity limit C^{max}.

$$c_i \leq C^{max}, \ \forall i \in V \tag{4}$$

In the quantum computing context, all operations must be reversible. As a result, it is necessary to store the load at each preceding node in order to compute the current load. The load at node i is thus calculated as the sum of the loads of the previous nodes, conditioned by the y_i variables, plus the load associated with node i itself. As previously defined, $y_{i-1} = 1$ indicates that node $i-1$ marks the end of a sub-tour, implying that the next node and — thus next subtour— is served by a new vehicle.

$$c_i = q_{P_i} + (c_{i-1} \times (1 - y_{i-1})) \tag{5}$$

Time window constraint This constraint behaves similarly to the capacity constraint. The visiting time at node i, denoted by t_i, must not exceed the latest acceptable arrival time, denoted by b_i.

$$t_i \leq b_{P_i}, \ \forall i \in V \tag{6}$$

As with the capacity constraint, it is necessary to store the visiting times of all previous nodes in order to compute the arrival time at node i. In this case, when $y_i = 1$, the current time is reset, as each sub-tour is assumed to be performed by a separate vehicle. Therefore, the computation of the time window constraint follows a similar logic to that of the capacity constraint. The key difference lies in the use of a maximum: the arrival time at node i is defined as the maximum between the earliest acceptable time a_i and the cumulative travel time.

$$t_i = \max\left(a_{P_i}, T_{f,P_i} \times y_{i-1} + \left(\left(t_{i-1} + T_{P_{i-1},P_i}\right) \times (1 - y_{i-1})\right)\right) \tag{7}$$

This gives us the following model :

$$\min_{P,y} \sum_{i=1}^{n-1} \left[D_{P_i,P_{i+1}} \left(1 - y_i\right) + \left(D_{P_i,depot} + D_{depot,P_{i+1}} \right) y_i \right] + D_{depot,P_1} + D_{P_n,depot}$$

$$\text{s.t.} \quad P_i \neq P_j, \quad \forall (i,j) \tag{8a}$$

$$c_i \leq C^{max}, \quad \forall i \tag{8b}$$

$$t_i \leq b_{P_i}, \quad \forall i \tag{8c}$$

$$c_i = q_{P_i} + (c_{i-1} \times (1 - y_i)) \tag{8d}$$

$$t_i = \max \left(a_{P_i}, T_{f,P_i} \times y_i + \left(\left(t_{i-1} + T_{P_{i-1},P_i} \right) \times (1 - y_i) \right) \right) \tag{8e}$$

$$P_i \in [\![1, n]\!], \forall i [\![1, n]\!] \tag{8f}$$

2.2 Quantum Search

The quantum computing paradigm has led to the development of several noteworthy algorithms. Among them, Grover's search algorithm [6] is particularly relevant in the context of operations research, as it enables the identification of solutions within an unstructured search space. The only requirement is to construct an oracle function that returns 1 for all states satisfying the problem's constraints, and 0 otherwise. The algorithm amplifies the amplitudes—and consequently the probabilities—of the desirable states, starting from a uniform superposition over all possible configurations. Theoretically, Grover's algorithm achieves a success probability approaching 1 after $O(\sqrt{N/M})$ [3] iterations, where N is the size of the search space and M is the number of valid solutions. As a result, the algorithm's complexity is directly influenced by the number of decision variables used in the problem formulation.

To perform function minimization, we adopt the quantum search-based optimization algorithm introduced by Dürr and Høyer [5]. The principle of the algorithm is as follows:

1. Set an initial threshold value k and add a constraint to the Grover search that only considers solutions with a cost less than k.
2. If a valid solution is found, update k to this solution's cost.
3. Repeat the process until no better solution can be found.

We define the oracle function O as follows :

$$O(|P\rangle, |y\rangle, k) = \begin{cases} 1 & \text{if } (l(|P\rangle, |y\rangle) < k) \\ & \quad \wedge (P_i \neq P_j, \ \forall (i,j)) \\ & \quad \wedge (c_i \leq C^{max}, \ \forall i \in V) \\ & \quad \wedge (t_i \leq b_{P_i}, \ \forall i \in V) \\ 0 & \text{otherwise} \end{cases}$$

where l computes the objective function 2.

Let $\hat{O}_k$ denote the oracle operator for a given threshold k, and let $\hat{S}_{|\Psi\rangle}$ be the amplitude amplification operator introduced by Grover. Then, the Grover operator is defined as $\hat{G} = \hat{O}_k \hat{S}_{|\Psi\rangle}$. Applying $\hat{G}$ iteratively amplifies the amplitudes of all states that satisfy the constraints given in Equations (3), (4), and (6).

Algorithm 1 outlines the procedure used to implement and solve the model. We define $|AD\rangle$ as the register encoding the *All Different* constraint, $|\delta\rangle$ as the qubit for the distance constraint, and $|\kappa\rangle$ and $|\tau\rangle$ as the vectors encoding the capacity and time constraints, respectively. For instance, $|\kappa_i\rangle = 1$ if the accumulated load of the vehicle since the start of the subtour is valid: it equals 0 if P_i is the first node of a subtour, or $|c_{i-1}\rangle$ plus the demand at the ith node, provided this sum does not exceed the vehicle's capacity.

Algorithm 1: Grover Search CVRPTW Algorithm

Input: q the quantity vector, T the travelling time matrix, D the distance matrix, k the cost threshold
Output: P and y quantum register

$|P\rangle \leftarrow H^{\otimes n \log n}$
$|y\rangle \leftarrow H^{\otimes n}$
$|q\rangle \leftarrow |-\rangle$
$m \leftarrow 1$
while $m \leq \sqrt{2^{n \log n + n}}$ **do**
  for $i \leftarrow 1$ **to** n **do**
    for $j \leftarrow i + 1$ **to** n **do**
      $|AD\rangle \leftarrow P_i \neq P_j$
    $|c_i\rangle \leftarrow q_i$
    if $\neg\, |y_{i-1}\rangle$ **then**
      $|w\rangle \leftarrow |w\rangle + D_{P_{i-1},P_i}$
      $|c_i\rangle \leftarrow |c_{i-1}\rangle + |c_i\rangle$
      $|t_i\rangle \leftarrow \max(a_{P_i}, |t_{i-1}\rangle + T_{P_{i-1},P_i})$
    else
      $|w\rangle \leftarrow |w\rangle + D_{P_{i-1},P_{depot}} + D_{P_{depot},P_i}$
      $|t_i\rangle \leftarrow \max(a_{P_i}, T_{depot,P_i})$
    $|\kappa_i\rangle \leftarrow (|c_i\rangle \leq C^{max})$
    $|\tau_i\rangle \leftarrow (|t_i\rangle \leq b_{P_i})$
  $|\delta\rangle \leftarrow (|w\rangle \leq k)$
  $|q\rangle \leftarrow |AD\rangle \wedge \bigwedge_{i=0}^{n} |\kappa_i\rangle \wedge \bigwedge_{i=0}^{n} |\tau_i\rangle \wedge |\delta\rangle$
  Apply Grover operator to $|P\rangle \otimes |y\rangle$
  $m \leftarrow m + 1$

3 Implementation

In this section, we will discuss how to implement such algorithm on a gate based quantum computer. We assume that *multi-controlled X* gates also known as MCX gates are available.

P register There are various ways to represent a giant tour, which ultimately corresponds to a solution of the TSP. The implementation of constraints depends heavily on the chosen representation. For example, [12] introduces a successor-based representation that enables efficient cost evaluation. However, as noted in [14], it requires additional transformations to ensure that only valid solutions are marked.

In this work, we adopt a basic position-based representation, where $V \subset \mathbb{N}$ and $P_i \in V \ \forall i$. The uniqueness constraint can be easily enforced by ensuring that each register holds a distinct value.

3.1 Capacity and time constraints

Comparisons As described in the previous section, both the capacity and time window constraints ultimately reduce to comparisons, which can be efficiently implemented using a single ancilla qubit and components from the ripple-carry adder circuit introduced in [4].

Capacity constraint As discussed in Section 2.1, each value c_i representing the cumulative load at node P_i must be stored, which requires n dedicated quantum registers. To compute each c_i as defined in Equation 5, we must first encode the demand associated with node P_i. For this, we use a *conditional encoder*, a type of circuit that conditionally superposes multiple values onto a quantum register based on the value of one or more index registers. An implementation of such an encoder is described in paragraph 3 of the METHODS section in [14], it is denoted as F gate.

Depending on the value of the register $|P_i\rangle$, we use the encoder to assign the corresponding demand value to a load register. If $y_i = 0$, indicating that P_i is not the start of a new sub-tour, the previous accumulated load is added to this demand value. Conversely, if $y_i = 1$, the vehicle is assumed to have returned to the depot and restarted its route, so the cumulative load is reset. Figure 2 illustrates this implementation, which directly corresponds to lines 9, 12, and 17 of Algorithm 1.

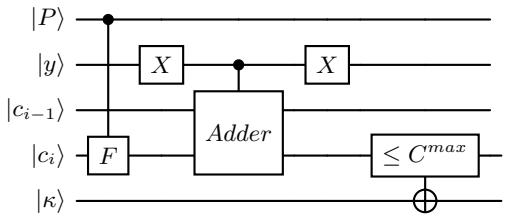


Fig. 2. Quantum circuit for capacity constraint

Time window constraint The time window constraint is structurally similar to the capacity constraint, with the key difference being the need to compute a max() function. This is required to determine whether the vehicle arrives within the allowed time window. The arrival time at node P_i is computed by comparing two values: the earliest allowed delivery time a_{P_i} and the current travel time. The maximum of these two values is selected as the effective arrival time. This comparison and assignment can be implemented using a basic comparator. Figure 3 depicts this implementation, which directly maps to lines 13, 16, and 18 of Algorithm 1. The $Enc()$ gate denotes the encoding of a value into a register.

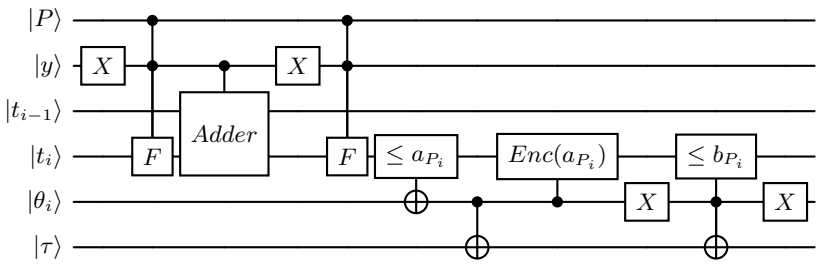


Fig. 3. Quantum circuit for time constraint

3.2 Complexity

Space complexity In this paper, we adopt a position-based encoding of the solution, where customer P_i is interpreted as the ith customer visited in the giant tour. This representation requires $n \log n$ qubits to encode the tour sequence. Additionally, we introduce n qubits to store the binary variables y_i, which represent the splitting decisions.

A significant number of qubits is also required to enforce the constraints. Specifically, $\mathcal{O}(n^2)$ qubits are needed to guarantee tour validity, since there are $n(n-1)/2$ pairwise inequality constraints to ensure that each customer appears exactly once.

For the capacity constraints, the cumulative load must be computed and stored at each step along the tour. This requires $n(\log d_{\max} + 1)$ qubits, where $d_{\max}$ denotes the maximum vehicle load capacity. In addition, n extra qubits are needed to store boolean results indicating whether the load at each step violates the capacity constraint. The total qubit requirement for capacity is therefore: $n(\log d_{\max} + 1) + n = \mathcal{O}(2n + n \log d_{\max})$.

Regarding the time window constraints, a similar number of registers is required. The arrival time at each node must be stored using n registers of size $\log t_{\max} + 1$, where $t_{\max}$ is the maximum allowed time. Since the effective arrival time is computed as the maximum between the accumulated travel time and

the customer's earliest delivery time, n additional boolean qubits are used to indicate which value is selected. Moreover, n further boolean qubits are needed to store whether the time window is respected. The total number of qubits for time-related constraints is thus: $n(\log t_{\max} + 1) + 2n = \mathcal{O}(3n + n \log t_{\max})$.

Finally, the cost evaluation requires $\log w_{max}$ where $w_{\max}$ represents an upper bound on the solution's cost.

Putting all components together, the total number of qubits required by the algorithm is: $\mathcal{O}(n^2 + n \log n + 6n + n \log d_{\max} + n \log t_{\max} + \log w_{max})$. However, the number of decision variables remains relatively small, namely $n \log n + n$, as only n qubits are added compared to the most compact exact quantum formulation for the TSP [12]. For instance, the example presented in 1.3 would require at least 147 qubits to solve.

To the best of our knowledge, no oracle-based formulations of the CVRPTW exist for comparison. A few QUBO-based formulations have been proposed: route-based formulations, where the number of routes grows exponentially and must therefore be generated using heuristics [8], and time-expanded formulations [13], where the decision variables represent every possible arc for each vehicle between nodes at different times. In the former case, the number of decision variables depends on the number of routes considered, while in the latter it is significantly larger than in our formulations, even before accounting for potential slack variables. In classical computing, the state of the art relies on route-based formulations, where routes generation is handled via column generation [9].

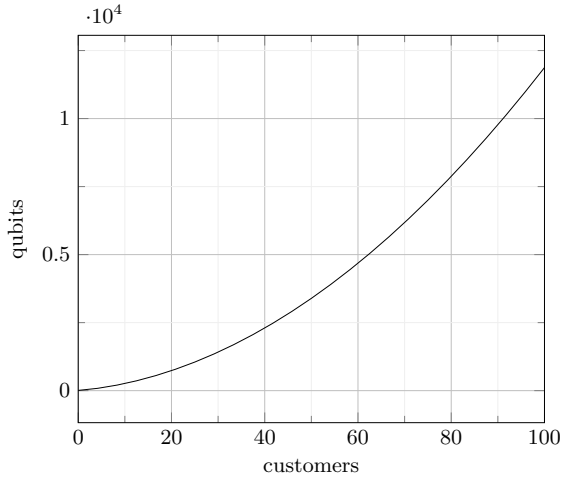


Fig. 4. Number of qubits relative to the number of customers for the oracle with the maximum capacity and number of time windows both fixed to 8.

Time complexity We use the number of MCX gates as a unit for time complexity analysis. For each gate set, we do not account for the inverse (uncomputation) circuits required to *free* ancilla qubits, as they mirror the forward operations and are already covered by the $\mathcal{O}$ notation. There are three categories of constraints contributing to the overall circuit depth.

First, the *all-different* constraint, ensuring uniqueness of the tour, can be implemented using $\mathcal{O}(n^2)$ MCX gates.

Second, the *capacity constraint* involves several components. The conditional encoder requires $\mathcal{O}(n \log d_{\max})$ MCX gates. We use Cuccaro et al.'s adder [4], which is efficient for this representation and contributes $\mathcal{O}(6 \log d_{\max})$ MCX gates. The comparison circuit can also be implemented using a modified version of the same adder. Since these operations are applied at each node, the total cost for capacity-related operations becomes: $\mathcal{O}\left(n^2 \log d_{\max} + 12n \log d_{\max}\right)$.

Third, the *time window constraint* has a similar structure to the capacity constraint, with the difference that both the encoder and the comparison are applied *twice* per node. This results in a total cost of: $\mathcal{O}\left(2n^2 \log d_{\max} + 24n \log d_{\max}\right)$.

Finally, for cost evaluation, we require n *conditional matrix encoders* and adders to compute the tour cost. These operations contribute a total of: $\mathcal{O}(n^3 \log n + 6n \log n)$. This component is relatively expensive and could potentially be optimized using a phase-based cost encoding via the Quantum Fourier Transform (QFT). However, such an approach would require translating our representation into a form compatible with QFT-based evaluation, which would in turn increase both qubit requirements and gate complexity [14].

In order to achieve a success probability close to 1, the Grover operator $\hat{G}$ must be applied $\mathcal{O}(\sqrt{N})$ times, where N is the size of the search space. In our case, $N = 2^{n \log n + n}$, which dominates the overall time complexity.

Conclusion

The proposed model allows one to tackle a complex routing problem, namely the CVRPTW. By adopting a split-inspired modeling approach on quantum computers, the model introduces only a linear additive overhead in the number of decision qubits compared to standard quantum TSP encodings while being optimal. Consequently, the time complexity is $\mathcal{O}(\sqrt{2^{n \log n + n}})$. Although large-scale implementations remain limited by current quantum hardware due to the gate complexity, this research lays the groundwork for future advancements in quantum logistics optimization.

References

1. (PDF) Applying the ANT System to the Vehicle Routing Problem. In: ResearchGate. https://doi.org/10.1007/978-1-4615-5775-3˙20
2. Beasley, JE.: Route first—Cluster second methods for vehicle routing. Omega **11**(4), 403–408 (Jan 1983). https://doi.org/10.1016/0305-0483(83)90033-6

3. Boyer, M., Brassard, G., Høyer, P., Tapp, A.: Tight Bounds on Quantum Searching. Fortschritte der Physik **46**(4-5), 493–505 (1998). https://doi.org/10.1002/(SICI)1521-3978(199806)46:4/5¡493::AID-PROP493¿3.0.CO;2-P
4. Cuccaro, S.A., Draper, T.G., Kutin, S.A., Moulton, D.P.: A new quantum ripple-carry addition circuit (Oct 2004). https://doi.org/10.48550/arXiv.quant-ph/0410184
5. Durr, C., Hoyer, P.: A Quantum Algorithm for Finding the Minimum (Jan 1999). https://doi.org/10.48550/arXiv.quant-ph/9607014
6. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. pp. 212–219. STOC '96, Association for Computing Machinery, New York, NY, USA (Jul 1996). https://doi.org/10.1145/237814.237866
7. Koskosidis, Y.A., Powell, W.B., Solomon, M.M.: An Optimization-Based Heuristic for Vehicle Routing and Scheduling with Soft Time Window Constraints. Transportation Science **26**(2), 69–85 (1992)
8. Lucas, A.: Ising formulations of many NP problems. Frontiers in Physics **2** (2014). https://doi.org/10.3389/fphy.2014.00005
9. Pessoa, A., Sadykov, R., Uchoa, E., Vanderbeck, F.: A generic exact solver for vehicle routing and related problems. Mathematical Programming **183**(1), 483–523 (Sep 2020). https://doi.org/10.1007/s10107-020-01523-z
10. Prins, C.: A simple and effective evolutionary algorithm for the vehicle routing problem. Computers & Operations Research **31**(12), 1985–2002 (Oct 2004). https://doi.org/10.1016/S0305-0548(03)00158-8
11. Shor, P.W.: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM Journal on Computing **26**(5), 1484–1509 (Oct 1997). https://doi.org/10.1137/S0097539795293172
12. Srinivasan, K., Satyajit, S., Behera, B.K., Panigrahi, P.K.: Efficient quantum algorithm for solving travelling salesman problem: An IBM quantum experience (May 2018). https://doi.org/10.48550/arXiv.1805.10928
13. Vargas, A., Shukla, P., Allmendinger, R., Jaeger, A.: On Solving the Capacitated Vehicle Routing Problem with Time Windows using Quantum Annealing. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 1979–1983. GECCO '24 Companion, Association for Computing Machinery, New York, NY, USA (Aug 2024). https://doi.org/10.1145/3638530.3664139
14. Zhu, J., Gao, Y., Wang, H., Li, T., Wu, H.: A Realizable GAS-based Quantum Algorithm for Traveling Salesman Problem (Dec 2022). https://doi.org/10.48550/arXiv.2212.02735

IQNN-CS: Interpretable Quantum Neural Network for Credit Scoring

Abdul Samad Khan[1], Nouhaila Innan[2,3], Aeysha Khalique[1], and Muhammad Shafique[2,3]

[1] Lahore University of Management Sciences, Pakistan
[2] eBRAIN Lab, Division of Engineering, New York University Abu Dhabi (NYUAD), Abu Dhabi, UAE
[3] Center for Quantum and Topological Systems (CQTS), NYUAD Research Institute, NYUAD, Abu Dhabi, UAE
24120006@lums.edu.pk, nouhaila.innan@nyu.edu, aeysha.khalique@lums.edu.pk, muhammad.shafique@nyu.edu

Abstract. Credit scoring is a high-stakes task in financial services, where model decisions directly impact individuals' access to credit and are subject to strict regulatory scrutiny. While Quantum Machine Learning (QML) offers new computational capabilities, its black-box nature poses challenges for adoption in domains that demand transparency and trust. In this work, we present IQNN-CS, an interpretable quantum neural network framework designed for multiclass credit risk classification. The architecture combines a variational QNN with a suite of post-hoc explanation techniques tailored for structured data. To address the lack of structured interpretability in QML, we introduce Inter-Class Attribution Alignment (ICAA), a novel metric that quantifies attribution divergence across predicted classes, revealing how the model distinguishes between credit risk categories. Evaluated on two real-world credit datasets, IQNN-CS demonstrates stable training dynamics, competitive predictive performance, and enhanced interpretability. Our results highlight a practical path toward transparent and accountable QML models for financial decision-making.

Keywords: Quantum Machine Learning · Credit Scoring · Interpretability · Quantum Finance

1 Introduction

Credit scoring is a foundational task in financial services, directly influencing decisions related to loan approvals, credit limits, and interest rates [1]. Its impact is not only financial but also societal, affecting individuals' access to economic opportunities. As such, credit scoring systems must meet two core requirements: high predictive performance and strict compliance with fairness and transparency regulations, such as the European Union's General Data Protection Regulation (GDPR) and the Basel III banking regulations. In this context,

interpretability becomes essential, not as an optional enhancement, but as a mandatory feature for deployment in regulated environments.

Quantum Neural Networks (QNNs) have recently emerged as promising models for structured learning tasks, owing to their potential to encode and process high-dimensional data using fewer resources [2, 3]. Their expressive power enables them to represent complex decision boundaries for domains like finance, where patterns in structured tabular data can be subtle and interdependent.

However, existing QNN-based models often prioritize performance over interpretability, which is crucial for sensitive decision-making contexts such as credit scoring. Moreover, the development of interpretable QNNs remains in its infancy. Unlike classical models, where attribution methods and explanation techniques are well-studied, the quantum learning community lacks standardized tools to reason about QNN decisions, especially in multiclass classification settings. In this work, we ask a central question:

> *How can QNNs be designed for real-world credit scoring applications, where achieving high accuracy is not sufficient, and interpretability is a core requirement?*

To address this, we propose **IQNN-CS**, an interpretable architecture of QNNs specifically tailored for multiclass credit scoring. Our goal is not to demonstrate quantum supremacy or to outperform classical machine learning models, but rather to explore how QNNs can be made suitable for high-stakes financial applications through interpretability-aware design.

The main contributions of this paper are as follows:

- We propose a hybrid classical–quantum pipeline where a variational QNN performs classification on structured financial data, with interpretability achieved using adapted classical and quantum techniques.
- We introduce Inter-Class Attribution Alignment (ICAA), a new metric that quantifies attribution divergence across predicted classes, enabling structured analysis of model reasoning in multiclass tasks.
- We evaluate IQNN-CS on two real-world credit datasets, focusing on robustness, attribution behavior, and interpretability, while emphasizing deployment feasibility in regulated financial settings.

2 Background and Related Work

2.1 QML for Credit Scoring

Quantum Machine Learning (QML) has been increasingly applied in the financial sector [4], covering applications such as loan eligibility prediction [5], credit default classification [6], market prediction [7, 8], fraud detection [9–12], and other financial applications [13, 14]. Some of these tasks are naturally formulated as classification problems, while others are formulated as combinatorial optimization problems. Within this broad spectrum, credit scoring has received limited but growing attention in QML research.

When it comes to credit scoring, only a few works have directly addressed the problem using quantum models. The systemic quantum score model introduced a quantum kernel-based approach aimed at improving generalization in low-data and imbalanced settings, demonstrating performance benefits in a production-grade financial scenario [15]. A separate study explored a hybrid quantum–classical model for credit scoring among small- and medium-sized enterprises, highlighting reductions in training time by embedding a quantum layer into a classical neural network [16]. Both approaches treated credit scoring as a binary classification task and focused on performance rather than transparency or explainability.

Other research efforts have framed credit scoring as an optimization problem. One study utilized a QUBO formulation to optimize scoring card thresholds and combinations, employing both quantum annealing and classical heuristics to enhance bank profitability under credit risk constraints [17]. A related work investigated cost-effective feature selection for mobile credit scoring using a quantum-inspired evolutionary algorithm based on quantum gates and representations, achieving reduced feature costs and competitive accuracy [18].

While these studies highlight the potential of quantum and quantum-inspired models for credit scoring, they either treat the task as binary classification or focus on optimization formulations. Crucially, they overlook the interpretability requirements critical for real-world deployment in regulated financial environments. In contrast, we address credit scoring as a multiclass classification problem and focus on the interpretability of quantum models, an aspect that has not been covered in prior work.

2.2 Interpretability in QML

Interpretability in QML is increasingly important for deploying models in sensitive domains such as finance. However, most classical explainability methods do not directly extend to quantum models due to their probabilistic nature and circuit-based representations.

Initial efforts, such as Q-LIME, adapt local explanation techniques to quantum settings by approximating the influence of features on predictions [19]. Other works combine quantum autoencoders with classifiers and apply LIME and SHAP for interpretability in quantum representation learning [20].

To address global explainability, frameworks like QuXAI use Q-MEDLEY to trace feature attributions through quantum feature maps [21]. Visual tools such as QGrad-CAM further enable fine-grained class-specific explanations using variational quantum circuits [22]. Despite these advances, interpretability for structured, multiclass QML tasks, such as credit scoring, remains largely unexplored. Our work addresses this gap through a domain-specific, interpretable QNN framework.

3 Methodology

The IQNN-CS framework is structured as a five-stage pipeline designed to deliver interpretable and accurate credit scoring using a hybrid quantum-classical architecture (see Fig. 1 and Appendix Algorithm 1).

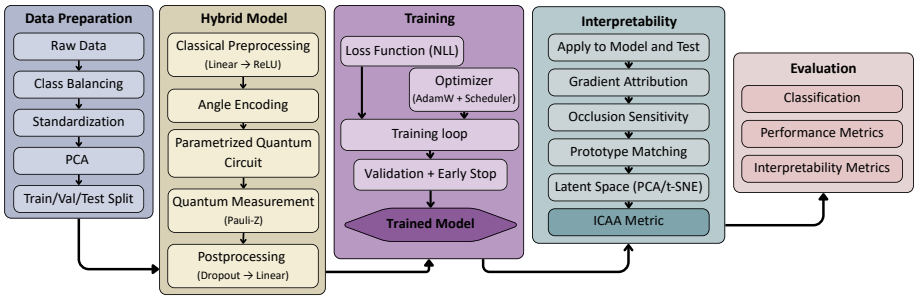


Fig. 1. Overview of the IQNN-CS pipeline.

3.1 Data Preparation

The process begins with preprocessing two benchmark credit datasets containing numerical and categorical financial attributes. To address class imbalance, we apply undersampling to Dataset 1 and SMOTE to Dataset 2, depending on the severity of imbalance. All numerical features are standardized by transforming each feature x_j to $x'_j = (x_j - \mu_j)/\sigma_j$, where μ_j and σ_j denote the empirical mean and standard deviation. To reduce dimensionality and ensure compatibility with quantum hardware constraints, Principal Component Analysis (PCA) is applied, resulting in quantum-compatible input vectors. PCA projects the data $\mathbf{x}$ onto a lower-dimensional subspace using orthogonal components $\mathbf{w}_k$ as $z_k = \mathbf{w}_k^\top \mathbf{x}$, for $k = 1, \ldots, d$. The resulting representation aligns with the available number of qubits. Stratified sampling is used to divide the dataset into training, validation, and test sets, preserving class distribution.

3.2 Hybrid Quantum-Classical Model

The model architecture consists of three sequential components. First, a classical preprocessing block transforms the PCA-compressed input $\mathbf{z} \in \mathbb{R}^d$ into an encoded latent vector $\mathbf{h}_{cl1} \in \mathbb{R}^{d'}$ through fully connected layers with ReLU activations. This output $\mathbf{h}_{cl1}$ is then passed to the quantum layer, where the features are encoded onto N_Q qubits, typically initialized to the $|0\rangle^{\otimes N_Q}$ state, using angle encoding, where each input feature ϕ_j (from $\mathbf{h}_{cl1}$) rotates a qubit via an operation like $R_P(\phi_j) = e^{-i\phi_j P/2}$, with P being a Pauli operator (e.g., $\sigma_x, \sigma_y, \sigma_z$). This is followed by a multi-layer entangling variational quantum circuit (VQC), $U(\boldsymbol{\theta})$, parameterized by trainable angles $\boldsymbol{\theta}$. Expectation values of Pauli-Z observables, $\langle \sigma_z^{(k)} \rangle = \langle \psi_0 | U^\dagger(\boldsymbol{\theta}) \sigma_z^{(k)} U(\boldsymbol{\theta}) | \psi_0 \rangle$ for each qubit k, are measured to extract quantum features. These features are concatenated and forwarded to the classical postprocessing head. This final block, which includes dropout regularization and linear projections, produces logits $\mathbf{o} \in \mathbb{R}^C$ over the C credit risk classes. The full system is end-to-end differentiable and implemented via Pen-

nyLane's `TorchLayer`, enabling seamless integration with PyTorch's automatic differentiation capabilities.

3.3 Training Procedure

The model is trained using the negative log-likelihood (NLL) loss function, given for a batch of M samples as $L_{NLL} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$, where $y_{i,c}$ is the ground-truth indicator for class c and $\hat{y}_{i,c}$ is the predicted probability from the softmax output. Optimization is performed with AdamW, which introduces decoupled weight decay. Learning rate scheduling, such as cosine annealing, is employed to stabilize convergence. Quantum gradients with respect to variational parameters $\boldsymbol{\theta}$ are computed using the parameter-shift rule: $\frac{\partial\langle\hat{O}\rangle}{\partial\theta_k} = \frac{1}{2s}\left(\langle\hat{O}\rangle_{\theta_k+s} - \langle\hat{O}\rangle_{\theta_k-s}\right)$, with $s = \pi/2$ for single-qubit gates. Early stopping is applied based on validation loss to prevent overfitting, and the best-performing model checkpoint is retained.

3.4 Post-Hoc Interpretability

Interpretability is assessed using four complementary techniques (see Appendix Algorithm 2). First, gradient-based attribution methods, including saliency maps, gradient×input, integrated gradients, and SmoothGrad, estimate the importance of each input feature x_j through $\partial L/\partial x_j$ or related formulations. Second, prototype matching is performed in the quantum feature space by extracting activation vectors $\mathbf{a}_{QNN}(x)$ and computing cosine similarity with training instances: $S_C(\mathbf{u},\mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$. This allows retrieval of similar historical examples for a given test input. Fourth, we introduce the ICAA metric. For each class c, let $\mathbf{A}_c(x_{inst})$ be its attribution vector. The ICAA matrix is defined by: $ICAA_{ij}(x_{inst}) = S_C(\mathbf{A}_i(x_{inst}), \mathbf{A}_j(x_{inst}))$, capturing the pairwise similarity in feature attributions across classes. Additionally, we assess the region of indecision by perturbing inputs and measuring the variance in attribution vectors.

3.5 Evaluation Metrics

The final model performance is evaluated using accuracy and macro F1-score. Accuracy is computed as $\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN}$, where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. while the macro F1-score aggregates class-wise F1-scores: $F1_c = \frac{2P_cR_c}{P_c+R_c}$, $\quad$ Macro F1 $= \frac{1}{C}\sum_{c=1}^{C} F1_c$, where P_c and R_c are precision and recall for class c. Confusion matrices offer class-level diagnostic insights. Finally, we visualize interpretability outputs, gradient maps, and matched prototypes, along with latent space projections of QNN activations using t-SNE. These projections are color-coded by true labels to assess the separability of quantum features.

4 Results

4.1 Experimental Settings

Experiments were conducted on two benchmark credit scoring datasets: Dataset 1 [23] and Dataset 2 [24]. All models were implemented using PyTorch and

PennyLane [25], with quantum circuits simulated via `default.qubit`. Training was executed on a CPU (Apple M3, 16 GB RAM) without GPU acceleration. Each dataset was split into training (70%), validation (15%), and test (15%) sets using stratified sampling. The hybrid model combined six qubits and four `StronglyEntanglingLayers` with classical dense layers and dropout. Optimization was performed over 50 epochs using the AdamW optimizer (learning rate 0.01) and class-weighted cross-entropy loss. Full configuration details are provided in Appendix Table 2.

4.2 Training and Convergence

Fig. 2 shows the training, validation, and test curves for both datasets. For Dataset 1, accuracy curves in (a) exhibit rapid improvement within the first few epochs, stabilizing above 98% across all splits. The corresponding loss curves in (b) show a sharp decrease followed by a smooth plateau, with minimal discrepancy between training and validation, indicating efficient convergence and strong generalization. In contrast, Dataset 2 exhibits more unstable training behavior.

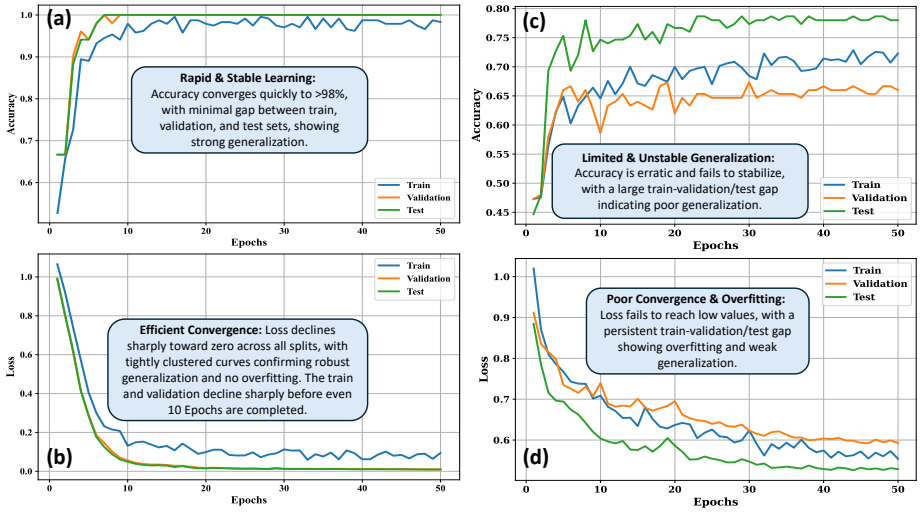


Fig. 2. Accuracy and loss curves for IQNN-CS. (a) accuracy and (b) loss for Dataset 1; (c) accuracy and (d) loss for Dataset 2.

Accuracy trends in (c) plateau early, with validation and test performance remaining below 80%. The loss curves in (d) reveal a slower and noisier decline, particularly in the validation set. The wider gap between training and validation suggests either overfitting or greater dataset complexity due to feature diversity and class imbalance.

4.3 Classification Performance

Table 1 summarizes the classification results for both datasets. On Dataset 1, the model achieved perfect performance, with 100% accuracy and F1-score. This

confirms the model's ability to separate risk levels cleanly in well-structured data. For Dataset 2, overall accuracy was 77.3%, with variable class-wise metrics. The model showed high recall for the Low class (0.97) but low precision (0.64), suggesting misclassifications. The High class had strong precision (0.95) but lower recall (0.67), reflecting under-identification of high-risk cases. These results highlight Dataset 2's complexity and imbalance.

Table 1. Classification performance on both Datasets.

Class	Dataset 1			Dataset 2		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Low	1.00	1.00	1.00	0.64	0.97	0.77
Average	1.00	1.00	1.00	0.73	0.84	0.78
High	1.00	1.00	1.00	0.95	0.67	0.79
Accuracy (%)		**100**			**77.3**	

4.4 Interpretability Analysis

To assess the transparency and trustworthiness of the IQNN-CS model, we conduct a comprehensive interpretability analysis. This includes examining latent quantum geometry, attribution consistency, example influence, class attribution alignment, and softmax entropy behavior. We compare findings across both datasets to understand how dataset quality impacts interpretability.

Feature Attribution Analysis: As illustrated in Fig. 3, saliency maps for a representative test instance show focused and interpretable patterns in Dataset 1. In contrast, Dataset 2 exhibits diffuse, noisy attributions that vary across runs, suggesting that the model relies less consistently on meaningful input features.

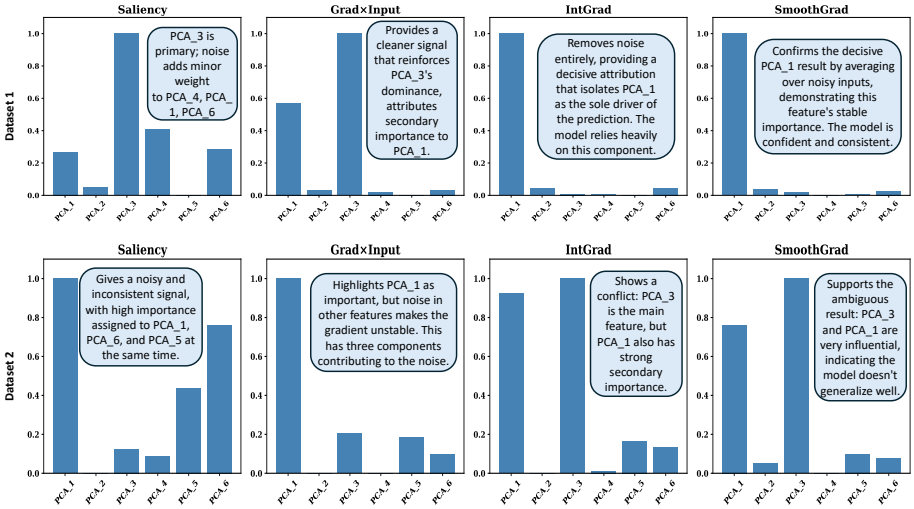


Fig. 3. Saliency maps for a selected test sample. Dataset 1 shows concentrated attributions; Dataset 2 appears more diffuse.

Quantum Representation Geometry: Latent embeddings extracted from the quantum layer are visualized using t-SNE as shown in Fig. 4. Dataset 1 shows well-separated manifolds, indicating robust internal representations. In contrast, Dataset 2 presents entangled clusters, particularly between the Average and High classes, reflecting the model's confusion in classification.

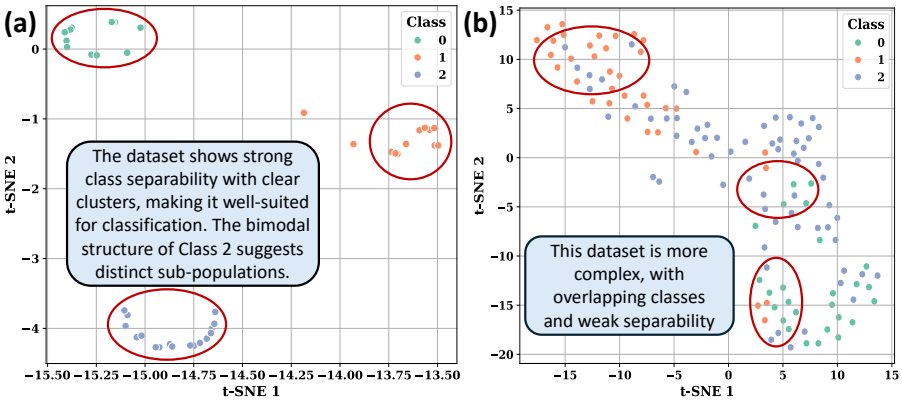


Fig. 4. t-SNE projection of quantum embeddings. (a) Dataset 1 yields a clear separation; (b) Dataset 2 shows an overlap.

Attribution Sensitivity: To assess how strongly predictions depend on top-ranked features, we occlude inputs and track confidence degradation. As shown in Fig. 5, Dataset 1 experiences sharp confidence drops, indicating reliance on a few informative features. In Dataset 2, degradation is gradual and less structured.

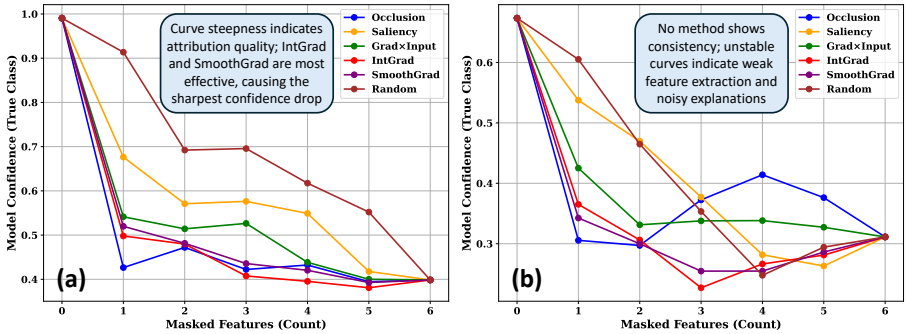


Fig. 5. Prediction confidence vs. occluded features. Left: Dataset 1 shows sharp drops; Right: Dataset 2 decays smoothly.

ICAA: We assess whether the model produces disentangled explanations for each class using ICAA. As shown in Fig. 6, Dataset 1 yields low inter-class attribution similarity, while Dataset 2 exhibits significant overlap, revealing a less distinct decision rationale.

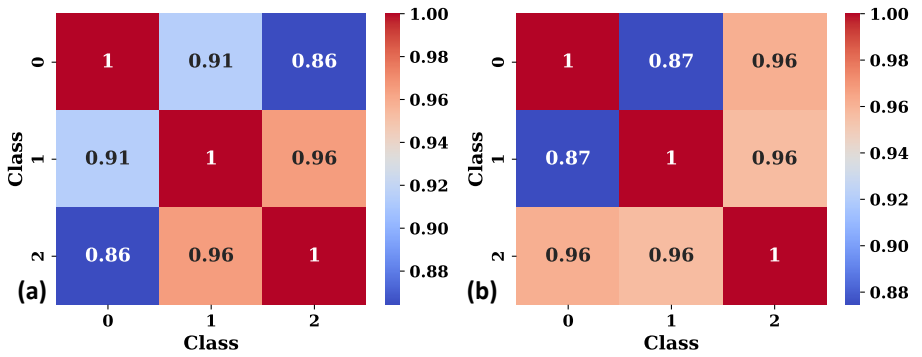


Fig. 6. ICAA matrices. Dataset 1 (a) shows clean class-wise separation; Dataset 2 (b) reveals overlapping attribution logic.

4.5 Discussion

Our experiments highlight a stark contrast in how the IQNN-CS model performs across two credit scoring datasets. While Dataset 2 proved more challenging, despite using the same model, its lower accuracy and less structured latent space suggest that the success of QNNs strongly depends on how well the data aligns with the model's inductive bias. Dataset 2 presented complex and overlapping feature spaces that the model struggled to disentangle. Interpretability tools helped uncover these differences. Occlusion and our proposed ICAA metric revealed that, in Dataset 2, attribution patterns overlapped, indicating that the model relied on similar features across classes, a potential source of misclassification. Gradient-based explanations were inconsistent, especially in noisy predictions, suggesting they should be used cautiously in quantum settings. Additionally, example-based attribution showed that Dataset 2 often aligned with incorrect classes, despite high-confidence predictions. This shows that a model's internal logic can break down even when output probabilities appear reliable.

In summary, these findings underscore that interpretability is not simply about explanation; it is a crucial diagnostic tool. Metrics like ICAA provide deeper insight into how well a model separates reasoning across classes, which is particularly important in high-stakes domains like finance.

5 Conclusion

We introduced IQNN-CS, a hybrid quantum-classical model for credit scoring that combines strong predictive performance with interpretable outputs. Across two real-world datasets, the model demonstrated high accuracy when the data structure aligned with its quantum encoding and revealed clear decision patterns through interpretability analysis. A key contribution of this work is the ICAA metric, which quantifies how distinctly the model reasons about different classes. It proved useful for identifying when the model's internal logic breaks down, even when standard metrics suggest good performance. Together with attribution sensitivity and example-based influence tracing, ICAA strengthens the tools available for understanding quantum models. IQNN-CS shows that accurate quantum models can also be transparent and trustworthy. As quantum

computing advances, such interpretability-first frameworks will be essential for deploying reliable AI in sensitive fields like finance and healthcare.

Acknowledgments

This work was supported in part by the NYUAD Center for Quantum and Topological Systems (CQTS), funded by Tamkeen under the NYUAD Research Institute grant CG008, and the Center for Cyber Security (CCS), funded by Tamkeen under the NYUAD Research Institute Award G1104.

References

1. L. Thomas *et al.*, *Credit scoring and its applications*. SIAM, 2017.
2. A. Abbas *et al.*, "The power of quantum neural networks," *Nature Computational Science*, 2021.
3. N. Innan *et al.*, "Next-generation quantum neural networks: Enhancing efficiency, security, and privacy," in *2025 IEEE 31st International Symposium on On-Line Testing and Robust System Design (IOLTS)*, pp. 1–4, IEEE, 2025.
4. J. Biamonte *et al.*, "Quantum machine learning," *Nature*, 2017.
5. N. Innan *et al.*, "Lep-qnn: Loan eligibility prediction using quantum neural networks," *arXiv preprint arXiv:2412.03158*, 2024.
6. A. Sengar *et al.*, "A comparative analysis of hybrid quantum neural networks in binary credit defaulting tasks," in *2023 IEEE MIT Undergraduate Research Technology Conference (URTC)*, IEEE, 2023.
7. P. K. Choudhary *et al.*, "Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction," *arXiv preprint arXiv:2503.15403*, 2025.
8. P. Pathak, V. Oad, A. Prajapati, and N. Innan, "Resource allocation optimization in 5g networks using variational quantum regressor," in *2024 International Conference on QCNC*, pp. 101–105, IEEE, 2024.
9. N. Innan *et al.*, "Financial fraud detection: a comparative study of quantum machine learning models," *International Journal of Quantum Information*, 2024.
10. N. Innan *et al.*, "Qfnn-ffd: Quantum federated neural network for financial fraud detection," in *2025 IEEE International Conference on QSW*, pp. 41–47, IEEE, 2025.
11. N. Innan *et al.*, "Financial fraud detection using quantum graph neural networks," *Quantum Machine Intelligence*, vol. 6, no. 1, p. 7, 2024.
12. M. E. Alami *et al.*, "Comparative performance analysis of quantum machine learning architectures for credit card fraud detection," *arXiv preprint arXiv:2412.19441*, 2024.
13. M. Kashif *et al.*, "Evaluating quantum amplitude estimation for pricing multi-asset basket options," *arXiv preprint arXiv:2509.09432*, 2025.
14. N. Innan *et al.*, "Quantum portfolio optimization with expert analysis evaluation," *arXiv preprint arXiv:2507.20532*, 2025.
15. J. Mancilla *et al.*, "Empowering credit scoring systems with quantum-enhanced machine learning," *arXiv preprint arXiv:2404.00015*, 2024.
16. N. Schetakis *et al.*, "Quantum machine learning for credit scoring," *Mathematics*, vol. 12, no. 9, p. 1391, 2024.
17. B. Zhang *et al.*, "Credit scoring card combination optimization model based on qubo," in *Proceedings of the International Conference on Machine Learning, Pattern Recognition and Automation Engineering*, 2024.
18. C. M. Chen *et al.*, "Quantum optimized cost based feature selection and credit scoring for mobile micro-financing," *Computational Economics*, 2024.
19. L. Pira and C. Ferrie, "On the interpretability of quantum neural networks," *Quantum Machine Intelligence*, vol. 6, no. 2, p. 52, 2024.
20. A. Kottahachchi K.D and I. Khalil, "Qrlaxai: quantum representation learning and explainable ai," *Quantum Machine Intelligence*, 2025.
21. S. Barua *et al.*, "Quxai: Explainers for hybrid quantum machine learning models," *arXiv preprint arXiv:2505.10167*, 2025.
22. H.-Y. Lin *et al.*, "Quantum gradient class activation map for model interpretability," in *2024 IEEE Workshop on SiPS*, pp. 165–170, IEEE, 2024.
23. "Credit score classification dataset." https://www.kaggle.com/datasets/sujithmandala/credit-score-classification-dataset.
24. "Credit score classification dataset." https://www.kaggle.com/datasets/parisrohan/credit-score-classification.
25. V. Bergholm *et al.*, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," *arXiv preprint arXiv:1811.04968*, 2018.

Appendix

A1 IQNN-CS Training Procedure

The IQNN-CS architecture is trained using a hybrid quantum-classical pipeline as detailed in Algorithm 1.

Algorithm 1 IQNN-CS Training Procedure

Require: Raw Datasets (D_{raw1}, D_{raw2}), NumEpochs (N_{epochs}), BatchSize (B), LearningRate (η), PCA Dimensions (d_{PCA1}, d_{PCA2}), Qubit Counts (N_{Q1}, N_{Q2})
Ensure: Trained IQNN-CS model ($M_{IQNN-CS}$)
Initialize the IQNN-CS model structure ($M_{IQNN-CS}$):
  Define classical pre-processing network $M_{classical_pre}$ using d_{PCA} input dimensions.
  Define Quantum Neural Network M_{QNN} with N_Q qubits, specified layers, and encoding method.
  Define classical post-processing network $M_{classical_post}$ for the specified number of output classes.
Assemble $M_{IQNN-CS}$ by sequentially connecting $M_{classical_pre} \rightarrow M_{QNN} \rightarrow M_{classical_post}$.
for all dataset D_{raw} in $\{D_{raw1}, D_{raw2}\}$ **do**
  ▷ Configure dataset-specific parameters: d_{PCA}, N_Q
  Perform preprocessing on D_{raw} to obtain D_{proc}:
    Apply class balancing technique (e.g., Undersampling, SMOTE).
    Apply feature standardization (e.g., StandardScaler).
    Apply PCA to reduce dimensionality to d_{PCA}.
  Split D_{proc} into training D_{train}, validation D_{val}, and test D_{test} sets.
  Initialize optimizer (e.g., AdamW) with $M_{IQNN-CS}$ parameters and learning rate η.
  Initialize learning rate scheduler (e.g., StepLR or CosineAnnealingLR).
  for $epoch = 1 \rightarrow N_{epochs}$ **do**
    Set $M_{IQNN-CS}$ to training mode.
    for all batch (X_b, y_b) in D_{train} **do**
      Clear gradients in the optimizer.
      Obtain predictions $\hat{y}_b$ via a forward pass of X_b through $M_{IQNN-CS}$.
      Calculate loss (e.g., Negative Log-Likelihood) between $\hat{y}_b$ and y_b.
      Compute gradients via hybrid backpropagation through $M_{IQNN-CS}$.
      Update model parameters using the optimizer.
    end for
    Adjust learning rate using the LR scheduler.
    Set $M_{IQNN-CS}$ to evaluation mode.
    Calculate validation loss $loss_{val}$ on D_{val} using $M_{IQNN-CS}$.
    if validation loss $loss_{val}$ meets early stopping criteria **then**
      break ▷ Cease training if validation performance degrades or stagnates
    end if
  end for
end for
return Trained $M_{IQNN-CS}$

A2 Interpretability Pipeline

The interpretability module, summarized in Algorithm 2, combines gradient-based saliency maps, occlusion analysis, example-based attribution, ICAA, and latent space visualization. These methods are executed on selected test instances, allowing insight into feature influence and consistency across representations.

A3 Experimental Settings

The experimental configuration used across datasets is outlined in Table 2. This setup ensures consistency for reproducibility and fair interpretability comparisons.

A4 Extended Interpretability Analysis

Attribution Stability and Regions of Indecision: To assess robustness, we introduced Gaussian noise and computed the standard deviation of saliency

maps. Most samples showed stable attribution under perturbation. However, Sample 12 in Dataset 2 showed significantly higher variance (Table 3), indicating unreliable class evidence.

Algorithm 2 Post-hoc Interpretability Analysis

Require: Trained IQNN-CS model ($M_{IQNN-CS}$), Preprocessed Test Data (D_{test}), Set of test instances ($X_{interpret} \subseteq D_{test}$)
Ensure: Interpretability outputs (Attribution Maps, ICAA Matrix, Plots, etc.)
Set $M_{IQNN-CS}$ to evaluation mode.
for all test instance x_{inst} in $X_{interpret}$ **do**
▷ **1. Gradient-Based Attribution**
For each target class c:
Compute gradient-based attribution $A_c(x_{inst})$ for x_{inst} towards class c using $M_{IQNN-CS}$.
Store or visualize $A_c(x_{inst})$ (e.g., as feature heatmaps).
▷ **2. Occlusion Analysis**
Perform occlusion analysis on x_{inst} with $M_{IQNN-CS}$ to obtain prediction probability drop curve P_{drop_curve}.
▷ This involves iteratively masking features of x_{inst} and recording the drop in prediction probability.
Store or visualize P_{drop_curve}.
▷ **3. Example-Based Attribution (using QNN Activations)**
if QNN activations for training data are not precomputed **then**
Extract QNN activations Act_{train} from M_{QNN} for all instances in D_{train}.
end if
Extract QNN activation act_{inst} from M_{QNN} for instance x_{inst}.
Calculate cosine similarity scores Sim_{scores} between act_{inst} and all activations in Act_{train}.
Identify influential training examples based on Sim_{scores}.
▷ **4. Inter-Class Attribution Alignment (ICAA)**
Let $A_0(x_{inst}), A_1(x_{inst}), \ldots$ be the attribution vectors for x_{inst} for each class (obtained from gradient-based attribution).
Calculate $ICAA_{ij} \leftarrow \text{CosineSimilarity}(A_i(x_{inst}), A_j(x_{inst}))$ for all pairs of classes (i, j).
Store the resulting $ICAA_{matrix}$ for x_{inst}.
end for
▷ **5. Latent Space Geometry Visualization (on a subset of D_{test})**
Select a subset X_{subset} from D_{test}.
Extract QNN activations Act_{QNN_subset} from M_{QNN} for all instances in X_{subset}.
Apply a dimensionality reduction technique (e.g., t-SNE or PCA) to Act_{QNN_subset}.
Plot the reduced-dimension embeddings, coloring them by true class labels.
return All generated interpretability maps, matrices, plots, and insights.

Table 2. Experimental Configuration for IQNN-CS

Component	**Setting**
Quantum Framework	PennyLane (default.qubit)
Number of Qubits	6
Quantum Layers	4 (StronglyEntanglingLayers)
Embedding Strategy	AngleEmbedding
Classical Layers	Linear → ReLU → Linear + Dropout
Preprocessing	StandardScaler, OneHotEncoder, SMOTE
Dimensionality Reduction	PCA (6 components)
Batch Size	16
Optimizer	AdamW
Learning Rate	0.01
Scheduler	StepLR
Loss Function	CrossEntropy (class-weighted)
Epochs	50
Validation Strategy	70-15-15 train-val-test split (stratified)
Random Seed	42 (NumPy + PyTorch)

Table 3. Standard deviation of saliency maps under 20 random Gaussian perturbations for selected test samples.

Sample	Dataset 1 (std)	Dataset 2 (std)	Indecisive?
3	0.0493	0.0651	No
7	0.1570	0.0395	No
12	0.1426	**0.2797**	Yes

Influence of Training Examples on Test Samples: We analyze the influence of training samples on test instance 7 using test-to-train cosine similarity. As shown in Table 4, in Dataset 1, the most influential samples align with the test sample's true class, indicating coherent internal representations. In contrast, Dataset 2 shows high-similarity points from multiple classes, suggesting ambiguity in the learned decision boundary.

Table 4. Cosine similarity between test sample 7 and top influential training examples in both datasets. Dataset 1 shows high similarity to same-class examples, while Dataset 2 yields a mix of same and different-class influences, suggesting less coherent internal representations.

Dataset	Test Sample	Train Index	Label	Cosine Sim.	Observation
1	7	146	2	0.9989	Same class as test
1	7	181	2	0.9976	Same class as test
1	7	160	2	0.9976	Same class as test
1	7	22	2	0.9930	Same class as test
1	7	94	2	0.9416	Same class as test
2	7	503	1	1.0000	Different class
2	7	19	1	1.0000	Different class
2	7	433	1	1.0000	Different class
2	7	588	1	1.0000	Different class
2	7	173	2	1.0000	Same class as test

Prediction Confidence and Attribution Consistency Fig. 7 shows that Dataset 1 had sharper softmax distributions, indicating high-confidence predictions. Dataset 2 yielded broader distributions, consistent with noisy convergence and less reliable model explanations. While the attribution similarity matrix (Fig. 8) reveals well-separated patterns in Dataset 1, while Dataset 2 lacks structure, underscoring inconsistency in explanation reliability.

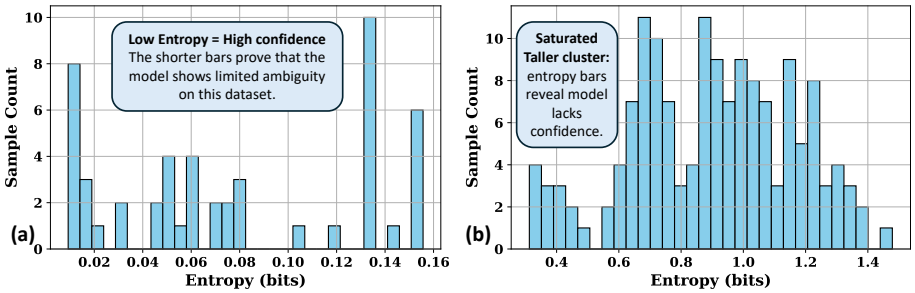


Fig. 7. Softmax entropy distributions. Dataset 1 (a) exhibits lower uncertainty compared to Dataset 2 (b).

Method Evaluation Summary: Table 5 synthesizes the comparative effectiveness of various interpretability techniques across both datasets. The evaluation is based on how well each method provides meaningful, class-consistent, and stable explanations aligned with model behavior.

In Dataset 1, most interpretability methods yielded coherent and high-confidence outputs. Occlusion and ICAA emerged as the most reliable techniques, offering sharply localized and class-discriminative explanations. Gradient-based methods such as saliency maps and integrated gradients also performed well, whereas SmoothGrad shows limited value due to noise sensitivity. In Dataset 2, the utility

of nearly all methods degraded, consistent with earlier findings of unstable training dynamics and less confident predictions. ICAA and occlusion retained partial utility but suffered from reduced clarity. Example-based attributions revealed inconsistencies in training-test relationships, and attribution stability flagged specific instances of model indecision. Overall, the summary table highlights that interpretability reliability is closely tied to training convergence quality and dataset structure. Robust interpretability requires not only method design but also stable and semantically meaningful latent representations from the model itself.

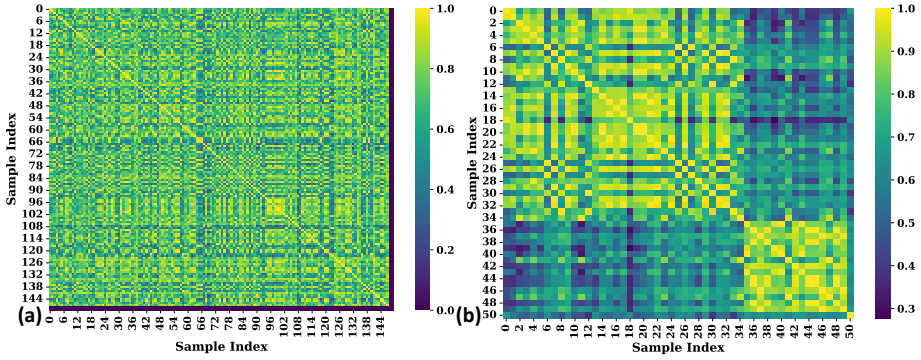


Fig. 8. Attribution similarity matrices across test samples for both Datasets: Dataset 1 (a) and Dataset 2 (b).

Table 5. Interpretability method utility across both datasets.

Method	Dataset 1 Utility	Dataset 2 Utility
Saliency	High	Medium
Gradient × Input	Moderate	Low
Integrated Gradients	High	Medium
SmoothGrad	Low	Low
Occlusion	Very High	Medium
ICAA	Very High	Medium
Example-Based	High	Low
Indecision Detection	Clean (all stable)	Useful (flagged Sample 12)

Quantum Geometric Learning: Encoding and Classification in Kendall Shape Spaces

Rasha Friji[2,3][0009−0004−4506−8815], Mehdi Houas[1], Behjet Boussofara[2], and Mourad Ben Ammar[2]

[1] Talan Group
[2] Talan Tunisia, Charguia1, Tunisia
[3] Cristal Lab, ENSI University, Tunisia
racha.friji@talan.com , mehdi.houas@talan.com, behjet.boussofara@talan.com, mourad.benammar@talan.com

Abstract. Kendall shape spaces provide a powerful geometric framework for modeling shapes independently of rotation, translation, and scale. These spaces, which arise naturally in statistical shape analysis and morphometrics, are structured as quotient Riemannian manifolds, such as complex projective spaces. In parallel, quantum machine learning (QML) has shown promise in tackling high-dimensional data and structured learning problems, yet its extension to non-Euclidean data domains remains largely unexplored. In this work, we propose a theoretical framework that enables QML algorithms to operate on Kendall shape spaces. We define a quantum encoding scheme that maps centered and normalized shape configurations into quantum states via amplitude encoding on the complex projective manifold. Furthermore, we introduce a quantum kernel function grounded in the geodesic distance between shapes, enabling quantum-enhanced classification and clustering algorithms. This approach opens the door to geometric quantum learning on structured manifolds and offers new perspectives for the efficient quantum processing of shape-based data. We discuss the feasibility of implementing this framework on NISQ hardware and highlight its potential for future applications in quantum shape analysis and beyond.

Keywords: QML · Kendall Shape Space · Quantum Kernel · Quantum Encoding.

1 Introduction

Quantum machine learning (QML) has emerged as a promising paradigm that aims to leverage quantum computing capabilities to enhance classical learning algorithms in terms of expressivity and computational efficiency [1, 2]. Recent advances in QML have shown potential speed-ups in supervised and unsupervised learning tasks, particularly in the context of kernel methods [3], variational circuits [4], and quantum embeddings [5]. However, most QML applications have been limited to data represented in Euclidean or Hilbert spaces, with little exploration of learning tasks involving non-Euclidean geometric structures.

In parallel, Kendall shape spaces provide a rigorous mathematical formalism for analyzing geometric shapes, especially in biological and computer vision domains [6]. These spaces arise from factoring out translation, rotation, and scale from point configurations, resulting in manifolds such as complex or real projective spaces (e.g., CP^{k-2} for 2D shapes). The geometric and statistical structure of these manifolds supports tasks such as shape clustering, classification, and regression through intrinsic methods [7, 15].

Bridging QML with shape analysis is a compelling yet underexplored direction. Recent works have begun addressing the generalization of machine learning to manifolds [8–11], and interest in quantum geometric representations has grown [12, 13], but a quantum framework for learning on shape spaces has not yet been proposed. In this work, we present a theoretical framework that enables quantum learning on Kendall shape spaces. We focus on:

1. Developing an encoding scheme for shape configurations into quantum states based on their projection onto shape manifolds.
2. Defining a quantum kernel that captures the geometric similarity between shapes using quantum state overlaps or geodesic-based distances.

This paper is structured as follows. Section 2 provides background on Kendall shape theory and its manifold structure. Section 3 outlines the proposed quantum encoding scheme and learning model. Section 4 discusses implementation feasibility and theoretical insights. Section 5 presents potential applications and research perspectives.

2 Background

2.1 Kendall Shape Spaces

Kendall shape theory provides a foundational mathematical framework for the statistical analysis of shapes, where a shape is defined as all geometrical information about an object that remains after filtering out location, scale, and rotation [6]. In what follows, k denotes the number of landmark points in the shape configuration. For a configuration of k landmark points in R^m, the pre-shape space is obtained by translating the configuration to the origin and scaling it to unit norm. The resulting pre-shapes lie on a unit hypersphere in R^{mk}. The shape space is then the quotient of this hypersphere under the action of the rotation group $SO(m)$, giving rise to a manifold with non-Euclidean geometry [14].

In the planar case ($m = 2$), the shape space of k-point configurations corresponds to the complex projective space CP^{k-2}, which is a smooth Riemannian manifold equipped with a natural metric structure—the Procrustes distance, derived from geodesic distances on the sphere [7]. This metric allows one to compute intrinsic means, variances, and perform statistical inference in shape spaces, forming the basis of statistical shape analysis [16].

2.2 Applications of Shape Analysis

Kendall shape analysis has been widely applied in biology (e.g., morphological variation in organisms), medicine (e.g., anatomical structure comparison), and computer vision (e.g., gesture and gait recognition) [10, 11, 17]. These applications rely on understanding shape variability and similarity through intrinsic measures, which remain consistent under transformations that do not affect the shape itself. Traditional machine learning methods have been extended to shape spaces via tangent space approximations, kernel methods, and intrinsic regression models [18]. However, these approaches often face scalability and representation challenges in high-dimensional shape datasets. This motivates the exploration of quantum machine learning paradigms for shape spaces.

2.3 Quantum Learning and Geometric Structures

Recent developments in quantum machine learning have explored the use of quantum-enhanced models for data embedded in structured spaces. For example, quantum kernel methods exploit the inner product structure of Hilbert spaces to capture nonlinear patterns in data [19], and variational quantum circuits can act as universal function approximators in hybrid quantum-classical architectures [20]. While these methods typically assume Euclidean input spaces, there is growing interest in quantum representations of geometric data, such as graphs, manifolds, and topological spaces [13]. Despite this interest, no theoretical framework currently exists for embedding and processing data from Kendall shape spaces within quantum machine learning pipelines. Addressing this gap could open new avenues for shape classification, clustering, and pattern discovery using quantum resources. In the next section, we introduce a novel quantum encoding scheme for Kendall shape configurations and propose learning models adapted to their intrinsic geometry.

3 Theoretical Proposition: Quantum Learning over Kendall Shape Spaces

In this paper, we propose a novel theoretical framework for applying quantum machine learning to data residing in Kendall shape spaces, with a focus on the planar case (CP^{k-2}) for k-landmark shapes. Figure 1 depicts the full framework of the proposed QML framework on Kendall shape space composed of three main processing blocks: 1- the mapping in Kendall shape space of the input landmarks configuration, 2- the encoding quantum state and 3- the quantum classifier. These blocks are described in the following sections.

3.1 Quantum Embedding of Shape Configurations

Let $X \in R^{2\times k}$ be a configuration of k landmarks in 2D. We assume that X has been preprocessed to remove translation (centered), scaling (unit norm),

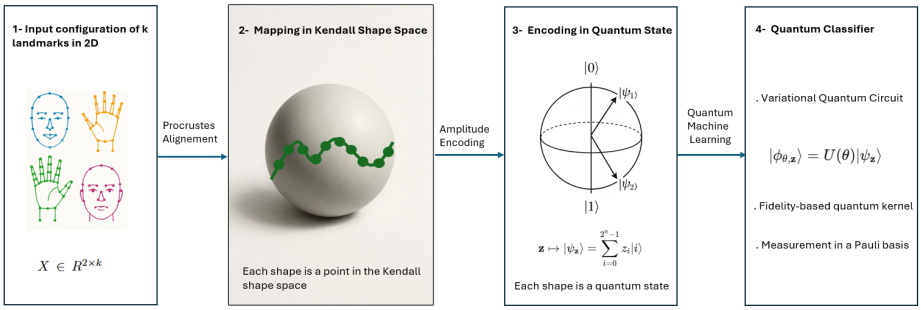


Fig. 1. Illustration of the full workflow of the proposed quantum machine learning framework on Kendall shape spaces. 1- The input configuration of 2D landmarks (representing a face, an action, an organ etc depending on the use case), 2- Mapping in Kendall shape space, 3- Encoding of each shape as a quantum state using amplitude encoding and 4- Quantum Classifier.

and rotation (Procrustes alignment). This yields a point on the Kendall shape manifold — a nonlinear Riemannian manifold of shapes modulo similarity transformations.

We propose to encode this point into a quantum state via the two following steps: complex representation and quantum state preparation. In the first step, we map the configuration X to a complex vector $z \in C^k$ such that $\sum_i |z_i|^2 = 1$. In the second step, we normalize the shape vector and loads it into the amplitudes of a n-qubit state using amplitude encoding (Equation 1).

$$\mathbf{z} \mapsto |\psi_\mathbf{z}\rangle = \sum_{i=0}^{2^n-1} z_i|i\rangle \tag{1}$$

Having encoded shapes as quantum states that respect the geometry of Kendall shape spaces, we can leverage these embeddings to define similarity measures between shapes in the quantum domain. The inner product structure of the Hilbert space naturally leads to the construction of a quantum kernel, which forms the foundation for quantum classification algorithms on shape data.

3.2 Quantum Shape Kernel

We define in equation 2 a quantum shape kernel $K_q(X, Y)$, where $|\psi_X\rangle$ and $|\psi_Y\rangle$ are the quantum embeddings of two shape configurations. This kernel corresponds to the fidelity between quantum states and aligns with the **cosine similarity** in Kendall space, which has a geometric interpretation in terms of the angle between complex pre-shapes.

$$K_q(X, Y) = |\langle\psi_X|\psi_Y\rangle|^2 \tag{2}$$

We conjecture that this kernel is positive semi-definite- a crucial property for kernels used in machine learning- and can be implemented via a SWAP

test [23] on a quantum device. Moreover, it preserves geodesic distances up to isometric transformations, allowing downstream quantum classifiers to operate in alignment with the intrinsic geometry of the shape manifold.

3.3 Variational Quantum Classifier on Shape Space

We design a variational quantum circuit (VQC) $U(\theta)$, parameterized by a set of angles θ, acting on quantum states $|\psi_X\rangle$. After shapes are encoded into quantum states, the VQC acts as a parametrized quantum model that learns to separate shape classes based on these embeddings. It replaces classical classifiers (like SVM or neural networks) with a trainable quantum model. It applies a sequence of quantum gates with tunable parameters θ, optimized through classical feedback loops. The final measurement output (in a Pauli basis or projection onto specific output states) corresponds to a classification label.

Given that the input quantum states encode shapes while respecting their intrinsic manifold structure, and since the VQC operates directly in this Hilbert space, the model benefits from the nonlinear separation capacity of quantum kernels. This alignment with the shape's geodesic geometry offers better generalization in low-data regimes, especially where classical flattening approaches (e.g., PCA) fail.

4 Feasibility Study

This section evaluates the feasibility of implementing the proposed quantum machine learning framework over Kendall shape spaces, considering both theoretical compatibility and current technological constraints.

4.1 Quantum Encoding Overhead

Encoding shape data into quantum states involves preparing a quantum system into a state $|z\rangle \in C^k$, where k is the number of landmarks. This can be achieved using amplitude encoding, which requires $\log_2(k)$ qubits, but in practice, state preparation complexity depends on the method. Classical-to-quantum encoding may require $\mathcal{O}(k)$ gate operations approximate or parametrized state preparation circuits can reduce gate depth but may introduce encoding noise. For small k (e.g., $k \leq 16$), encoding is feasible on Noisy Intermediate-Scale Quantum (NISQ devices), with potential for simulation on quantum emulators. So, for low-dimensional shape configurations, such as those common in biological morphometrics and medical imaging, encoding costs are therefore tractable.

4.2 Circuit Architecture and Complexity

The proposed variational quantum classifier (VQC) architecture operates on encoded shape states with several practical considerations. First, the qubit requirement scales logarithmically with the number of landmarks k, needing only

$\lceil \log_2(k) \rceil$ qubits; for example, with $k = 8$, just 3 qubits suffice, making it compatible with current quantum hardware. Second, to maintain trainability, moderate-depth ansätze such as hardware-efficient circuits or layered entanglers are employed. Third, while VQCs can suffer from barren plateaus—regions of vanishing gradients that hinder training, especially as circuit depth or input dimensionality increases, the use of structured and geometrically meaningful embeddings, as proposed here, may help mitigate this challenge [21]. Together, these features render the classifier well-suited for prototype implementations on today's quantum devices or simulators.

4.3 Classical vs Quantum Tradeoffs

When considering the integration of quantum models into shape analysis pipelines, it is essential to assess the tradeoffs between classical and quantum approaches across several dimensions. Table 1 summarizes key differences between classical and quantum paradigms for shape analysis. It contrasts their respective strengths in terms of expressivity, computation, generalization in low-data regimes, and compatibility with geometric invariants. Notably, while classical models are well-supported by existing toolchains and scalable architectures, quantum models offer the potential for superior generalization and richer geometric encodings but at the cost of current hardware limitations and algorithmic maturity.

Table 1. Key differences between classical and quantum paradigms for shape analysis.

Criteria	Classical (tangent space methods)	Proposed Quantum Approach
Geometry handling	Linearization introduces approximation	Intrinsic via Hilbert space embedding
Representation	Vectors in tangent space	Rays in complex projective Hilbert space
Scalability	Challenging in high dimensions	Requires qubit-efficient encoding
Interpretability	Established via geodesics and PCA	Requires definition of quantum observables

5 Perpectives and use cases

5.1 Perspectives

Leveraging quantum machine learning on Kendall shape spaces opens new avenues for both foundational and applied research. The alignment between Kendall's shape geometry and quantum Hilbert spaces invites exploration of quantum

Riemannian geometry, quantum analogues of Procrustes distances, and quantum principal geodesic analysis. Extending fidelity-based kernels to task-specific quantum kernels informed by shape classes or anatomical priors may enhance model expressivity. Quantum generative models, such as quantum GANs or Born machines, could enable shape synthesis with uncertainty quantification and augmentation on shape manifolds.

5.2 Use Cases

This quantum shape learning framework applies broadly where geometric structure is key. In medical imaging and morphometry, it supports shape-based classification from limited data, improving analysis of organ morphology and surgical outcomes. In biomechanics and anthropometry, it enhances robustness in classifying body shapes and gait patterns from noisy landmarks. Robotics benefits via improved shape understanding for grasping and perception tasks. Archaeology and evolutionary biology gain powerful tools for clustering fossil morphologies, reconstructing phylogenies, and restoring fragmented artifacts through invariant quantum shape embeddings.

Conclusion

In this work, we present a theoretical framework integrating quantum machine learning with Kendall shape analysis to develop geometric learning algorithms on shape manifolds. By embedding shape configurations from Kendall's shape space into quantum Hilbert spaces, our approach enables quantum-enhanced shape classification, similarity assessment, and generative modeling. Although theoretical, we discuss feasibility, outline implementation strategies, and illustrate applications. Future work will focus on experimental benchmarking of this framework against classical methods. Overall, this proposal lays the foundation for models that combine geometric invariants with quantum advantages in representation and generalization.

References

1. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**(7671), 195–202 (2017). https://doi.org/10.1038/nature23474
2. Schuld, M., Petruccione, F.: Supervised Learning with Quantum Computers. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96424-9
3. Schuld, M., Bocharov, A., Svore, K. M., Wiebe, N.: Circuit-centric quantum classifiers. Physical Review A **101**(3), 032308 (2021). https://doi.org/10.1103/PhysRevA.101.032308
4. Havlíček, V., Córcoles, A. D., Temme, K., et al.: Supervised learning with quantum-enhanced feature spaces. Nature **567**(7747), 209–212 (2019). https://doi.org/10.1038/s41586-019-0980-2

5. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. Nature Physics **10**(9), 631–633 (2014). https://doi.org/10.1038/nphys3029
6. Kendall, D. G.: Shape manifolds, Procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society **16**(2), 81–121 (1984). https://doi.org/10.1112/blms/16.2.81
7. Dryden, I. L., Mardia, K. V.: Statistical Shape Analysis with Applications in R. Wiley, Chichester (2016). https://doi.org/10.1002/9781119072492
8. Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: Going beyond Euclidean data. IEEE Signal Processing Magazine **34**(4), 18–42 (2017). https://doi.org/10.1109/MSP.2017.2693418
9. Pennec, X.: Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision **65**(1), 81–110 (2018). https://doi.org/ff10.1007/s10851-006-6228-4
10. Friji, R., Drira, H., Chaieb, F., Kchok, H., Kurtek, S.: Geometric deep neural network using rigid and non-rigid transformations for human action recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2021, pp. 12611–12620. IEEE, Piscataway (2021). https://doi.org/10.1109/TPAMI.2023.3291663
11. Hosni, N., Ben Amor, B.: A geometric convnet on 3D shape manifold for gait recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0. IEEE, Piscataway (2020). https://doi.org/10.1109/CVPRW50498.2020.00434
12. Larkin, J., Sornborger, A. T., Cao, Y.: Quantum manifold learning and geometry estimation. arXiv:2301.05707 (2023). https://arxiv.org/abs/2301.05707
13. Gyurik, C., van Vreumingen, D., Dunjko, V.: Structural risk minimization for quantum linear classifiers. In: Quantum 7, 893. Quantum, Vienna (2023). https://doi.org/10.22331/q-2023-01-13-893
14. Le, H., Kendall, D. G.: The Riemannian structure of Euclidean shape spaces: A novel environment for statistics. Annals of Statistics **21**(3), 1225–1271 (1993). https://doi.org/10.1214/aos/1176349259
15. Barbaresco, F.: Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Frechet Median. In: Nielsen, F., Bhatia, R. (eds.) Matrix Information Geometry, pp. 199–255. Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-30232-9_9
16. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. Statistica Sinica **20**(1), 1–100 (2010).
17. Srivastava, A., Klassen, E.: Functional and Shape Data Analysis. Springer, Cham (2016). https://doi.org/10.1007/978-1-4939-4020-2
18. Sommer, S., Lauze, F., Nielsen, M.: Optimization over geodesics for exact principal geodesic analysis. In: Scale Space and Variational Methods in Computer Vision, LNCS, vol. 6667, pp. 110–121. Springer, Cham (2011). https://doi.org/10.1007/s10444-013-9308-1
19. Schuld, M., Bocharov, A., Svore, K. M., Wiebe, N.: Circuit-centric quantum classifiers. Physical Review A **101**(3), 032308 (2021). https://doi.org/10.1103/PhysRevA.101.032308
20. Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M.: Parameterized quantum circuits as machine learning models. Quantum Science and Technology **4**(4), 043001 (2019). https://doi.org/10.1088/2058-9565/ab4eb5

21. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. Nature Communications **9**(1), 4812 (2018). https://doi.org/10.1038/s41467-018-07090-4
22. Kent, J. T.: The Fisher–Bingham distribution on the sphere. Journal of the Royal Statistical Society: Series B (Methodological) **44**(1), 71–80 (1982). https://doi.org/10.1111/j.2517-6161.1982.tb01189.x
23. M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2010. (Chapter 9, Exercise 9.10 describes the SWAP test).

Quantum-Enhanced Consensus Clustering through Quantum Annealing and QAOA

Daniele Franch[1][0009-0003-4820-8900], Rui Wang[1,2][0000-0003-4445-9473], Amer Delilbasic[1,2][0000-0001-7845-5193], Kristel Michielsen[2,3,4][0000-0003-1444-4262], and Gabriele Cavallaro[1,2,4][0000-0002-3239-9904]

[1] University of Iceland, 102 Reykjavík, Iceland
[2] Forschungszentrum Jülich, 52428 Jülich, Germany
[3] RWTH Aachen University, 52056 Aachen, Germany
[4] AIDAS, 52428 Jülich
daf9@hi.is

Abstract. Consensus clustering integrates multiple partitions into a unified solution, enhancing stability across clustering algorithms. However, ensuring global consistency remains computationally demanding. Quadratic Unconstrained Binary Optimization (QUBO) formulations provide a natural framework for this task, making consensus clustering amenable to quantum optimization methods such as quantum annealing and the Quantum Approximate Optimization Algorithm (QAOA).
We extend the established QUBO formulation for correlation clustering with a refinement protocol that allows for improvement of suboptimal solutions. To address current quantum hardware limitations, we implement a subproblem decomposition strategy that iteratively resolves constraint violations on manageable subsets.
Validation on benchmark and real-world datasets demonstrates competitive accuracy compared to established methods.

Keywords: Consensus Clustering · QUBO · Quantum Annealing · QAOA

1 Introduction

Consensus clustering, also known as cluster ensemble or clustering aggregation, is a widely used approach to improve the robustness and stability of unsupervised learning by combining multiple partitions into a single consensus solution [1,2,3]. Applications range from bioinformatics [4] to document analysis [5], where heterogeneous algorithms often produce divergent results. Many formulations of the problem are combinatorial, and several important ones, such as those based on correlation clustering [6,7], are NP-hard. Classical heuristics can produce reasonable results, but they scale poorly and typically lack performance guarantees.

A natural way to capture consensus clustering is QUBO formulations. These formulations are directly compatible with quantum optimization techniques such

as quantum annealing (QA)[8] and QAOA[9]. However, their cubic growth in transitivity constraints limits scalability on current hardware, and approximate solutions produced by noisy or resource-limited solvers often violate consistency.

In this work, we build on the established QUBO formulation for consensus clustering [7], which encodes pairwise agreement and transitivity constraints in a quadratic framework. Our main contribution is a refinement protocol with monotonicity guarantees that ensures approximate solutions can only improve or remain unchanged, never deteriorate. To address hardware limitations, we employ subproblem decomposition that restricts optimization to subsets of variables involved in constraint violations, iteratively updating until consistency is achieved.

2 QUBO Formulation

We build on the standard QUBO formulation for consensus clustering [7], where clustering is encoded as a pairwise decision problem. Given a dataset of n elements and m base clusterizations, we define s_{ij} and d_{ij} as the counts of how often elements i and j appear in the same or different clusters, respectively, and set $b_{ij} = d_{ij} - s_{ij}$ to capture their disagreement. Consider the upper triangular matrix $X = (x_{ij})$, where x_{ij} denotes a binary variable equal to 1 if i and j are assigned to the same cluster in the consensus, and 0 otherwise. The initial objective function is:

$$H_{\text{pair}}(X) = \sum_{i<j} b_{ij}\, x_{ij}, \tag{1}$$

which favors agreement with the base clusterings.

To ensure consistency in the clustering assignment, we must enforce transitivity: if i and j are clustered together, and j and k are clustered together, then i and k should also be clustered. Violations occur then when $(x_{ij}, x_{ik}, x_{jk}) \in \{(0,1,1),(1,0,1),(1,1,0)\}$. These infeasible triplets are penalized through the introduction of two auxiliary slack variables $s_1^{(i,j,k)}, s_2^{(i,j,k)} \in \{0,1\}$ for each triplet (i,j,k). The penalties are defined as:

$$P_1^{(i,j,k)} = p_1\big(x_{ij} + x_{ik} + x_{jk} + s_1^{(i,j,k)} - 2s_2^{(i,j,k)} - 1\big)^2, \tag{2}$$

$$P_2^{(i,j,k)} = p_2\, s_1^{(i,j,k)} s_2^{(i,j,k)}, \tag{3}$$

with $p_2 > p_1 > t$, where t is chosen large enough to make violations prohibitively costly. The complete QUBO objective is therefore:

$$H(X) = \sum_{i<j} b_{ij}\, x_{ij} + \sum_{i<j<k} \Big[P_1^{(i,j,k)} + P_2^{(i,j,k)}\Big]. \tag{4}$$

Solutions with no violations correspond to valid clusterings. The formulation scales quadratically with the number of elements and does not require setting the final number of clusters in advance.

2.1 Refinement of Approximate Solutions

Approximate solutions obtained from noisy or resource-limited solvers, or even classical heuristics, may satisfy feasibility constraints but still be suboptimal with respect to the linear objective. To enhance such solutions, we consider the following refinement procedure.

Let $x^{\text{app}} = \{x_{ij}^{\text{app}}\}$ be a feasible approximate assignment (or clustering). We define the *disagreement set*

$$\mathcal{D}(x^{\text{app}}) = \{(i,j) : \text{sign}(b_{ij}) = \text{sign}(2x_{ij}^{\text{app}} - 1)\},$$

which contains the indices where the sign of the coefficients b_{ij} opposes the current assignment x_{ij}^{app}. Equivalently, this set identifies where x^{app} differs from the unconstrained solution $x^{\text{unp}} = \Theta(-B)$, where Θ is the Heaviside step function, obtained by ignoring feasibility constraints.

Given any non-empty subset $S \subseteq \mathcal{D}(x^{\text{app}})$, we define a modified coefficient matrix:

$$b_{ij}^* = \begin{cases} -b_{ij}, & (i,j) \in S, \\ b_{ij}, & \text{otherwise}, \end{cases}$$

denoted collectively as $B^* = \{b_{ij}^*\}$. Let x^{imp} be an optimal solution under the modified coefficients B^*. This procedure encourages the solution to the problem with the modified matrix B^* to align with x^{app}.

We now demonstrate that this refinement is monotonic: the original objective cannot worsen.

Proof. Let x^{imp} be an optimal solution for the problem with the modified matrix B^*. By optimality, we have:

$$\sum_{i<j} b_{ij}^* \, x_{ij}^{\text{imp}} \leq \sum_{i<j} b_{ij}^* \, x_{ij}^{\text{app}}. \tag{5}$$

We can then express b_{ij}^* in terms of b_{ij} as $b_{ij}^* = b_{ij} - 2\mathbf{1}_{(i,j)\in S} b_{ij}$, where $\mathbf{1}_{(i,j)\in S}$ is the indicator function. Substituting this expression into Eq. (5) and rearranging the terms, we obtain:

$$\sum_{i<j} b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{imp}}) \geq 2 \sum_{(i,j)\in S} b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{imp}}). \tag{6}$$

We can now rewrite the right-hand side by adding and subtracting the unconstrained solution x^{unp}:

$$2 \sum_{(i,j)\in S} b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{imp}}) = 2 \sum_{(i,j)\in S} b_{ij}\big[(x_{ij}^{\text{app}} - x_{ij}^{\text{unp}}) + (x_{ij}^{\text{unp}} - x_{ij}^{\text{imp}})\big]. \tag{7}$$

By definition, the unconstrained solution takes the value 0 if $b_{ij} > 0$ and 1 if $b_{ij} < 0$ and since $(i,j) \in S$ implies $x_{ij}^{\text{app}} \neq x_{ij}^{\text{unp}}$, the difference $x_{ij}^{\text{app}} - x_{ij}^{\text{unp}}$ is either $+1$ or -1.

Considering both cases, for all $(i,j) \in S$ we have

$$b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{unp}}) = \begin{cases} b_{ij}, & b_{ij} > 0, \\ -b_{ij}, & b_{ij} < 0, \end{cases} = |b_{ij}| \geq 0.$$

The second term on the right-hand side of Eq. (7) satisfies the trivial bound:

$$2 \sum_{(i,j)\in S} b_{ij}(x_{ij}^{\text{unp}} - x_{ij}^{\text{imp}}) \geq -2 \sum_{(i,j)\in S} |b_{ij}|,$$

because each summand is bounded below by $-|b_{ij}|$. Combining these bounds, we conclude that the right-hand side of (6) is non-negative:

$$2 \sum_{(i,j)\in S} b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{imp}}) \geq 0.$$

Substituting back into (6) we obtain:

$$\sum_{i<j} b_{ij}(x_{ij}^{\text{app}} - x_{ij}^{\text{imp}}) \geq 0,$$

or equivalently,

$$\sum_{i<j} b_{ij}\, x_{ij}^{\text{imp}} \leq \sum_{i<j} b_{ij}\, x_{ij}^{\text{app}}.$$

This shows that the refinement step never increases the original linear objective. The argument is general for any linear objective with an additive penalty that vanishes on feasible assignments.

3 Methodology

The empirical evaluation of our quantum-enhanced consensus clustering framework pursues three main objectives: first, to empirically validate the monotonicity property established in Section 2.1; second, to compare the QUBO-based consensus clustering against both individual clustering methods and established ensemble approaches; and third, to investigate scalability and refinement strategies when hardware limitations prevent solving the full problem.

3.1 Datasets and Experimental Setup

We selected two benchmark datasets for our experiments: the Iris dataset and the Wine dataset, both accessed through scikit-learn [10]. Additionally, we validated our approach on the reduced EuroSAT satellite imagery dataset[11,12] to demonstrate scalability to real-world scenarios. The Iris dataset was used both in its full version and in reduced subsets of 10, 15, 20, and 25 samples where the latter are used for empirically demonstrating the refinement property. The datasets are small enough to allow evaluation under classical consensus clustering frameworks while providing meaningful validation of our approach.

3.2 Base Clustering Ensemble Construction

To construct diverse ensembles of base clusterings, we employed heterogeneous algorithms representing distinct clustering paradigms. For the benchmark datasets (Iris and Wine), we used centroid-based partitions via K-Means[13] with $k \in \{2, 3, 4\}$, hierarchical partitions via Agglomerative Clustering[14] with $k \in \{2, 3, 4\}$, and graph-based partitions via Spectral Clustering[15] for $k \in \{2, 3, 4\}$. Density-based partitions with DBSCAN [16] using neighborhood radius $\varepsilon \in \{0.3, 0.5, 0.7\}$ were additionally included for the Iris dataset. For the EuroSAT dataset, we employed K-Means with $k \in \{5, 8, 14\}$ using different random seeds, Agglomerative Clustering with $k = 10$, and Spectral Clustering with nearest neighbors affinity ($k = 10$).

3.3 Evaluation Metrics and Baseline Comparisons

We assessed clustering quality using both external and internal evaluation metrics. External indices measure agreement with ground-truth labels and include the Adjusted Rand Index (ARI)[17] and Normalized Mutual Information (NMI)[1]. The Silhouette coefficient[18] serves as our internal evaluation metric, assessing cluster quality independently of ground-truth labels. Additionally, we employ the Mean ARI metric, computed as the average ARI between the final consensus clustering and each of the m base clustering methods, to quantify how well the consensus solution captures agreement across all input methods.

We compared the QUBO-based approach against several established consensus clustering methods. For the Cluster-based Similarity Partitioning Algorithm (CSPA)[1] with METIS and the Meta-CLustering Algorithm (MCLA)[1], we used implementations from the consensus clustering package available at[19]. Additionally, we implemented agglomerative and spectral clustering variants of CSPA, Non-negative Matrix Factorization (NMF)-based consensus clustering[20], and the Da-sm and Da-cr methods proposed by Cohen et al.[21].

3.4 QUBO Solution Methods

We solved the QUBO formulation using multiple optimization approaches. For Quantum Annealing, we used D-Wave's Advantage_system4.1 with 3500 reads, keeping all other parameters at their default values. As a classical baseline, we employed Simulated Annealing[22,23]. Solutions obtained with both optimization approaches were identical on our benchmark problems.

To test the QUBO on gate-model quantum hardware, we implemented a QAOA solver in Qiskit. We instantiate a QAOA Ansatz directly from the QUBO cost operator and sweep the circuit depth $p \in \{1, \ldots, 5\}$. Parameters are trained with a multi-start SPSA routine: each batch/round starts from the previous angles and adds at least one random seed (60–100 iterations per start), retaining the best expectation. Execution uses Aer in two modes: (i) noiseless `AerSimulator` and (ii) noisy (Aer with the `FakeTorino` noise model), with 4,096 shots unless stated otherwise.

3.5 Scalability and Refinement Protocol

To address hardware scalability limitations, we adopted a submatrix decomposition strategy combined with iterative refinement. The process works as follows: starting from the global bias matrix, we identify transitivity violations across all element triplets. We then extract a subset of 15 elements, chosen as a safe bound for the quantum annealer's capacity, focusing on elements involved in violations.

The subproblem is solved iteratively, as in [24]: violations within the submatrix are detected, penalized, and the problem is resolved until no violations remain in the subset. The resulting assignment updates the global bias matrix by flipping the sign of entries where the submatrix solution conflicts with the global configuration. We repeat this process until convergence, yielding the final solution $\mathbf{x} = \Theta(-\mathbf{B})$. For QAOA, we mirror this refinement: in each round, we partition violating constraints into disjoint batches (5 qubits per batch) that fit a fixed qubit budget, run QAOA per batch, and accept a batch only if its local cost decreases or its violation count drops until no batch is accepted or all violations are cleared.

4 Results

This section presents the empirical evaluation of the QUBO consensus clustering framework across benchmark and real-world datasets. The QUBO-based approach was tested using both quantum annealing and simulated annealing, with both methods producing identical solutions on benchmark datasets. Additionally, QAOA was evaluated specifically on the EuroSAT dataset to demonstrate the approach's compatibility with gate-model quantum hardware.

4.1 Empirical validation of the refinement protocol

To provide empirical evidence that the refinement of an approximate solution is feasible, we constructed four reduced versions of the Iris dataset containing 10, 15, 20, and 25 elements, respectively. For each dataset, four feasible but approximate solutions (i.e., valid assignments without transitivity violations) were generated. For these cases, we applied the refinement strategy described in Section 2.1. As shown in Fig. 1, across all datasets every point lies on or below the line $E_i = E_f$, empirically confirming that the refinement procedure may improve but never worsens the energy and may improve the quality of approximate solutions. Many points lie on the diagonal $E_i = E_f$, showing no improvement. This likely occurs because $|D \setminus S|$ is too small, leaving insufficient disagreement pairs available for refinement.

The main purpose of refinement is to simplify the problem by reducing constraint violations, deliberately settling for approximate solutions rather than pursuing optimality. Despite the lack of theoretical guarantees when using subproblem decomposition, we tested this approach on the full Iris dataset to evaluate its practical effectiveness on problems exceeding hardware limits.

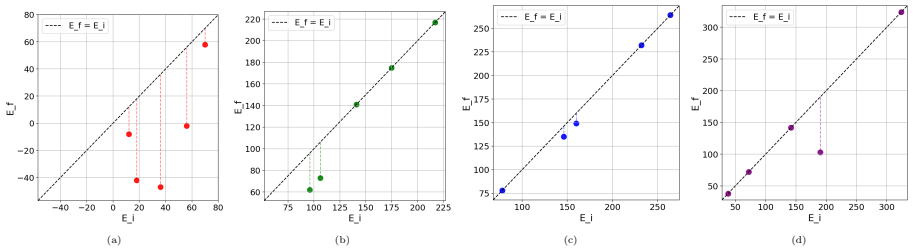


Fig. 1: Final energy versus initial energy for the four reduced Iris datasets: 10 elements (Fig. 1a), 15 elements (Fig. 1b), 20 elements (Fig. 1c), and 25 elements (Fig. 1d). Each point represents an approximate solution refined using the proposed protocol. All points lie on or below the diagonal $E_i = E_f$, showing that the refinement procedure never increases energy and can improve solution quality.

For this experiment, we considered two different starting solutions: the true labels and a random solution close to the optimal one. In both cases, we tested configurations where 50, 100, 200, and 300 pairs from the disagreement set are left unchanged, i.e., pairs in $D \setminus S$, with each larger set containing all pairs from the smaller configurations. This setup allows us to analyze how the number of fixed pairs influences the refinement process and simplifies the problem. The refinement results for both starting solutions are reported in Table 1. Overall, the

Table 1: Refinement results for different starting solutions and numbers of unchanged pairs. E_f denotes the final energy after refinement.

	True Labels			Near-optimal solution		
$\lvert D \setminus S \rvert$	E_f	Violations	Subproblems	E_f	Violations	Subproblems
0	-24101	0	0	-28024	0	0
50	-24101	1798	12	-28002	4310	28
100	-24452	3002	45	-28656	7976	43
200	-24439	4336	39	-28896	14300	105
300	-24624	4274	29	-28024	18102	98
$\lvert D \rvert$	-28963	26452	121	-28963	26452	121

final energy generally improves as more pairs are left unchanged, though some fluctuations are observed. Starting from the true labels, the procedure successfully reduces the number of subproblems to roughly one-third of those required by the full solution, while a substantial gap in solution quality compared to the complete solution remains. When starting from the threshold-based solution, the refinement behavior is more variable. Notably, for $|D \setminus S| = 50$, the final energy is worse than the initial solution, likely due to the decomposition into subproblems. This suggests that, although the method can still be effective, its performance is not guaranteed in all instances.

In general, the refinement method's effectiveness depends both on the quality of the initial solution and on the number of pairs left unchanged. It proves more beneficial when starting farther from the optimal solution, while small numbers of fixed pairs offer limited computational savings. The approach remains restricted to smaller datasets until hardware improvements allow scaling to larger problems.

4.2 Comparison with Consensus Clustering Methods

We benchmarked our approach against representative consensus clustering methods as described in Section 3.3. For methods requiring the number of clusters k as input, we set k equal to the ground-truth number of clusters. Evaluation metrics include standard external indices (ARI, NMI), the Silhouette score, and the QUBO cost function (Eq. (4)).

Table 2: Comparison of consensus clustering methods on Iris and Wine datasets.

Dataset	Method	Mean ARI	ARI	NMI	Silhouette	QUBO objective
Iris	CSPA	0.688	0.705	0.705	0.501	-26508
	CSPA Agglomerative	0.731	0.770	0.770	0.554	-30748
	CSPA Spectral	0.731	0.770	0.770	0.554	-30748
	NMF	0.469	0.661	0.661	0.405	-21100
	MCLA	0.758	0.786	0.786	0.542	-30320
	Da-sm	0.744	0.777	0.777	0.553	-30690
	Da-cr	0.731	0.770	0.770	0.554	-30748
	QUBO	0.731	0.770	0.770	0.554	-30748
Wine	CSPA	0.391	0.399	0.399	0.531	-28150
	CSPA Agglomerative	0.368	0.416	0.416	0.565	-34283
	CSPA Spectral	0.368	0.416	0.416	0.565	-34283
	NMF	0.368	0.416	0.416	0.565	-34283
	MCLA	0.359	0.420	0.420	0.557	-30162
	Da-sm	0.368	0.416	0.416	0.565	-34283
	Da-cr	0.368	0.416	0.416	0.565	-34283
	QUBO	0.368	0.416	0.416	0.565	-34283

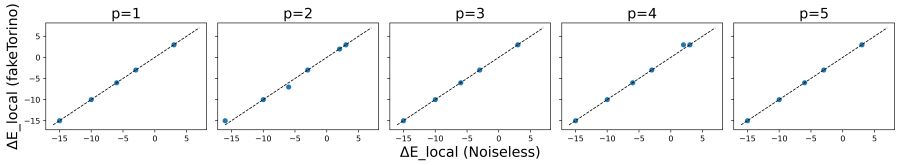
Table 2 shows that our QUBO solver matches the best consensus baselines on both Iris and Wine. On Iris it achieves strong agreement with the base clusterers and competitive external metrics; on Wine, where methods perform more similarly, it remains on par across criteria. Notably, QUBO consistently attains the lowest value of its own objective, confirming alignment with the intended optimization. The main drawback is runtime from solving many subproblems, which requires multiple iterations. This efficiency–optimality trade-off should lessen as quantum hardware permits larger subproblems, improving scalability and applicability.

4.3 EuroSAT Dataset Validation

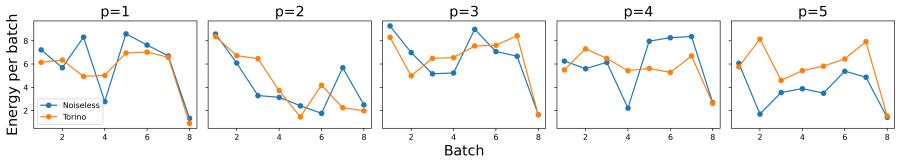
We evaluated our approach on a subset of 50 EuroSAT satellite images, which include 10 land-use classes, using feature embeddings from a pre-trained ResNet18 [25].

Noise Robustness of the Batched QAOA To evaluate the performance under realistic noise conditions, we assess the noise robustness of the batched QAOA by comparing simulations with a realistic gate-level noise model (`FakeTorino`)

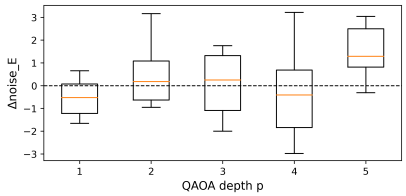
to an ideal noiseless simulator (`Aer`) for circuit depths $p \in \{1, \ldots, 5\}$. For each batch, we recorded the batch Hamiltonian expectation value E_{batch}, the local energy change ΔE_{local}, and the noise-induced energy inflation, which was quantified as $\Delta E_{\text{noise}} = E_{\text{batch}}^{\text{Torino}} - E_{\text{batch}}^{\text{Noiseless}}$, calculated for each batch at a fixed circuit depth. Remarkably, the acceptance rate for local updates remained iden-



(a) **Local decision stability.** Comparison of ΔE_{local} on the `FakeTorino` backend versus the noiseless simulator for $p = 1, \ldots, 5$.



(b) **Batch energy trajectories.** Evolution of the per-batch E_{hat} for the noiseless (blue) and `FakeTorino` (orange) simulations. The curves track closely across batches and depths, with deviations increasing slightly for $p = 5$.



(c) **Noise-induced energy inflation.** Boxplots of the energy difference ΔE_{noise} by depth.

Fig. 2: Impact of simulated gate-level noise on the batched QAOA refinement. The analysis shows that acceptance statistics are close on both backends, as the local energy-based decisions are robust to the effects of noise from the `FakeTorino` model.

tical (0.625) for both the noisy and noiseless simulations across all depths. This stability arises because, as shown in Fig. 2(a), the noise model preserves the sign and relative ordering of ΔE_{local}. The energy expectation trajectories in Fig. 2(b) track each other closely, with minor deviations increasing with circuit depth. The resulting energy inflation, ΔE_{noise}, is centered near zero for $p \leq 4$ and exhibits a slight positive shift at $p = 5$ that is insufficient to alter acceptance behavior (Fig. 2(c)). In summary, our numerical simulations demonstrate that for the problem scale and batching scheme considered, the local decisions within the batched QAOA framework are robust to a realistic gate-level noise model. The `FakeTorino` backend introduces a modest, depth-dependent energy inflation that is insufficient to alter the algorithm's acceptance mechanism. However, the listed results are only for one device model and problem size; deeper circuits,

fewer shots, or different mixers could increase inflation. Real hardware may add calibration that the FakeTorino model does not capture. Therefore, exploring hardware execution and depth-aware optimizers remains future work.

Clustering Performance on EuroSAT Table 3 summarizes the clustering performance for classical and quantum-based methods on the 50-image EuroSAT subset. Metrics include the Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Silhouette score. For QUBO, solutions were obtained via quantum annealing, while for QAOA, results correspond to batched QAOA refinement with circuit depth $p = 2$.

Table 3: Clustering performance on the EuroSAT subset using classical and quantum-based methods.

Method	ARI	Mean ARI
NMF	0.393	0.654
Agglomerative	0.548	0.731
Spectral	0.393	0.656
QUBO (QA)	0.552	0.727
Batched QAOA	0.553	0.729

The performances of batched QAOA and QUBO are comparable, whereas classical methods perform at a similar or lower level, highlighting the effectiveness of quantum-based optimization for clustering on the EuroSAT subset.

5 Conclusions

This study extends earlier QUBO formulations [7] by introducing a refinement procedure that can improve approximate clustering solutions. While decomposition into subproblems is necessary due to current hardware limits and does not guarantee convergence to the global optimum, the workflow effectively combines quantum exploration with classical refinement.

Experiments show that QUBO-based clustering performs comparably to classical ensemble methods, with refinement ensuring transitivity. Similarly, batched QAOA demonstrates robust local decision-making even under realistic gate-level noise, with modest energy inflation insufficient to alter acceptance behavior. These results indicate that quantum solvers can contribute to clustering pipelines.

Future work will explore alternative penalty formulations, larger datasets, and hardware-specific optimizations, including noise-aware strategies for QAOA.

Acknowledgments. This work is supported by the Icelandic Centre for Research, project *"Hybrid Quantum-Classical Workflows for EO"*, RANNÍS grant number: 2511078-051 (`http://en.rannis.is/`).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alexander Strehl and Joydeep Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3:583–617, 2002.
2. Sandro Vega-Pons and José Shulcloper. A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25:337–372, 05 2011.
3. Fanrong Meng, Xuejiao Tong, and Zhixiao Wang. A clustering-ensemble approach based on voting. In Hepu Deng, Duoqian Miao, Jingsheng Lei, and Fu Lee Wang, editors, *Artificial Intelligence and Computational Intelligence*, pages 421–427, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
4. Stefano Monti, Pablo Tamayo, Jill Mesirov, and Todd Golub. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52:91–118, 07 2003.
5. Carla Brodley. Solving cluster ensemble problems by bipartite graph partitioning. 09 2004.
6. Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine Learning*, 56:89–113, 2004.
7. Moses Charikar, Venkatesan Guruswami, and Anthony Wirth. Clustering with qualitative information. *Journal of Computer and System Sciences*, 71(3):360–383, 2005. Learning Theory 2003.
8. Aleta Berk Finnila, Maria A Gomez, C Sebenik, Catherine Stenson, and Jimmie D Doll. Quantum annealing: A new method for minimizing multidimensional functions. *Chemical physics letters*, 219(5-6):343–348, 1994.
9. Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014.
10. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
11. Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 204–207. IEEE, 2018.
12. Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.
13. J MacQueen. Multivariate observations. In *Proceedings ofthe 5th Berkeley Symposium on Mathematical Statisticsand Probability*, volume 1, pages 281–297, 1967.
14. Joe H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.
15. Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001.
16. Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Data Mining*, 1996.

17. Lawrence J. Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.
18. Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
19. Gregory Giecold. Cluster_ensembles. `https://github.com/GGiecold-zz/Cluster_Ensembles`, 2016. Accessed: September 2025.
20. Daniel Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In T. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2000.
21. Eldan Cohen, Avradip Mandal, Hayato Ushijima-Mwesigwa, and Arnab Roy. Ising-based consensus clustering on specialized hardware. In *Advances in Intelligent Data Analysis XVIII: 18th International Symposium on Intelligent Data Analysis, IDA 2020, Konstanz, Germany, April 27–29, 2020, Proceedings*, page 106–118, Berlin, Heidelberg, 2020. Springer-Verlag.
22. Scott Kirkpatrick, C Daniel Gelatt Jr, and Mario P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
23. D-Wave Systems Inc. neal: Simulatedannealingsampler for dimod. GitHub repository, 2021. Implementation of simulated annealing sampler for general Ising model graphs; used with dimod — see "SimulatedAnnealingSampler".
24. Daniele Franch, Enrico Zardini, Enrico Blanzieri, and Davide Pastorello. Consensus ranking by quantum annealing, 2025.
25. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

Hybrid quantum classical algorithms: a cloud on-demand viewpoint

Aleksander Wennersteen[0009−0006−5486−0980], Kemal Bidzhiev, Mauro D'Arcangelo, Matthieu Moreau[0009−0001−7886−3426], Anton Quelle, Alexandre Dauphin[0000−0003−4996−2561], and Mourad Beji[0009−0008−7998−6357]

Pasqal, 24 Av. Emile Baudot, 91120 Palaiseau, France
aleksander.wennersteen@pasqal.com
mourad.beji@pasqal.com

Abstract. In the last decade, advances in quantum technologies have allowed for the rapid development of industrialized quantum processing units. These new devices exploit the laws of quantum mechanics to perform complex calculations. Quantum processing units require new ways of thinking and programming. In particular, these new algorithms will be hybrid, with part of the computation performed on classical high-performance computing hardware and part on the dedicated quantum hardware. At Pasqal, we have developed a cloud platform hosting a neutral atom quantum processing unit (QPU) operating in the analog paradigm and a series of hybrid quantum classical algorithms that cover applications such as quantum optimization, quantum machine learning and quantum simulation. In this paper, we will show how this platform is used during the execution of real workloads.

1 Introduction

Neutral atom based quantum computers are gaining prominence as a leading Quantum Processing Unit (QPU) modality due to their scalability, programmability, and robustness. At Pasqal, we have developed and deployed cloud-accessible[1] QPUs based on neutral atom architectures that operate in the analog mode [12]. However, a purely quantum computational paradigm is insufficient. This has led to the emergence of hybrid quantum-classical algorithms as a practical and scalable approach.

Hybrid quantum-classical algorithms leverage the strengths of classical high-performance computing (HPC) systems alongside quantum processors. Moreover, integrating QPU and emulator backends through cloud access, offering the choice of backend to the user, allows efficient workflows during the whole development and execution process.

In this paper, we present how Pasqal's cloud platform supports this hybrid paradigm, providing an integrated architecture for algorithm design, benchmarking, and execution. The tools available on the cloud allow the user to:

[1] Pasqal Cloud Service available at: https://portal.pasqal.cloud

- prepare sequences that can be sent to the QPU using both traditional and graphical programming SDKs in Pulser and Pulser Studio,
- benchmark the algorithms on system size from 50 to 100 qubits with a suite of state-of-the art emulators,
- benchmark the resilience of the algorithms to noise with the noisy emulators,
- Monitor your jobs through our all in one platform: Follow the progress of your jobs in real-time, download or visualize their results when they are ready, retry or cancel your jobs when needed, extract logs and add per-time-step observables for emulators to verify the simulations validity,
- easily switch between the emulators and the QPU backends.

After introducing the architecture and the tools, we show the usage of the cloud with a series of applications.

2 Platform architecture

In this section, we discuss the full architecture of our platform. We describe briefly how the HPC cluster is integrated with cloud services to provide our quantum device emulation, and cloud infrastructure, including orchestration, scheduling, and dispatching of jobs. The emulator which is used herein, which we will refer to as `EMU-TN` as introduced in [3], and more recently a suite of statevector and tensor network emulators [2].

Orchestration and scheduling The infrastructure, illustrated in Fig. 1, is split between public cloud services orchestrated through Kubernetes, a HPC cluster hosted in a private data center running a Slurm scheduler, and the QPU located in Pasqal's datacenter. We next describe their integration for providing quantum computing as a service.

It is possible to run the same workloads on a Kubernetes pod alongside the other services, either as a long-running service responsible for receiving and executing a job or scheduled using Kubernetes compatible schedulers. However, a downside of these approaches is that it would be challenging to integrate the on-premise HPC cluster with the Kubernetes scheduler. Not scheduling the jobs on the cluster, which would mean just running a job execution service in a Kubernetes pod, would put limitations both on how effectively the resources in the cluster are used, and ultimately on the performance of the HPC applications. In this work we use a custom, internal, scheduler for our program, it is the same scheduler as for QPU jobs, to dispatch jobs.

Job dispatching To connect our two clusters, we used the Slurm REST API [20] to submit jobs to Slurm from the regular backend service. A "Slurm connector" service was developed, later merged into "ManGo" to have a single scheduler, to act as a bridge between the existing Pasqal cloud Services and the HPC cluster. The service creates a control script on-demand for the job which is submitted to Slurm as the batch script. This script then takes care of executing all

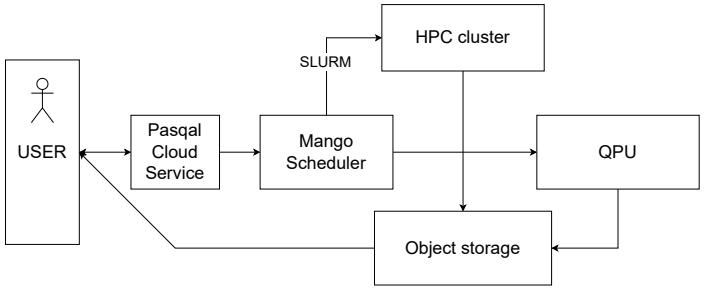


Fig. 1. Architecture of the cloud platform providing quantum computing as a service. The cloud services are deployed on a Kubernetes cluster and use a microservice architecture, each microservice exposing a public REST API to end users. The HPC cluster where the application runs runs, consists of 10 DGX A100 systems from NVIDIA and uses Slurm, an HPC cluster management and job scheduling system, for resource management and job submission.

the higher-level logic of the job. The service is also responsible for setting up the job with appropriate resources and inserting all necessary information for future communication with the cloud service. An alternative would have been to extend the Kubernetes cluster using the bridge pattern [17]. However, this would not have allowed us to manage the emulators in the same way as the QPU without significant development.

Upon completion of the job the results, as well as the input, are uploaded to an object storage. The full results include the logs from the emulator, bitstrings and upon request arbitrary Pauli-string observables and the quantum state. In order to aid in the exploration of the job, the individual programs can be loaded into Pulser Studio[2], a Pulser graphical user interface version that can visualize the program. For smaller programs (up to 12 atoms) the system can be simulated in the browser. For larger programs Pulser Studio can submit jobs to the cloud platform described.

The internal logic of the job submission endpoint validates the pulse sequence and the emulation configuration. The validation step checks that the request and the underlying data such as the sequence and emulator configuration are correctly formatted, that the requested resources are accessible to the user and, for a real device, that the sequence is physically valid. It then dispatches the request to the appropriate scheduler. In Code Sample 1, we show an example of sending a serialized pulse sequence to Pasqal Cloud Services for execution. To build the sequences we have used `Pulser` [21], an open source package for designing neutral atom QPU sequences. To communicate with cloud services we have used the `PasqalCloud` `Pulser` extension.

```
1  import numpy as np
2  import pulser
```

[2] Available at: https://pulserstudio.pasqal.cloud/

```python
from pulser_pasqal import PasqalCloud, EmuMPSBackend

connection = PasqalCloud(
    username=USERNAME, # Your username on the Pasqal Cloud Platform
    project_id=PROJECT_ID, # The ID of the billed project
    password=PASSWORD, # The password for your account
)

# Build a device object to describe the specs of our "Fresnel" QPU
device = connection.fetch_available_devices()["FRESNEL"]

# Create a register with 2 atoms with automatic calibration
reg = pulser.Register({"q0": (-5, 0), "q1": (5, 0)}).
    with_automatic_layout(device)
seq = pulser.Sequence(reg, device)
seq.declare_channel("rydberg_global", "rydberg_global")
t = seq.declare_variable("t", dtype=int)

amp_wf = pulser.BlackmanWaveform(t, np.pi)
det_wf = pulser.RampWaveform(t, -5, 5)
seq.add(pulser.Pulse(amp_wf, det_wf, 0), "rydberg_global")

# Run the program on a real QPU
# backend = pulser.backends.QPUBackend(seq, connection=connection)
# Choose the latest emulator, Emu-MPS
backend = EmuMPSBackend(seq, connection=connection)

# Parameteraized job execution with job_params
job_params = [
    {"runs": 100, "variables": {"t": 1000}},
    {"runs": 50, "variables": {"t": 2000}},
]
results = backend.run(job_params=job_params, wait=True)
print([job.bitstring_counts for job in results])
```

Code Sample 1. Running QPU or emulation jobs through the cloud service. The cloud connection is configured the program is created using Pulser the chosen backend selected and the final parameter values as inserted before the job is finally executed and the results are retrived.

3 Applications

In this section, we discuss an application of the framework, motivated by typical tasks that can be performed with a Neutral-Atom Quantum Processor addressing the ground-Rydberg basis [12]. The interactions between atoms causes the quantum system to evolve according to the following Hamiltonian, that is remi-

niscent of the Ising model Hamiltonian:

$$\hat{H}(t) = \sum_{\text{atom } i} \left(\frac{\Omega(t)}{2} \hat{\sigma}_i^x - \delta(t) \hat{n}_i + \sum_{j<i} \frac{C_6}{|\mathbf{r}_{ij}|^6} \hat{n}_i \hat{n}_j \right), \tag{1}$$

where $\hat{n}_i = |1_i\rangle\langle 1_i| = (\mathbb{I} - \hat{\sigma}_i^z)/2$ is the projector into the one-qubit excited state of the i-th atom, C_6 is a constant that depends on the properties of the targeted excited state, and we have we have set the phase $\phi = 0$.

As shown in (1), the Hamiltonian can be split into two parts: the global single-qubit interactions driven by programmable lasers through the Rabi frequency $\Omega(t)$ and the detuning $\delta(t)$; and the atom-atom interactions that are programmable via the position of the atoms in the register.

Two promising areas of applications to explore are graph-based problems [8,9,6,13,7,16] and quantum simulation [5,15,25,1,4,18,11]. These are seen as particularly promising on the neutral atom analog QPUs compared to other qubit modalities and paradigms such as digital superconducting or ion-trap QPUs. The difference comes down to two distinguishing features:

1. The ability to encode graph structures in the register by placing the atoms in the form of the encoded graph, utilizing the atom-atom interactions in the Hamiltonian (1).
2. The natural time-evolution of the Hamiltonian, unlike in the digital case where said Hamiltonian must be Trotterized to an acceptable precision with a possibly large number of gates, each of which will have an error rate.

In order to further democratize quantum computing, along with the cloud platform outlined herein, we are developing open-source software to provide reusable solvers abstracting specific projects, such as Maximum Independent Set (MIS) solver [23] and Quantum Evolution Kernel (QEK) solver [24]. The MIS problem is explored further in the following subsection. For an introduction to QEK see [13]. The aforementioned libraries come with out-of-the-box support for Pasqal's cloud platform.

After the quantum evolution and readout, one obtains a bitstring of 0's and 1's representing excited and ground-state qubits, respectively. To calculate the expectation value of a given observable, many cycles need to be performed in order to generate enough statistics. The output of the emulator provides bitstrings that simulate the actual output of the QPU.

Maximum Independent Set on Unit Disk graphs We present a typical use case of neutral atom devices for two-dimensional registers. The possibility of placing atoms in arbitrary 2D configurations allows to solve certain classes of hard graph problems on the QPU. An interesting application is solving the Maximum Independent Set (MIS) problem on Unit Disk (UD) graphs [19].

A graph is an object with nodes and connections between nodes. A UD graph is a graph where two nodes are connected if an only if they are closer than a certain minimal distance. By representing the nodes of a graph with neutral

atoms and matching the Rydberg blockade radius with the UD graph minimal distance, one can establish a direct mapping where a connection exists between atoms that are within a blockade distance of each other. The quantum evolution of such a system is naturally restricted to those sectors of the Hilbert space where excitations of connected atoms are forbidden. These configurations translate to independent sets of a graph, *i.e.* subsets of nodes that are not directly connected to each other. Driving the quantum evolution in such a way as to produce as many excitations as possible, one would obtain with high probability as the outcome of a measurement an independent set of high cardinality, representing a good candidate solution to the MIS problem.

The best pulse sequence to find the MIS of a graph is based on adiabatic pulses. As shown in [10], however, for arbitrary graphs it is often necessary to optimize the pulse shape in order to find the MIS with a high enough probability. We present in Fig. 2 a one-parameter family of pulses where the parameter t_c controls at which time the detuning goes from negative to positive. Rather than sweeping over the free parameter sequentially, one can send multiple simultaneous jobs for each value to be tested, and finally select the parameter giving the best result. Parallelizing this task is a direct advantage of the cluster infrastructure, and the cloud service provides a dedicated interface in which all jobs can be organized and retrieved, saving large amounts of time. We perform a sweep over 9 parameters for a pulse of 5 μs on a graph with 30 nodes. The results are shown in Fig. 2, where the optimal zero-crossing for the detuning is found to be at around 1.6 μs.

While the example shown here is centered on pulse parameters, other tasks suited for parallelization are also implementable, e.g. averages over noise and disorder realizations. Being able to quickly set up and run studies of this kind for systems of many qubits opens the door for creative algorithm design that is at the same time compatible with a QPU. By attaching `NoiseModel`s to the job submission in Code Sample 1, this submission can easily be extended to also include a study of the implementations noise-robustness by running noisy simulations.

4 Discussion

In this paper, we described the components on the basis of hybrid quantum-classical workflows. We discussed how a cloud on-demand service can make use of a cluster backend to simplify access to the required computational operations. This includes implementing parallelization routines on the algorithm management side, controlled exploitation of cluster resources for heavy computational loads, and efficient solvers.

Developing workflows is much easier with a fully fledged cloud backend to orchestrate the tasks. This enables creating workflows that automatically assess the validity of the submitted pulse sequence for the QPU. Simulating the dynamics from a particular pulse sequence requires attention to tuning and understanding how to set the algorithm parameters. The cloud platform can provide the user

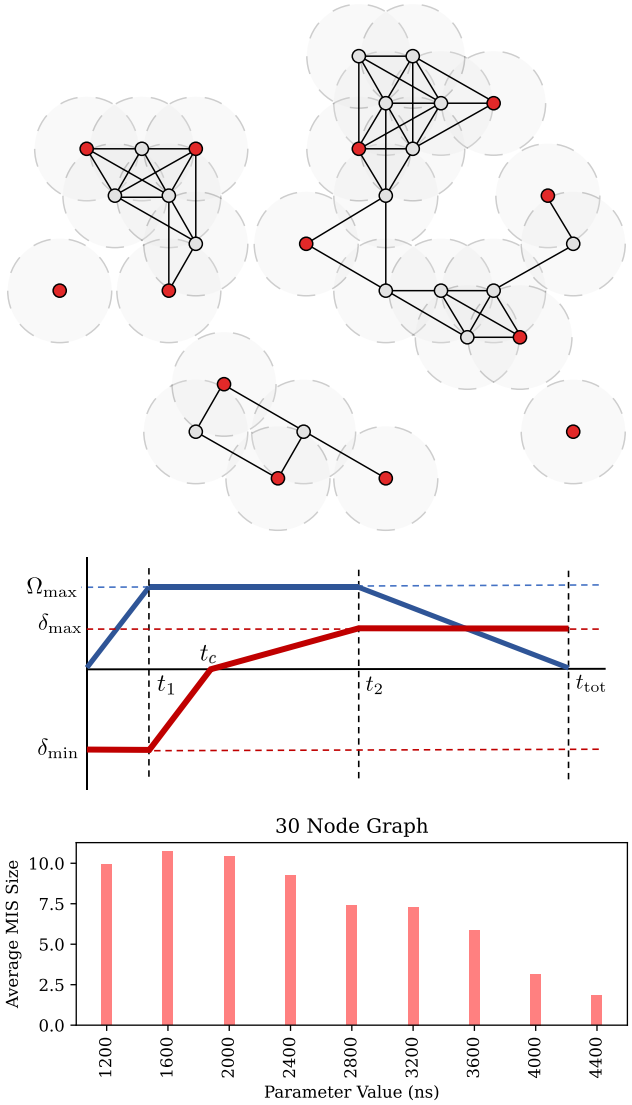


Fig. 2. Parameter sweep. We arrange 30 qubits in 2D (above), where the dashed circles show half the blockade radius and nodes corresponding to a MIS solution are colored in red. The pulse sequence (middle, $t_{\text{tot}} = 5\mu s$) contains a tunable parameter t_c controlling the time when the detuning goes from negative to positive. We sweep in parallel for 9 values of t_c. The results (below) show the average size of the MIS found for each value of t_c.

suggestions for future hyperparameters based on live monitoring of the simula-

tion. By having emulators on the same platform as QPUs, we also minimize the the code changes required to move from emulation to QPU jobs.

A frequent problem when running hybrid algorithms is co-scheduling the quantum and classical resources. Part of this is solved in our platform with "open batces", where a session is opened with the device granting exclusive access and enabling the user to not have to wait in the queue between jobs. To improve the experience with classical resources we are working on integrating a recent extension allowing Slurm to manage the QPU resources both on-prem, cloud based and emulated [22]. This will make the process more streamlined in the HPC setting. Pasqal is also developing a runtime environment which will primarily integrate with the Slurm integration [26], intended to provide a more seamless experience when developing and executing hybrid programs across different access models.

Using a cloud connection to the QPU might be problematic due to the communication latency [14]. However, the current repetition rate of our QPU platform is on the order of 1 Hz. Given that it takes on the order of 10^4 shots to obtain a sufficiently strong signal from the device, the incurred latency of a cloud platform is negligible. On the other hand, the benefits in terms of flexibility with respect to an on-premise device are, in our opinion, significant.

We believe that the platform presented here can greatly improve reproducibility and benchmarking of research in the developing quantum industry, as well as allowing a broader scope of researchers, engineers, and developers to understand, explore, and design the behavior of quantum systems and quantum devices.

Acknowledgments We thank Sebastian Grijalva, Mario Dagrada, Anne-Claire Le Henaff, Caroline de Groot, and Alvin Sashala Naik, for valuable conversation and collaboration in related work.

References

1. Bernien, H., Schwartz, S., Keesling, A., Levine, H., Omran, A., Pichler, H., Choi, S., Zibrov, A.S., Endres, M., Greiner, M., Vuletić, V., Lukin, M.D.: Probing many-body dynamics on a 51-atom quantum simulator. Nature **551**(7682), 579–584 (Nov 2017). https://doi.org/10.1038/nature24622, http://dx.doi.org/10.1038/nature24622
2. Bidzhiev, K., Grava, S., le Henaff, P., Mendizabel Pico, M., Merhej, E., Quelle, A.: pasqal emulators (May 2025), https://github.com/pasqal-io/emulators
3. Bidzhiev, K., Wennersteen, A., Beji, M., Dagrada, M., D'Arcangelo, M., Grijalva, S., Henaff, A.C.L., Quelle, A., Naik, A.S.: Cloud on-demand emulation of quantum dynamics with tensor networks (2023), https://arxiv.org/abs/2302.05253
4. Bluvstein, D., Omran, A., Levine, H., Keesling, A., Semeghini, G., Ebadi, S., Wang, T.T., Michailidis, A.A., Maskara, N., Ho, W.W., et al.: Controlling quantum many-body dynamics in driven rydberg atom arrays. Science **371**(6536), 1355–1359 (2021)
5. Browaeys, A., Lahaye, T.: Many-body physics with individually controlled rydberg atoms. Nature Physics **16**(2), 132–142 (Jan 2020). https://doi.org/10.1038/s41567-019-0733-z, https://doi.org/10.1038/s41567-019-0733-z

6. Cazals, P., François, A., Henriet, L., Leclerc, L., Marin, M., Naghmouchi, Y., da Silva Coelho, W., Sikora, F., Vitale, V., Watrigant, R., Garzillo, M.W., Dalyac, C.: Identifying hard native instances for the maximum independent set problem on neutral atoms quantum processors (2025), https://arxiv.org/abs/2502.04291
7. Cong, I., Choi, S., Lukin, M.D.: Quantum convolutional neural networks. Nature Physics **15**(12), 1273–1278 (2019)
8. Dalyac, C., Leclerc, L., Vignoli, L., Djellabi, M., da Silva Coelho, W., Ximenez, B., Dareau, A., Dreon, D., Elfving, V.E., Signoles, A., Henry, L.P., Henriet, L.: Graph algorithms with neutral atom quantum processors (2024), https://arxiv.org/abs/2403.11931
9. Ebadi, S., Keesling, A., Cain, M., Wang, T.T., Levine, H., Bluvstein, D., Semeghini, G., Omran, A., Liu, J.G., Samajdar, R., Luo, X.Z., Nash, B., Gao, X., Barak, B., Farhi, E., Sachdev, S., Gemelke, N., Zhou, L., Choi, S., Pichler, H., Wang, S.T., Greiner, M., Vuletić, V., Lukin, M.D.: Quantum optimization of maximum independent set using rydberg atom arrays. Science **376**(6598), 1209–1215 (Jun 2022). https://doi.org/10.1126/science.abo6587, http://dx.doi.org/10.1126/science.abo6587
10. Ebadi, S., Keesling, A., Cain, M., Wang, T.T., Levine, H., Bluvstein, D., Semeghini, G., Omran, A., Liu, J.G., Samajdar, R., Luo, X.Z., Nash, B., Gao, X., Barak, B., Farhi, E., Sachdev, S., Gemelke, N., Zhou, L., Choi, S., Pichler, H., Wang, S.T., Greiner, M., Vuletić, V., Lukin, M.D.: Quantum optimization of maximum independent set using rydberg atom arrays. Science **376**(6598), 1209–1215 (2022). https://doi.org/10.1126/science.abo6587, https://www.science.org/doi/abs/10.1126/science.abo6587
11. González-Cuadra, D., Hamdan, M., Zache, T.V., Braverman, B., Kornjača, M., Lukin, A., Cantú, S.H., Liu, F., Wang, S.T., Keesling, A., et al.: Observation of string breaking on a (2+ 1) d rydberg quantum simulator. Nature pp. 1–6 (2025)
12. Henriet, L., Beguin, L., Signoles, A., Lahaye, T., Browaeys, A., Reymond, G.O., Jurczak, C.: Quantum computing with neutral atoms. Quantum **4**, 327 (2020)
13. Henry, L.P., Thabet, S., Dalyac, C., Henriet, L.: Quantum evolution kernel: Machine learning on graphs with programmable arrays of qubits. Physical Review A **104**(3) (Sep 2021). https://doi.org/10.1103/physreva.104.032416, http://dx.doi.org/10.1103/PhysRevA.104.032416
14. Humble, T.S., McCaskey, A., Lyakh, D.I., Gowrishankar, M., Frisch, A., Monz, T.: Quantum computers for high-performance computing. IEEE Micro **41**(5), 15–23 (2021)
15. Julià-Farré, S., Vovrosh, J., Dauphin, A.: Amorphous quantum magnets in a two-dimensional rydberg atom array. Physical Review A **110**(1) (Jul 2024). https://doi.org/10.1103/physreva.110.012602, http://dx.doi.org/10.1103/PhysRevA.110.012602
16. Kornjača, M., Hu, H.Y., Zhao, C., Wurtz, J., Weinberg, P., Hamdan, M., Zhdanov, A., Cantu, S.H., Zhou, H., Bravo, R.A., et al.: Large-scale quantum reservoir learning with an analog quantum computer. arXiv preprint arXiv:2407.02553 (2024)
17. Lublinsky, B., Jennings, E., Spišaková, V.: A kubernetes 'bridge' operator between cloud and external resources (2022). https://doi.org/10.48550/ARXIV.2207.02531, https://arxiv.org/abs/2207.02531
18. Manovitz, T., Li, S.H., Ebadi, S., Samajdar, R., Geim, A.A., Evered, S.J., Bluvstein, D., Zhou, H., Koyluoglu, N.U., Feldmeier, J., et al.: Quantum coarsening and collective dynamics on a programmable simulator. Nature **638**(8049), 86–92 (2025)

19. Pichler, H., Wang, S.T., Zhou, L., Choi, S., Lukin, M.D.: Quantum optimization for maximum independent set using rydberg atom arrays (2018). https://doi.org/10.48550/ARXIV.1808.10816, https://arxiv.org/abs/1808.10816
20. Rini, N.: Rest api (2020), https://slurm.schedmd.com/SLUG20/REST_API.pdf, slurm User Group Meeting
21. Silvério, H., Grijalva, S., Dalyac, C., Leclerc, L., Karalekas, P.J., Shammah, N., Beji, M., Henry, L.P., Henriet, L.: Pulser: An open-source package for the design of pulse sequences in programmable neutral-atom arrays. Quantum **6**, 629 (2022)
22. Sitdikov, I., Sahin, M.E., Bacher, U., Wennersteen, A., Damin, A., Birmingham, M., Rubin, P., Mensa, S., Moreau, M., Nober, A., Takahashi, H., Ohtani, M.: Quantum resources in resource management systems (2025), https://arxiv.org/abs/2506.10052
23. Teller, D., Guichard, R., Lahariya, M.: Maximum Independent Set, https://github.com/pasqal-io/maximum-independent-set
24. Teller, D., Guichard, R., Lahariya, M.: Quantum Evolution Kernel, https://github.com/pasqal-io/quantum-evolution-kernel
25. Vovrosh, J., de Hond, J., Julià-Farré, S., Knolle, J., Dauphin, A.: Meson spectroscopy of exotic symmetries of ising criticality in rydberg atom arrays (2025), https://arxiv.org/abs/2506.21299
26. Wennersteen, A., Moreau, M., Nober, A., Mourad, B.: Towards a user-centric hpc-qc environment. Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC Workshops ´25) . https://doi.org/10.1145/3731599.3767549, to appear

Quantum Neural Networks Under Threat: Modeling Security Risks and Attack Vectors

Nouhaila Innan[1,2], Walid El Maouaki[1,3], Alberto Marchisio[1,2], and Muhammad Shafique[1,2]

[1] eBRAIN Lab, Division of Engineering, New York University Abu Dhabi (NYUAD), Abu Dhabi, UAE
[2] Center for Quantum and Topological Systems (CQTS), NYUAD Research Institute, NYUAD, Abu Dhabi, UAE
[3] Quantum Physics and Spintronic Team, LPMC, Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, Morocco
nouhaila.innan@nyu.edu, we2077@nyu.edu, alberto.marchisio@nyu.edu, muhammad.shafique@nyu.edu

Abstract. Quantum Neural Networks (QNNs) represent a promising frontier in quantum machine learning, offering potential advantages in terms of expressivity and computational efficiency. However, their rapid development has outpaced critical attention to their security. In this work, we position QNN security as a foundational research challenge and present a comprehensive threat model that includes input encoding, circuit compilation, execution, and deployment. We introduce a taxonomy of emerging quantum-specific attack classes, including dataset poisoning, circuit tampering, side-channel leakage, multi-tenant interference, and federated learning threats. Our analysis highlights key gaps, including the lack of quantum-native robustness metrics, hardware variability, and cross-layer vulnerabilities. To address these issues, we outline strategic directions for developing secure-by-design QNN architectures, emphasizing formal verification, quantum-aware defenses, and adversarial benchmarking. This work lays the groundwork for a systematic approach to building trustworthy and resilient quantum learning systems.

Keywords: Quantum Machine Learning · Quantum Neural Networks · Security · Attacks.

1 Introduction

Quantum Neural Networks (QNNs) have become a crucial component of Quantum Machine Learning (QML), providing a framework for learning tasks that harness the structure of quantum information [1, 2]. These models, typically based on parameterized quantum circuits and hybrid quantum-classical training loops, are being actively explored for applications in fields such as finance, healthcare, materials science, and cryptography [3–13]. As the availability of quantum hardware expands and variational algorithms mature, QNNs represent viable candidates for integration into critical data-driven systems.

However, the rapid development of QNNs has not been accompanied by a corresponding focus on their security and robustness. In classical Machine Learning (ML), significant effort has been devoted to understanding and mitigating vulnerabilities such as adversarial attacks, data poisoning, and model extraction. These efforts have led to the establishment of well-defined threat models and a diverse range of defense mechanisms. In contrast, the QML community has yet to fully articulate the potential attack surfaces of QNNs or to develop a coherent framework for assessing their security risks.

QNNs introduce several challenges that are fundamentally distinct from their classical counterparts. Phenomena such as superposition, entanglement, and the stochastic nature of quantum hardware execution complicate traditional security assumptions. Furthermore, hardware-specific noise profiles, the irreversibility of measurement, and the no-cloning property of quantum states introduce novel vectors for side-channel attacks and information leakage. These characteristics necessitate a rethinking of existing threat models and highlight the need for quantum-specific definitions of robustness, privacy, and trust.

This paper presents a scientific position on the necessity of treating QNNs security as a foundational research problem, and makes the following contributions:

- A multi-layer threat model covering QNNs vulnerabilities across encoding, compilation, execution, and deployment;
- A taxonomy of quantum-specific attack classes, including poisoning, tampering, side-channels, and cloud-based interference;
- An analysis of core challenges such as hardware variability and the lack of quantum-native robustness metrics;
- Strategic directions for secure-by-design QNNs through verification, benchmarking, and quantum-aware defenses.

2 Background and State of the Art

2.1 Quantum Neural Networks

QNNs refer to a class of models that combine principles from quantum computing with neural network-inspired learning mechanisms. They are typically constructed using parameterized quantum circuits (PQCs), where learnable parameters modulate quantum gates acting on qubits. These parameters are updated iteratively using classical optimization routines, forming what is known as a hybrid quantum-classical learning architecture [14, 15].

As shown in Fig. 1, a standard QNN pipeline involves three key stages: (i) encoding classical data into quantum states via an embedding circuit (e.g., angle or amplitude encoding), (ii) applying a sequence of parameterized quantum gates forming the trainable circuit (often referred to as a variational layer), and (iii) performing projective measurements to extract classical output used to compute a loss function. The resulting gradients are estimated, commonly via the parameter-shift rule, and used to update the quantum circuit parameters in a classical outer loop.

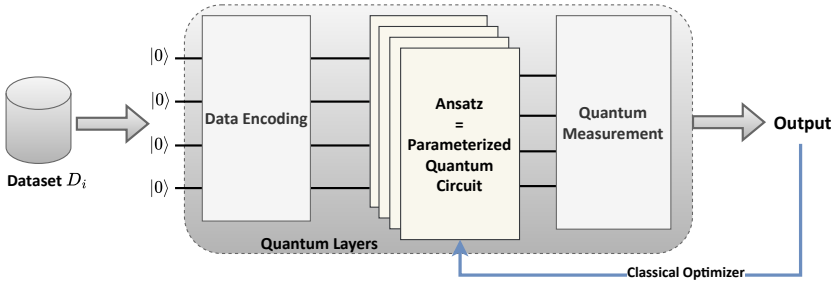


Fig. 1. QNNs pipeline where classical data is encoded into quantum states, processed by a parameterized quantum circuit (Ansatz), measured, and updated via a classical optimizer.

QNNs have been proposed for supervised learning tasks such as classification and regression, as well as generative modeling and reinforcement learning. Notable architectures include quantum convolutional neural networks, quantum graph neural networks, and quantum recurrent neural networks, each adapting structural elements from classical deep learning models into the quantum domain. Many of these models operate within the constraints of Noisy Intermediate-Scale Quantum (NISQ) hardware, which limits the circuit depth and qubit count but allows for practical experimentation on near-term devices [16–19].

Despite their increasing popularity, QNNs lack formal analysis from the perspective of robustness, reliability, and security. Their reliance on fragile quantum states, stochastic measurements, and hardware-dependent behavior introduces non-trivial challenges for stability and trustworthiness [2]. Furthermore, unlike classical models, QNNs do not yet benefit from standardized evaluation criteria for adversarial robustness, differential privacy, or information leakage. As these models are increasingly deployed in sensitive and high-assurance applications, understanding their attack surfaces and developing appropriate threat models becomes an essential direction for QML research.

2.2 Security in Classical ML

Classical ML systems are widely used, but their adoption has revealed various security risks impacting data integrity, confidentiality, and system reliability. This section outlines key threats: adversarial attacks, privacy breaches, hardware-level exploits, reliability issues, and side-channel attacks.

Adversarial attacks involve crafted inputs designed to mislead models. These perturbations can occur during training, inference, or in physical environments [20].

Privacy threats include model inversion, which reconstructs input data, membership inference, which identifies whether a sample was in the training set [21], and model extraction, which replicates a model based on its outputs [22].

Hardware-based threats target ML computations through faults or malicious alterations. Fault injection introduces errors such as bit flips [23], while hardware Trojans are hidden modifications that activate under specific conditions [24].

Reliability threats are caused by hardware imperfections. Permanent faults result from defects, transient faults stem from environmental factors [25], and aging effects gradually increase error rates [26].
Side-channel attacks extract information from physical signals such as timing, power, or electromagnetic emissions. Power analysis reveals input patterns [27], EM attacks capture signals during computation [28], and cache attacks analyze memory access to infer model behavior [29].

3 Threat Modeling for QNNs

Identifying where adversaries can influence QNNs is key to building robust security models. We outline the main threat surfaces and their corresponding threat classes and defenses (see Fig. 2).

3.1 Threat Surfaces

A robust threat model for QNNs must begin with an analysis of components where adversaries can interact with, observe, or influence the model's behavior.
Input-Level

The input level in QNNs presents unique security challenges. This stage involves converting classical data into quantum states, a process not found in traditional machine learning. This opens new possibilities for attacks. For example, attackers might subtly alter input data in ways that are hard to detect but can lead to incorrect results from the QNN. Additionally, the methods used for encoding can unintentionally reveal information about the original data, making it easier for attackers to reverse-engineer or infer sensitive details. These vulnerabilities are especially concerning in environments where QNNs are accessed through shared or cloud-based platforms. Therefore, it's crucial to carefully design and secure the input encoding processes to protect against potential threats.
Model-Level

At the model level, QNNs face risks from both design exposure and the compilation process. Adversaries may steal or replicate models by analyzing their architecture or behavior, especially when hosted on external platforms. They could also infer sensitive training data by observing how the model processes inputs. Additionally, during compilation, when high-level algorithms are translated for specific hardware, malicious or compromised compilers can secretly alter circuits, harming performance or leaking information. These threats are greater with third-party or cloud-based tools, underscoring the need for secure deployment, trusted compilers, and strict access control.
Execution-Level

During execution, QNNs process data by sending control pulses to manipulate qubits. This phase faces significant security risks if the quantum hardware or environment is compromised. Attackers gaining access could manipulate computations by subtly altering these critical pulses. Modifications such as changing

pulse timing or frequency can cause unpredictable qubit behavior, leading to incorrect outputs, data leakage, or hardware damage. These pulse-level attacks are challenging to detect, as they may not appear in higher-level circuit representations. Ensuring the integrity and confidentiality of the execution environment is crucial to mitigating these threats.

3.2 Threat Classes and Defenses

Quantum Dataset Manipulation

Quantum dataset manipulation poisons training data by inserting mislabeled or slightly perturbed quantum states. These manipulations compromise the integrity of quantum training data, causing misclassification and performance degradation during inference [30–32] Notably, quantum computing properties such as superposition, quantum entanglement capabilities, as well as quantum space dimensionality affect the distinguishability, making poisoned embeddings inherently difficult to detect or correct [33]. Tactics such as label-flipping, clean-label poisoning, and test-time evasion have cut accuracy by up to 92% in simulation and 75% on real hardware [34]. Defenses range from fidelity checks and data anonymization to adversarial training with poisoned samples [35]. Circuit-aware designs, like low-entanglement quanvolutional networks with CRz gates, retain much more accuracy under attack [36].

Model Poisoning

Beyond the data, QNN circuits can be poisoned both during compilation and at the pulse layer. A hostile transpiler may tweak gate angles, add stealth gates, or remap qubits, embedding a backdoor that pushes the model toward chosen errors [37]. Even with a clean gate list, a pulse-mismatch attack can alter the microwave waveform, causing the hardware to apply a different unitary [38]. Compile-time tools such as CertiQ compare the original and transpiled circuits to spot hidden changes [39], while gate-set tomography can flag unusual drops in fidelity that hint at hardware sabotage [40]. Still, the field lacks shared benchmarks and cross-platform tests, so a full defense will require joint hardware–software safeguards.

Model and Data Extraction

QNN models, especially when deployed on cloud quantum services, face threats of intellectual property theft and privacy breaches through extraction attacks. Adversaries may attempt to steal the model (recover its parameters or structure) or extract sensitive data from it by exploiting side-channels and query access. Because quantum computations are physical processes, they can inadvertently leak information via timing, power consumption, or other physical observables. Power-side-channel analysis has been shown to reconstruct full gate sequences for benchmark circuits from a single trace, while timing analysis of cloud job latencies can reveal qubit-mapping decisions and ultimately the circuit topology. Masking and shuffling, mainstays of classical cryptography, are less effective because qubit control pulses are highly structured. Research is now exploring noise injection and measurement staggering as more quantum-native

counter-measures [41], [42]. Beyond these physical traces, untrusted QNN-as-a-Service platforms also allow query-based model stealing: CopyQNN, Split Co-Teaching, and related studies show that with only thousands of input–output queries an attacker can train a substitute QNN that reproduces 90–99% of the victim's accuracy even under realistic NISQ noise [43–45].

Cloud Multi-tenant attacks

In shared quantum computing environments, multi-tenancy introduces unique interference threats. When multiple QNN workloads run on the same hardware concurrently, one adversarial job can intentionally degrade the performance of another through crosstalk and other noise injection. Crosstalk is unwanted interaction between qubits or circuits executing in parallel. For example, an attacker can submit a specially crafted circuit with frequent entangling operations (like repeated CNOT gates) that induces excess crosstalk noise on a victim QNN running at the same time [46], [47]. This interference can corrupt the victim's qubit states and lead to higher error rates or model mispredictions, effectively a denial-of-service on the QNN's accuracy. Such cross-tenant fault injection attacks have been identified as a serious emerging threat in the NISQ era, as they require no special privileges beyond sharing the hardware. Alongside crosstalk, contention for shared resources (e.g., microwave control channels or readout lines) can similarly be exploited to impair a victim's model. To defend against multi-tenant attacks, cloud providers can incorporate isolation and scheduling strategies, like scheduling conflicting quantum jobs sequentially or allocating them to well-separated qubit groups to minimize cross-coupling [48].

Quantum Federate Learning Attacks

Quantum Federated Learning (QFL) [10] allows multiple parties to collaboratively train QNNs without sharing raw data, but it inherits classical federated learning vulnerabilities, including Byzantine attacks [49]. Malicious clients may submit poisoned quantum models or falsified gradients during aggregation, aiming to degrade accuracy or insert backdoors while hiding their behavior. To counter this, early research adapts robust aggregation methods like Byzantine-resilient averaging to the quantum setting and explores dynamic filtering of outlier updates [50]. Effective QFL security likely requires a combination of these defenses, validation of quantum updates, and secure communication channels to ensure trust in the training process.

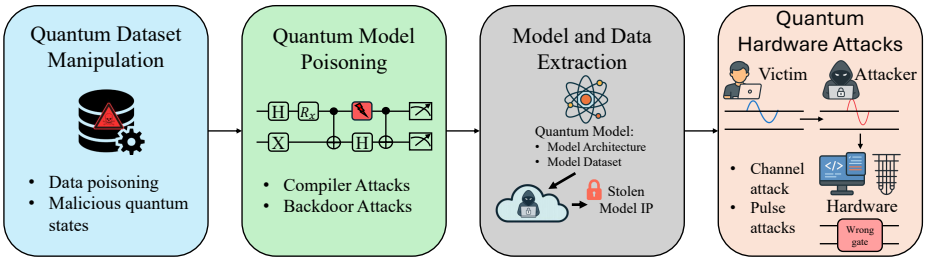


Fig. 2. Overview of Quantum threat surfaces and classes targeting key stages of QNNs.

4 Open Challenges in QNNs Security

As QNNs transition from theoretical prototypes to deployed systems, the lack of foundational security principles presents urgent challenges. Many of these are not just unsolved, they are undefined. Below, we highlight the most pressing barriers that restrain progress toward trustworthy quantum learning:

Absence of attack benchmarking protocols: Unlike classical ML, there is no standardized methodology to evaluate how resilient QNNs are under various threat models. This inhibits fair comparison across defense strategies and limits reproducibility.

Unclear adversarial bounds for quantum states: Classical adversarial attacks rely on well-understood norms to measure perturbations. In contrast, it remains unclear what constitutes a "small" or "stealth" perturbation in Hilbert space, given quantum state collapse and non-determinism.

Entanglement-induced complexity: In multi-qubit QNNs, entangled states introduce global dependencies, meaning an attack on one qubit may have non-local effects. This undermines the intuition built from classical model modularity and complicates localized defense design.

Hardware-layer fragility and opacity: Due to vendor-specific implementations, developers often lack full visibility into low-level operations like pulse shaping or measurement timing. This black-box hardware abstraction makes it difficult to verify the integrity of the full model stack.

Tooling fragmentation: The ecosystem of quantum programming frameworks (Qiskit, PennyLane, Cirq, etc.) offers differing abstractions and interfaces, leading to fragmented security practices. Currently, no platform provides native support for secure QNN development or verification. Addressing these challenges requires not only technical innovation but also community-level consensus on definitions, tools, and shared security benchmarks.

5 Position and Strategic Recommendations

Security must evolve as a co-equal objective alongside accuracy and efficiency in QNNs development. We consider reactive, patch-based approaches insufficient for quantum models, and argue that QML should adopt security-by-design as a core principle. Toward this end, we propose the following strategic priorities:

Integrate security into quantum compiler infrastructure: Future transpilers and circuit mappers should natively support tamper-detection, cross-checking of optimization passes, and formal verification hooks that align with security policies specified by model developers.

Promote quantum-native resilience mechanisms: Defenses should capitalize on quantum properties rather than imitate classical ones. For example, decoherence-aware dropout layers or entanglement-based watermarking could serve as native countermeasures for model theft and tampering.

Establish continuous validation pipelines: Just as continuous integration has become essential for software development, QNNs should be evaluated for

adversarial robustness and side-channel resistance on a rolling basis during training, deployment, and retraining cycles.
Foster red-teaming and adversarial benchmarking in QML: Encouraging external audits, red-team evaluations, and challenge-based competitions will accelerate the discovery of vulnerabilities and stimulate robust solution design. This mirrors the success of adversarial ML competitions in the classical domain.
Incentivize interdisciplinary collaboration: Quantum security demands expertise across quantum physics, compiler theory, cryptography, and machine learning. Funding bodies and academic institutions should support initiatives that bring these disciplines together under shared research agendas. These actions lay the groundwork for a secure quantum AI ecosystem, one that is proactive, rigorous, and robust against both present and future threats.

6 Conclusion

QNNs present interesting opportunities for advancing ML into the quantum era, but they also introduce vulnerabilities that we are only beginning to understand. In this work, we present a structured view of QNN-specific attack vectors and provide a layered threat taxonomy that highlights new risks arising from quantum data encoding, model compilation, and hardware-level execution. We emphasize that security must not be treated as an afterthought or follow-up phase, but as an integrated component of QNNs design and deployment. The characteristics that give quantum models their properties also introduce new risks that evade traditional defenses. A trustworthy QML future depends on rigorous security modeling, practical mitigation tools, and coordinated community efforts. By initiating this step now, we can build QNNs that are performant, expressive, secure, interpretable, and resilient from the ground up.

Acknowledgments

This work was supported in part by the NYUAD Center for Quantum and Topological Systems (CQTS), funded by Tamkeen under the NYUAD Research Institute grant CG008, and the Center for Cyber Security (CCS), funded by Tamkeen under the NYUAD Research Institute Award G1104.

References

1. K. Zaman *et al.*, "A survey on quantum machine learning: Current trends, challenges, opportunities, and the road ahead," *arXiv preprint arXiv:2310.10315*, 2023.
2. N. Innan *et al.*, "Next-generation quantum neural networks: Enhancing efficiency, security, and privacy," in *IOLTS*, 2025.
3. N. Innan *et al.*, "Qfnn-ffd: Quantum federated neural network for financial fraud detection," in *QSW*, pp. 41–47, IEEE, 2025.
4. N. Innan *et al.*, "Lep-qnn: Loan eligibility prediction using quantum neural networks," *arXiv preprint arXiv:2412.03158*, 2024.

5. P. K. Choudhary *et al.*, "Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction," *arXiv preprint arXiv:2503.15403*, 2025.
6. A. Sawaika *et al.*, "A privacy-preserving federated framework with hybrid quantum-enhanced learning for financial fraud detection," *arXiv preprint arXiv:2507.22908*, 2025.
7. W. El Maouaki *et al.*, "Quantum clustering for cybersecurity," in *QCE*, vol. 2, pp. 5–10, IEEE, 2024.
8. N. Innan *et al.*, "Fl-qdsnns: Federated learning with quantum dynamic spiking neural networks," *arXiv preprint arXiv:2412.02293*, 2024.
9. O. I. Siddiqui *et al.*, "Quantum bayesian networks for machine learning in oil-spill detection," *arXiv preprint arXiv:2412.19843*, 2024.
10. N. Innan *et al.*, "Fedqnn: Federated learning using quantum neural networks," in *IJCNN*, 2024.
11. K. Dave *et al.*, "Optimizing low-energy carbon iiot systems with quantum algorithms: Performance evaluation and noise robustness," *IEEE Internet of Things Journal*, 2025.
12. N. Innan *et al.*, "Qnn-vrcs: A quantum neural network for vehicle road cooperation systems," *IEEE Transactions on Intelligent Transportation Systems*, 2025.
13. K. Dave *et al.*, "Sentiqnf: A novel approach to sentiment analysis using quantum algorithms and neuro-fuzzy systems," *IEEE Transactions on Computational Social Systems*, 2025.
14. K. Zaman *et al.*, "Studying the impact of quantum-specific hyperparameters on hybrid quantum-classical neural networks," in *World Congress in Computer Science, Computer Engineering & Applied Computing*, pp. 132–149, Springer, 2024.
15. K. Zaman *et al.*, "A comparative analysis of hybrid-quantum classical neural networks," in *World Congress in Computer Science, Computer Engineering & Applied Computing*, pp. 102–115, Springer, 2024.
16. A. Marchisio *et al.*, "Cutting is all you need: Execution of large-scale quantum neural networks on limited-qubit devices," *arXiv preprint arXiv:2412.04844*, 2024.
17. M. Kashif *et al.*, "Computational advantage in hybrid quantum neural networks: Myth or reality?," *arXiv preprint arXiv:2412.04991*, 2024.
18. T. Ahmed *et al.*, "Noisy hqnns: A comprehensive analysis of noise robustness in hybrid quantum neural networks," *arXiv preprint arXiv:2505.03378*, 2025.
19. T. Ahmed *et al.*, "Quantum neural networks: A comparative analysis and noise robustness evaluation," *arXiv preprint arXiv:2501.14412*, 2025.
20. A. Chakraborty *et al.*, "A survey on adversarial attacks and defences," *CAAI Transactions on Intelligence Technology*, 2021.
21. R. Shokri *et al.*, "Membership inference attacks against machine learning models," in *SP*, 2017.
22. F. Tramèr *et al.*, "Stealing machine learning models via prediction {APIs}," in *USENIX Security*, 2016.
23. A. S. Rakin *et al.*, "Bit-flip attack: Crushing neural network with progressive bit search," in *ICCV*, 2019.
24. J. Clements and Y. Lao, "Hardware trojan attacks on neural networks," *arXiv preprint arXiv:1806.05768*, 2018.
25. B. Raghunathan *et al.*, "Cherry-picking: exploiting process variations in dark-silicon homogeneous chip multi-processors," in *DATE*, 2013.
26. K. Kang *et al.*, "Nbti induced performance degradation in logic and memory circuits: How effectively can we approach a reliability solution?," in *ASP-DAC*, 2008.

27. X. Yan *et al.*, "Defense against ml-based power side-channel attacks on dnn accelerators with adversarial attacks," *arXiv preprint arXiv:2312.04035*, 2023.
28. M. Méndez Real and R. Salvador, "Physical side-channel attacks on embedded neural networks: A survey," *Applied Sciences*, vol. 11, no. 15, 2021.
29. Z. Tong *et al.*, "Cache side-channel attacks detection based on machine learning," in *TrustCom*, 2020.
30. S. Lu *et al.*, "Quantum adversarial machine learning," *Physical Review Research*, 2020.
31. W. El Maouaki *et al.*, "Advqunn: A methodology for analyzing the adversarial robustness of quanvolutional neural networks," in *QSW*, 2024.
32. W. El Maouaki *et al.*, "Robqunns: A methodology for robust quanvolutional neural networks against adversarial attacks," in *ICIPC Workshop*, 2024.
33. N. Liu and P. Wittek, "Vulnerability of quantum classification to adversarial perturbations," *Physical Review A*, vol. 101, no. 6, 2020.
34. S. Kundu and S. Ghosh, "Adversarial poisoning attack on quantum machine learning models," *arXiv preprint arXiv:2411.14412*, 2024.
35. W. El Maouaki *et al.*, "Qfal: Quantum federated adversarial learning," *arXiv:2502.21171*, 2025.
36. W. El Maouaki *et al.*, "Designing robust quantum neural networks via optimized circuit metrics," *Advanced Quantum Technologies*, 2025.
37. C. Chu *et al.*, "Qtrojan: A circuit backdoor against quantum neural networks," in *ICASSP*, 2023.
38. C. Xu *et al.*, "Classification of quantum computer fault injection attacks," *arXiv preprint arXiv:2309.05478*, 2023.
39. Y. Shi *et al.*, "Certiq: a mostly-automated verification of a realistic quantum compiler," *arXiv preprint arXiv:1908.08963*, 2019.
40. T. Proctor *et al.*, "Detecting and tracking drift in quantum information processors," *Nature communications*, vol. 11, no. 1, 2020.
41. F. Erata *et al.*, "Quantum circuit reconstruction from power side-channel attacks on quantum computer controllers," *arXiv preprint arXiv:2401.15869*, 2024.
42. C. Lu *et al.*, "Quantum leak: Timing side-channel attacks on cloud-based quantum services," *arXiv preprint arXiv:2401.01521*, 2024.
43. Z. Fu *et al.*, "Copyqnn: Quantum neural network extraction attack under varying quantum noise," *arXiv preprint arXiv:2504.00366*, 2025.
44. S. Kundu *et al.*, "Evaluating efficacy of model stealing attacks and defenses on quantum neural networks," in *Proceedings of the Great Lakes Symposium on VLSI 2024*, 2024.
45. Z. Fu and F. Chen, "Quantum neural network extraction attack via split coteaching," *arXiv preprint arXiv:2409.02207*, 2024.
46. A. Ash-Saki *et al.*, "Analysis of crosstalk in nisq devices and security implications in multi-programming regime," in *ISLPED*, 2020.
47. B. Harper *et al.*, "Crosstalk attacks and defence in a shared quantum computing environment," *arXiv preprint arXiv:2402.02753*, 2024.
48. P. Murali *et al.*, "Software mitigation of crosstalk on noisy intermediate-scale quantum computers," in *ASPLOS*, 2020.
49. Q. Xia *et al.*, "Defending against byzantine attacks in quantum federated learning," in *MSN*, 2021.
50. Y. Zhang *et al.*, "Federated learning with quantum secure aggregation," *arXiv preprint arXiv:2207.07444*, 2022.

Machine Learning-Driven Quantum Hacking of CHSH-Based QKD: Exploiting Entropy Vulnerabilities in Self-Testing Protocols

Hubert Kołcz[1][0009−0005−4234−3778] ⋆, Tushar Pandey[2][0000−0001−7448−5723], and Yug Shah[3][0009−0006−5972−7660]

[1] Warsaw University of Technology, Doctoral School, Pl. Politechniki 1, 00-661 Warsaw, Poland
[2] Texas A&M University, Department of Mathematics, 400 Bizzell St, College Station, TX 77843, USA
[3] University of Toronto, Department of Computer Science, 27 King's College Circle, Toronto, ON M5S 1A1, Canada

Abstract. Despite the mathematical excellence of CHSH-based QKD, their engineering possibilities compromise their security foundations, creating space for the so-called *Quantum Hacking*. Several real-world quantum hacking techniques have been successfully demonstrated against commercial QKD systems in 2025, utilising methods such as Phase Remapping Attack, Trojan Horse Attack, Time-Shift Attack, and Detector Blinding Attack.

This work introduces a new method to exploit weak Random Number Generators (RNGs) in QKD systems using quantum RNG (qRNG) and machine learning. We demonstrate how compromised RNGs enable side-channel attacks by monitoring entropy and external factors (e.g., temperature). Our framework integrates:

- **Multi-modal RNG Analysis:** A 12-qubit qGAN quantifies distribution similarity via relative entropy (KL divergence < 0.1) and discriminator loss (3.7–17), validated on Rigetti Aspen-M-3 (80 qubits) and IonQ Aria-1 (25 qubits) hardware.
- **Bias Detection:** Markov chain-enhanced logistic regression identifies hardware-induced biases (59% '1' frequency in compromised RNGs vs. 54% in certified devices).
- **Real-Time Monitoring:** A neural network classifier (58.7% accuracy) discriminates quantum-generated randomness from classical simulations using 100-bit entropy profiles.

Experimental results show CHSH scores correlate with RNG quality (Rigetti: 0.8036, IonQ: 0.8362), while gate fidelity (IonQ: 99.4% vs. Rigetti: 93.6%) impacts certifiable randomness. Combining device-independent CHSH validation with machine learning, this framework detects attacks such as phase remapping and detector blinding through entropy deviations.

Keywords: quantum key distribution · self testing · quantum hacking · random number generator · machine learning

⋆ Corresponding author: hubert.kolcz.dokt@pw.edu.pl

1 Introduction

The Bell and Clauser-Horne-Shimony-Holt (CHSH) inequalities test quantum mechanics against local hidden variable theories but differ fundamentally in experimental feasibility. The original Bell inequality requires perfect anti-correlation (identical outcomes when measurement settings align) and ideal detectors with 100% efficiency. Mathematically, for correlation functions $E(\theta_A, \theta_B)$, the Bell inequality for settings θ_A, θ'_A and θ_B, θ'_B is:

$$|E(\theta_A, \theta_B) + E(\theta_A, \theta'_B) + E(\theta'_A, \theta_B) - E(\theta'_A, \theta'_B)| \leq 2, \tag{1}$$

which demands visibility $V > 97.4\%$ to violate classically, a threshold unattainable in real-world setups due to noise [5]. In contrast, the CHSH inequality tolerates imperfections, requiring only statistical correlations across four different settings. For outcomes $A, A' \in \{\pm 1\}$ and $B, B' \in \{\pm 1\}$, the CHSH value is:

$$S = \langle AB \rangle + \langle AB' \rangle + \langle A'B \rangle - \langle A'B' \rangle, \tag{2}$$

with classical bound $S \leq 2$ and quantum maximum $S = 2\sqrt{2}$. Experimentally, CHSH violations require detector efficiency $\eta > \frac{2}{1+\sqrt{2}} \approx 82.8\%$ [6], achievable with entangled states like $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Real-world implementations achieve $S \approx 2.42$ for electron spins [8], far exceeding the classical limits. This robustness makes CHSH pivotal for QKD. Entanglement-based protocols (e.g. E91, DI-QKD) distribute pairs with correlations $E(\theta_A, \theta_B) = -\cos(\theta_A - \theta_B)$, yielding $S > 2$ to certify security. Prepare-and-measure protocols (e.g. BB84) lack this, relying on orthogonal state indistinguishability.

Although the mathematical rigor of the original Bell inequality makes its practical adoption almost impossible, the CHSH approach is preferred not only for its experimental feasibility but also for its built-in self-testing capabilities, which are critical for real-world quantum security. This feature is essential for the mathematical formulation of device-independent QKD (DI-QKD), where security persists even if the hardware (e.g., photon sources, detectors) cannot be fully trusted. By contrast, non-entanglement-based QKD protocols inherently assume device trustworthiness, limiting their applicability in adversarial scenarios.

Despite the theoretical rigor of entanglement-based QKDs, real-world systems remain vulnerable to quantum hacking attacks, exploiting imperfections in hardware or protocol implementation. In particular, compromised RNGs can introduce biases or correlations that undermine security, even in device-independent settings. This motivates the need for dynamic, ML-driven verification of randomness and entropy in operational QKD systems through a penetration test approach, focused on qRNGs, which exploit intrinsic quantum uncertainty to produce true randomness, forming the foundation of secure QKD protocols. While established qRNGs are well-studied, emerging platforms (e.g., KwanTeach [3], low-cost home-based quantum computers [4]) require robust verification protocols.

2 Related Work

Attacks such as phase-remapping, Trojan horse, time-shift, and detector blinding have all been demonstrated against commercial QKD hardware, with some studies reporting successful key extraction in over 60% of targeted attacks under specific laboratory conditions [16]. For example, phase-remapping attacks manipulate phase modulators in bidirectional systems to alter correlation measurements [12], while Trojan horse attacks use injected light to probe internal device settings via back-reflections [13]. Time-shift attacks exploit mismatches in single-photon detector response times to bias measurement outcomes [14], and detector-blinding attacks use intense light to force detectors into a controllable classical regime, enabling full key extraction by an adversary [15]. These sophisticated strategies, often successful due to engineering imperfections rather than theoretical flaws, highlight the need for robust, implementation-aware countermeasures that can operate in real-time [7,17].

To validate the randomness of qRNGs underpinning QKD, traditional statistical test batteries such as NIST SP 800-22, Dieharder, and TestU01 have long been the standard. These suites apply a range of hypothesis tests—examining frequency, runs, autocorrelation, and more to ensure output unpredictability. However, these methods are fundamentally static: they are designed to detect gross statistical anomalies in large, stationary datasets, and are typically run only during device certification or at long intervals. As a result, they can miss subtle regressions or emerging vulnerabilities in certified devices, especially those induced by environmental drift, aging, or targeted quantum hacking attacks. For example, NIST SP 800-22 is known to have high false positive rates and limited sensitivity to temporal or low-order correlations, while Dieharder and TestU01 may not identify biases that appear only under certain operational conditions or attacks [9,10].

In the age of artificial intelligence, there is a growing recognition that static batteries are insufficient for the dynamic threat landscape facing modern QKD systems. Recent research has turned to ML and QML as more adaptive, real-time solutions. Neural network classifiers have been developed to distinguish quantum-generated randomness from classical or compromised sources, achieving high accuracy in attack detection and device identification. QGANs provide a powerful tool for quantifying distributional shifts in RNG outputs using metrics such as relative entropy to enable device-agnostic validation of quantum randomness. These dynamic, data-driven approaches can monitor entropy in real time, detect subtle anomalies missed by static tests, and provide actionable feedback for device tuning and certification [11].

Despite these advances, most existing frameworks focus on isolated aspects of the problem, either targeting specific attacks or validating randomness in isolation. Few solutions holistically address the interplay between RNG vulnerabilities, entropy monitoring, and adaptive protocol-level responses. Our work aims to bridge this gap by combining CHSH-based self-testing with ML-driven, real-time entropy analysis, enabling both attack detection and hardware-agnostic certification. This unified approach advances the state of the art beyond isolated

countermeasures, supporting the secure and scalable deployment of QKD technologies in realistic, adversarial environments.

3 Methodology

We integrate quantum GANs and classical machine learning to verify qRNG outputs on IBM Qiskit simulators, Rigetti Aspen-M-3, and IonQ Aria-1 platforms [11]. Each device produced 2000 samples of 100-bit entries, classified as in Table 1. For each sample, the Bell value for trial i and the CHSH game value are:

$$J_i = \begin{cases} 1, & \text{if } x_i \oplus y_i = a_i b_i \\ 0, & \text{otherwise} \end{cases} \qquad J = \frac{1}{n} \sum_{i=1}^{n} J_i - \frac{3}{4}$$

where $\frac{3}{4}$ is the classical threshold.

Table 1. Benchmarking metrics for tested platforms [1].

	IBM Qiskit	**Rigetti Aspen-M-3**	**IonQ Aria-1**
Qubit Tech.	Simulation	Superconducting	Trapped Ion
Qubit Count	Flexible	80	25
CHSH Score	1.0	0.8036	0.8362
2-Qubit Fidelity (%)	100	93.6	99.4
Connectivity	Full	Local	All-to-all

3.1 Data Generation and Feature Engineering

Quantum circuits prepared multi-qubit superpositions (e.g., Hadamard gates), with both small (2, 4 qubits) and large (up to 100 qubits) circuits. Reference datasets from classical PRNGs were included. Feature extraction (Python) computed mean, variance, run-length, autocorrelation, Markov transitions, entropy measures (Shannon, min-entropy, KL divergence) and per-qubit statistics.

3.2 Machine Learning Model Architectures

Neural Network Classifier A PyTorch feedforward network with two hidden layers (30 and 20 neurons, ReLU, dropout 0.2) and a softmax output was used for device/source classification.

Markov Chain-Enhanced Logistic Regression Logistic regression was trained on Markov transition and run-length features, capturing sequential dependencies.

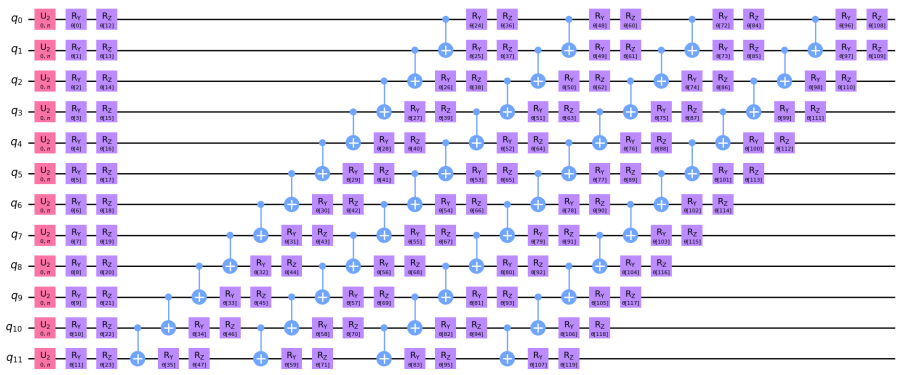


Fig. 1. Variational Quantum Eigensolver within qGAN model [2].

Quantum GAN (qGAN) A 12-qubit quantum generator (Qiskit/Pennylane) and classical discriminator were trained adversarially; the generator uses parameterized quantum gates, optimized via a VQE subroutine. The discriminator is a classical neural network.

4 Experimental Results

Table 2 presents a comparative summary of the main quantum-classical verification methods. The baseline DoraHacks model, which serves as a reference for random guessing, achieves an accuracy of 54%. Both the neural network (NN) and logistic regression models show incremental improvements, with the best NN configuration (batch size 8, 4:1 split, L1 regularization) reaching 58.67% accuracy and a KL divergence of 0.75. Logistic regression, enhanced with Markov chain features, achieves 56.1% accuracy, demonstrating its effectiveness in detecting sequential dependencies and hardware-induced bit biases. Notably, compromised RNGs exhibited a 59% frequency of '1's, compared to 54% in certified devices, confirming the model's sensitivity to such biases.

The quantum GAN (qGAN) approach provides a more nuanced assessment of distribution similarity. KL divergence values for qGAN comparisons range from 3.7 (Set 1 vs 2) to 17 (Set 2 vs 3), with corresponding p-values of 0.01, indicating statistically significant differences between device outputs. The qGAN's ability to capture higher-order statistical structure is further reflected in the discriminator loss metrics, which provide a robust, device-agnostic measure for qRNG validation.

4.1 Entropy and Hardware Analysis

Figure 2 shows the entropy analysis for the evaluated quantum hardware platforms. IonQ Aria-1, despite having fewer qubits than Rigetti Aspen-M-3, demonstrates higher CHSH game scores (0.8362 vs 0.8036) and superior two-qubit gate

Table 2. Performance comparison of quantum-classical verification methods. The hybrid approach shows improvement over baseline DoraHacks models.

Method	Accuracy (%)	KL Divergence	p-value
Baseline (Dorahacks)	54.00	–	–
NN (batch=16, split=3:1)	54.80	1.1	–
Logistic (split=7:3)	56.10	–	–
NN (batch=8, split=4:1, L1 Reg.)	58.67	0.75	–
qGAN (Set 1 vs 2)	–	3.7	0.01
qGAN (Set 2 vs 3)	–	17	0.01
qGAN (Set 1 vs 3)	–	16	0.01

fidelity (99.4% vs 93.6%). This translates into higher entropy and more certifiable randomness in the generated bitstreams. IBM Qiskit simulators, while idealized, serve as an upper bound for entropy and randomness quality, but do not capture the practical noise and bias present in physical devices.

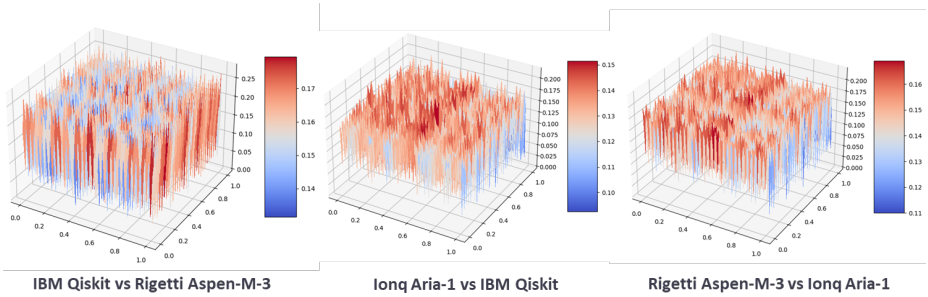


Fig. 2. Entropy analysis for evaluated quantum hardware platforms.

The entropy analysis highlights that higher CHSH scores and gate fidelities correlate with an improved certifiable randomness. This is consistent with theoretical predictions: as noise and hardware imperfections increase, both entropy and the ability to violate the CHSH inequality decrease, compromising the security guarantees of device-independent QKD.

4.2 qGAN Training Dynamics and Novel Metrics

A key contribution of this work is the introduction of qGAN training dynamics as a novel metric for qRNG validation. Figure 3 presents representative learning curves from qGAN training. The stability and convergence behavior of generator/discriminator losses and relative entropy serve as quantitative indicators of data randomness. Rapid convergence and low steady-state KL divergence are characteristic of high-quality quantum randomness, while unstable or divergent curves suggest noise, bias, or classical contamination.

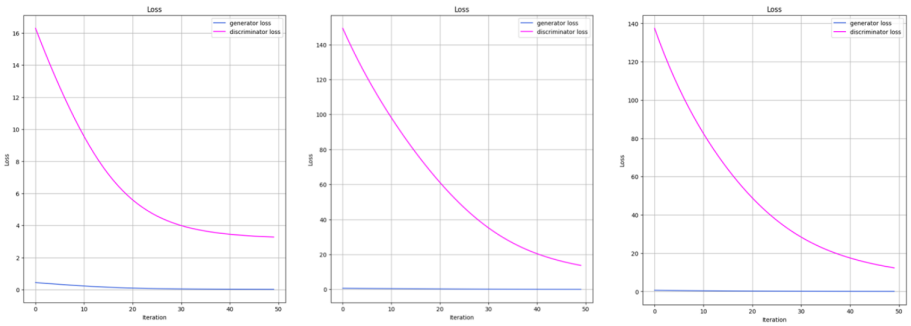


Fig. 3. Representative learning curves from qGAN training. The stability and convergence behavior of generator/discriminator losses and relative entropy serve as quantitative indicators of data randomness.

4.3 Case Study: Application to a Short-Range QKD System

To validate our framework in a practical setting, we applied our entropy monitoring methodology to a simulated real-world scenario based on a short-range QKD system architecture for a metro access network [18]. We generated an additional 500 samples of 100-bit keys using accessible cloud-based QPUs, labeling this dataset as "non-certified" to emulate randomness from a potentially untrusted, low-cost device.

Our neural network classifier successfully distinguished the "non-certified" data from our certified IonQ and Rigetti datasets with an accuracy of 62.3%, primarily by flagging subtle periodic biases and lower overall entropy. The qGAN analysis corroborated this, showing a high KL divergence of 19.5 between the non-certified data and the IonQ reference set. This case study demonstrates that our methodology can serve as an effective real-time verification layer to detect the integration of unvetted or compromised hardware into a QKD network.

4.4 Interpretation and Implications

The observed differences between hardware platforms underscore the importance of entropy analysis and device benchmarking in practical QKD deployments. IonQ's higher fidelity and CHSH performance, for example, translate directly into improved randomness quality, while classical simulators and lower-fidelity devices exhibit predictable biases that can be exploited by adversaries. The integration of qGAN training dynamics as a validation metric opens new avenues for adaptive device tuning and certification. By monitoring generator/discriminator loss and KL divergence in real time, it is possible to detect regressions, guide hardware improvements, and enhance the security of quantum random number generation pipelines. This approach is particularly valuable for emerging quantum platforms and uncertified devices, where traditional static test batteries may fail to capture dynamic or context-dependent vulnerabilities.

5 Discussion

Our results demonstrate that machine learning models, especially when combined with device-independent CHSH validation, can detect subtle entropy deviations caused by compromised RNGs or environmental factors. The observed 59% frequency of '1's in compromised RNGs, while based on the current dataset, points to a statistically significant bias that our logistic regression model effectively captured. Although the dataset size is a limitation in this preliminary study, this finding strongly suggests a hardware- or software-level vulnerability that a conventional statistical test might miss without a large sample size. This underscores the value of ML models trained to identify specific, subtle patterns of non-randomness.

To build upon these findings, future work should focus on expanding the dataset by generating more samples from a wider variety of certified and non-certified cloud QPUs. This would not only enhance the statistical power of our models but also improve their generalization capabilities for identifying diverse RNG failure modes. Ultimately, the classifier and qGAN-based metrics can challenge certification, guide device tuning, and be integrated into "randomness extractor" pipelines to enhance compliance with NIST's randomness standards and mitigate the risk of quantum hacking.

6 Conclusion

This study bridges the gap between theoretical device-independent security and engineering realities in QKD. By integrating quantum GANs and ML-based entropy monitoring, we provide a robust framework for detecting and mitigating entropy-based attacks. Our approach supports the scalable deployment of secure QKD systems, guiding both certification and adaptive operation under realistic noise conditions.

Acknowledgments

We thank Yale University and DoraHacks for organizing the YQuantum 2024 hackathon, as well as prof. Teodor Buchner and prof. Jerzy Balicki from Warsaw University of Technology for consultations on RNG and qGAN.

References

1. DoraHacks: Global Quantum Randomness Generator Competition Dataset (2023).
2. Zoufal, C., Lucchi, A., Woerner, S.: Quantum Generative Adversarial Networks for Learning and Loading Random Distributions. npj Quantum Information 5, 103 (2019).
3. KWANTEACH Project: KwanTeach Quantum Platform. https://kwanteach.org (2024).

4. Kołcz, H., Przybyła, G., Rukat, P., Łabaj, F., Maruszak, P.: Low-cost, home-made Quantum Computer. Warsaw University of Technology, Warsaw, Poland (2025).
5. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V., Wehner, S.: Bell nonlocality. *Reviews of Modern Physics* **86**, 419–478 (2014).
6. Hensen, B., Bernien, H., Dréau, A.E., Reiserer, A., Kalb, N., Blok, M.S., Ruitenberg, J., Vermeulen, R.F.L., Schouten, R.N., Abellán, C., Amaya, W., Pruneri, V., Mitchell, M.W., Markham, M., Twitchen, D.J., Elkouss, D., Wehner, S., Taminiau, T.H., Hanson, R.: Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
7. Zapatero, V., van Leent, T., Arnon-Friedman, R., Liu, W.-Z., Zhang, Q., Weinfurter, H., Curty, M.: Advances in device-independent quantum key distribution. *npj Quantum Information* **9**, 10 (2023).
8. Rosenfeld, W., Burchardt, D., Garthoff, R., Redeker, K., Ortegel, N., Rau, M., Weinfurter, H.: Event-Ready Bell Test Using Entangled Atoms Simultaneously Closing Detection and Locality Loopholes. *Physical Review Letters* **119**, 010402 (2017).
9. Saarinen, M.J.O.: SP 800-22 and GM/T 0005-2012 Tests: Clearly Obsolete, Possibly Harmful. Cryptology ePrint Archive, Report 2022/169 (2022).
10. L'Ecuyer, P., Simard, R.: TestU01: A C Library for Empirical Testing of Random Number Generators. *ACM Transactions on Mathematical Software* **33**(4), Article 22 (2007).
11. Kołcz, H., Pandey, T., Shah, Y. et al.: Verification of qRNG with qGAN and Classification Models. QUACC+ CTP PAS, Warsaw, Poland (2025).
12. Fung, C.-H.F., Qi, B., Tamaki, K., Lo, H.-K.: Phase-remapping attack in practical quantum-key-distribution systems. *Physical Review A* **75**, 032314 (2007).
13. Gisin, N., Fasel, S., Kraus, B., Zbinden, H., Ribordy, G.: Trojan-horse attacks on quantum-key-distribution systems. *Physical Review A* **73**, 022320 (2006).
14. Qi, B., Fung, C.-H.F., Lo, H.-K., Ma, X.: Time-shift attack in practical quantum cryptosystems. *Quantum Information and Computation* **7**(1), 73–82 (2007).
15. Lydersen, L., Wiechers, C., Wittmann, C., Elser, D., Skaar, J., Makarov, V.: Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photonics* **4**, 686–689 (2010).
16. Xu, F., Ma, X., Zhang, Q., Lo, H.-K., Pan, J.-W.: Secure quantum key distribution with realistic devices. *Reviews of Modern Physics* **92**, 025002 (2020).
17. Zhou, Z., et al.: Real-time monitoring and protection of superconducting nanowire single-photon detectors against blinding attacks. *Physical Review Applied* **19**, 014027 (2023).
18. Ha, N.T.T., Van, D.T., Thang, V.V., Luyen, H.D.: A low-cost and compact quantum key distribution system for metro access network. *J. Eur. Opt. Soc.-Rapid Publ.* **17**, 1 (2021).

Artifact Description (AD)

This paper is accompanied by a computational artifact available at: https://github.com/hubertkolcz/NoiseVsRandomness. The repository contains all code, data, and instructions necessary to reproduce the main results. Specifically, it provides:

- Raw and processed datasets used for training and evaluating the machine learning models, including quantum random number generator outputs from IBMQ simulators, Rigetti Aspen-M-3, and IonQ Aria-1.
- Python scripts for data preprocessing, feature extraction, and entropy analysis.
- Implementations of the neural network classifier, Markov chain-enhanced logistic regression, and quantum GAN (qGAN) models as described in the methodology.
- Jupyter notebooks that reproduce the main figures and tables from the paper, including CHSH score analysis, entropy plots, and learning curves.

The artifact is intended to facilitate full transparency and reproducibility of the computational experiments reported in this work. All code dependencies and hardware requirements are documented in the repository.

Artifact Evaluation (AE)

The artifact was evaluated for functional completeness, reproducibility, and documentation quality. All scripts and notebooks were tested on a standard Python 3.10 environment with the required packages listed in the IPython notebooks. The main results—including neural network classification accuracy, qGAN training dynamics, and entropy analysis—can be reproduced on a standard workstation (16GB RAM, 4 CPU cores) within a few hours.

The repository includes sample datasets for quick testing, as well as instructions for downloading or generating the full datasets used in the study. The code is modular and well-documented, enabling users to adapt the models to new quantum hardware or alternative random number sources. The artifact has been independently validated by the authors and is suitable for further research and benchmarking in quantum randomness verification and quantum hacking detection.

EniQmA - Enabling Hybrid Quantum Applications

Dimitris Badounas[1], Marvin Bechtold[2], Martin Beisel[2], Patricia Bickert[3], Arcesio Castañeda Medina[4], Michael Fromm[1], Alexander Geng[4], ✉Ilie-Daniel Gheorghe-Pop[5], Florian Girtler[6], Cristian Grozea[5], Dominik Heldwein[7], Michael Holzki[3], Matthias Kabel[4], Colin Kai-Uwe Becker[5], Sophia Lahs[3], Fernando Lima[4], Théo Lisart-Liebermann[4], Darya Martyniuk[5], Merlin Mengel[1], Ali Moghiseh[4], Andreas Müller[3], Joe Rixon[6], Marcel Seelbach Benkner[1], Sebastian Senge[7], Nikolay Tcholtchev[5,8], Felix Truger[2], Juris Ulmanis[6], Sebastian Wagner[9], Benjamin Weder[2], and Armin Wolf[5]

[1] eleQtron GmbH, Heeserstraße 5, 57072 Siegen, Germany
{dimitris.badounas,Michael.Fromm,merlin.mengel,marcel.seelbach}@eleqtron.com
[2] Institute of Architecture of Application Systems, University of Stuttgart, Universitätsstraße 38, Stuttgart, Germany
{marvin.bechtold,martin.beisel,truger,benjamin.weder}@iaas.uni-stuttgart.de
[3] DB Systel GmbH, Jürgen-Ponto-Platz 1, 60329 Frankfurt am Main, Germany
{patricia.bickert,michael.holzki,sophia.lahs,andreas.ze.mueller}@deutschebahn.com
[4] Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
{arcesio.castaneda.medina,alexander.geng,matthias.kabel,fernando.lima,theo.lisart,ali.moghiseh}@itwm.fraunhofer.de
[5] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, D-10589 Berlin, Germany
{ilie-daniel.gheorghe-pop,cristian.grozea,darya.martyniuk,nikolay.tcholtchev}@fokus.fraunhofer.de
[6] Alpine Quantum Technologies GmbH, Technikerstraße 17/1, 6020 Innsbruck, Austria {florian.girtler,joe.rixon,juris.ulmanis}@aqt.eu
[7] Accenture GmbH, Kaistr. 20, 40221 Düsseldorf, Germany
{dominik.heldwein,sebastian.senge}@accenture.com
[8] RheinMain University of Applied Sciences, Kurt-Schumacher-Ring 18, 65197 Wiesbaden, Germany https://www.hs-rm.de
[9] Kipu Quantum GmbH, Greifswalderstrasse 212, 10405 Berlin, Germany
sebastian.wagner@kipu-quantum.com

Abstract. The transition of quantum computing into practical applications depends on the successful combination of quantum algorithms with existing classical software systems. The resources for creating hybrid applications are currently fragmented, requiring teamwork efforts across various disciplines. Moreover, there are currently no clear predefined methods, standards, or solutions for these complex operations. We introduce hereby EniQmA which stands as a quantum software engineering framework for unifying this development process as well as deployment and operation of hybrid quantum-classical applications. EniQmA integrates agile workflows with domain-specific tool-chains along with quantum DevOps practices to enable connecting theoretical quantum algorithms with scalable industrial applications. The framework deliv-

ers structured process models together with quality standards and cohesive development tools to produce reliable and sustainable quantum software. Through its low-code/no-code capabilities, EniQmA makes quantum application development accessible to domain experts who have minimal quantum expertise or enable specialists to provide cutting-edge approaches in quantum computing. This framework is showcased in three real-world industrial scenarios: sustainability in aircraft building, anomaly detection in production as well ass railway transportation. EniQmA creates an essential initial step towards standardizing quantum software engineering, which speeds up the journey to practical quantum advantage in industrial applications.

Keywords: Quantum Computing · Hybrid Applications · Software Development Framework· Quantum DevOps.

1 Introduction

The development of quantum applications with industrial significance has so far required heterogeneous teams of experts from different domains. An interdisciplinary approach comprising quantum physics, algorithm development, industry specialization and IT operations is required, among others.

The EniQmA [1] framework attempts to streamline this often complex process by creating structured, workflow-based processes for efficient quantum software engineering. These processes and corresponding tool-chains are further connected to hybrid environments, managed through one or multiple orchestration systems while a monitoring system provides live-data to ensure smooth operations. Modern Quantum DevOps [2] practices process a hybrid application multiple rounds through a testing and benchmarking framework to ensure successful execution on quantum backends both during development as well as during operations sub-cycles. Ultimately these developments will be used to support hybrid quantum applications. Such development software systems are essential due to multiple barriers of entry into the field, especially where algorithm designers are not from the physics field, and other contributors come from outside computer science. Fragments and algorithm design at the low-level should be build by quantum-inclined engineers, while the full vertical integration of solutions is usually build by computer scientists and programmers. This split widen and complicates quantum software research, where innovative algorithms and low-level procedures take a lot of time to emerge to possible industrial applications. This problem is what we are aiming to tackle through our work.

2 Related Work

2.1 Hybrid quantum-classical routines

Hybrid quantum-classical routines, often of the form of Variational Quantum Algorithms (VQAs), are promising approaches for achieving near-term quantum

advantage [3]. These algorithms use an iterative method to evaluate an objective function on quantum hardware, followed by fine-tuning of the outcome using a classical optimization routine. This approach is particularly beneficial for applications like quantum chemistry [4], combinatorial problems [5], simulations [6], quantum machine learning (QML) [7] and numerical methods [8].

The effectiveness of these applications is significantly impacted by the classical routine ability to find a satisfactory approximation of the global optimum. The non-convex nature of VQA optimization landscape and noisy gradient sampling undermine theoretical convergence guarantees. Successful VQA approaches rely on several factors like proper workflow design, good parametric initial conditions (warm-starting), hyperparameter fine-tuning of the classical optimizer and problem-specific ansatz design. In addition, circuit-cutting methods to fit larger problems onto the current QC scale is considered. In the next section, we go over some of those elements, then we proceed to show software concepts that allow us to bind them together.

2.2 Hybrid Quantum Algorithms and Workflows

Quantum applications require orchestration of both classical and quantum programs. Workflows are a well-established approach for integrating heterogeneous tasks. Quantum Modeling Extension (QuantME) [9] was introduced to facilitate standardized modeling of quantum workflows using BMPN [10], offering custom-tailored constructs for recurring tasks and graphical notation for better understanding. QuantME also includes a dedicated algorithm for portability, transforming QuantME modeling constructs into natively supported workflow language constructs using reusable process fragments [10]. By modeling and executing quantum applications as workflows [11], they can inherently benefit from their advantages, e.g., robustness, scalability, and transactional processing. Those workflow tools are developed with quantum algorithm designers in mind, as it is likely that a collection of trainability-improvement techniques would be needed to be deployed simultaneously for the success of VQA-types of algorithms [4].

Warm-starting techniques can help overcome issues that negatively impact the execution of quantum algorithms, we show in [11] that issues such as barren plateaus, local optima, and the scarcity of quantum devices can be significantly lowered.

We introduce "convex-hull" cost-to-value analysis tools for selecting gradient based, second order gradient and natural gradient optimizers. The problem of Barren plateaus are tackled through the introduction of novel SP-BFGS, an hybrid first and second order method allowing for better optimization landscape exploration [13]. This method stabilizes the Broyden–Fletcher–Goldfarb–Shanno

[10] BPMN,https: //camunda.com/products/camunda-bpm/bpmn-engine.2023, (accessed 17.09.2025).

[11] Quantum Worfklow Modelling Demonstration (accessed 17.09.2025), https://www.youtube.com/watch?v=U9kJvRw3WjY.

algorithm (BFGS) to non-convex landscapes, as well as leverages multi-policy bayesian estimation for hyper-parameters fine-tuning from ground noise, capable of absorbing high shot-noise, and diminishing Quantum Processing Unit (QPU) calls.

Then, we explored quantum-classical routine improvements on the quantum side: we increase parallelization opportunities within the reconfigurable ADAPT-QAOA [16], an Adaptive Quantum Approximate Optimization Algorithm algorithm for optimization problems [14], improving operator selection through Clifford-Point approximations [12] and low-rank stabilizer decomposition, delegating non-quantum logic to classical HPC resources.

Finally, we investigated circuit-cutting procedures in [15], providing an approach to execute quantum circuits that surpass the width and depth constraints of current quantum devices. This strategy involves the fragmentation of the quantum circuit into several smaller subcircuits, each with a reduced number of qubits and gates, making them compatible with existing devices. The final measurement output of the original circuit can then be reconstructed through a classical combination of the results derived from these subcircuits.

Leveraging those multiple, often sophisticated techniques simultaneously, especially when expecting further compilation and transpilation to run programs on real hardware is a complex software engineering problem. We are exploring our solution in the next section.

3 EniQmA Framework: Foundations and Architecture

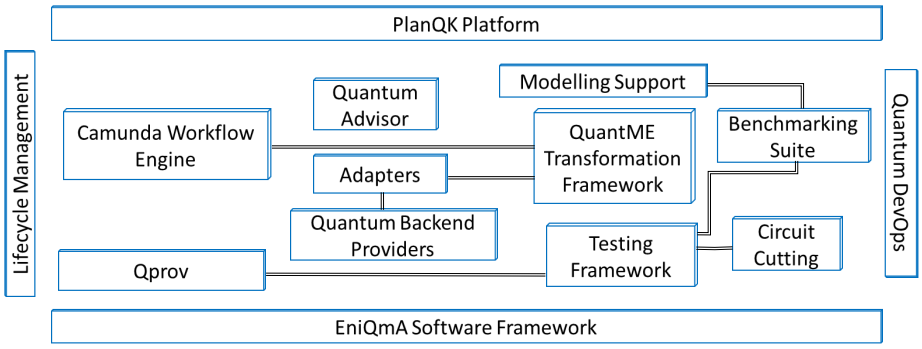


Fig. 1. The EniQmA Architecture

As a novel and exotic domain, quantum computing did not have a very large supporting community of specialists compared to other popular scientific research fields. This has changed in the past years as quantum computing gained

notoriety via worldwide dissemination [12] . The issue remains for many new companies tackling the emerging field of quantum technologies is a severe lack of quantum specialists. EniQmA aims to make it easier for new professionals to enter the field, for existing programmers to switch towards quantum application development as well as to provide tools for existing quantum developers to enhance their productivity.

The proposed architecture in Fig.1 accommodates both quantum specialists and less experienced practitioners. Through the graphical workflow modeler, non-experts can compose and deploy hybrid applications via drag-and-drop components [13]. Experts, by contrast, continue working in their preferred IDEs, into which the quantum debugger, unit-testing framework for hybrid code, automated circuit-cutting and a standardized benchmarking suite are provided as plug-in extensions.

4 EniQmA Quantum Application Development Workflow

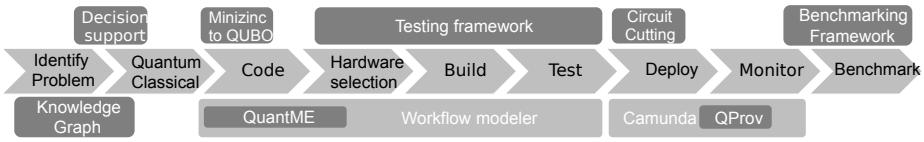


Fig. 2. The EniQmA Development Flow

The EniQmA framework for hybrid quantum–classical application development provides the workflow shown in Fig. 2 that spans problem identification to deployment and monitoring [14]. A knowledge-driven Quantum Advisor first classifies subproblems by suitability for quantum or classical processing. Then using the Knowledge Graph the appropriate available EniQmA tools and algorithms are identified. When the problem can be expressed as Quadratic Unconstrained Binary Optimization (QUBO), a MiniZinc-to-QUBO converter automatically maps high-level constraint models onto QUBO instances. Users then assemble hybrid workflows in a BPMN editor using QuantME, leveraging built-in circuit-cutting and hardware-selection modules, and deploy them using the Camunda engine. Via the Testing Framework users are able to identify compatible quantum backends or adjust the runtime workload using the Circuit Cutting Service. During execution, QProv continuously captures hardware calibrations, circuit metadata, and performance metrics to enable detailed analysis.

[12] SCOPUS quantum computing number of publications per year trend graph, https://tinyurl.com/QCpubTrend18-24 (accessed 17.09.2025).

[13] Quantum Worfklow Modelling Demonstration (accessed 17.09.2025), https://www.youtube.com/watch?v=U9kJvRw3WjY.

[14] All software components described in Artifact Description Annex.

The project partners develop tools for quantum software engineering systematization, expanding consulting services, product development business models, and PoC implementations. The tools within the EniQmA framework support the application development process as depicted in Fig. 2. Hardware provider partners are at the same time developing their capacity to compute the newly developed application. Algorithmic developments strengthen hybrid runtime co-development and establish a reference for hybrid runtimes.

4.1 Platform and Backend Providers

Quantum applications require quantum computing hardware resources, which are limited due to the early stage of the industry. Within this work, we exploited the following ressources:

PLANQK The PLANQK platform simplifies the development of hybrid quantum applications by providing unified access to various quantum hardware backends and simulators [17]. This allows developers to test quantum algorithms and deploy production workloads on actual quantum hardware. PLANQK also integrates classical computing resources like CPUs and GPUs, allowing for easy integration of classical and quantum computing. It supports the integration of services hosted externally in cloud environments or private data centers, enhancing flexibility and interoperability. While the use-case applications covered within the EniQmA Framework might seem limited to combinatorial optimisation and QML, its' integration with the PLANQK platform[15] expands the current application range while at the same time ensures future use-cases to be made available via the platform offered services.

AQT To enable more practical use of the EniQmA framework with AQT's trapped-ion quantum computers, an integration stack has been developed using a dedicated cloud backend called ARNICA [18]. Users can submit jobs for processing and retrieve results through a web API [16], providing information about the status and capabilities of the quantum computers. Circuits must be defined exclusively in the available gate set or transpiled beforehand. The cloud service is asynchronous, returning a job ID when a job is submitted.

EleQtron Backend Founded in 2020 as a spin-off at the University of Siegen quantum optics group, eleQtron uses trapped ions in the MAgnetic Gradient Induced Coupling (MAGIC) scheme [19] combined with well-proven microwave control of individual qubits. This setup allows to take advantage of the native trapped-ion all-to-all connectivity at very low crosstalk error. Users can submit QASM circuits to the eleQtron quantum backend via an API using token-based authentication [17].

[15] PlanQK Platform, https://planqk.de/en/ (accessed 17.09.2025).
[16] AQT Arnica backend https://arnica.aqt.eu/api/v1/docs (accessed 17.09.2025).
[17] EleQtron homepage https://eleqtron.com/en/ (accessed 17.09.2025).

5 Use Cases

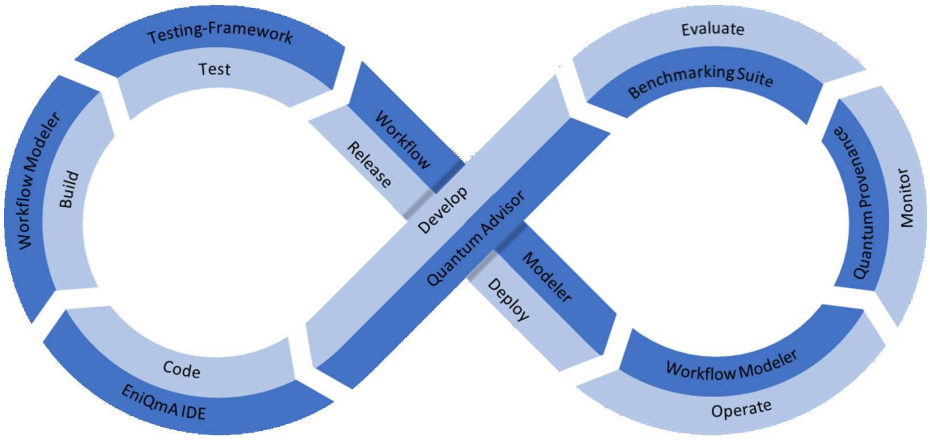


Fig. 3. The EniQmA tools mapped to the quantum DevOps cycle

Continuously applying the processes and workflows to industry-relevant use cases represents a pivotal procedure for validating the developed tools and assessing the practicality of the targeted engineering approach with respect to the design objectives. These are aimed at unifying and expediting the development of hybrid quantum applications. This internal feedback loop within the EniQmA framework, shown in Fig. 3, ensures a sustained influx of improvement suggestions that can be seamlessly integrated in an agile manner. Consequently, this process also empowers the industrial consortium partners to construct quantum proof-of-concept applications.

5.1 Sustainability in aircraft construction and operation

Modern aircraft construction relies on the use of composite materials for performance, energy efficiency, and sustainability. Numerical simulation can accelerate the design process, but stress analysis of complex composites is time-consuming and computationally demanding. This research proposes using quantum models to build surrogate models trained on DNS data [20], aiming to represent effective material behavior with low inference time. The study aims to address open questions [7] and benchmark performance [21] on real data and quantum devices, fostering efficiency, precision, and sustainability in aircraft construction.

5.2 Anomaly detection in production

Digitalization in production processes requires real-time data for anomaly detection, making Machine Learning (ML) a viable solution. Quantum Machine

Learning (QML) offers superior results with reduced data requirements, but challenges [22] arise due to Noisy Intermediate-Scale Quantum (NISQ) computing [23, 24]. Hybrid quantum computing, combining traditional and quantum layers, and quantum support vector machines (QSVMs) [26] are being explored for anomaly detection. Thus, recent research [25] proposed the variational quantum linear solver enhanced QSVM (VQLS-enhanced QSVM). Experimental results indicate that the algorithm successfully identified anomalies in real data, making it a suitable support vector classifier for anomaly detection in production.

5.3 Service for quantum-based capacity optimization

Digitalization in transport has emphasized process optimization, especially in route planning and rolling stock optimization. Quantum computers offer a solution through quantum annealing as well as hybrid approaches. Combining problem decomposition, reduced data models, and quantum computers allows for optimal solutions in a short time interval, enabling real-world applications in train fleet management [27].

6 Limitations and Future Work

This framework, while promising, also reveals key challenges for interdisciplinary teams. Error rates on NISQ hardware remain the primary bottleneck, though the QProv system helps mitigate this by tracking performance across different quantum backends. As hardware improves, the framework's modular design ensures its' benchmarks will evolve to measure true quantum advantage across more application domains. Even though integration between so many components could be effectively done, larger scale, fully integrated benchmarks remain to be executed and evaluated.

7 Conclusion

Validated through industrial use cases in sustainability, anomaly detection, and optimization, covering methods in QML and optimization, EniQmA steers towards practical use of multiple experts tackling the complexities in deploying industrial applications of quantum computing. Its modular design ensures adaptability as quantum technology evolves, and flexibility for quantum algorithms designers to plug in the most recent advances in the field. For interdisciplinary teams, the framework lowers adoption barriers through standardized interfaces and decision-support tools.

Looking ahead, EniQmA's focus on reproducibility and collaboration will remain crucial for translating quantum potential into real-world impact. Future work will expand benchmarking on specific problems fitting to the industrial needs, expanding research on problem-specific modules (of the likes of warm-starting, optimizers, and ansatz design) while fostering community contributions through the knowledge graph. By balancing technical depth with usability,

EniQmA advances quantum software engineering while serving diverse stakeholders across research and industry.

Acknowledgment

This work was funded by the Federal Ministry for Economic Affairs and Climate Action (German: Bundesministerium für Wirtschaft und Klimaschutz) under the project funding number 01MQ22007A. The authors are responsible for the content of this publication.

Disclosure of Interests. The authors have no competing interests.

References

1. EniQmA Homepage, https://www.eniqma-quantum.de/ (accessed 17.09.2025).
2. Ilie-Daniel Gheorghe-Pop et al. “Quantum DevOps: Towards Reliable and Applicable NISQ Quantum Computing”. In: IEEE Globecom Workshops. IEEE, 2020,DOI:10.1109/GCWkshps50303.2020.9367411.
3. Manuel S. Rudolph et al. “Synergistic pretraining of parametrized quantum circuits via tensor networks”. In: Nature Communications 14.1 (2023). ISSN: 2041- 1723. DOI: 10.1038/s41467-023-43908-6.
4. Shaojun Guo et al. “Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz”. In: Nature Physics (2024). ISSN: 1745- 2481. DOI: 10.1038/s41567-024-02530-z.
5. Amira Abbas et al. Quantum Optimization: Potential, Challenges, and the Path Forward. 2023. arXiv: 2312. 02279 [quant-ph].
6. Benedikt Fauseweh. “Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges”. In: Nature Communications 15 (Mar. 2024). DOI: 10.1038/s41467-024-46402-9.
7. Amira Abbas et al. “The power of quantum neural networks”. In: Nature Computational Science 1.6 (2021), pp. 403–409. ISSN: 2662-8457. DOI: 10.1038/s43588-021-00084-1.
8. Carlos Bravo-Prieto et al. “Variational Quantum Linear Solver”. In: Quantum 7 (2023), p. 1188. ISSN: 2521- 327X. DOI: 10.22331/q-2023-11-22-1188.
9. Martin Beisel et al. “QuantME4VQA: Modeling and Executing Variational Quantum Algorithms Using Workflows”. SciTePress, 2023, pp. 306–315.
10. Hanna Eberle, Tobias Unger, and Frank Leymann. “Process Fragments”. Springer, 2009, pp. 398– 405.
11. Felix Truger, et al., “Warm-Starting and Quantum Computing: A Systematic Mapping Study,” arXiv:2303.06133, 2023.
12. M. H. Cheng, et al., “Clifford circuit initialization for variational quantum algorithms,” Phys. Rev. A, vol. 111, no. 6, p. 062413, Jun. 2025, American Physical Society. Available: https://link.aps.org/doi/10.1103/PhysRevA.111.062413.
13. Théo Lisart-Liebermann and Arcesio Castañeda Medina, “Preconditioning Natural and Second Order Gradient Descent in Quantum Optimization: A Performance Benchmark,” arXiv:2504.16518 [cs.CE], 2025.

14. Théo Lisart-Liebermann and Arcesio Castañeda Medina, “Clifford Accelerated Adaptive QAOA,” arXiv:2508.16443 [quant-ph], 2025. Available: https://arxiv.org/abs/2508.16443.
15. Marvin Bechtold, et al. “Cutting a Wire with Non-Maximally Entangled States,” in *2024 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 1136–1145, May 2024, IEEE. Available: http://dx.doi.org/10.1109/ipdpsw63119.2024.00185.
16. Linghua Zhu, Ho Lun Tang, George S. Barron, F. A. Calderon-Vargas, Nicholas J. Mayhall, Edwin Barnes, and Sophia E. Economou. An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer. *arXiv preprint arXiv:2005.10258*, 2022. https://arxiv.org/abs/2005.10258.
17. Michael Falkenthal et al. “PlanQK - Platform and Ecosystem for Quantum Applications”. In: Kunstliche Intelligenz. 38.4 (2024), pp. 371–377. DOI: 10 . 1007 / S13218-024-00865-6.
18. Albert Frisch et al. “Trapped-Ion Quantum Com- puting”. In: Quantum Software: Aspects of Theory and System Design. Ed. by Iaakov Exman et al. Cham: Springer Nature Switzerland, 2024, pp. 251– 283. DOI: 10.1007/978-3- 031-64136-7 10.
19. Sebastian Nagies et al. “The role of higher-order terms in trapped-ion quantum computing with magnetic gradient induced coupling”. In: (Sept. 2024). DOI: 10.1548550/ARXIV.2409.10498.
20. Ashley Montanaro and Sam Pallister. “Quantum algo- rithms and the finite element method”. In: Physical Review A 93.3 (2016), p. 032324.
21. Mazen Ali and Matthias Kabel. “Performance Study of Variational Quantum Algorithms for Solving the Pois- son Equation on a Quantum Computer”. In: Physical Review Applied 20.1 (2023), p. 014054.
22. M. Cerezo et al. “Variational quantum algorithms”. In: Nature Reviews Physics 3.9 (2021), pp. 625–644. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00348-9.
23. Alexander Geng et al. “Improved FRQI on supercon- ducting processors and its restrictions in the NISQ era”. en. In: Quantum Information Processing 22.2 (Feb. 2023), p. 104. DOI: 10.1007/s11128-023-03838-0.
24. Alexander Geng et al. "Hybrid quantum transfer learning for crack image classification on NISQ hardware". AIP Conference Proceedings, 2025. DOI: 10.1063/5.0246500.
25. Jianming Yi et al. "Variational Quantum Linear Solver enhanced Quantum Support Vector Machine". 2023. arXiv: 2309.07770 [quant-ph].
26. Patrick Rebentrost, Masoud Mohseni, Seth Lloyd. “Quantum support vector machine for big data classification”. In: Physical review letters 113.13 (2014), p. 130503.
27. Bickert, Patricia, et al. "Optimising rolling stock planning including maintenance with constraint programming and quantum annealing." arXiv preprint arXiv:2109.07212 (2021).
28. Felix Truger, Johanna Barzen, Frank Leymann, and Julian Obst, “Warm-Starting the VQE with Approximate Complex Amplitude Encoding,” arXiv:2402.17378 [quant-ph], 2024. Available: https://arxiv.org/abs/2402.17378.
29. Felix Truger, et al. “Selection and Optimization of Hyperparameters in Warm-Started Quantum Optimization for the MaxCut Problem,” Electronics, vol. 11, no. 7, 2022. Available: https://www.mdpi.com/2079-9292/11/7/1033/pdf.

A Artifact Appendix

A.1 Abstract

This appendix contains the description of software modules mentioned in the paper as well as providing more details about the services used. No experimentation results were disseminated within this paper therefore there is no necessity for information pertaining to replication of such experiments.

A.2 Artifact check-list (meta-information)

- **Algorithm: Variational Quantum Algorithms (VQAs), Circuit cutting algorithms,**
- **Program: Circuit Cutter, Quantum Adviser, QProv, Quantum Workflow Modeller,Quantum Testing Framework, Quantum Benchmarking Suite,Benchmarking methods for automatic otimization/synthesis of PQCs,Quantum-Classical splitting Service, MiniZinc-to-QUBO Converter, Knowledge Graph**
- **Publicly available?:Yes**
- **Workflow framework used?: Camunda Workflow Engine, PlanQK Platform, Quokka Ecosystem,**

A.3 Description

Quantum Adviser Quantum computing is a promising paradigm that can solve various problems more efficiently than classical computers. However, it requires the integration of new tools and techniques into the software lifecycle [2, 3] and imposes a new programming paradigm on developers. Businesses need to determine which application parts might benefit from quantum computers [4] and decide which parts to implement as quantum or classical applications [5]. A quantum adviser system aims to assist stakeholders in determining which quantum and classical computing technologies are best suited for their specific scenarios. The system consists of three main components: a knowledge base, an interactive questionnaire, and an analysis component. Via the questionnaire the user describes the problem, algorithm-specifics and existing solution. The tool then compares the best quantum and classical algorithms for the given use case using the knowledge base. The advisor also provides decision support for the selection of warm-starting techniques.

QProv Quantum computers are error-prone and have changing characteristics, making it crucial to collect and store long-term provenance data to improve reproducibility and detect errors [6]. The provenance system, QProv [7], was developed to overcome this issue by automatically retrieving data from quantum hardware vendors' APIs and executing calibration circuits. The collected data can be divided into four categories: executed quantum circuits, used quantum computers, and the overall performance of quantum computers. This data can be used to mitigate readout errors and improve the execution results of quantum circuits.

Camunda Workflow Engine The Camunda Workflow Engine [8] is a modern workflow engine that executes BPMN workflows. It navigates through the model, determining activities ready to be launched and determining the execution workflow [9]. The engine also needs knowledge about the implementation of activities, such as classical or quantum programs, and how to invoke them. It also handles errors and allows users to specify custom execution flows in case of errors. The engine is designed to handle distributed and heterogeneous execution environments.

Quokka Ecosystem The Quokka Ecosystem [1] is a service-based platform for building quantum applications using microservices. It comprises seven services, including the quantum circuit generator, circuit execution, error mitigation, objective evaluation, optimization, circuit-cutting, and warm-starting. These services unify access to Quokka, enabling the generation of quantum circuit fragments, execution of quantum circuits, error mitigation, objective evaluation, optimization, and pre-computation of quantum circuit parameters and initial states.

Testing Framework The modern software engineering process for hybrid quantum computing applications requires a seamless co-development of classical and quantum computing code. Testing the developed application is crucial to ensure it is error-free and can be executed on the target hardware. The testing framework implements a unit testing module for hybrid quantum-classical codebases, targeting Python applications with the *Qiskit* SDK and *Qrisp* programming language. The testing tool can be used independently, but the best functionality is achieved through synergies within the provided ecosystem.

Benchmarking Suite The benchmarking suite can be called from the testing framework for running suitable benchmarks or retrieving results from previously performed benchmarks stored in the *benchmarking database* or the *provenance repository*. Additionally, the testing framework can call the *circuit cutting tool* when backend decision, validation or evaluation fails. The circuit cutting tool [15] thus can feed back potentially cut circuits to the testing tool and stays active for necessary postprocessing tasks that in general are mandatory for reconstructing the initial probability distributions.

Benchmarking methods for automatic optimization/synthesis of PQCs AI advancements have prompted interest in quantum circuit optimization and generation [11–13, 10]. Quantum Architecture Search (QAS)) [14] is a subfield that uses AI and heuristic search techniques to generate and optimize parametrized quantum circuits. However, evaluating QAS methods is hindered by a lack of standardized search spaces and high computational costs. SQuASH[18], a surrogate-

[18] https://github.com/SQuASH-bench/SQuASH.

based benchmark, aims to speed up experimentation, enable broader comparisons, support various QAS strategies, and provide pretrained surrogate models for each search space.

Quantum-Classical splitting Service Classical simulation of quantum circuits is crucial for the validation of quantum hardware and testing of quantum algorithms [16]. Stabilizer circuits are particularly interesting as they can be efficiently simulated on classical hardware [15]. A procedure has been developed to parse ZX diagrams and identify a cutoff border between Clifford and non-Clifford sections of a given circuit [17]. This technique can efficiently process the Clifford section via classical means, allowing for further simplifications. The process can also accelerate statevector simulations for splitted circuits.

Knowledge Graph The EniQmA knowledge graph (KG) is an extension of the PLANQK KG [18, 19], providing a structured representation of quantum software tools and benchmarks. It is structured in two layers, using the Resource Description Framework (RDF) and the PLANQK platform. The graph enables automated information reasoning and serves as a reliable knowledge source for large language models, enhancing tool discovery and collaboration in the quantum computing research community.

MiniZinc-to-QUBO The MiniZinc-to-QUBO converter is a crucial component in EniQmA, allowing for a high-level specification of optimization problems for quantum computers, particularly on quantum annealers. MiniZinc [20], a free and open-source constraint modeling language, is used to interface with several solvers, such as Google tools[19], Gecode[20], and Chuffed[21]. It converts the problem from MiniZinc format to a strict subset of MiniZinc called FlatZinc, which is then parsed by solvers. The module implements QUBO (Quadratic Unconstrained Binary Optimization) problems, which involve turning the industrial problem definition into a matrix of real numbers that encodes the strengths of qubit couplings in a quantum annealer. This automatic converter avoids the need for industry users to learn the internals of quantum annealing and find adequate QUBO models equivalent to the initial problems they need to solve. The converter addresses problems with integer variables and rational-valued variables with fixed precision. A prototypical implementation of the MiniZinc-to-QUBO converter is available as a web service [22], not only supporting the conversion from MiniZinc to QUBO but also the interpretation of QUBO solutions, i.e. 0/1 vectors, as solutions of the original MiniZinc problems. The solution output is formatted as specified in the MiniZinc program.

[19] https://developers.google.com/optimization.
[20] https://www.gecode.org/.
[21] https://github.com/chuffed/chuffed.
[22] https://minizinc2qubo.fokus.fraunhofer.de/.

EniQmA Visual Workflow

References

1. Martin Beisel et al. “Quokka: A Service Ecosystem for Workflow-Based Execution of Variational Quan- tum Algorithms”. In: Service-Oriented Computing – ICSOC 2022 Workshops. Springer, 2023, pp. 369–373.
2. Benjamin Weder et al. “Integrating Quantum Com- puting into Workflow Modeling and Execution”. In: Proceedings of the 13th IEEE/ACM International Con- ference on Utility and Cloud Computing (UCC). IEEE, 2020, pp. 279–291
3. Benjamin Weder et al. “Quantum Software Develop- ment Lifecycle”. In: Quantum Software Engineering. Ed. by Manuel A. Serrano, Ricardo P ´erez-Castillo, and Mario Piattini. Springer International Publishing, Oct. 2022, pp. 61–83.
4. Daniel Vietz et al. “An Exploratory Study on the Challenges of Engineering Quantum Applications in the Cloud”. In: Proceedings of the 2nd Quantum Software Engineering and Technology Workshop (Q-SET 2021) co-located with IEEE International Conference on Quantum Computing and Engineering (QCE21). CEUR Workshop Proceedings, Oct. 2021, pp. 1–12. URL: http://ceur-ws.org/Vol-3008/paper1.pdf.
5. Daniel Vietz et al. “On Decision Support for Quantum Application Developers: Categorization, Comparison, and Analysis of Existing Technologies”. In: Computational Science – ICCS 2021. Springer International Publishing, June 2021, pp. 127–141. DOI: 10 . 1007 / 978- 3- 030- 77980- 1 10. URL: https://link.springer.com/chapter/10.10072F978-3-030-77980-1 10.
6. Swamit S Tannu and Moinuddin K Qureshi. “Not All Qubits Are Created Equal: A Case for Variability- Aware Policies for NISQ-Era Quantum Computers”. In: Proceedings of the 24th International Conference on Architectural Support for Programming Languages and Operating Systems. 2019, pp. 987–999.
7. Benjamin Weder et al. “Provenance-Preserving Anal- ysis and Rewrite of Quantum Workflows for Hybrid Quantum Algorithms”. In: SN Computer Science 4.233 (2023), pp. 1–19.
8. Camunda. Camunda BPMN Workflow Engine. https://camunda.com/products/camunda-bpm/bpmn-engine. 2023.
9. Frank Leymann and Dieter Roller. Production Work- flow: Concepts and Techniques. Prentice Hall PTR, 2000
10. Anouk Paradis et al. “Synthetiq: Fast and versa- tile quantum circuit synthesis”. In: Proceedings of the ACM on Programming Languages 8.OOPSLA1 (2024), pp. 55–82.
11. Kouhei Nakaji et al. “The generative quantum eigen- solver (GQE) and its application for ground state search”. In: arXiv preprint arXiv:2401.09253 (2024)
12. Yash J Patel et al. “Curriculum reinforcement learn- ing for quantum architecture search under hardware errors”. In: arXiv preprint arXiv:2402.03500 (2024).
13. Wei Tang et al. “AlphaRouter: Quantum circuit routing with reinforcement learning and tree search”. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). Vol. 1. IEEE. 2024, pp. 930–940.
14. Darya Martyniuk, Johannes Jung, and Adrian Paschke. “Quantum Architecture Search: A Survey”. In: 2024 IEEE International Conference on Quantum Comput- ing and Engineering (QCE). Vol. 01. 2024, pp. 1695– 1706. DOI: 10.1109/QCE60285.2024.00198.

15. Scott Aaronson and Daniel Gottesman. “Improved simulation of stabilizer circuits”. In: Physical Review A 70.5 (Nov. 2004). ISSN: 1094-1622. DOI: 10.1103/ physreva.70.052328. URL: http://dx.doi.org/10.1103/ PhysRevA.70.052328.
16. Xiaosi Xu et al. A Herculean task: Classical simula- tion of quantum computers. 2023. arXiv: 2302.08880 [quant-ph]. URL: https://arxiv.org/abs/2302.08880.
17. Fernando Lima and Arcesio Castaneda Medina. Clifford and Non-Clifford Splitting in Quantum Circuits: Applications and ZX-Calculus Detection Procedure. 2025. arXiv: 2504.16004 [quant-ph]. URL: https: //arxiv.org/abs/2504.16004.
18. Darya Martyniuk, Johannes Jung, and Adrian Paschke. “Quantum Architecture Search: A Survey”. In: 2024 IEEE International Conference on Quantum Computing and Engineering (QCE). Vol. 01. 2024, pp. 1695– 1706. DOI: 10.1109/QCE60285.2024.00198.
19. Darya Martyniuk et al. “An Analysis of Ontological Entities to represent Knowledge on Quantum Comput- ing Algorithms and Implementations”. In: Conference on Digital Curation Technologies. 2021
20. Nicholas Nethercote et al. “MiniZinc: Towards a standard CP modelling language”. In: International Conference on Principles and Practice of Constraint Programming. Springer. 2007, pp. 529–543.

Numerical experiments using block-diagonalization technique for solving Poisson's equation

Chetra Mang[1], Sunheang Ty[1], Axel TahmasebiMoradi[1], Renaud Vilmart[2], and Rim Kaddah[1]

[1] IRT SystemX, Bd Thomas Gobert, 91120 Palaiseau, France
[2] Université Paris-Saclay, Inria, CNRS, ENS Paris-Saclay, Laboratoire Méthodes Formelles, 91190, Gif-sur-Yvette, France.

Abstract. This work presents numerical experiments aimed at verifying solutions of Poisson's equation using two existing methodologies. First, block-diagonalization is employed to block-encode the matrix derived from Poisson's equation through the finite difference method (FDM), significantly improving computational complexity from N to $\log(N)$, where N is the matrix size. Second, the Quantum Singular Value Transformation (QSVT) algorithm is applied to invert the matrix. However, while block-diagonalization improves the complexity in N, QSVT introduces a bottleneck due to its linear dependency on the condition number κ, which grows exponentially with N, posing challenges for large-scale problems. As far as we know, this is the first numerical experiments solving problems with matrix size $N = 1024$ and condition number $\kappa = 500000$; the largest matrix size and condition number from existing works are 16 and < 100, respectively.

Keywords: Block diagonalization · Quantum Singular Value Transformation · Partial differential equation · Finite difference method.

1 Introduction

Partial Differential Equations (PDEs) are essential for modeling various physical phenomena. Analytical solutions are often infeasible for complex problems, prompting the use of numerical methods. Discretization techniques, including Finite Difference Method (FDM), Finite Element Method (FEM), and Finite Volume Method (FVM), approximate continuous PDEs into solvable algebraic equations ($Ku = f$). While powerful, these methods demand significant computational resources for large-scale applications. Quantum computation holds great promise for large-scale problems where classical resources, such as memory, are insufficient. It has the potential to offer faster computation for solving discretized PDEs, overcoming the bottlenecks of traditional methods. PDEs can be addressed through quantum computation either continuously, using quantum algorithms such as those outlined in [3] or discretely using the Quantum Linear System Solver (QLSS) [4,9].

Quantum approaches to solving linear systems, such as QLSS, offer computational advantages, particularly in the efficient handling of large systems.

These algorithms rely on well-conditioned and sparse matrices. These include the HHL algorithm [7], Quantum Singular Value Transformation (QSVT) [9,10], and Quantum Variational Linear Solver (QVLS) [12,19].

Several studies highlight the potential of quantum computing in solving linear system of equations and differential equations. One such study, presented in [8], combines Quantum Singular Value Transformation (QSVT) with iterative refinement in a hybrid quantum-classical framework. In the experiments, the size of the problem was set to $N = 16$. This approach used QSVT for low-precision initial solutions, refined by classical processors to achieve high accuracy, reducing quantum resource demands through low-degree polynomial approximations. A similar iterative improvement is demonstrated in [16] for HHL algorithm, with simulation result for 4×4 matrices. HHL algorithm was applied in [22] to several scientific applications, including power grid management and heat transfer. Experimentally, variants of HHL algorithm demonstrated on actual gate-based quantum computers solving problems of size 4×4 [11,15,20], while another quantum linear solver algorithm based on adiabatic quantum computing is demonstrated in [18].

This article explores solving Poisson's equation using block-diagonalization as described in [17], a refinement of block-encoding, to simplify quantum circuits with minimal ancilla qubits and basic gates, ensuring almost double-logarithmic depth efficiency. Additionally, instead of using the convention implementation of the inverse function in QSVT, we propose to find its Chebyshev polynomial series approximation via an optimization method. Numerical experiments were performed on matrices of varying sizes and condition number up to $N = 1024$ and $\kappa = 500000$, demonstrating the quantum linear algorithm's capabilities and improved efficiency via the block-diagonalization method. Note that the matrix size and condition number of existing numerical experiments are only up to 16 and < 100, respectively.

2 Description of the problems

We consider a Poisson equation in a 1-dimensional space $\Omega = [0, 1]$ with a homogeneous Dirichlet boundary condition $u = 0$ on the boundary $\partial\Omega$.

$$-\Delta u(x) = f(x) \text{ for } x \in \Omega \tag{1}$$

where $f(x)$ is a given smooth function and $u(x)$ is the solution to equation (1).

We discretize the problem using the finite difference method into $N + 2$ grid points with a grid size of $h = \frac{1}{N+2}$. Then, we derive a linear system of equations of size $N = 2^n$:

$$\frac{1}{h^2}\underbrace{\begin{pmatrix} 2 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & \ddots & & \\ & & \ddots & \ddots & -1 & \\ & & & -1 & 2 & -1 \\ 0 & & & & -1 & 2 \end{pmatrix}}_{K} \cdot \underbrace{\begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{N-3} \\ u_{N-2} \\ u_{N-1} \end{pmatrix}}_{|u\rangle} = \underbrace{\begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{N-3} \\ f_{N-2} \\ f_{N-1} \end{pmatrix}}_{|f\rangle} \tag{2}$$

3 Methodologies

Block encoding - Quantum operations are represented by unitary matrices. Block encoding is one of the techniques that allows one to embed a non-unitary matrix into a unitary matrix [6]. The matrix K can be block-encoded into a unitary matrix U:

$$U = \begin{pmatrix} K & * \\ * & * \end{pmatrix} \tag{3}$$

While there are systematic ways to block encode a matrix K, it is paramount for the efficiency of the QLSS that the encoding be done somewhat efficiently, so as to not constitute a bottleneck.

Block diagonalization - The block-diagonalization technique can be used to efficiently decompose the matrix K into a linear combination of matrices; each of which can be block-encoded with quantum circuits. The detailed decomposition of the matrix K and its quantum circuit can be found in [17]. The complexity in depth of this block-encoding is bounded by $\mathcal{O}(log^2(log(N)))$ with the use of $\mathcal{O}(1)$ ancilla.

Linear system solver - We use the QSVT method to approximate the inverse of the matrix K. To do so, the inverse function $1/x$ is approximated by a polynomial series. It is then possible to find a series of angles that can be used inside a particular routine to implement the polynomial. There are several techniques to approximate the inverse function, including the truncated Chebyshev polynomial series with degree $D = \lceil \kappa \log(\frac{\kappa}{\epsilon}) \rceil$ to approximate the Child's function $\frac{1-(1-x^2)^b}{x}$ [1] where $b = \lceil \kappa^2 log(\frac{\kappa}{\epsilon}) \rceil$, and the Remez's algorithm [5]. As the condition number κ increases, the required number of terms in the Chebyshev series also grows. However, the influence of the error ϵ is relatively minor, as it only appears logarithmically in the formula as $\log(1/\epsilon)$. To be practical for this preliminary numerical experiment, we propose to find the approximated odd Chebyshev polynomial series via the optimization of Chebyshev polynomial series coefficients

with the following cost function:

$$C(w) = \left\| \sum_{i \in \{0,\dots,d\}} w_i cos((2i+1)\theta) - \frac{1}{2\kappa cos(\theta)} \right\|_{L^2} \tag{4}$$

where $w = \{w_0, ..., w_d\}^T$ and $\theta \in [0, acos(\frac{1}{\kappa})]$.

The proposed technique could significantly reduce the Chebyshev polynomial series terms to reach a certain precision.

State preparation f - Complexity of arbitrary state preparation of any vector is bounded by $\mathcal{O}(2^n)$ [21]. Here, we aim to generate some arbitrary vectors f with an $\mathcal{O}(1)$ complexity to save resources for the linear system solving part. In this case, we generate the vector f by applying the gate $R_Y(\theta_i)$ on each qubit $i \in \{0, ..., n-1\}$ with θ_i randomly chosen between $[0, \pi]$.

4 Numerical experiments

4.1 Numerical experiment setup

To perform numerical experiments, the block diagonalization to block encode the matrix K, and the QSVT method to invert the matrix have been implemented using the Pennylane simulator platform [13]. To employ QSVT, we first apply the optimization technique to determine the approximated Chebyshev polynomial series, and then use this polynomial with pyqsp [14] to find the phase factors to be used in the QSVT routine.

To efficiently employ the QSVT method, we generate a dictionary of phase factors corresponding to different values of the condition number κ from 20 to 500000 with a tolerance of 1% error.

First, for a given vector f, we solve the problem with different matrix sizes 2^n for $n = \{2, 3, 5, 7, 10\}$ to numerically analyze the precision of the solution via QSVT. Then, we solve the problem with matrix size $N = 32$ for different randomly generated vectors f.

4.2 Numerical results

QSVT numerical analysis - We present in Tab.1 the relationship of Chebyshev polynomial degree with respect to the condition number using the optimization function in 4, and the Child's function [1]. For $\kappa = 7500$, the polynomial degree using the optimization technique is 13 times smaller than that of the Child's function while its approximation error is still satisfied. For $\kappa = 500000$, the approximation error through the optimization technique is not fulfilled due the lack of polynomial terms; however, its degree is nearly 400 times smaller than that using the Child's function with the same error. Even if the optimization technique can achieve a very small degree compared to the Child's function, it is still a heuristic method. We cannot always obtain the desired error.

Table 1: Chebyshev polynomial degree via optimization technique and Child's function

κ	20	50	450	7500	500000
ϵ	0.009	0.007	0.004	0.006	0.062
Degree via optimization technique	50	132	1380	8160	22334
Degree via Child's function:$\kappa log(\frac{\kappa}{\epsilon})$	154	443	5233	105289	7967887

We investigate (Fig. 1) the evolution of the condition number κ in terms of the matrix size. κ rapidly increases as the matrix size increases. For the matrix size of $N = 1024$, the condition number is relatively large $\kappa \approx 400000$. This shows that the QSVT method is not suitable for solving large matrices from discretized PDE problems (at least without preconditioning). Figure 2 also illustrates the evolution of the relative error of the quantum solution with respect to the reference (classical) solution. For the matrix size $N = 1024$, the relative error is largely deviated from the tolerance because of the target error of the approximated Chebyshev polynomial series using the optimization technique. It should be noted that we did not reach a sufficient number of terms in the series to achieve the desired level of error.

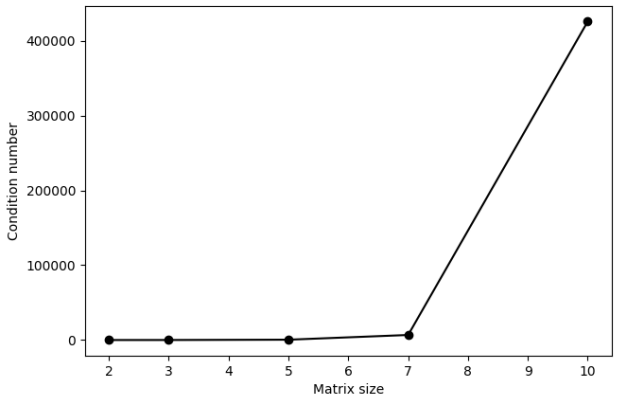


Fig. 1: Condition number κ for different matrix size 2^n for $n = \{2, 3, 5, 7, 10\}$.

PDE solutions - We present in Fig. 3 the relative error for the problem of matrix size $N = 32$ with different vectors f. Among the solutions corresponding to 50 randomly generated vectors f, 90% of the solutions provide an error less than the tolerance. Nevertheless, the other 10% of the solutions are only 0.25% deviated from the tolerance. Figure 4 illustrates the comparison between the

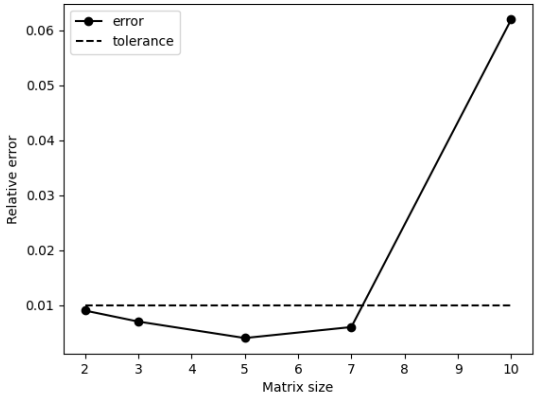


Fig. 2: Relative error ϵ for different matrix size 2^n for $n = \{2, 3, 5, 7, 10\}$.

reference and quantum solutions for the problem with matrix size of $N = 32$. The overall behavior of the quantum solution is well fit to the reference solution.

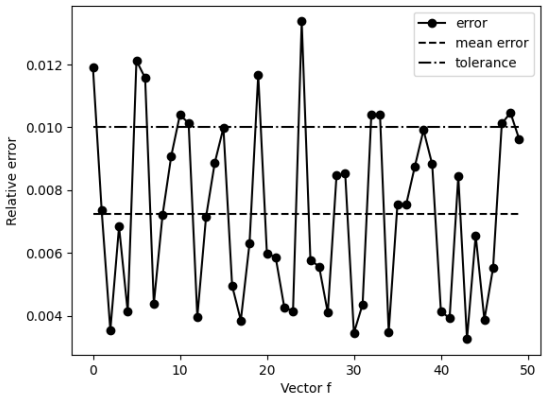


Fig. 3: Relative errors for the matrix size $N = 32$ for different vectors f.

5 Conclusions and discussions

We conducted numerical experiments of Poisson's equation by solving the associated linear system of equations using block-diagonalization and the Quantum

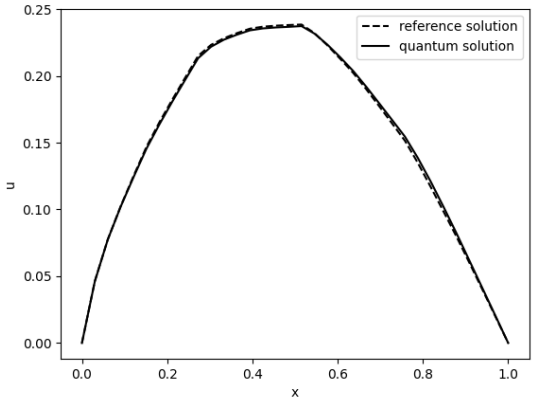


Fig. 4: Reference and quantum solution for the matrix size $N = 32$ for a given vector f.

Singular Value Transformation (QSVT) solver. These preliminary numerical experiments reveal certain limitations when applying the QSVT solver to linear systems derived from Poisson's equation. Specifically, for problems of large size, the condition number of the resulting matrix becomes exponentially large. Hence, the complexity of QSVT grows exponentially as well. To address this issue, the mixed-precision classical-quantum algorithm proposed in [8] offers a promising solution by reducing the number of Chebyshev terms needed through enhanced solution precision. Additionally, leveraging efficient quantum preconditioning techniques, such as those outlined in [2], could significantly mitigate the condition number.

References

1. Childs, A.M., Kothari, R., Somma, R.D.: Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. SIAM Journal on Computing **46**(6), 1920–1950 (2017)
2. Clader, B.D., Jacobs, B.C., Sprouse, C.R.: Preconditioned quantum linear system algorithm. Physical review letters **110**(25), 250504 (2013)
3. Dalzell, A.M., McArdle, S., Berta, M., Bienias, P., Chen, C.F., Gilyén, A., Hann, C.T., Kastoryano, M.J., Khabiboulline, E.T., Kubica, A., et al.: Quantum algorithms: A survey of applications and end-to-end complexities. arXiv preprint arXiv:2310.03011 (2023)
4. Dervovic, D., Herbster, M., Mountney, P., Severini, S., Usher, N., Wossnig, L.: Quantum linear systems algorithms: a primer. arXiv preprint arXiv:1802.08227 (2018)
5. Dong, Y., Meng, X., Whaley, K.B., Lin, L.: Efficient phase-factor evaluation in quantum signal processing. Physical Review A **103**(4), 042419 (2021)

6. Gilyén, A., Su, Y., Low, G.H., Wiebe, N.: Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing. pp. 193–204 (2019)
7. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Physical review letters **103**(15), 150502 (2009)
8. Koska, O., Baboulin, M., Gazda, A.: A mixed-precision quantum-classical algorithm for solving linear systems. arXiv preprint arXiv:2502.02212 (2025)
9. Low, G.H., Chuang, I.L.: Optimal hamiltonian simulation by quantum signal processing. Physical review letters **118**(1), 010501 (2017)
10. Low, G.H., Chuang, I.L.: Hamiltonian simulation by qubitization. Quantum **3**, 163 (2019)
11. Morgan, J., Ghysels, E., Mohammadbagherpoor, H.: An enhanced hybrid hhl algorithm. Physics Letters A **532**, 130181 (2025)
12. Patil, H., Wang, Y., Krstić, P.S.: Variational quantum linear solver with a dynamic ansatz. Physical Review A **105**(1), 012423 (2022)
13. Pennylane: Pennylane (2025), https://pennylane.ai/
14. pyqsp: pyqsp (2025), https://github.com/ichuang/pyqsp?tab=readme-ov-file
15. Sævarsson, B., Chatzivasileiadis, S., Jóhannsson, H., Østergaard, J.: Quantum computing for power flow algorithms: Testing on real quantum computers. arXiv preprint arXiv:2204.14028 (2022)
16. Saito, Y., Lee, X., Cai, D., Asai, N.: An iterative improvement method for hhl algorithm for solving linear system of equations. arXiv preprint arXiv:2108.07744 (2021)
17. Ty, S., Vilmart, R., TahmasebiMoradi, A., Mang, C.: Double-logarithmic depth block-encodings of simple finite difference method's matrices (2024), https://arxiv.org/abs/2410.05241
18. Wen, J., Kong, X., Wei, S., Wang, B., Xin, T., Long, G.: Experimental realization of quantum algorithms for a linear system inspired by adiabatic quantum computing. Physical Review A **99**(1), 012320 (2019)
19. Xu, X., Sun, J., Endo, S., Li, Y., Benjamin, S.C., Yuan, X.: Variational algorithms for linear algebra. Science Bulletin **66**(21), 2181–2188 (2021)
20. Yalovetzky, R., Minssen, P., Herman, D., Pistoia, M.: Solving linear systems on quantum hardware with hybrid hhl++. Scientific Reports **14**(1), 20610 (2024)
21. Zhang, X.M., Li, T., Yuan, X.: Quantum state preparation with optimal circuit depth: Implementations and applications. Physical Review Letters **129**(23), 230504 (2022)
22. Zheng, M., Liu, C., Stein, S., Li, X., Mülmenstädt, J., Chen, Y., Li, A.: An early investigation of the hhl quantum linear solver for scientific applications. arXiv preprint arXiv:2404.19067 (2024)

A Hybrid Quantum-Classical Algorithm
for Fault Tree Analysis Via Finding Vertex Separators

Pedro Henrique Pons Fiorentin[1][0009−0003−9255−0674][1], Mohamed Hibti[2][0009−0007−1334−7192], and Rola Saidi[1][0009−0009−1394−3247][2]

[1] Amazon ph.pfiorentin@gmail.com
[2] EDF Lab Paris-Saclay, PERICLES, France
{mohamed.hibti,rola.saidi}@edf.fr

Abstract. Fault Tree Analysis (FTA) is one of the main problems for reliability studies, in particular for the probabilistic safety assessment of complex systems such as aircrafts or nuclear power plants. The extraction of Minimal Cut Sets (MCS) from a fault tree can be reduced to the problem of finding vertex separators in S-T directed acyclic graphs (S-T DAGs). Prior work introduced a quantum algorithm to identify these s-t vertex separators, but faces scalability issues in larger instances. In this paper, we present a novel hybrid quantum-classical algorithm to solve the problem of st-connectivity using a divide and quantum strategy. Large s–t DAG instances are partitioned into many subgraphs that fit the qubit capacity of current quantum hardware; each subgraph is solved on a quantum device, yielding partial solutions. These solutions are then recombined classically to produce all global vertex separators of the original s-t DAG, which are mapped back to the MCS of the original fault tree. The results showed that we were able to solve larger instances that have not been solved using a classical algorithm previously. We also quantify the effect of realistic noise by comparing results from noisy versus noise-free simulation, and report on prototype runs using IBM's publicly available quantum hardware.

Keywords: S-t connectivity · Fault-Tree Analysis · Minimal Cut Sets · Divide and Quantum· Hybrid Quantum Classical Algorithm

1 Introduction

Probabilistic Safety Assessment [14] is a systematic approach used to evaluate the safety and reliability of complex engineering systems, particularly in high-risk industries such as nuclear power, aerospace, chemical processing, and transportation. It is a risk-based analysis method that quantifies the probability of hazardous events and their potential consequences.

Fault Tree Analysis (FTA) is the main method for addressing PSA. It has been used since 1961 for visual displaying, analyzing and evaluating failure paths in a system [13]. Its objective is to identify all the combinations of basic events that lead to an undesired event, called the failure event, and sum up their probabilities. There are many algorithms that were used for FTA within commercial or

academic software applications; based on Boolean fusion over modules with approximation procedures for large instances (Riskspectrum, CAFTA, XFTA [15]). Others use Binary Decision Diagram (BDD)-like algorithms [16], [11] or/and approximations mixed with BDD or Ternary Decision Diagram (TDD) algorithms [12]. These algorithms are now challenged by the size of the instances and the frequent occurrence of non-rare events due to the implementation of external hazards such as seism.

Some quantum approaches have been investigated for the MCS problem, which mainly in the gate-based paradigm. One method uses a Quantum Fault Tree (QFT) model [1], where the FT is encoded into a quantum circuit, and applies Quantum Amplitude Amplification (QAA) to enhance the probability of sampling minimal cut sets [2]. Another approach reformulates the problem into a Boolean Satisfiablitiy problem and uses a Grover-based solver, with divide-and-conquer strategies [3]. A third method frames the problem as a vertex separator problem (VSP) in a directed graph [10]. In [18], it is shown that a quantum query algorithm can outperform any classical algrorithm for VSP. While promising, these methods remain limited by the small number of qubits and high noise levels of today's Noisy Intermediate-Scale Quantum (NISQ) computers.

Consequently, this work explores a hybrid quantum-classical algorithm for the MCS problem, in an attempt to tackle the number of qubit limitations.

2 FTA as the Vertex Separator Problem

Boolean formula evaluation reduces to st-connectivity problems [5], [6]. In [9], the problem of enumerating the cut-sets in a FT's boolean formula is converted to the VSP by representing the FT as a graph; the problem is then solved with a quantum algorithm, which is referred to as Quantum-VSP. In the VSP, the goal is to find all combinations of vertices that, if removed from the graph, disconnect a source vertex **S** from a terminal vertex **T**. These separators are equivalent to the cut-sets of the fault tree. The VSP is NP-Hard [7].

More formally, given a directed acyclic graph (DAG) $G = (V, E)$, where V is the set of vertices and E is the set of directed edges, and two distinguished vertices S (the source) and T (the target) in V, enumerate all minimal vertex separators $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ for G, such that:

1. Each C_i is a minimal vertex set that, when removed from G, disconnects all directed paths from S to T in the resulting subgraph $G'_i = (V \setminus C_i, E'_i)$.
2. No proper subset of C_i is a separator, i.e., C_i is minimal in disconnecting S and T.

In [10], the proposed quantum algorithm applies what we call movement oracles for the VSP to incrementally create a state with the superposition of all vertex separators that are not necessarily minimal. Each vertex v has an associated movement oracle that "moves" it to its successors *Succ(v)*. The runtime of the associated quantum circuit is linear, using $O(n + m)$ quantum gates and $n+2$ qubits, where n is the number of vertices and m the number of edges in the

DAG. In order to extract only the minimal separtors, a naive post-processing filtering step with complexity $O(N^2)$ is required, where N is the number of cuts. It consists of going through the list of separators and removing those that include other separators.

3 Hybrid quantum VSP algorithm

The approach of [10] to solve VSP problem requires a linear number of qubits to process an S-T graph, which prohibits the algorithm from scaling in current quantum computers with limited numbers of qubits. The question that arises from this is: can this approach be explored under the constraints of lower qubit availability?

This section introduces a hybrid quantum-classical approach for applying the VSP algorithm to subgraphs whose size is bounded by the number of available qubits N_q, enabling the construction of a global solution from smaller, tractable subgraphs.

The hybrid algorithm follows a divide and conquer approach, where the subdivided problems are processed by quantum computers and the global solution is found by a classical computer by joining subsolutions, as seen in Figure 1. The question is then: can we build the MCS for the whole graph from these sub-

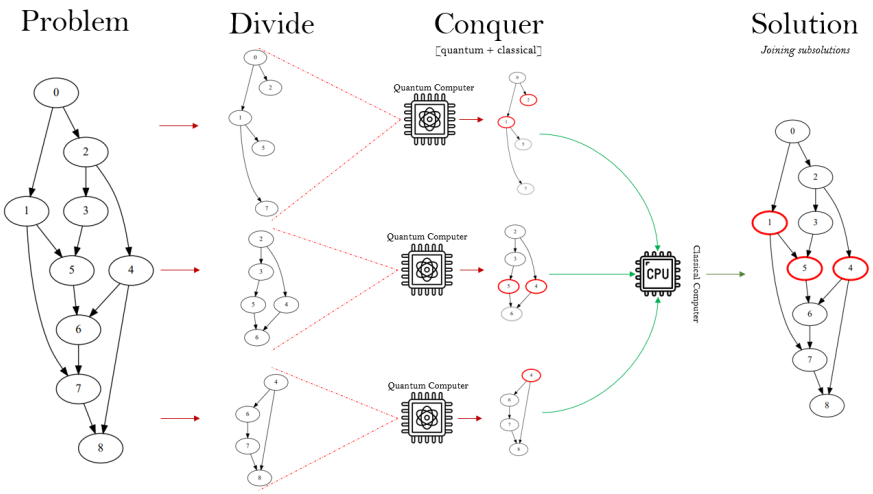


Fig. 1: Flow of hybrid VSP algorithm

solutions and how computationally expensive is it? To answer the question of whether it is possible or not, the intuition of how to join subsolutions is explained in section 4.

4 Intuition of the algorithm

To introduce the logic of the hybrid algorithm, we first present the basic intuition for it through examples of simple graphs. Consider the first example in Figure 2, an S-T graph with two vertex-disjoint paths. In this scenario, the minimal separators for this graph can be obtained by performing a cartesian product of sets A and B. This fundamental principle remains consistent for any number of parallel, vertex-disjoint paths that could be added. If a third vertex-disjoint path C were added, the MCS would be $A \times B \times C$. When looking at an S-T graph

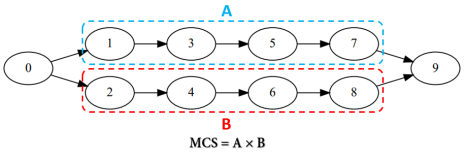


Fig. 2: S-T DAG with two vertex-disjoint paths.

with only a single path such as in Figure 3, each element within set A represents an MCS in itself. From an algorithmic perspective, this corresponds to executing a Cartesian product between set A and an empty set. To extend this logic to our

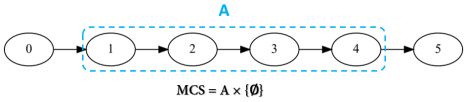


Fig. 3: Simple S-T DAG.

method, first let G be a DAG, and let v be a vertex in G. We define $S_G(v)$ as the subgraph of G that includes all vertices which topologically succeed vertex v. In other words, $S_G(v)$ consists of all vertices in G that are reachable from v through a directed path.

Having this in mind, we reexamine the graph $G = S_G(0)$ in Figure 2. It becomes clear that the minimal separators of G, i.e. $MCS(S_G(0))$, are equivalent to the cartesian product of the minimal separators of the subgraphs $S_G(1)$ and $S_G(2)$, or $MCS(S_G(1)) \times MCS(S_G(2))$. More generally, the MCS for a source vertex s can be obtained by the cartesian product between the minimal separators of the subgraphs induced by each of $succ(s)$; we define this intermediate set of MCS as $MCS(succ(s))$.

In the pseudocode below, this operation of cartesian products between different sets of MC is defined through a FOLD as $MCS(v) := FOLD(\times, \{\emptyset\}, MCS(succ(v)))$. The FOLD is applied by iteratively folding, or accumulating, the sets of mini-

mal separators using the cartesian product ($\times$) as the combining operator. The folding for the graph in Figure 2 would be as follows:

1. Starting with a set $\{\emptyset\}$ and the first set $\{\{1\},\{3\},\{5\},\{7\}\}$, we perform the cartesian product: $\{\emptyset\} \times \{\{1\},\{3\},\{5\},\{7\}\} = \{\{1\},\{3\},\{5\},\{7\}\}$.
2. To finish the folding, we take the result from step 1 and perform a cartesian product with the next set: $\{\{1\},\{3\},\{5\},\{7\}\} \times \{\{2\},\{4\},\{6\},\{8\}\}$.

The hybrid algorithm builds upon this logic, generating solutions through the exploration of local separators. Through this intuition, a global solution can be achieved, even if it includes non-minimal MCS that require subsequent filtering. This holds true as long as the algorithm operates in a topological sequence.

Algorithm 1: hybridCutSearch

Input: G - Directed S-T graph
S_G - Source vertex of G
N_q - Number of qubits available
$quantum_VSP$ - Quantum subroutine to enumerate MCS
Output: $CUTS$ - Separators of a graph

$CUTS_DICT := \{\}$
$Q := [S_G]$
$processed_nodes := \{\}$
while Q *is not empty* **do**
  $S \leftarrow Q.pop(0)$
  G', $processed_v \leftarrow buildSubgraph(G, S, N_q)$
  $processed_nodes := processed_nodes \cup processed_v \cup S$
  `// Build dictionary with cuts starting from each subgraph source`
  $CUTS_DICT[S] \leftarrow quantum_VSP(G')$

  `// Add unprocessed vertices to Q. An unprocessed vertex is a vertex whose successors were not included in the subgraph or whose degree is >= 2.`
  $Q := Q + \{G' \setminus processed_nodes\}$
end

`// Now process the cuts in reverse topological order`
for S *in* $CUTS_DICT.reversed()$ **do**
  $new_cuts := set()$
  for C *in* $CUTS_DICT[S]$ **do**
    $cuts_to_cross := \{v \cup CUTS_DICT[v] : v \in C\}$
    $new_cuts := FOLD(\times, \{\emptyset\}, cuts_to_cross)$
  end
  $CUTS_DICT[S] := new_cuts$
end
return $CUTS_DICT[S_G]$

Algorithm 2: buildSubgraph

Input: $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ - Original graph
s - Source of the subgraph
N_q - Number of qubits available
Output: G' - Subgraph of G
$\mathbb{P}$ - Fully processed vertices

Initialize an empty subgraph $\mathbb{G}' = (\mathbb{V}' = \emptyset, \mathbb{E}' = \emptyset)$
Initialize queue $Q = [s]$

while $|Q| > 0$ *and* $|\mathbb{V}'| < N_q$ **do**
  $v \leftarrow Q.pop(0)$
  $\mathbb{V}' = \mathbb{V}' \cup \{v\}$
  if $|\mathbb{V}'| + |Succ(v)| < N_q$ **then**
    if $inDeg(v) = 1$ **then**
      $\mathbb{P} := \mathbb{P} \cup \{v\}$
    Q += $Succ(v)$
end

$E' = \{(a, b) \in \mathbb{E} : a \in \mathbb{V}' \wedge b \in \mathbb{V}'\}$
Return $\mathbb{G}' = (\mathbb{V}', \mathbb{E}')$, $\mathbb{P}$

5 Evaluation of the method

This section covers the benchmarks executed to empirically evaluate the performance of the hybrid VSP method. The quantum subroutines execution were simulated with the library Qiskit assuming small devices of 6 qubits to guarantee correct results, as currently available quantum computers have too high of a noise rate.

Two kinds of benchmarks were executed, the first to evaluate the limits and running times of the novel method, and the second to compare its performance to another classical method. The goals of the benchmarks were to test the algorithm on randomly generated graphs of increasing size and analyze its performance, and compare the algorithm's running time against a recent classical method from the literature.

All graph instances generated were S-T DAGs of arbitrary size. We limit all vertices to have a maximum of 4 outgoing edges, following a standard in PSA models which often include 4 levels of redundancy at most. Figure 4 illustrates an instance that was evaluated and is tractable by the hybrid method under 30 minutes: To contextualize this performance, we compare our method to a recursive cutset enumerating algorithm published in 2012 [4]. This method's time complexity was not reported, but favorable benchmarks were shown in the paper. The classical method's performance is first plotted by itself in Figure 5 and then compared with the hybrid method in Figure 6.

The same method of instance generation was followed. The method evaluation was stopped once running times began surpassing 10 minutes, hence the cutoff point at 15 vertices. An exponential trend can also be seen for the recur-

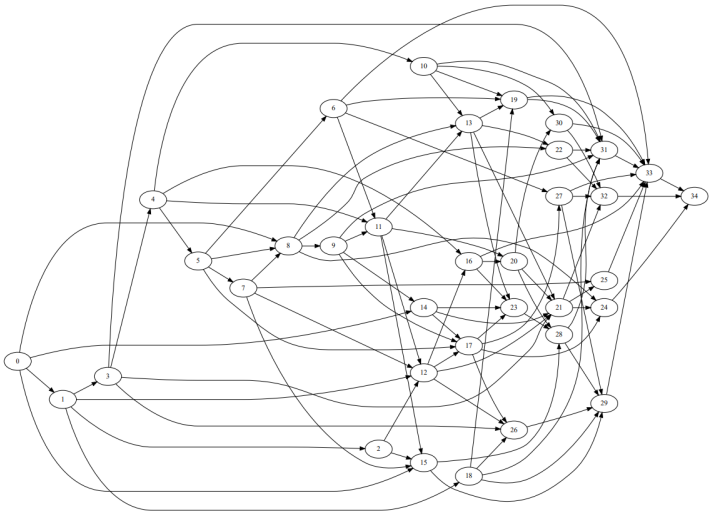


Fig. 4: Randomly generated instance with 34 vertices that have been solved with our hybrid algorithm but could not be solved with the classical one.

sive algorithm with a very predictable performance, but the method does not work well for larger instances:

The classical algorithm has a quickly degrading performance once the instance size hits 13 vertices, which our method can still easily handle. This weak performance is naturally not convincing, and the large gap between the methods was particularly surprising (Figure 6).

We remark that future work should include new benchmarks for the S-T DAG VSP problem with the hybrid algorithm and algorithms from the literature. Although research was done to find different classical implementations, to the best to our knowledge there were no other readily available implementations that could be used for comparisons in this benchmark.

To the best of our knowledge, these benchmarks show that this new hybrid algorithm is capable of dealing with relatively complex instances given the exponential nature of the problem. Additionally, to enhance this algorithm's practical applicability, a potential direction of future work is to implement it in a more performant programming language than Python.

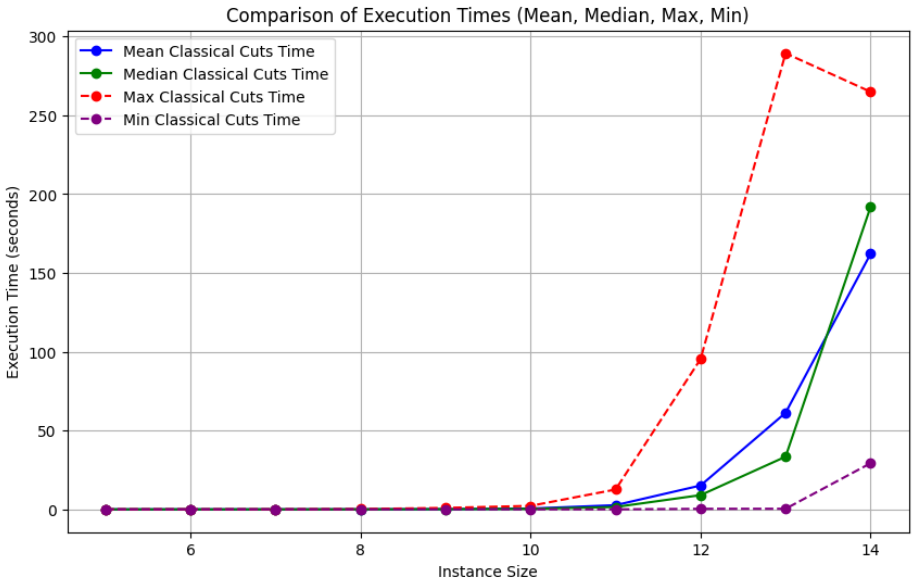


Fig. 5: Performance of classical recursive algorithm over increasing instance sizes.

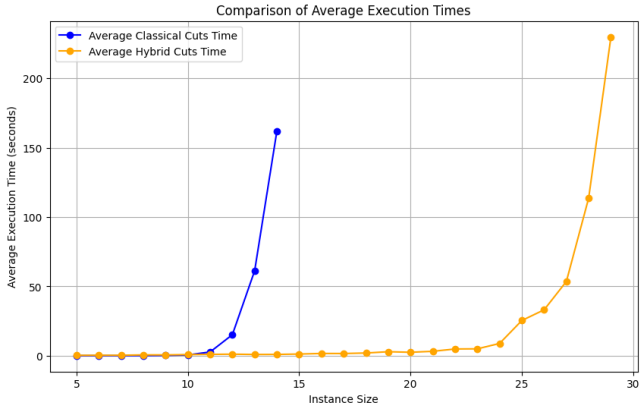


Fig. 6: Comparison of the hybrid algorithm running time against a recursive classical method.

6 Performance assessment on Noisy Intermediate-Scale Quantum Devices (NISQ)

In the previous section, only Qiskit's classical simulators were used for the execution of the quantum subroutine to avoid the high error rates in quantum computers. However, in this section, we provide an evaluation of the algorithm on current Noisy Intermediate-Scale Quantum (NISQ) devices.

The experiments were performed using the version 1.2.0 of the IBM Qiskit framework. The tests utilized Qiskit's noisy simulator (FakeSherbrooke), as well as the quantum computers IBM Sherbrooke (127 qubits) and IBM Brisbane (127 qubits). To better assess the quality of the results from real quantum hardware, the Hellinger distance [17] was used as a common performance metric. Moreover, error mitigation techniques, specifically the Mthree(M3)[3] IBM's measurement errors mitigation tool [19] were used with only very slight improvement.

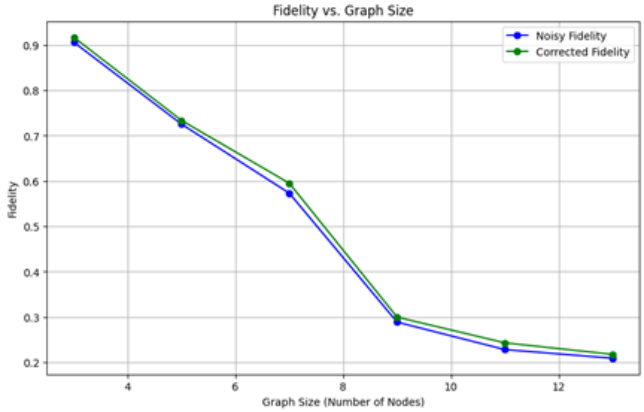


Fig. 7: Fidelity for MCS algorithm in function of graph sizes before and after Error Mitigation

7 Conclusion

This work's main contribution lies in proposing a novel hybrid framework for the S-T Vertex Separator Problem, which can make use of any new quantum subroutines that solve VSP. This hybrid framework enables the use small-scale quantum devices for solving PSA problems. By comparing this new method against a recent classical method for minimal cutset enumeration, there is a performance advantage with the hybrid algorithm. We remark that these results against a single method should be seen with caution and conclude that more benchmarks should be performed against different approaches from the literature.

[3] https://qiskit-extensions.github.io/mthree/

All things considered, this hybrid framework serves as a foundation upon which subsequent improvements can be made, ultimately leading to more efficient computation in PSA. While practical implementation on current quantum hardware remains a challenge due to noise levels and decoherence, this work underscores the potential of quantum computing for PSA. Future work may focus on discovering more efficient quantum subroutines for the vertex separator problem and improving the filtering processes for minimal separator identification.

References

1. G. San Martín Silva, T. Parhizkar, E. López Droguett, *Quantum Fault Trees*, Department of Civil & Environmental Engineering, University of California, Los Angeles, and The B. John Garrick Institute for the Risk Sciences, UCLA.
2. G. San Martín Silva and E. López Droguett, *Quantum Fault Trees and Minimal Cut Sets Identification*, The B. John Garrick Institute for the Risk Sciences, Center for Reliability Science and Engineering, Department of Civil and Environmental Engineering, UCLA.
3. M. Rennela, S. Brand, A. Laarman, and V. Dunjko, *Hybrid divide-and-conquer approach for tree search algorithms*, arXiv preprint arXiv:2007.07040, 2021.
4. M Benaddy and M Wakrim. Cutset enumerating and network reliability computing by a new recursive algorithm and inclusion exclusion principle. *International Journal of Computer Applications (0975- 8887)*, 2012.
5. M. Hibti. *What if we revisit evaluation of PSA models with network algorithms?* ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis, 2013.
6. Stacey Jeffery and Shelby Kimmel. *Quantum algorithms for graph connectivity and formula evaluation,* arXiv, 1704.00765, 2019. URL https://arxiv.org/abs/1704.00765.
7. Fanz Rendl and Renata Sotirov. The min-cut and vertex separator problem. *Computational Optimization and Applications*, 69(1):159–187, September 2017.
8. Kloks Ton and Kratsch Dieter. Listing All Minimal Separators of a Graph, SIAM J. *Comput.*, 27:605–613, 1998.
9. Ahmed Zaiou. *Quantum machine learning approaches for graphs and sequences: application to nuclear safety assessment.* PhD thesis, 2022. Thèse de doctorat dirigée par Mateï, Basarab Informatique Paris 13 2022.
10. Ahmed Zaiou, Younès Bennani, Mohamed Hibti, and Basarab Matei. Quantum approach for vertex separator problem in directed graphs. In Ilias Maglogiannis, Lazaros Iliadis, John Macintyre, and Paulo Cortez, editors, *Artificial Intelligence Applications and Innovations*, pages 495–506, Cham, 2022. Springer International Publishing.
11. W. S. Jung. ZBDD algorithm features for an efficient Probabilistic Safety Assessment. *Nuclear Engineering and Design*, 2009. ISSN 00295493. https://doi.org/10.1016/j.nucengdes.2009.05.005.
12. W. S. Jung. Zero-suppressed ternary decision diagram algorithm for solving non-coherent fault trees in probabilistic safety assessment of nuclear power plants. *Nuclear Engineering and Technology*, 56(6):2092–2098, 2024. ISSN 1738-5733. https://doi.org/https://doi.org/10.1016/j.net.2024.01.017. URL https://www.sciencedirect.com/science/article/pii/S1738573324000184.
13. C. A. Ericson. Fault tree analysis: A history. In *17th International System Safety Conference*, 1999.

14. USAEC. *Reactor safety study: An assessment of accident risks in the U.S. Commercial Nuclear Power Plants Report WASH 1400 1975. U.S. Atomic Energy Commission, Washington D.C.* Eyrolles, 1975.
15. Rauzy, Antoine. An Open-PSA Fault Tree engine. *PSAM/ESREL 2011*, 2012.
16. Arboost Technologies. *Aralia User Manual.* Arboost, 2004.
17. Subhayan Dasgupta and Travis S. Humble, *Characterizing the Reproducibility of Noisy Quantum Circuits*, Entropy, vol 24, num 2, 2022, MDPI, https://doi.org/10.3390/e24020244
18. Jeffery, S., Kimmel, S., & Piedrafita, A. Quantum Algorithm for Path-Edge Sampling. In O. Fawzi, & M. Walter, 18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023).
19. P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, "Scalable Mitigation of Measurement Errors on Quantum Computers," *PRX Quantum*, vol. 2, no. 4, p. 040326, Nov. 2021. [Online]. Available: https://doi.org/10.1103/PRXQuantum.2.040326

Monte Carlo Particle Transport on Quantum Computers

Noe Olivier[1] and Michel Nowak[1]

cortAIx Labs, Thales Research and Technology, 91120 Palaiseau, France
noe.olivier@thalesgroup.com
michel.nowak@thalesgroup.com

Abstract. Monte Carlo particle transport codes are well established on classical hardware and are considered as the reference tool for nuclear applications. In a growing number of domains, the design of algorithms is progressively shifting towards the field of quantum computing, where theoretical speedups over their classical counterparts are expected. In some of these domains, Monte Carlo methods have already been converted to a quantum computing friendly setup where the expected and observed gain in complexity is quadratic. In this work, we address the particle transport problem in view of a first implementation on these architectures. We propose a quantum algorithm to model the particle transport based on discrete-time quantum walks and compare our approach to Monte Carlo and deterministic classical numerical schemes. The proof-of-concept algorithm is applied to a 10-qubit small-size problem by simulating its behavior on both ideal and noisy quantum computers based on the qiskit framework in view of reproducing classical results.

Keywords: Monte Carlo · Particle transport · Quantum walks · Quantum algorithm

1 Introduction

Monte Carlo particle transport codes [1,2,3] are well established within numerous fields of research. They are used as reference calculations in the field of nuclear energy, and more generally for radiation protection. Continuous developments on classical computers still succeed at improving the convergence rate of these codes, in particular by leveraging GPU capabilities [4]. However running those codes on HPC clusters remains costly and resource demanding.

As quantum computers are emerging, researchers develop and implement algorithms that can leverage these architectures in a more and more efficient way. Boosted by promising theoretical speedups, these developments are now guided by real-world applications in many fields. Quantum friendly Monte Carlo methods have already been proposed in a wide range of fields, starting from finance [5,6], to theoretical [7] and practical integration [8], with the motivation of a quadratic speedup.

In this work, we aim to explore the possibility of computing particle transport on quantum computers. In Section 2, we give a brief introduction to the particle transport problem and present associated classical methods. In Section 3, we describe our proposition for a quantum algorithm to address the particle transport problem in heterogeneous media. Finally, in Section 4 we present and discuss the numerical results obtained from ideal and noisy simulations on small-size problems and compare them to classical ones.

2 Particle Transport

2.1 Problem Definition

In this paper, we address the particle transport problem, which is described by the transport equation. Let's consider an infinitesimal region of phase space positioned at $\boldsymbol{r}$, of direction $\boldsymbol{\Omega}$ and energy E. The exact transport equation that governs the angular flux ψ can be expressed as in page 47 of [9]

$$\Omega \cdot \nabla\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E) + \Sigma_t(\boldsymbol{r}, E)\psi(\boldsymbol{r}, \boldsymbol{\Omega}, E) = \int_{4\pi} \Sigma_s(\boldsymbol{r}, \boldsymbol{\Omega'} \to \boldsymbol{\Omega}, E' \to E)\psi(\boldsymbol{r}, \boldsymbol{\Omega'}, E')dE'd^2\Omega' + S_0(\boldsymbol{r}, \boldsymbol{\Omega'}, E) \quad (1)$$

where Σ_t and Σ_s are respectively the total and scattering cross sections of the medium, and S_0 the source. It can be simplified to the diffusion equation [9], with Σ_a the absorption cross section and D the diffusion coefficient:

$$-\nabla \cdot D(\boldsymbol{r})\nabla\phi(\boldsymbol{r}) + \Sigma_a(\boldsymbol{r})\phi(\boldsymbol{r}) = S_0(\boldsymbol{r}) \quad (2)$$

and governs the scalar flux ϕ.

2.2 Monte Carlo Method

Consider a probability distribution that carries information about the system of interest, from which we can sample trajectories. Let $\mathbb{T}$ be the set of possible trajectories. A batch of Monte Carlo sampling consists in generating a set of trajectories $\mathcal{S} \in \mathbb{T}^N$ where N is the number of trajectories sampled.

In order to sample such trajectories, we first initialize particles at specified source points [9]. Each particle then experiments an exponential flight before a next collision or absorption. The flight length ρ between two collisions is sampled from $\pi_{\text{flight}}(\rho) = \Sigma_t e^{-\Sigma_t \rho}$.

We consider that particles are either absorbed or scattered. The scattering phenomenon is composed of two steps: the selection of the isotope on which the particle collides, and the computation of the outgoing scattering angle. For simplicity, we consider that the scattering is isotropic and that the materials are homogenized into macroscopic cross sections. The probability of scattering is characteristic to the medium and is given by $p_s = \Sigma_s/\Sigma_t$.

These steps are repeated until the particle gets absorbed by the material. We can score the flux by counting the number of collisions in a region of interest

$$\phi_{\mathcal{S}}(\mathbf{r}) = \frac{1}{N} \sum_{\mathcal{T} \in \mathcal{S}} \sum_{\mathbf{r}' \in \mathcal{T}} \frac{\mathbb{1}_r(\mathbf{r}')}{\Sigma_t}. \tag{3}$$

In complex radiation shielding problems, the time to solution for such calculations can reach several days on a supercomputer and might be reduced considerably by applying relevant quantum computing tools.

2.3 Finite Differences Method

Deterministic solvers, along with simplifications of the model, are often used in order to achieve a faster result than the classical Monte Carlo method. To this end, we consider equation 2 and replace the gradient of the flux by its finite differences approximation. We follow section 3.9.2 of [10] to derive the numerical scheme and implement it in multiple dimensions. The advantage of the deterministic approach is that we get the response in a reasonable amount of time and with reasonable computer resources. However, it requires the end-user of the code to be aware of the simplifications made to the transport equation.

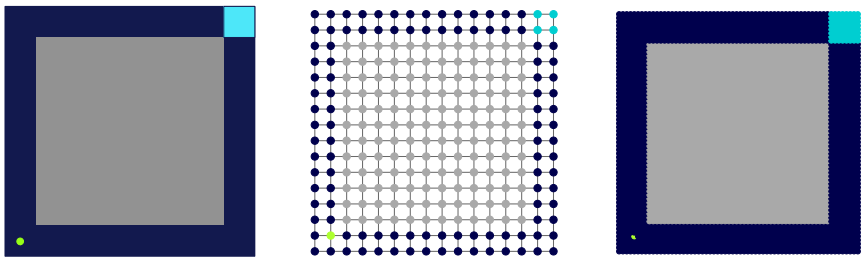
3 Quantum Approach

The Monte Carlo method remains the reference in the field of particle transport, thus making an investigation of a strategy to implement it on a quantum computer highly valuable. In this section, we propose an extension of particle transport numerical methods to the field of quantum computing. This extension comes with various simplifications and intends to model as faithfully as possible the dynamics of the transport with quantum computing tools.

Let's consider the problem of particle transport in a heterogeneous domain represented in Fig 1 where both media have the same total cross section $\Sigma_t = 1\text{cm}^{-1}$ but differ in their scattering cross section: $\Sigma_s = 0.1\text{cm}^{-1}$ for the obstacle (center), $\Sigma_s = 0.9\text{cm}^{-1}$ for the propagation medium (outside). Although the graphs are two-dimensional for the purpose of this paper, the approach can be generalized by considering that the domain in the z-direction is unbounded.

Generally, a particle transport code relies on numerous calls to the geometry in order to account for the position of the materials, boundary conditions and local materials characteristics. In particular, the probabilities of interaction vary depending on the position of a particle, and on the isotopes present at this location.

In that regard, we propose in this paper to rely on the coin-based discrete-time quantum walk formalism [11,12] on a grid to simulate the particle transport problem.



(a) Domain geometry (b) 16×16 nodes, 8 qubits (c) 64×64 nodes, 12 qubits

Fig. 1: Domain geometry and its discretized versions. The geometry studied (a) is discretized (b,c) according to the size of the position qubits registers: $2^{n_x} \times 2^{n_y}$. The different colors correspond to different media (grey and dark blue) or help identify the source (green) and the detector (cyan).

3.1 General Encoding

The modeling of the diffusion of the particles across the domain consists in a coin-based discrete-time quantum walk. The general structure of the proposed algorithm is given in Fig 2.

The time evolution of the process is discretized in a number of iterations, each of which is composed of two operations. The coin operator acts on a coin register and defines directions of propagation to be applied. The shift operator modifies the position register according to the state of the coin register.

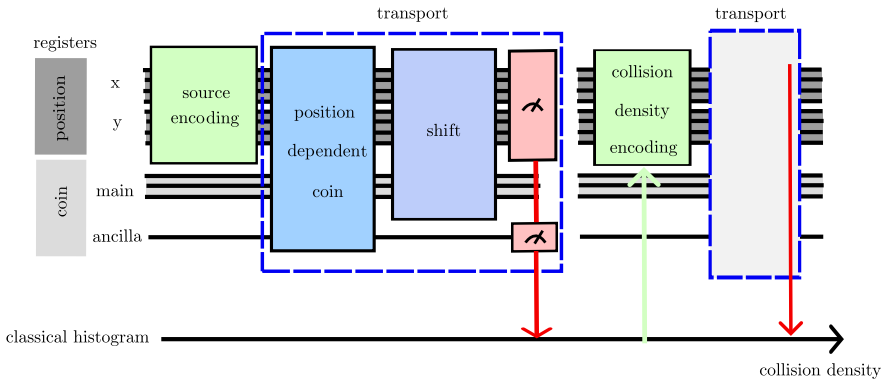


Fig. 2: Diagram of the quantum circuit for the measured quantum walk. The source encoding block consists in a single layer of X-gates for a punctual source, while the collision density encoding relies on the method from [13].

Let's define a position quantum register of $n_{\text{pos}} = n_x + n_y$ qubits, allowing the state representation of all the nodes of a two-dimensional grid of dimensions $2^{n_x} \times 2^{n_y}$ as depicted in Fig 1b. The quantum state corresponding to a single node is described by $\left|q_{n_{\text{pos}}-1} \dots q_1, q_0\right\rangle = \left|x_{n_x-1} \dots x_1 x_0 y_{n_y-1} \dots y_1 y_0\right\rangle$. This position register encodes the distribution of particles across the geometry in the states amplitudes. Similarly, we define a coin register that encodes the distribution of directions of propagation of the particles.

3.2 Algorithm Description

Source Initialization. As for the classical methods, the first step of the algorithm is the initialization of the particle trajectories. This is done by preparing a uniform superposition state over the position states that matches the source nodes. Initially, we consider a punctual source, efficiently implemented with a single layer of Pauli X gates.

Position-Dependent Coin Operator. Most works on coin-based quantum walks consider global coins, such as the Hadamard or Grover coins, which act on the coin register independently from the state of the qubits in the position register. However, the probabilities of interaction within the geometry are local and the behavior of the particles depends strongly on their location. Therefore, we choose to rely on a local coin operator, in Fig 3, in order to transport the particles according to the local cross sections. We consider the "naive" position-dependent coin operator introduced by [14] for our implementation despite its exponential scaling in the number of position qubits, due to the definition of one coin operator per position state. This paper also presents another implementation of a local coin operator which allows a linear-depth circuit at the cost of an exponential number of ancillary qubits.

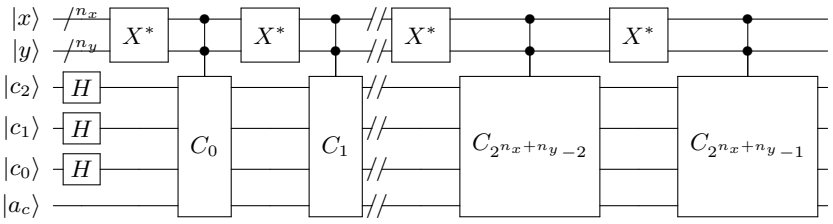


Fig. 3: Quantum circuit for the position-dependent coin, as proposed in [14]. X^* are layers of Pauli X gates to ensure $|x, y\rangle = |i\rangle$ and apply C_i. H gates are required to perform the desired operation described by the matrix $C_{|x,y\rangle}$.

In our algorithm, the coin register is composed of three qubits to account for the four possible directions of propagation and the absorption process using the

convention described in Table 1. As a first simplification in our coin operator design, we choose to focus on macroscopic cross sections such that we average the cross sections locally over all isotopes. Let's consider both scattering and absorption effects and their associated local cross sections $\Sigma_{\mathrm{s}}^{x,y}$ and $\Sigma_{\mathrm{a}}^{x,y}$ respectively. For simplicity, let's omit the indices x, y in the following notations. The coin operator is designed to result in the coin state $|\circlearrowleft\rangle$ with probability $p_{\mathrm{a}} = \Sigma_{\mathrm{a}}/(\Sigma_{\mathrm{a}} + \Sigma_{\mathrm{s}})$ and in either of the four other states with probability $p_{s,i} = \Sigma_{\mathrm{s},i}/(\Sigma_{\mathrm{a}} + \Sigma_{\mathrm{s}})$ respectively, where $i \in \{\rightarrow, \uparrow, \leftarrow, \downarrow\}$. Numerical results are obtained assuming that all scattering directions occur with the same probability $p_{\mathrm{s},i} = p_{\mathrm{s}}/4$ with $p_{\mathrm{s}} = \Sigma_{\mathrm{s}}/(\Sigma_{\mathrm{a}} + \Sigma_{\mathrm{s}})$.

Coin state $\lvert c_2 c_1 c_0\rangle$	Directions
$\lvert 100\rangle$	$\lvert\rightarrow\rangle$
$\lvert 110\rangle$	$\lvert\uparrow\rangle$
$\lvert 101\rangle$	$\lvert\leftarrow\rangle$
$\lvert 111\rangle$	$\lvert\downarrow\rangle$
$\lvert 0\cdot\cdot\rangle$	$\lvert\circlearrowleft\rangle$

Table 1: Coin-states qubit representation.

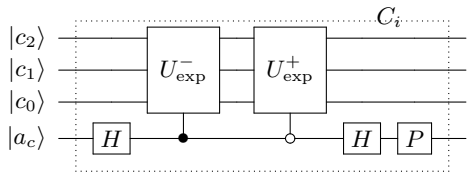


Fig. 4: Quantum circuit for the diagonal non-unitary coin operator, as proposed in [15].

The coin design results in the construction of an operator represented by a diagonal (but non-unitary) matrix $C_{|x,y\rangle}$ of dimension 8×8 for a three-qubit coin register. It can be implemented using a single ancillary qubit [15] after a first layer of hadamard gates as shown in Fig 4. Nevertheless, the successful application of the operator requires a postselection on the ancilla qubit after each step of the walk.

$$C_{|x,y\rangle} = \begin{pmatrix} M_{A|x,y\rangle} & 0 \\ 0 & M_{S|x,y\rangle} \end{pmatrix}$$

where the local absorption and scattering matrices are given by

$$M_A = \mathrm{diag}\left(\sqrt{p_{\mathrm{a}}}, \sqrt{p_{\mathrm{a}}}, \sqrt{p_{\mathrm{a}}}, \sqrt{p_{\mathrm{a}}}\right)/4, \quad M_S = \mathrm{diag}\left(\sqrt{p_{\mathrm{s},\rightarrow}}, \sqrt{p_{\mathrm{s},\leftarrow}}, \sqrt{p_{\mathrm{s},\uparrow}}, \sqrt{p_{\mathrm{s},\downarrow}}\right)$$

We choose to implement reflective boundary conditions by modifying the local coin matrices on the edge of the geometry such that the probability of going in a prohibited direction is uniformly spread between all remaining scattering directions without altering the local absorption probability.

Shift Operator. The chosen implementation of the shift operator, in Fig 5, is based on the generalised inverter gates implementation [16] using a series of X and Toffoli gates such that the position states amplitudes are updated

according to the scattering and absorption probabilities. Since c_2 is one of the control qubits, absorbed particles are not propagated but maintained at their previous location. This model is chosen despite its failure to faithfully represent the physical distribution of particles over multiple steps of the quantum walk: absorbed particles will eventually be propagated.

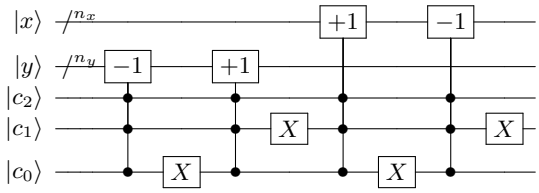


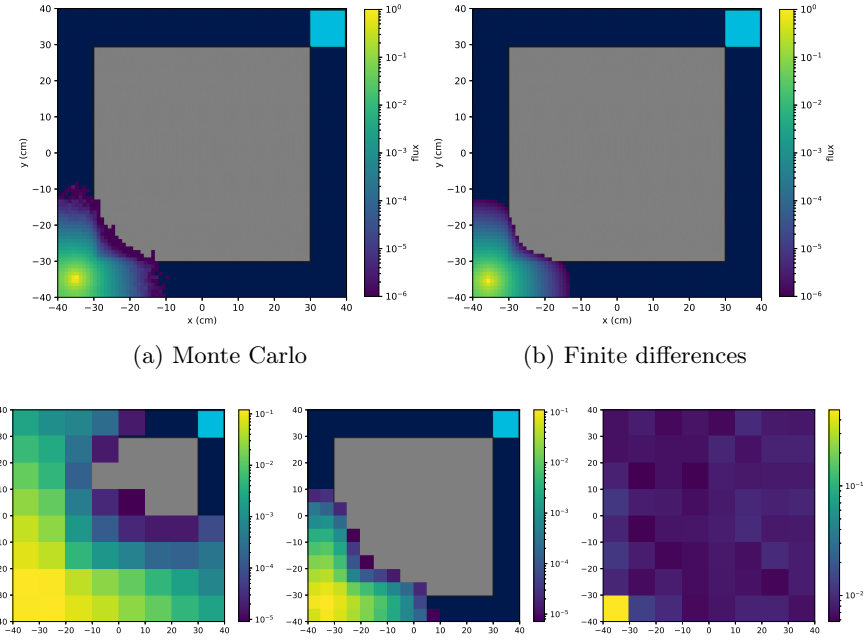
Fig. 5: Quantum circuit for the shift operator, using the toffoli-based "increment" and "decrement" operators proposed in [16].

Measured Quantum Walk. For the purpose of reproducing classical results, we overcome this limitation by constructing a circuit and measuring it for single steps of the quantum walk [17], as seen in Fig 2. The postselection on the ancillary coin qubit gives a set of successful shots for which the circuit is well implemented while a second postselection on c_2 reduces this set to that of scattered particles. Thus, the source for the next step is no longer punctual and can be implemented from this distribution of scattered particles using the state preparation method proposed in [13] at the cost of $O(2^n)$ circuit depth.

4 Results and Discussion

The measured quantum walk (MQW) algorithm is implemented using qiskit 1.4.2 [18] and QiskitRuntimeService from qiskit-ibm-runtime 0.38.0. We use IBM's tools to simulate our circuits with the BasicSimulator backend for ideal simulations and the FakeMarrakesh backend emulating the 156-qubit ibm_marrakesh Heron r2 QPU.

Fig 6 represents the collision density obtained numerically from the different classical and quantum methods. In particular, the finite difference result is cropped once the flux is below 10^{-6}. The Monte Carlo scheme considers 500.000 particles and both the ideal and noisy quantum experiments use 20.000 initial shots at each time step. For the quantum approach, the collision density is computed as the normalized sum of distributions resulting from the circuits runs. These distributions must first be normalized such that they are represented by the same number of scattered shots.



(a) Monte Carlo (b) Finite differences

(c) MQW (ideal, 8×8) (d) MQW (ideal, 16×16) (e) MQW (noisy, 8×8)

Fig. 6: Comparison of the collision flux. Results are obtained via classical methods (a) and (b), and the proposed quantum approach in ideal (c,d) settings after 29 steps and noisy (e) settings after 1 step. The noisy emulation is performed on qiskit's FakeMarrakesh backend.

We can observe that the collision density generated from the ideal simulations of the quantum algorithm on both 8×8 (c) and 16×16 (d) discretized domains successfully match the geometry since most collisions occur in the propagation medium, therefore showing an expected similarity with classical results on that regard. Only few particles manage to propagate through the absorbing medium due to the definition of the source, which only represents scattered particles in the measured quantum approach. In contrast, although the numerical experiment on IBM's FakeMarrakesh backend (e) encodes the punctual source distribution with reasonable errors, it fails to apply the quantum walk step as the result is a quasi-uniform superposition state over all positions. This result was expected due to the significantly large depth of the circuit describing the quantum walk step (Fig 2) after transpilation into the native gates of the machine (more than 1.1×10^6 gates). This main bottleneck mainly comes from the position-dependent coin operator, a repetition of $O(2^n)$ multi-controlled operations including $n+1$ control qubits (Fig 3 and Fig 4). To a smaller extent, the depth of the circuit is also impacted by the series of toffoli-based increment and decrement operators

in the shift (Fig 5) controlled by all coin qubits, and by the implementation of the source state on the position register. Further investigation is necessary to optimize the circuits implementation to reduce its depth in view of facilitating future implementations on real hardware.

Additionally, we can observe that even in the ideal simulations, particles propagate much farther than when the discretization mesh is refined. In practice, when the mesh size becomes greater than the flight distance, scattering should also occur within a single cell. The scattering probability should therefore depend on the mesh size. In this proof-of-concept algorithm, this constant scattering probability corresponds to the probability of going from one cell to the next and does not allow scattering within a single cell (common in practice) and does not allow flights of more than one cell (rare event in practice): if the coin state is $|\circlearrowleft\rangle$ then particles are supposed to be absorbed and they are no longer considered in the next step. By overcoming this bottleneck, the results are expected to be competitive with the classical results presented here.

As a first proof of concept algorithm, we can validate the encoding of the problem on a given geometry. Nevertheless, to make this algorithm more complete, a possible perspective is to adapt the coin and shift operators such that the flight distance does not remain constant but could be sampled according to its exponential distribution. One research direction consists in increasing the size of the coin register to allow multi-cell flights. Furthermore, the coin matrix definition must also be adapted to allow scattered particles to remain within their initial cell when the flight distance is small enough. These two achievements would enable a complementary quantitative comparison to classical methods and expected competitive results. Finally, while the current bottleneck for runs on real hardware remains the large depth of the circuit, further works are necessary to identify an efficient way to adapt the measured quantum approach to a fully quantum algorithm that would not require intermediate measurements in order to leverage quantum features of the algorithm and potentially recover an expected quadratic speedup.

Acknowledgments. As part of the MetriQs-France program, this work is supported by France 2030 under the French National Research Agency grant number ANR-22-QMET-0002. We acknowledge the use of IBM Qiskit services for this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum team.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Emeric Brun, F Damian, CM Diop, E Dumonteil, FX Hugot, C Jouanne, YK Lee, F Malvagi, A Mazzolo, O Petit, et al. Tripoli-4®, cea, edf and areva reference monte carlo code. In *SNA+ MC 2013-Joint International Conference on Supercomputing in Nuclear Applications+ Monte Carlo*, page 06023. EDP Sciences, 2014.

2. Paul K Romano and Benoit Forget. The openmc monte carlo particle transport code. *Annals of Nuclear Energy*, 51:274–281, 2013.
3. Joel A Kulesza, Terry R Adams, Jerawan Chudoung Armstrong, Simon R Bolding, Forrest B Brown, Jeffrey S Bull, Timothy Patrick Burke, Alexander Rich Clark, Robert Arthur Art Forster III, Jesse Frank Giron, et al. Mcnp® code version 6.3. 0 theory & user manual. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2022.
4. John Tramm, Paul Romano, Patrick Shriwise, Amanda Lund, Johannes Doerfert, Patrick Steinbrecher, Andrew Siegel, and Gavin Ridley. Performance portable monte carlo particle transport on intel, nvidia, and amd gpus. *arXiv preprint arXiv:2403.12345*, 2024.
5. Patrick Rebentrost, Brajesh Gupt, and Thomas R Bromley. Quantum computational finance: Monte carlo pricing of financial derivatives. *Physical Review A*, 98(2):022321, 2018.
6. Kazuya Kaneko, Koichi Miyamoto, Naoyuki Takeda, and Kazuyoshi Yoshino. Quantum speedup of monte carlo integration with respect to the number of dimensions and its application to finance. *Quantum Information Processing*, 20(5):185, 2021.
7. Ashley Montanaro. Quantum speedup of monte carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, Sep 2015.
8. Kwangmin Yu, Hyunkyung Lim, and Pooja Rao. Practical numerical integration on nisq devices. In *Quantum Information Science, Sensing, and Computation XII*, volume 11391, page 1139106. International Society for Optics and Photonics, 2020.
9. Mireille Coste-Delclaux, Cheikh Diop, Anne Nicolas, and Bernard Bonin. La neutronique, 2013.
10. Alain Hébert. *Applied reactor physics*. Presses inter Polytechnique, 2009.
11. Julia Kempe. Quantum random walks: an introductory overview. *Contemporary Physics*, 44(4):307–327, 2003.
12. M. Szegedy. Quantum speed-up of markov chain based algorithms. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–41, 2004.
13. Mikko Mottonen, Juha J Vartiainen, Ville Bergholm, and Martti M Salomaa. Transformation of quantum states using uniformly controlled rotations. *arXiv preprint quant-ph/0407010*, 2004.
14. Ugo Nzongani, Julien Zylberman, Carlo-Elia Doncecchi, Armando Pérez, Fabrice Debbasch, and Pablo Arnault. Quantum circuits for discrete-time quantum walks with position-dependent coin operator. *Quantum Information Processing*, 22(7):270, 2023.
15. Julien Zylberman, Ugo Nzongani, Andrea Simonetto, and Fabrice Debbasch. Efficient quantum circuits for non-unitary and unitary diagonal operators with space-time-accuracy trade-offs. *arXiv preprint arXiv:2404.02819*, 2024.
16. Konstantinos Georgopoulos, Clive Emary, and Paolo Zuliani. Comparison of quantum-walk implementations on noisy intermediate-scale quantum computers. *Phys. Rev. A*, 103:022408, Feb 2021.
17. A Didi and E Barkai. Measurement-induced quantum walks. *Physical Review E*, 105(5):054108, 2022.
18. Ali Javadi-Abhari, Matthew Treinish, Kevin Krsulich, Christopher J. Wood, Jake Lishman, Julien Gacon, Simon Martiel, Paul D. Nation, Lev S. Bishop, Andrew W. Cross, Blake R. Johnson, and Jay M. Gambetta. Quantum computing with qiskit, 2024.

Pattern Matching Using Sliding Windows on Quantum Computers

Y. Ferhi[1 [0009−0000−7290−1121]], A. Geng[1 [0000−0002−9955−4809]], M. Kiefer-Emmanouilidis[2,3 [0009−0005−7065−3511]] and A. Moghiseh[1 [0000−0001−9126−3495]]

[1] Fraunhofer Institute ITWM, Kaiserslautern, Germany
[2] DFKI, Kaiserslautern, Germany
[3] RPTU Kaiserslautern-Landau, Kaiserslautern, Germany

Abstract. Quantum hardware is progressing rapidly, but many high-performance algorithms are not yet feasible due to data embedding overhead and the limitations of NISQ devices. Quantum pattern matching algorithms, which on average outperform classical methods [26], rely on binary quantum embeddings and are not practical until we have larger qubit counts and fault-tolerant quantum computers. Although recent progress has reduced the overhead of amplitude encoding, there is still a lack of intermediate pattern matching algorithms that are efficient in terms of qubit usage. In this work, we introduce a sliding window template matching approach and explore suitable quantum primitives, including the standard and destructive swap tests, as well as the classical shadow method and the usage of a quantum artificial neuron, all of which have been performed, and evaluate their performance on NISQ devices. We integrate the swap test into a hybrid quantum support vector machine (QSVM) with a sliding window approach, improving noise resilience in quantum image template matching. We also illustrate, that our proposed methods are hardware-efficient, easy to implement, and suitable for near-term quantum devices.

Keywords: Quantum Pattern Matching · Quantum Computing · Quantum Image Processing · IBM Quantum Experience.

1 Introduction

Detecting structured patterns in noisy images is a central problem in image processing and computer vision, with applications, for example, in astrophysics [22], and biomedical imaging [15]. Classical approaches – notably template-based convolution – are reliable, but their performance degrades rapidly at low signal-to-noise ratios [17]. Quantum computing algorithms, in principle, could speed up certain calculations [26,31]. Moreover, encoding data as quantum states [1,13,20], can exploit interference and entanglement to estimate similarities via overlap measurements [23,24,28]. These methods are becoming increasingly feasible in

practice and applicable in their current form [26]. The applications span from automating defect detection in micro-fabricated components [16], real-time recognition of linear infrastructures or anomalies in drone and satellite captures [3], crack detection and segmentation in concrete surfaces [14,32], detecting vascular patterns, tumors, or lesions on MRI and CT scans [8] to identifying molecular binding motifs in microscopy imagery [5].

In this work, we use the *swap test* – both in its standard and destructive forms [7,10] – as a direct pattern matching primitive. Traditionally, swap tests estimate fidelity between unknown quantum states [9], or are used in quantum communication [4]. Here, each image patch is prepared as a quantum state, and its overlap with a template state is evaluated, where a high overlap probability indicates a pattern's presence.

We use a classical convolution baseline, applying a sliding window whose size matches that of the padded template. Additionally, we also incorporate the classical shadow framework to estimate fidelity without full tomography [33]. We compare these methods to quantum artificial neuron (QAN), which can be used as a phase-based quantum similarity test [34]. Finally, we integrate the swap tests into a hybrid Quantum Support Vector Machine (QSVM) sliding-window algorithm, using quantum overlaps to build the kernel matrix for a classical SVM classifier.

2 Quantum Overlap Tests

The presented pattern matching algorithm is based on the similarity provided by the Uhlman fidelity, commonly used in quantum information [6,27] $F(\rho,\sigma) = \left(\mathrm{Tr}\,\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}\right)^2$, which simplifies to

$$F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2,$$

for pure states $|\psi\rangle$ and $|\phi\rangle$. Classical fidelity estimation requires complete knowledge of the density matrices. In contrast, quantum techniques such as the swap test allow fidelity to be approximated with an error tolerance ϵ, requiring $\mathcal{O}(1/\epsilon^2)$ repetitions (shots) to achieve the desired accuracy.

This work evaluates quantum primitives for fidelity estimation in clean and noisy template matching tasks. We assess their performance across various image types, analyzing error rates and sensitivity to features such as brightness and shape. As base, we use four varying methods – standard and destructive swap test, QAN, classical shadow, and QSVM.

2.1 Standard and Destructive Swap Test

We explore quantum overlap estimation, beginning with the standard [27] and destructive swap test to evaluate the fidelity [29], where the destructive version uses fewer qubits than the standard one, and the same precision by the equal number of shots. We use the amplitude encoding method QPIE [35] to encode

Algorithm 1 Fidelity Estimation with Classical Shadows

Input: Quantum state initializer `psi_init`, amplitudes `amps_p`, `amps_q`, number of qubits n, number of shots s
Output: Estimated fidelity between $|\psi_p\rangle$ and $|\psi_q\rangle$

`fidelities` $\leftarrow$ empty list
for $i = 1$ to s **do**
  $b \leftarrow$ randomly sampled Pauli basis of length n
  Construct circuit qc_ϕ on n qubits and n classical bits
  Apply `psi_init(amps_p)` on qc_ϕ
  Apply basis change corresponding to b on qc_ϕ
  Measure all qubits of qc_ϕ in computational basis
  Simulate qc_ϕ with 1 shot, extract bitstring r
  Compute projector P_r^b using `snapshot_projector(b, r)`
  Construct statevector $|\psi_q\rangle$ using `psi_init(amps_q)`
  Compute fidelity estimate $f_i \leftarrow \langle\psi_q| P_r^b |\psi_q\rangle$
  Append $\mathrm{Re}(f_i)$ to `fidelities`
end for
median of `fidelities`

the classical pixel information of both the image and the patch into quantum states. The number of required qubits is $2 \cdot \log_2(p)$, or $2 \cdot \log_2(p) + 1$ if an ancilla qubit is used, where p denotes the number of pixels in the pattern padded to the next power of two. We use Qiskit's [21] Initialize function for that.

2.2 Classical Shadow Method

With the classical shadow [18,19] framework, as given in Algorithm 1, we can efficiently estimate fidelity using only a small number of randomized measurements. This approach is useful for near-term quantum devices, where full state tomography is infeasible.

The classical shadow procedure involves applying a randomly chosen unitary U (from the Clifford group [27]) to the quantum state, then measuring in the computational basis, and finally recording the outcome b and the unitary U used. By repeating this process N times, we obtain a set of measurement results $\{(U_i, b_i)\}_{i=1}^N$. From this data, we can reconstruct an estimator $\hat{\rho}$ that approximates the original state ρ.

Suppose we have two unknown pure quantum states, ρ and σ. By generating classical shadows $\{\hat{\rho}_i\}$ for ρ and $\{\hat{\sigma}_j\}$ for σ, we can estimate the fidelity using

$$\hat{F} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \mathrm{Tr}(\hat{\rho}_i \hat{\sigma}_j),$$

which approximates $\mathrm{Tr}(\rho\sigma)$ [6].

2.3 Approach Using a Quantum Artificial Neuron (QAN)

The algorithm proceeds in three principal stages [11, 12, 34]. First, an N-qubit register is initialized in the state $|0\rangle^{\otimes N}$, upon which a unitary transformation U_i is applied such that

$$U_i|0\rangle^{\otimes N} = |\psi_i\rangle. \tag{1}$$

In other words, the initial state $|0\rangle^{\otimes N}$ is transformed into the encoded input vector $|\psi_i\rangle$. This state preparation can be achieved either by employing an $m \times m$ unitary matrix whose first column corresponds to the input vector $\tilde{i}$, or by retrieving the state $|\psi_i\rangle$ directly from quantum memory.

In the second stage, a unitary operator U_w is constructed to satisfy

$$U_w|\psi_w\rangle = |1\rangle^{\otimes N} = |m-1\rangle, \tag{2}$$

thereby mapping the reference state $|\psi_w\rangle$, which encodes the weight vector $\tilde{w}$, to the computational basis state $|m-1\rangle$.

Applying both unitary operators U_w and U_i to the initialized state yields

$$U_w U_i|0\rangle^{\otimes N} = \sum_{j=0}^{m-1} c_j|j\rangle \equiv |\varphi_{i,w}\rangle, \tag{3}$$

where $c_j \in \mathbb{C}$. Using Equation (2), the scalar product between the input and weight states can then be expressed as

$$\langle\psi_w|\psi_i\rangle = \langle\psi_w|U_w^\dagger U_w|\psi_i\rangle = \langle m-1|\varphi_{i,w}\rangle = c_{m-1}. \tag{4}$$

This establishes a direct correspondence between the scalar product of the input and weight vectors and the coefficient c_{m-1} of the final state $|\varphi_{i,w}\rangle$, up to a normalization factor:

$$\tilde{w} \cdot \tilde{i} = m \langle\psi_w|\psi_i\rangle .$$

Thus, the desired overlap required for pattern matching is encoded in the amplitude c_{m-1}. For additional details on the construction of the unitary operators, we refer to [11, 12, 34].

2.4 Quantum Support Vector Machine for Cross Detection

A Quantum Support Vector Machine (QSVM) [30] is a hybrid quantum–classical algorithm that leverages a quantum computer to compute a kernel matrix K in a potentially exponential–dimensional Hilbert space, which is then used by a classical SVM [25] algorithm for classification or regression tasks. In our application, we try to make the algorithms more robust to noise. Three training sets were prepared: 200 noiseless perfect crosses, 300 noisy crosses, and 500 random negative samples. For each pair (x_i, x_j) of image patches x_i and x_j, we compute the quantum kernel entry $K_{ij} = |\langle\psi(x_i)|\psi(x_j)\rangle|^2$ using a swap test (standard or destructive).

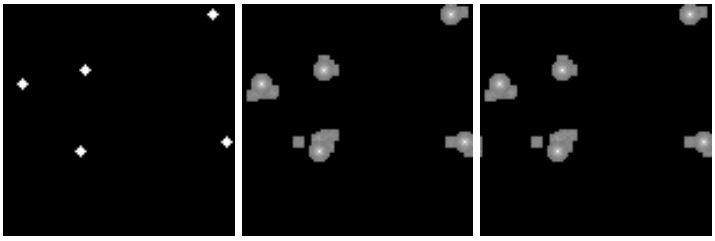


Fig. 1: 100×100 image to be analyzed (imperceptible noise) (left), standard (middle), and destructive swap test result (right).

3 Experimental Results

In the following, we present the results of our experiments, beginning with the introduction of a practical exponential threshold method for the swap tests, followed by a detailed analysis of each method on selected examples. All experiments were conducted on an Intel® Xeon® Gold 6442Y CPU (24 cores, 48 threads) operating at a base frequency of 2.6 GHz. The system included 500 GB of system memory.

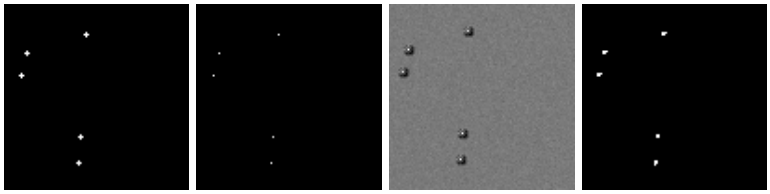
Both the standard and the destructive swap test perform comparably, as illustrated, for example, in Figure 1, for detecting a 5×5 cross pattern with gray value 1 embedded in a 100×100 image. where we added noise by changing the values of random pixels values in the range $[0, 1)$.

However, noise can interfere with the swap test, leading to artificially high overlap values in noisy regions that visually resemble the target pattern. As a result, the reconstructed image does not reliably indicate the true position of the pattern. To mitigate this issue, we introduce a probabilistic threshold on the measured swap test outcomes. The threshold is defined as

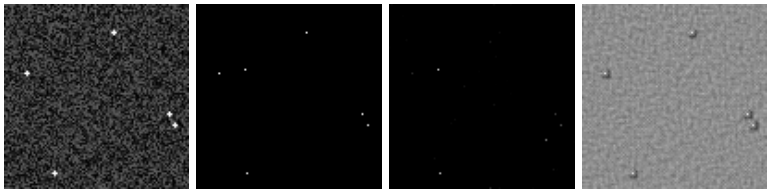
$$\tau(r, p) = \exp\big(p(r-1)\big), \tag{5}$$

where r denotes the raw swap test result normalized to values in the range $[0, 1]$ and p is a scaling parameter. This exponential form sharpens the separation between true matches and overlaps by strongly suppressing values below $r = 1$, while still retaining sensitivity near the upper range of the similarity measure. Based on experimental observations for patterns larger than 3×3, swap tests between simple patterns (such as crosses, triangles, or rhombi) and random noisy images of the same size consistently produced similarity values between 0.6 and 0.75. Thus, the threshold effectively filters out these cases, while preserving matches for identical or closely related patterns. This corresponds to a practical parameter choice of $p = 100$.

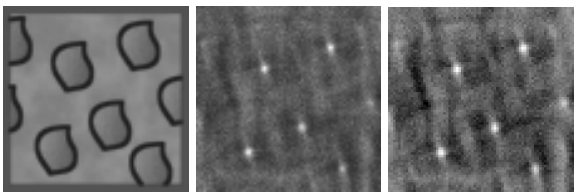
Figure 2 shows the results of the four methods described in Section 2. Since the standard and destructive swap tests yield similar outcomes, we report only the results of the standard swap test. Figure 2a illustrates the performance on a noise-free 100×100 image containing five 3×3 white crosses. As shown,



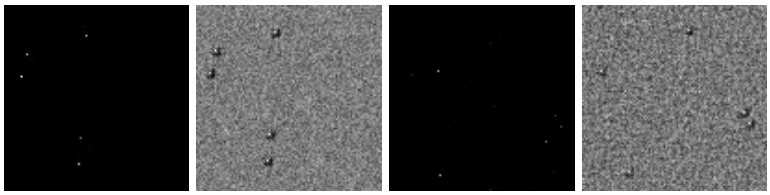
(a) Non-noisy 100×100 images with crosses. In order: Analyzed image, standard swap test, QAN's algorithm, and classical shadow (200 shots).



(b) Noisy 100×100 images with crosses. In order: Analyzed image, QSVM, swap test with exponential threshold, QAN's algorithm.



(c) 74×74 sample image. In order: analyzed image, swap test, QAN's algorithm. For a more complex pattern than a simple cross, despite the noise, the swap test needs no threshold nor QSVM to detect the pattern without false positives.



(d) In order: swap test and QAN's algorithm for non-noisy image, as in (a), then for the noisy image, as in (b) (with threshold for the swap test). The results on real hardware are not deviating much from the simulated ones.

Fig. 2: Our test results with 1024 shots, if not mentioned differently. The expected result in all samples here is a bright point in the center of each pattern. For QAN's algorithm, no threshold method is applied. (a)-(c) are simulated on Qiskit's `qasm_simulator`, (d) on IBM's real backend `ibm_aachen`.

all methods successfully detected the crosses. In the classical shadow method, however, the result exhibits an uncertainty region approximately equal in size to the pattern itself. Because generating quantum circuits for each iteration is computationally intensive, simulations quickly encounter memory limitations.

Therefore, we reduced the number of iterations in the classical shadow method to 200. Nevertheless, processing the entire 100×100-pixel noisy image, as done in the other methods (see Figure 2b), remained impractical within a reasonable computational time-frame. This observation suggests that the classical shadow method could become significantly more efficient if parallelized, for instance by executing all randomized circuits at once for each patch.

Figure 2b highlights the differences between the methods under noisy conditions. The QSVM demonstrates clearer detections and fewer artifacts than both swap test variants. The single false positive detected corresponds to a noise-generated cross of lower brightness. The swap test with exponential threshold, in contrast, produces six false positives: one caused by the low-intensity cross and five additional detections unrelated to the target pattern. These false positives arise due to normalization effects in amplitude encoding, where patches dominated by bright noise mimic the template. Future refinements, such as tuning the parameter p in Equation (5) for different pattern classes or adopting alternative noise models, could further mitigate these false detections.

The QAN algorithm yields no false positives and demonstrates strong noise resilience, both in simulations using Qiskit's `qasm_simulator` and on IBM's real backends (e.g., `ibm_aachen`; see Figure 2d). The hardware results closely match the simulated ones, and the patterns can be successfully identified in the image.

Finally, increasing the pattern size and complexity naturally reduces sensitivity to noise, as random fluctuations are increasingly unlikely to replicate larger templates (e.g., a 23×23 pattern) compared to smaller ones (e.g., a 3×3 cross). In these cases, the true positions of the larger patterns remain clearly identifiable. Elevated probabilities may still appear near image boundaries, where truncated patches resemble the target; however, these values remain consistently lower than those of correct matches. As demonstrated in Figure 2c, both the swap test and the QAN algorithm successfully identify the larger patterns under such conditions.

4 Conclusion and Discussion

The choice of the optimal quantum primitive for pattern matching depends strongly on the application context. The phase-encoding approach, as employed in QAN's algorithm, is sensitive to brightness disparities between the pattern and the image; consequently, it does not detect randomly generated low-intensity crosses. This method offers rapid execution and robustness to noise, reliably identifying only the desired pattern. In contrast, swap tests can be used in two ways: with an exponential threshold to detect visually similar simple patterns (at the cost of false positives, e.g., six detections among the 10 000 patches in Figure 2b), or without a threshold for sufficiently complex patterns. Increasing the number of shots further improves swap test accuracy by reducing the rate of false positives. The single false positive found by the QSVM, however, cannot be avoided, as the corresponding cross is genuinely present in the data. Unlike QAN, both swap tests and the QSVM identify shapes resembling the target pattern, includ-

ing partial or incomplete matches. Because amplitude encoding represents the total brightness of each extracted patch, swap tests are inherently insensitive to illumination variations while requiring minimal quantum hardware resources. Importantly, their practicality is supported by experiments on real hardware (e.g., `ibm_aachen`), which produced results consistent with those obtained in simulation.

Training a QSVM kernel for classification enables finer control over which patterns are accepted or rejected as matches. While the kernel training time remains moderate (on the order of a few hours), it is not negligible. Moreover, excessively large kernels can lead to overfitting. Comparable performance can often be achieved with smaller training sets, for example, 300 samples. Although no universal rule exists to determine the optimal kernel size and avoid over- or underfitting, established approaches such as the Nyström method [2] can be employed to reduce dataset size and computational cost.

SVMs are particularly appealing when the same kernel can be reused across multiple images—for instance, when detecting the same pattern in a sequence of frames. In other contexts, where noise resilience is paramount, QAN's method or swap tests with exponential thresholding may provide sufficient accuracy. By contrast, fidelity estimation with classical shadows proves impractical for this task: although only a few measurements are required, each necessitates generating a separate circuit, resulting in significantly lower efficiency compared to other approaches.

In summary, we have demonstrated how hardware-efficient quantum algorithms for pattern matching can be effectively applied on current NISQ devices. Each primitive presents distinct trade-offs, and its suitability depends on the complexity of the target pattern, the noise characteristics of the data, and the computational resources available.

Acknowledgment

This work was supported in part by the Quantum Initiative Rheinland-Pfalz (QUIP). It was also partially funded by the Federal Ministry for Economic Affairs and Climate Action through the EniQmA project (funding number 01MQ22007A).

References

1. Araujo, I.F., Blank, C., Araújo, I.C.S., da Silva, A.J.: Low-Rank Quantum State Preparation. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **43**(1), 161–170 (2024). `https://doi.org/10.1109/TCAD.2023.3297972`
2. Bach, F.R., Jordan, M.I.: Predictive low-rank decomposition for kernel methods. In: Proceedings of the 22nd international conference on Machine learning - ICML '05. pp. 33–40. ACM Press, Bonn, Germany (2005). `https://doi.org/10.1145/1102351.1102356`
3. Bozcan, I., Kayacan, E.: UAV-AdNet: Unsupervised anomaly detection using deep neural networks for aerial surveillance. In: 2020 IEEE/RSJ international conference

on intelligent robots and systems (IROS). pp. 1158–1164. IEEE (2020). https://doi.org/10.1109/IROS45743.2020.9341790
4. Buhrman, H., Cleve, R., Watrous, J., De Wolf, R.: Quantum Fingerprinting. Physical Review Letters **87**(16), 167902 (2001). https://doi.org/10.1103/PhysRevLett.87.167902
5. Caicedo, J.C., Cooper, S., Heigwer, F., Warchal, S., Qiu, P., Molnar, C., Vasilevich, A.S., Barry, J.D., Bansal, H.S., Kraus, O.: Data-analysis strategies for image-based cell profiling. Nature methods **14**(9), 849–863 (2017). https://doi.org/10.1038/nmeth.4397
6. Chen, R., Song, Z., Zhao, X., Wang, X.: Variational quantum algorithms for trace distance and fidelity estimation. Quantum Science and Technology **7**(1), 015019 (2021). https://doi.org/10.1088/2058-9565/ac38ba
7. Das, S., Zhang, J., Martina, S., Suter, D., Caruso, F.: Quantum pattern recognition on real quantum processing units. Quantum Machine Intelligence **5**(1), 16 (2023). https://doi.org/10.1007/s42484-022-00093-x
8. El-Deredy, W.: Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR in Biomedicine **10**(3), 99–124 (1997). https://doi.org/10.1002/(SICI)1099-1492(199705)10:3<99::AID-NBM461>3.0.CO;2-#
9. Fanizza, M., Rosati, M., Skotiniotis, M., Calsamiglia, J., Giovannetti, V.: Beyond the Swap Test: Optimal Estimation of Quantum State Overlap. Physical Review Letters **124**(6), 060503 (2020). https://doi.org/10.1103/PhysRevLett.124.060503
10. Garcia-Escartin, J.C., Chamorro-Posada, P.: swap test and Hong-Ou-Mandel effect are equivalent. Physical Review A **87**(5), 052330 (2013). https://doi.org/10.1103/PhysRevA.87.052330
11. Geng, A.: Application of hybrid quantum machine learning for image processing in the NISQ era. PhD Thesis, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (2024). https://doi.org/10.26204/KLUEDO/8207
12. Geng, A., Moghiseh, A., Redenbach, C., Schladitz, K.: A hybrid quantum image edge detector for the NISQ era. Quantum Machine Intelligence **4**(2), 15 (2022). https://doi.org/10.1007/s42484-022-00071-3
13. Geng, A., Moghiseh, A., Redenbach, C., Schladitz, K.: Improved FRQI on superconducting processors and its restrictions in the NISQ era. Quantum information processing **22**(2) (2023). https://doi.org/10.1007/s11128-023-03838-0
14. Geng, A., Moghiseh, A., Redenbach, C., Schladitz, K.: Hybrid quantum transfer learning for crack image classification on NISQ hardware. In: AIP Conference Proceedings. vol. 3182, p. 130001. AIP Publishing LLC (2025). https://doi.org/10.1063/5.0246500
15. Glumov, N.I., Kolomiyetz, E.I., Sergeyev, V.V.: Detection of objects on the image using a sliding window mode. Optics & Laser Technology **27**(4), 241–249 (1995). https://doi.org/10.1016/0030-3992(95)93752-D
16. Guijo, D., Onofre, V., Bimbo, G.D., Mugel, S., Estepa, D., Carlos, X.D., Adell, A., Lojo, A., Bilbao, J., Orus, R.: Quantum artificial vision for defect detection in manufacturing (2024). https://doi.org/10.48550/arXiv.2208.04988
17. Hashemi, N.S., Aghdam, R.B., Ghiasi, A.S.B., Fatemi, P.: Template Matching Advances and Applications in Image Analysis (2016). https://doi.org/10.48550/arXiv.1610.07231
18. Huang, H.Y.: Learning quantum states from their classical shadows. Nature Reviews Physics **4**(2), 81–81 (2022). https://doi.org/10.1038/s42254-021-00411-5

19. Huang, H.Y., Kueng, R., Preskill, J.: Predicting many properties of a quantum system from very few measurements. Nature Physics **16**(10), 1050–1057 (2020). https://doi.org/10.1038/s41567-020-0932-7
20. Iten, R., Colbeck, R., Kukuljan, I., Home, J., Christandl, M.: Quantum circuits for isometries. Physical Review A **93**(3), 032318 (2016). https://doi.org/10.1103/PhysRevA.93.032318
21. Javadi-Abhari, A., Treinish, M., Krsulich, K., Wood, C.J., Lishman, J., Gacon, J., Martiel, S., Nation, P.D., Bishop, L.S., Cross, A.W., Johnson, B.R., Gambetta, J.M.: Quantum computing with Qiskit (2024). https://doi.org/10.48550/ARXIV.2405.08810
22. Juvela, M.: Template matching method for the analysis of interstellar cloud structure. Astronomy & Astrophysics **593**, A58 (2016). https://doi.org/10.1051/0004-6361/201628727
23. Kankeu, I., Fritsch, S.G., Schönhoff, G., Mounzer, E., Lukowicz, P., Kiefer-Emmanouilidis, M.: Quantum-inspired Embeddings Projection and Similarity Metrics for Representation Learning (2025). https://doi.org/10.48550/arXiv.2501.04591
24. Liu, W., Li, Y.Z., Yin, H.W., Wang, Z.R., Wu, J.: Quantum multi-state Swap Test: an algorithm for estimating overlaps of arbitrary number quantum states. EPJ Quantum Technology **11**(1), 46 (2024). https://doi.org/10.1140/epjqt/s40507-024-00259-5
25. Mammone, A., Turchi, M., Cristianini, N.: Support vector machines. WIREs Computational Statistics **1**(3), 283–289 (2009). https://doi.org/10.1002/wics.49
26. Montanaro, A.: Quantum Pattern Matching Fast on Average. Algorithmica **77**(1), 16–39 (2017). https://doi.org/10.1007/s00453-015-0060-4
27. Nielsen, M.A., Chuang, I.L.: Quantum computation and quantum information. Cambridge university press (2010)
28. Nishimura, H.: A Survey: SWAP Test and Its Applications to Quantum Complexity Theory. In: Algorithmic Foundations for Social Advancement: Recent Progress on Theory and Practice, pp. 243–261. Springer Nature Singapore Singapore (2025). https://doi.org/10.1007/978-981-96-0668-9_16
29. Niu, S., Suau, A., Staffelbach, G., Todri-Sanial, A.: A hardware-aware heuristic for the qubit mapping problem in the NISQ era. IEEE Transactions on Quantum Engineering **1**, 1–14 (2020). https://doi.org/10.1109/TQE.2020.3026544
30. Rebentrost, P., Mohseni, M., Lloyd, S.: Quantum Support Vector Machine for Big Data Classification. Physical Review Letters **113**(13), 130503 (2014). https://doi.org/10.1103/PhysRevLett.113.130503
31. Sasaki, M., Carlini, A., Jozsa, R.: Quantum Template Matching. Physical Review A **64**(2), 022317 (2001). https://doi.org/10.1103/PhysRevA.64.022317
32. Srinivasan, A., Geng, A., Macaluso, A., Kiefer-Emmanouilidis, M., Moghiseh, A.: A comparative study of q-seg, quantum-inspired techniques, and u-net for crack image segmentation **92**(7), 321–331 (2025). https://doi.org/10.1515/teme-2025-0017
33. Struchalin, G., Zagorovskii, Y.A., Kovlakov, E., Straupe, S., Kulik, S.: Experimental Estimation of Quantum State Properties from Classical Shadows. PRX Quantum **2**(1), 010307 (2021). https://doi.org/10.1103/PRXQuantum.2.010307
34. Tacchino, F., Macchiavello, C., Gerace, D., Bajoni, D.: An artificial neuron implemented on an actual quantum processor. npj Quantum Information **5**(1), 1–8 (2019). https://doi.org/10.1038/s41534-019-0140-4
35. Yao, Xi-Wei et al.: Quantum Image Processing and Its Application to Edge Detection: Theory and Experiment (2017). https://doi.org/10.1103/PhysRevX.7.031041

Solving the 3D Heat Equation with VQA via Remeshing-Based Warm Starts

Samuel Donachie[1 [0009−0006−7033−2838]], Ulysse Remond[1,2 [0000−0001−5925−1224]], Arthur Mathorel[1 [0009−0009−1426−1205]], and Kyryl Kazymyrenko[1 [0000−0001−5429−4348]]

[1] EDF R&D, 7 Bd Gaspard Monge, 91120 Palaiseau, France
[2] LUX, Sorbonne University, 4 Pl. Jussieu, 75005 Paris France

Abstract. Quantum computing holds great promise for solving classically intractable problems such as linear systems and partial differential equations (PDEs). While fully fault-tolerant quantum computers remain out of reach, current noisy intermediate-scale quantum (NISQ) devices enable the exploration of hybrid quantum-classical algorithms. Among these, Variational Quantum Algorithms (VQAs) have emerged as a leading candidate for near-term applications. In this work, we investigate the use of VQAs to solve PDEs arising in stationary heat transfer. These problems are discretized via the finite element method (FEM), yielding linear systems of the form Ku=f, where K is the stiffness matrix. We define a cost function that encodes the thermal energy of the system, and optimize it using various ansatz families. To improve trainability and bypass barren plateaus, we introduce a remeshing strategy which gradually increases resolution by reusing optimized parameters from coarser discretizations. Our results demonstrate convergence of scalar quantities with mesh refinement. This work provides a practical methodology for applying VQAs to PDEs, offering insight into the capabilities and limitations of current quantum hardware.

Keywords: Variational Quantum Algorithm · Mesh Refinement · Quantum Computer Benchmark.

1 Introduction

In the current state of quantum computing, fully fault-tolerant quantum computers (FTQCs) remain out of reach [3]. Instead, we operate in the so-called NISQ (Noisy Intermediate-Scale Quantum) era, characterized by quantum devices that are both limited in qubit count and prone to noise. Despite these constraints, NISQ devices can still be leveraged for meaningful tasks [8].

As the timeline toward FTQC remains uncertain, both academic and industrial communities are actively exploring use cases that can provide value using today's imperfect quantum hardware. Surpassing classical solutions typically requires a large number of qubits —something that current hardware is not yet capable of. While classical solvers are powerful, they often struggle with scalability and can require extensive computational resources. This challenge, called the curse of dimensionality, refers to the growth in computational complexity as the number of variables and/or degrees of freedom increases. The typical size of challenging industrial simulations reaches 10^6 degrees of freedom. Quantum computing offers a potential way to overcome this limitation: with n qubits, one can represent 2^n states, enabling the compact encoding of high-dimensional vectors and matrices. With 50 qubits, we expect to show one day quantum supremacy on some of those complex problems [14]. Among the many problems of interest, solving linear equations plays a central role in scientific computing, engineering, and industry. In particular, many physical systems are modeled by Partial Differential Equations (PDEs), which are typically discretized

into large linear systems. Efficiently solving these is therefore currently critical for simulations in fields as diverse as fluid dynamics [20], structural mechanics [2], and electromagnetism [9].

One particularly active area is optimization, where quantum computing may already offer practical benefits through the use of Variational Quantum Algorithms (VQAs)[5]. These hybrid quantum-classical algorithms are well-suited for NISQ devices and have shown promise in addressing linear algebra problems [17,10,11,4,6]. In this work, we investigate the use of VQAs for solving PDE-derived linear systems for a 3D object, with a focus on designing ansätze that are both expressive and hardware-efficient. To this end, we also introduce a cascading remeshing strategy, aimed at improving the convergence and stability of VQA-based linear equation solvers. The goal of this strategy is to use a coarse quantum representation of a solution, requiring less qubits, as a warm start in a more refined mesh. This procedure creates tailored warm starts for PDE-solving VQAs.

Finally, recognizing that current quantum hardware may not yet support the full execution of such VQAs at scale, we propose to use the ideal final state obtained from our simulated VQA as a benchmark. Since this state corresponds to the solution of a well-defined physical problem (in our case, the 3D heat equation $\Delta u = f$), it provides a concrete, structured target for quantum hardware. We suggest that reproducing such physically meaningful states could serve as a valuable and application-driven benchmark for future quantum devices.

2 Problem and Method

2.1 Problem definition

Our objective here is to solve the heat equation in 3D : $\Delta u = f$ with u being the temperature field and f the heat source term. One common approach to solve Partial Differential Equations is to use the Finite Element Method (FEM), which discretizes the PDE into a linear system[17]. For instance, in the case of the heat equation, the system can be written as:

$$K\boldsymbol{u} = \boldsymbol{f} \tag{1}$$

Here, K is the stiffness matrix arising from the FEM discretization [2]. It encodes the relationships between the degrees of freedom in the system and reflects both the geometry of the domain and the material or physical properties involved. In the quantum setting, the solution vector $\boldsymbol{u}$ and the source term (charge vector) $\boldsymbol{f}$ are encoded as quantum states $|u\rangle$ and $|f\rangle$, enabling a compact representation and allowing the use of variational algorithms to approximate the solution. Putting this equation into its variational form leads us to define the energy E_c of a given trial state $|x\rangle$:

$$E_c(|x\rangle) = \frac{1}{2}\langle x| K |x\rangle - \frac{\langle x|f\rangle + \langle f|x\rangle}{2} \tag{2}$$

We use the energy E_c as a cost function, which is minimized by the solution vector $|u\rangle$. However, we're working with quantum states, which are normalized, whereas the true solution vector $\boldsymbol{u}$ is not necessarily normalized. We therefore introduce a parameterized quantum state $|\psi(\theta)\rangle$, a norm $r \in \mathbb{R}$, such as a trial state is expressed as $|x\rangle = r\,|\psi(\theta)\rangle$. The equation we need to optimize then becomes:

$$|u\rangle = argmin_{r,\theta}\left(\frac{1}{2}r^2\langle\psi(\theta)| K |\psi(\theta)\rangle - r\frac{\langle\psi(\theta)|f\rangle + \langle f|\psi(\theta)\rangle}{2}\right) = \Re(r_{opt}\,|\psi(\theta_{opt})\rangle) \tag{3}$$

Analytically minimizing this variational equation with respect to r, prior to θ, we obtain :

$$r_{opt} = \frac{\langle\psi(\theta)|f\rangle + \langle f|\psi(\theta)\rangle}{2\langle\psi(\theta)| K |\psi(\theta)\rangle} \tag{4}$$

Leading us to the final expression of the cost function we use in our VQA [17]:

$$C(\psi(\theta)) = -\frac{1}{8}\frac{(\langle\psi(\theta)|f\rangle + \langle f|\psi(\theta)\rangle)^2}{\langle\psi(\theta)|\,K\,|\psi(\theta)\rangle} \tag{5}$$

By finding the state that minimizes this equation, we obtain the solution to our problem. Variational Quantum Algorithms (VQAs) leverage a hybrid quantum-classical loop to perform this minimization: a classical optimizer proposes parameters θ for a parameterized quantum circuit (ansatz), which is executed on a quantum processor. The resulting state is measured, and the outcome of the cost function C seen in (5) is used by the classical optimizer to update the parameters. Performance depends on each component of the loop: the choice of ansatz, the classical optimizer, and the measurement strategy. As with classical optimizers and preconditioning, different tradeoffs between the complexities of the quantum computations regarding K, $|f\rangle$ and $|u\rangle$ are possible. As a comparatively low number of parameters can be stored to represent large states, QPUs provide a sizeable advantage with regard to the efficient generation and evaluation of high-dimensional states. It is also possible to interact with K and $|f\rangle$ using measurements, scaling polylogarithmically with system size [11,16]. It is worth noting that a high fidelity with the target state $|\langle\psi(\theta|u\rangle|^2$ does not necessarily imply convergence in energy 3.2.

2.2 The choice of the Ansatz

In a Variational Quantum Algorithm (VQA), an ansatz is a parameterized quantum circuit designed to generate trial quantum states. It serves as a flexible model that is potentially optimized to approximate the solution to a given problem. The expressiveness and structure of the ansatz critically affect the performance and accuracy of the algorithm [18]. Firstly, since one of the main goals of this project is to ensure compatibility with today's NISQ devices, the ansatz must be designed with hardware constraints in mind. This means using quantum gates with high fidelity and limiting the overall circuit depth. Typically, ansätze consist of alternating layers of rotation gates and layers of entangling gates. It is well known that entangling gates—especially those involving two or more qubits—are significantly more error-prone than single-qubit gates, with their implementation difficulty increasing rapidly with the number of qubits involved [1,13].

We employ the NLocal class from Qiskit (Qiskit version: 1.4.2) [15] to construct our ansatz.The circuit is constructed by alternating layers of parameterized rotation and entangling gates, with a final layer of rotation gates. In each entangling layer, a linear chain of CNOTs is applied, where every qubit except the last controls an X gate on the subsequent qubit. This pattern is repeated throughout the circuit, while the final layer contains only rotation gates. When two rotation gates are used, each is applied in a separate layer. Using this Qiskit class, we constructed and compared a limited selection of ansätze, each chosen based on specific criteria relevant to our problem. Since we are addressing a heat diffusion problem, it seems that the target quantum state should be real-valued. All numerical results reported in the next section are obtained from noiseless state-vector simulations where the optimizer provides parameters to our simulated PQC, producing the ideal resulting state. The cost function (5) is then computed from this state, and the optimizer adjusts the parameters accordingly, repeating the process until it found the optimal set. We used the NEWUOA[19] gradient-free optimization algorithm throughout this article. We studied an inhomogeneous problem to showcase that our model can handle arbitrary boundary conditions. Specifically, we imposed Dirichlet conditions, by fixing the temperature on two faces of the 3D object, and Neumann conditions consisting in a uniform heat flux on an entire third face and an additional flux over half of a fourth face. Then, using Qiskit's NLocal, we benchmark the following ansätze —on 6 qubits with 36 parameters— each relevant to a real-valued heat-diffusion solution:

Table 1: List of Ansätze candidates we considered

	Name	Rotation Layer	Entangling Layer	Comments
1	Ry-only	RY	—	This baseline illustrates the limitations of an ansatz without entanglement.
2	Ry-Cnot	RY	CNOT	The simplest real-valued ansatz: RY yields purely real amplitudes, and the we chose CNOT as the 2 qubit entangling gate.
3	Ry-Toffoli	RY	CCNOT	Replaces the CNOT with a more complex three-qubit gate to test whether deeper entanglement outweighs its higher error rate.
4	Ry-1Cnot	RY	CNOT	Rotation layers: RY(θ); exactly one CNOT layer inserted just before the final rotation layer. Provides sparse entanglement while keeping depth minimal.
5	RxRy-Cnot	RX then RY	CNOT	Adding RX enlarges the reachable Hilbert space potentially giving the optimizer better control through a more expressible ansatz [12], it introduces complex amplitudes.
6	RxRy-Toffoli	RX then RY	CCNOT	Same expressiveness as above but with a higher-order entangling gate.

Let us note that we could have used fewer parameters for Ansatz 1 and 4 (namely 6 and 12, respectively). Given the small number of parameters, this did not affect the results presented below. However, if we had chosen those ansätze, we would have used 6/12 parameters in the later stages.

3 Results

3.1 Cold start

To compare these ansätze on equal footing, we tuned each circuit so that the total number of trainable parameters was approximately the same, instead of fixing the number of layers. We then ran the optimization 100 times on 6 simulated noiseless qubits, each time starting from random initial parameters θ generated using a pseudo-random number generator (PCG64, from NumPy version: 1.23.5). It is worth noting that changing the number of layers or trainable parameters can sometimes have a drastic impact on the performance of an ansatz. We also note that we let the optimizer run until it could no longer find any improvement, using an absolute tolerance of 10^{-12} and a relative tolerance of 10^{-9}, or until it was forcefully stopped after 2 hours. We can interpret these thresholds as resource limits. This second condition was not triggered until we reached 12 qubits. In order to compare them, we used the accuracy of the energy as a metric of convergence. Energy accuracy is defined as $\frac{C(|\psi(\theta)\rangle)}{C(|u\rangle)} \times 100\%$ reflecting how closely the optimization converges to the true solution. This metric is designated as "Energy" in the tables.

Table 2: Ansätze scores on 6 qubits (averaged over 100 2h-optimization runs, each with 36 parameters).

	Ansatz	Mean Energy (%)	NEWUOA Iterations	Max Energy (%)	Min Energy (%)
1	Ry	88.00 % ± 0.32	452 ± 82	91.17 %	87.97 %
2	Ry-Cnot	86.7 % ± 6.7	2645 ± 1331	96.89 %	68.70 %
3	Ry-Toffoli	86.31 % ± 6.52	2430 ± 1203	96.70 %	71.92 %
4	Ry-1Cnot	86.1 % ± 11.0	1104 ± 525	93.38 %	50.75 %
5	Rx-Ry-Cnot	93.0 % ± 4.2	3887 ± 1875	97.14 %	54.33 %
6	Rx-Ry-Toffoli	93.0 % ± 1.7	5793 ± 2720	97.17 %	88.48 %

Compared to the 6 qubits-discretized target energy all ansatz achieve 85% and above mean energy accuracy. We observe that multi-qubit entangling gates beyond pairs do not provide a noticeable advantage at this scale Rather than requiring every run to succeed perfectly, we only need most runs to perform well and at least one to converge very well. Based on this, we can distinguish three performance tiers among the ansätze. On average, using 2 rotation gates seems to help but considering only the maximum accuracy, it doesn't have such an impact. These performance insights can help guide the choice of quantum hardware when deploying the VQA in practice. For the rest of the article, thanks to the scores we obtained, we chose to work with the ansätze 2 and 5. We won't be using Toffoli gates since we can see no clear advantage at this scale, moreover, they perform usually worse than the other existing gates on current hardware. We could also take the optimization time into account by considering the number of iterations required to reach the final result. Adding more gates tends to increase this number, and since the overall waiting time scales with the size of the system, this becomes an important factor to consider. Therefore, we need to be cautious when adding complexity to the circuit.

When examining these results, one might notice that some runs perform quite poorly, with minimum accuracies sometimes dropping to around 50%. This is due to local minima and/or barren plateaus, where the optimizer gets stuck in regions of vanishing gradients. This accuracy can still be considered high and textbook strategies like a restart could help enhance this score. However, in this case and for our sake, it might not be a major concern, as some runs still manage to perform very well. The problem arises as the system becomes bigger, requiring more qubits: barren plateaus become more prevalent and harder to avoid. Such random cold start-based runs with 9 to 15 qubits can be observed in Table 3. For 9 qubits we had 108 parameters, for 12 qubits we had 252 parameters and for 15 qubits, we had 330 parameters. We will keep the same number of parameters for the rest of the article.

Table 3: Ansätze scores with cold starts (averaged over 100 2h-optimization runs).

Qubits	Par	Ansatz	Mean Energy (%)	NEWUOA Iterations	Max Energy (%)	Min Energy (%)
9	108	Rx-Ry-Cnot	83.3 ± 5.5	8506 ± 7313	92.9	78.5
9	108	Ry-Cnot	23.6 ± 2.4	4106 ± 1308	29.4	17.6
12	252	Rx-Ry-Cnot	38.0 ± 32	11716 ± 4160	66.9	0.70
12	252	Ry-Cnot	4.0 ± 0.7	4625 ± 2965	5.3	2.5
15	330	Rx-Ry-Cnot	0.21 ± 0.44	5352 ± 714	2.71	0.02
15	330	Ry-Cnot	0.53 ± 0.03	4867 ± 796	0.57	0.25

The results show lower convergence when working with a more complex system of 9 qubits especially with the second ansatz Ry-Cnot. Using 12 qubits or more, the optimizer always fails to find a way to optimize the parameters and/or it stops after the 2-hour limit we imposed. We can notice a great difference between the two ansatz for 12 qubits. The Ry-Cnot ansatz showcases very low accuracy as it is twice as deep, and only outputs real states making it less expressible. The largest systems' low accuracy can also be explained by the optimizer getting stuck on a barren plateau and/or in local minima. In such cases, several classical strategies can accelerate convergence like dual-objective optimization, restarting the optimization process, or a warm start.

3.2 Uniform start

In a warm start, the initial state is already close to the target, which helps the optimizer find better parameters faster. Since we optimize parameters and not quantum states directly, we need to identify both a state near the objective and a parameterization within our ansatz that produces it. Our first strategy is the **Uniform start**. The parameters of the first layer set a

Hadamard-like transformation, by setting the parameters of R_y gates of the first layer to $\frac{\pi}{2}$ and the R_x gates to 0, while the remaining parameters are initialized randomly from [-0.01,0.01]:

$$R_y(\theta) = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}, \quad R_x(\theta) = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & -i\sin\left(\frac{\theta}{2}\right) \\ -i\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}, \quad R_y\left(\frac{\pi}{2}\right)|0\rangle = d\,|+\rangle \tag{6}$$

This prepares a near to equal superposition $\approx |+\rangle^{\otimes n}$, a more structured start than full randomness. In physical terms, it is like setting all components of the system close to the same temperature, rather than assigning random temperatures to each one. This uniform starting condition can help the optimizer avoid getting lost in barren plateaus by giving it a more meaningful direction to begin with.

Table 4: Ansätze scores with Uniform starts (averaged over 100 2h-optimization runs).

Qubits	Par	Ansatz	Mean Energy (%)	NEWUOA Iterations	Max Energy (%)	Min Energy (%)
9	108	Rx-Ry-Cnot	92.4 ± 0.77	5489 ± 3162	94.4	85.9
9	108	Ry-Cnot	92.3 ± 6.2	4727 ± 2381	94.2	33.4
12	252	Rx-Ry-Cnot	71.6 ± 23	4665 ± 2288	91	0.32
12	252	Ry-Cnot	83.7 ± 20	7409 ± 2583	92.2	11.1
15	330	Rx-Ry-Cnot	46.7 ± 2.90	5585 ± 594	51	35
15	330	Ry-Cnot	54.71 ± 5.02	4707 ± 240	63	44

The results improve with regard to the cold starts. Nevertheless, as the system becomes more complex, even this form of structured initialization begins to lose effectiveness. It still works better than a cold start; for 12 and 15 qubits, most of the runs that were doing well stopped because of the time running out and not because the optimizer couldn't find better parameters. In the 15-qubit case using the Ry–Cnot ansatz, the mean fidelity reaches 95% even though energy convergence is poor, highlighting that fidelity can be misleading as a performance metric. Moreover, convergences using fidelity disregard the physics of the object, leading to discontinuous states, and therefore cannot be used as a cost function [16]. It will be useful later on when evaluating the performances of a QPU's output. Usual warm starts are inherently problem-dependent, since they leverage specific knowledge about the target state or solution landscape. For instance, in combinatorial optimization, a warm start might be obtained by using a classical heuristic solution mapped into the quantum circuit's parameter space[7]. This start only induces a uniform state, meaning that it merely leverages broad information about the solution's smoothness. That said, it remains valuable in scenarios like ours, where we are primarily concerned with properties such as the temperature of an object—typically a uniform value across the object—rather than a more specialized configuration.

In that sense, Hadamard initialization can be better described as a structured cold start: it provides a more uniform and reproducible starting point than random parameters, but it still lacks the informed guidance that characterizes a true warm start. To overcome this limitation, we introduce the Cascade protocol.

3.3 The Cascade Protocol

To choose our ansatz, we compared the performance of various circuits, and all of them performed well on smaller systems (6 qubits). However, as we increase the number of qubits, the Hilbert space grows exponentially, and the optimizer struggles more. The quantum states become more complex, and even the number of parameters becomes harder to manage. In this regime, a warm start is currently one of the only viable strategies to get a VQA to converge effectively. Inspired by recent results on quantum solutions in structural mechanics [16] obtained by our

research group, we propose a novel strategy to improve the scalability of variational quantum algorithms (VQAs) by extending optimized solutions obtained on smaller, tractable systems as structured initializations for larger problems. The central idea is to use the solution of a lower-dimensional instance as a coarse-grained approximation, which can then be embedded and refined within a higher-dimensional Hilbert space. To illustrate this, consider a 6-qubit system whose 64 computational basis states can be naturally interpreted as a 4×4×4 3D grid:

$$|u\rangle_{6qb} = \sum_{\bar{x},\bar{y},\bar{z}} |z_{coarse} z_{fine} y_{coarse} y_{fine} x_{coarse} x_{fine}\rangle \text{ , where } \bar{w} = w_{coarse} w_{fine} \in \{0,1\} \times \{0,1\} \tag{7}$$

This smaller problem can typically be addressed efficiently either using classical techniques and quantum state preparation, or using a simpler VQA. To extend the solution to a more complex instance —such as a 9-qubit system representing an 8×8×8 grid— we construct an initialization that embeds the optimized 6-qubit state into the larger state space. Concretely, each basis state of the smaller system becomes embedded in a local 2×2×2 neighborhood of the larger configuration, effectively increasing the resolution of the space. This hierarchical strategy provides a structured and informed starting point, potentially enhancing convergence and solution quality in the variational optimization. This Cascade protocol creates a natural way to scale the problem up while preserving useful structure from the lower-dimensional solution. To visualize a bit better how it works, here is a remeshing of a 2D grid.

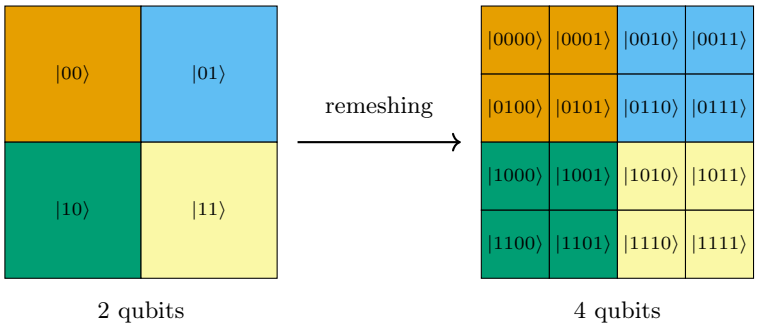


Fig. 1: Remeshing from a 2-qubit 2D-grid into a 4-qubit 2D-grid. Each 2-qubit state expands into a structured patch of finer resolution made of $2^2 = 4$ states.

Since we want to keep the same element geometry (a cube) we need to add 8 new states. In order to obtain those 8 new finite elements per old element, we need to introduce 3 additional qubits, since $2^3 = 8$. Suppose that an optimized solution has already been obtained for the 6-qubit instance of the problem. To take advantage of this solution when scaling up, we begin by initializing the full 9 qubit system in the state

$$|u\rangle_{6qb} \otimes |000\rangle \tag{8}$$

Next, we apply Hadamard gates to each of the 3 newly added qubits. This creates a uniform superposition over all their basis states, yielding:

$$|u\rangle_{6qb} |+\rangle |+\rangle |+\rangle = |u\rangle_{6qb} \otimes \frac{1}{\sqrt{8}} \sum_{i=0}^{7} |i\rangle \tag{9}$$

However, unlike in one-dimensional systems —where an additional qubit can simply be appended to the end of the register— in three-dimensional settings, the mapping between basis states and spatial positions is inherently more complex. In particular, the tensor product structure does not directly preserve spatial locality across axes. To ensure that newly added qubits are correctly positioned to encode higher-resolution features along each spatial dimension, it is necessary to introduce a sequence of SWAP gates. These gates effectively rearrange the qubits, allowing the additional degrees of freedom to control fine-grained resolution along the x, y, and z axes of the enlarged spatial grid. Concretely, we can think of the 6-qubit state as encoding coordinates with pairs of qubits representing coarse and fine positions (7) as $\bar{w} = w_{coarse}w_{fine} \in \{0,1\} \times \{0,1\}$. The procedure can be illustrated as $|\bar{z}\rangle|\bar{y}\rangle|\bar{x}\rangle \overset{+3}{\mapsto} |\bar{z}\rangle|\bar{y}\rangle|\bar{x}\rangle|000\rangle \overset{H}{\mapsto} |\bar{z}\rangle|\bar{y}\rangle|\bar{x}\rangle|+\rangle|+\rangle|+\rangle \overset{swap}{\mapsto} |\bar{z}\rangle|+\rangle|\bar{y}\rangle|+\rangle|\bar{x}\rangle|+\rangle$.
By carefully swapping, we ensure the new qubits replace the fine control qubits where higher resolution is needed. Nevertheless, this step is unnecessary as you could construct your circuit to leave unused wires in the right spots. Using this technique, we should get a warm start by remeshing the solution of a previous optimization with a lower resolution.

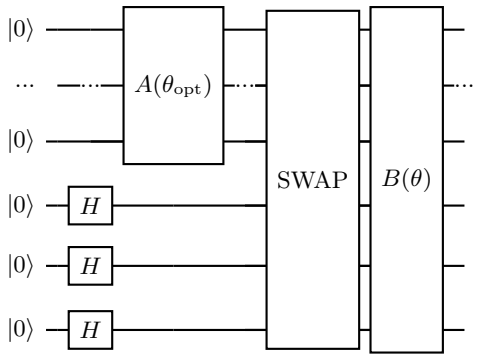


Fig. 2: Circuit of the remeshing cascade strategy using previous results $A(\theta_{\text{opt}})$ and superposition expansion via Hadamard gates and swaps.

We define the second ansatz $B(\theta)$ to be the composition of two NLocal ansätze: $Ansatz(\theta) \times Ansatz(0)^{\dagger}$, such that $B(0) = \mathbb{I}$. If it is possible to implement an identity gate $\mathbb{I}$ easily with a specific ansatz by simply setting the correct first parameters, this last part isn't needed. However, composing an ansatz in this way simplifies the process, as we only need to initialize all the parameters θ to 0 to get $\mathbb{I}$, but obviously complexifies greatly the circuit. We can now compare the results obtained using the Cascade strategy to the Uniform start and the cold start from 9 to 15 qubits on our 3D object.

Using this strategy, we ran again the optimizations with the two different ansätze. Let's note that the parameters chosen for the first part of the Cascade protocol are the one obtained through a previous optimization on a coarser system. As shown in the previous table, those parameters made us obtain the maximum accuracy recorded for each type of ansatz and were at least 91 % accurate in energy.

The Cascade protocol not only improves convergence but also enhances its consistency, as show by the significantly lower standard deviation of the energy accuracy. Notably, for the 15-qubit case, the protocol yields higher energy accuracy, successfully achieving our objective of constructing an effective warm-start strategy. Specifically, we observe a 15–20% (4,5) improvement compared to the Uniform start, which enables us to meaningfully visualize one of the layers of the

Table 5: Ansätze scores with the Cascade protocol (averaged over 100 2h-optimization runs).

Qubits	Par	Ansatz	Mean Energy (%)	NEWUOA Iterations	Max Energy (%)	Min Energy (%)
6 to 9	108	Rx-Ry-Cnot	90.0 ± 0.28	4456 ± 1907	90.2	89.6
6 to 9	108	Ry-Cnot	91.5 ± 0.897	10513 ± 955	92.0	84.6
9 to 12	252	Rx-Ry-Cnot	87.4 ± 2.04	5209 ± 870	88.3	66.3
9 to 12	252	Ry-Cnot	87.1 ± 2.37	7563 ± 2156	90.2	80.9
12 to 15	330	Rx-Ry-Cnot	72.1 ± 1.45	4377 ± 379	74.0	62
12 to 15	330	Ry-Cnot	62.5 ± 4.84	3725 ± 892	70.3	44

target 3D volume B.1. As expected, at 9 and 12 qubits the overhead of applying an additional layers of gates —including the cost of possible swapping— renders the strategy less relevant: the improvement at 9 qubits is negligible, while at 12 qubits the process becomes prohibitively slow relative to the modest gains. At 15 qubits, however, the benefits dominate, with faster convergence and higher accuracy.

4 Conclusion

In this work, we explored the use of Variational Quantum Algorithms (VQAs) for solving partial differential equations, focusing on the 3D heat equation discretized via FEM. We applied a noise-free simulation of a VQA to minimize a physically meaningful cost function related to thermal energy. After demonstrating the growing difficulty of such an optimization as system size increases, we introduced the Cascade protocol, a remeshing-based warm-start strategy that reuses optimized parameters from coarser discretizations to initialize larger systems, allowing us to converge with a 15 qubits system. Through extensive simulation, we benchmarked various ansätze and initialization strategies. Basic circuits like Ry-Cnot were gradually outperformed by more expressive configurations such as Rx-Ry-Cnot as system size increased. Moreover, since it's using fewer two-qubit gates, the latter are also more noise resistant. The Cascade protocol significantly improved convergence stability and accuracy at larger scales, surpassing both cold and Uniform start baselines. This strategy offers a scalable, physically motivated approach to training VQAs for PDEs, mitigating barren plateaus and improving overall performance. Moreover, this approach opens the possibility of investigating other initialization schemes, for instance selecting tailored parameter sets at the start of optimization, to evaluate their influence on VQA performance.

Even if current quantum hardware remains limited, we propose through this article a path towards a realistic and structured benchmark for evaluating quantum processing units. By comparing noise-free simulations, noisy models, and real QPU outputs, we highlight both the potential and the current limitations of those devices. In particular, fidelity and energy-based accuracy metrics suggest that small systems can already be meaningfully benchmarked using our approach as seen in B.2. Moving forward, we envision further improvements through problem-specific warm starts, optimization algorithms, new observables and hardware-aware ansatz design. As quantum hardware matures, the methodology introduced here may serve as a foundation for solving real-world PDEs and also assessing quantum advantage in scientific computing.

Acknowledgments. We would like to thank Joseph Mikael from EDF R&D for his many fruitful discussions. We are also particularly grateful to Heidi Nelson-Quillin (IonQ) for her valuable technical assistance and the whole IonQ team for their support.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdurakhimov, L., et al.: Technology and performance benchmarks of IQM's 20-qubit quantum computer (2024), https://arxiv.org/abs/2408.12433
2. Bathe, K.: Finite Element Procedures. Prentice Hall (2006), https://books.google.fr/books?id=rWvefGICfO8C
3. Battistel, F., Chamberland, C., Johar, K., Overwater, R.W.J., Sebastiano, F., Skoric, L., Ueno, Y., Usman, M.: Real-time decoding for fault-tolerant quantum computing: progress, challenges and outlook. Nano Futures **7**(3), 032003 (Aug 2023). https://doi.org/10.1088/2399-1984/aceba6
4. Cao, Y., Papageorgiou, A., Petras, I., Traub, J., Kais, S.: Quantum algorithm and circuit design solving the poisson equation (7 2012). https://doi.org/10.1088/1367-2630/15/1/013021
5. Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., McClean, J.R., Mitarai, K., Yuan, X., Cincio, L., Coles, P.J.: Variational quantum algorithms. Nature Reviews Physics **3**(9), 625–644 (Aug 2021). https://doi.org/10.1038/s42254-021-00348-9
6. ColibriTD: Solving partial differential equations on IBM quantum processors with a variational quantum algorithm. Tech. rep. (2025), https://docsend.com/view/9xsbyyi7pxfkgt7q
7. Egger, D.J., Mareček, J., Woerner, S.: Warm-starting quantum optimization. Quantum **5**, 479 (Jun 2021). https://doi.org/10.22331/q-2021-06-17-479
8. Emeriau, P.E.: AI Image Processing: Quandela's Quantum Leap in Day-to-Night Scene Translation (2024), https://www.quandela.com/resources/blog/ai-image-processing-quandelas-quantum-leap-in-day-to-night-scene-translation, december 2024
9. Hoole, S.R.H. (ed.): Finite Elements, Electromagnetics and Design. Elsevier, Amsterdam (1995)
10. Jaffali, H., de Araujo, J.B., Milazzo, N., Reina, M., de Boutray, H., Baumann, K., Holweck, F.: H-DES: a quantum-classical hybrid differential equation solver (2024), https://arxiv.org/abs/2410.01130
11. Liu, H., Wu, Y., Wan, L., Pan, S., Qin, S., Gao, F., Wen, Q.: Variational quantum algorithm for poisson equation (12 2020). https://doi.org/10.1103/PhysRevA.104.022418
12. Liu, Y., Baba, K., Kaneko, K., Takeda, N., Koyama, J., Kimura, K.: Analysis of parameterized quantum circuits: on the connection between expressibility and types of quantum gates (2024), https://arxiv.org/abs/2408.01036
13. Mai, X., Zhang, L., Yu, Q., Zhang, J., Lu, Y.: Scalable entangling gates on ion qubits via structured light addressing (2025), https://arxiv.org/abs/2506.19535
14. Pollinger, T., Craen, A.V., Offenhäuser, P., Pflüger, D.: Realizing joint extreme-scale simulations on multiple supercomputers—two superfacility case studies. pp. 1–17 (2024). https://doi.org/10.1109/SC41406.2024.00104
15. Qiskit Development Team: Qiskit: An Open-source Framework for Quantum Computing. Zenodo (2019). https://doi.org/10.5281/zenodo.2562111
16. Rémond, U., Emeriau, P.E., Lysaght, L., Ruel, J., Mikael, J., Kazymyrenko, K.: VQA for structural mechanical problems (2025), private communication
17. Sato, Y., Kondo, R., Koide, S., Takamatsu, H., Imoto, N.: Variational quantum algorithm based on the minimum potential energy for solving the poisson equation (6 2021). https://doi.org/10.1103/PhysRevA.104.052409
18. Sim, S., Johnson, P.D., Aspuru-Guzik, A.: Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Advanced Quantum Technologies **2**(12) (Oct 2019). https://doi.org/10.1002/qute.201900070
19. Zaslavski, A., Powell, M.: The NEWUOA software for unconstrained optimization without derivatives, pp. 255–297 (01 2006). https://doi.org/10.1007/0-387-30065-1_16
20. Zienkiewicz, O., Taylor, R., Nithiarasu, P.: The finite element method for fluid dynamics. In: The Finite Element Method for Fluid Dynamics (Seventh Edition), p. i. Butterworth-Heinemann, Oxford, seventh edition edn. (2014). https://doi.org/10.1016/B978-1-85617-635-4.00016-9

A Using our results as a Benchmark

Building on the results obtained through simulation, we introduce a benchmark strategy for quantum hardware based on physically meaningful quantum states. These states arise from the solution of a 3D heat equation and offer a concrete, interpretable target for evaluating quantum processors. While executing the full optimization process directly on quantum hardware remains challenging, we propose a more accessible benchmark: given a set of optimized parameters θ, how accurately can a quantum device reproduce the corresponding state $|\psi(\theta)\rangle$? This approach isolates the hardware's ability to generate complex quantum states without requiring full in-loop optimization. To assess this, we compare the output of the IonQ Aria 1 noise model —designed to emulate the behavior of the Aria 1 QPU— to both the ideal solution and the noiseless simulation. Specifically, we evaluate the accuracy with respect to the ideal solution's energy, the accuracy with respect to the optimized simulation's energy, and the fidelity between the quantum output and each of these reference states. As a baseline, we consider the effect of stochastic sampling noise. Using the best parameters previously obtained, we simulate 100 000 measurement shots to produce a reference stochastic state, which allows us to quantify the expected variance in metrics due to finite sampling alone.

Table 6: Simulation results with stochastic noise

Qubits	Ansatz	Energy		Fidelity	
		stoch/solution	stoch/simulation	stoch/solution	stoch/simulation
6	Rx-Ry-Cnot	93.3	99.9	96.9	99.8
6	Ry-Cnot	97.8	99.9	98.4	99.9
9	Rx-Ry-Cnot	89.3	97.2	97.8	99.8
9	Ry-Cnot	91.2	97.0	98.8	99.9
12	Rx-Ry-Cnot	40.2	46.2	97.1	98.7
12	Ry-Cnot	42.8	47.1	97.7	98.8

A clear trend emerges: as the number of qubits increases, the accuracy decreases. Since the simulations are noise-free, this decline can be attributed solely to stochastic sampling noise. The exponential growth of the Hilbert space with the number of qubits leads to a corresponding increase in the number of possible measurement outcomes, requiring more shots to obtain statistically reliable estimates. For example, in the 12 qubit case, the Hilbert space is made of 4096 basis states, and 100,000 shots begin to fall short of delivering accurate statistics across the entire state space. This effect becomes even more pronounced in the 15-qubit case, where the degradation in accuracy due to insufficient sampling is both more severe and more evident.

Table 7: Noisy simulation results

Qubits	Ansatz	Energy		Fidelity	
		noisy/solution	noisy/simulation	noisy/solution	noisy/simulation
6	Rx-Ry-Cnot	92.4	99.0	96.8	99.6
6	Ry-Cnot	93.5	95.7	96.3	97.2
9	Rx-Ry-Cnot	80.2	87.4	96.8	98.1
9	Ry-Cnot	64.4	68.5	92.4	93.1
12	Rx-Ry-Cnot	20.8	23.9	89.1	89.6
12	Ry-Cnot	7.65	8.41	64.3	64.6

We observe a clear difference between the two ansatz. To match the same number of total parameters, the Rx-Ry-CNOT ansatz uses half as many entangling layers as the Ry-CNOT ansatz. Since two-qubit gates have lower success rates, adding more entangling layers reduces performance, as seen above. Other factors, such as gate types and qubit count, also influence results. Given hardware costs and simulation limits, we restricted our comparison to 6-qubit circuits with the Ry-CNOT ansatz, running 10,000 shots to obtain the results shown in B.2.

B Imaging our results

B.1 The Cascade results

As explained above, we used the results obtained through the best optimization thanks to the Cascade warm-start, and applied matplotlib's Lánczos interpolation on the raw statevectors.

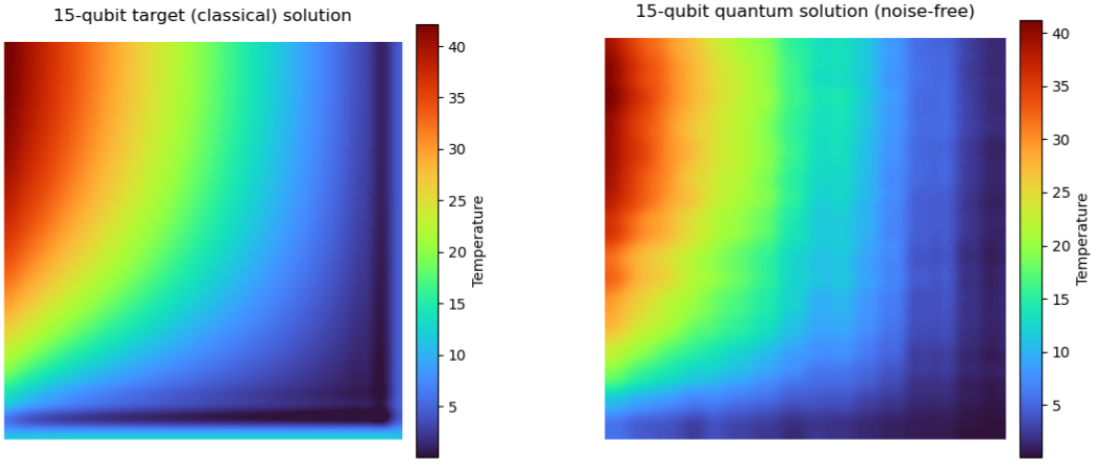


Fig. 3: Comparison between the target and the quantum solution for a 15-qubit system.

To visualize this 2D slice of the 3D volume, we used the quantum state obtained from the best optimization run (15 qubits, 330 parameters, Rx-Ry-Cnot Ansatz, with 74% energy accuracy). We can see how closely the two states match, by noticing the preservation of the general shape and gradients.

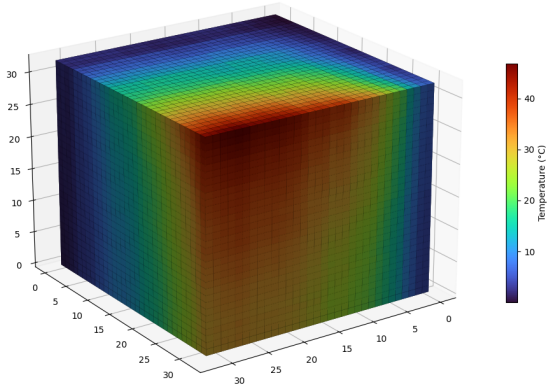


Fig. 4: 3D Heatmap of the 15 qubits noise-free simulation using the Cascade warm-start

B.2 Imaging the noisy simulation and QPU's output

We ran our PQC on IonQ's QPU Aria 1, allowing us to run it against their noise model, the stochastic noise and the noise-free simulation. We retrieved the following results :

Table 8: Performance of the Aria 1 noise model and QPU from IonQ for 6 qubits

Metric	Ideal	Stochastic	Noisy (Aria 1) simulation	Aria 1 (QPU)
Energy (state/solution)	96.8	97.1	93.5	92.2
Energy (state/simulation)	100	99.3	95.7	97.8
Fidelity (state/solution)	98.6	98.2	96.3	95.3
Fidelity (state/simulation)	100	99.8	97.2	96.3

These performances show us that all results are close to one another, and both the QPU output and the noisy simulation remain close to the ideal simulated state for such small qubit counts. Naturally, some minor differences and reduced accuracy can still be observed, likely due to noise—whether stochastic or arising from other sources. We can visualize those differences by imaging a slice of our 3D object (made with 6 qubits) as shown previously for 15 qubits.

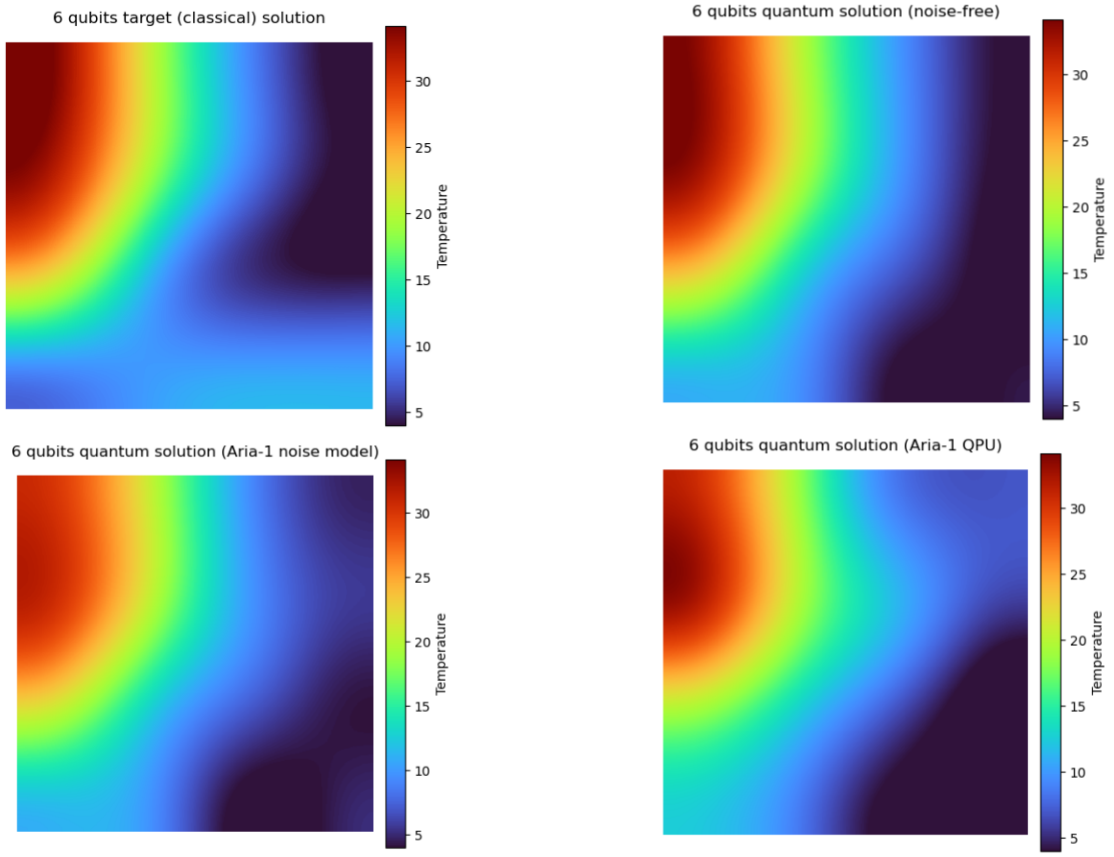


Fig. 5: Comparison between the classical target and quantum solutions for a 6-qubit system using Lánczos interpolation.

The noise-free simulation closely reproduces the overall structure, suggesting that the chosen ansatz captures the essential features of the problem. When introducing the Aria 1 noise model (third), we observe a shift in temperature as observed with the color distribution and the color bar, indicating the degradation caused by noise. The result from the real QPU (rightmost) exhibits similar features to the noisy simulation but with slightly higher variation, that is likely due to hardware imperfections or shot noise. Despite these limitations, the general shape and gradients are preserved, validating the approach for small systems under realistic condition.

K-ADAPT-VQE: Optimizing Molecular Ground State Searches by Chunking Operators

Tatiana A. Bespalova[1,2][0000−0003−3300−0843], Oumaya Ladhari[1,2][0009−0002−5318−7561], and Guido Masella[1][0000−0001−7108−2996]

[1] QPerfect, 23 Rue du Loess, 67000 Strasbourg, France
[2] Université de Strasbourg and CNRS, ISIS (UMR 7006), 67000 Strasbourg, France

Abstract. Classical simulation of molecular systems is fundamentally limited by exponential scaling, motivating the development of quantum algorithms such as the Variational Quantum Eigensolver (VQE). While the Adaptive VQE (ADAPT-VQE) improves upon VQE by dynamically constructing ansätze from an operator pool, its one-operator-at-a-time strategy can lead to high computational overhead. In this work, we introduce K-ADAPT-VQE, a generalization of ADAPT-VQE that selects operators in chunks of K per iteration, thereby accelerating convergence. Using the MIMIQ quantum simulator, we benchmark the method on BeH_2, LiH, and N_2, demonstrating that K-ADAPT-VQE consistently reduces the number of iterations and quantum function calls required to achieve chemical accuracy, with improvements of up to an order of magnitude relative to ADAPT-VQE. These findings indicate that chunked operator selection offers a scalable path for applying adaptive ansätze construction to increasingly complex molecular systems.

Keywords: Quantum Chemistry · Quantum Computing · VQA.

1 Introduction

Quantum chemistry is one of the most computationally demanding fields in science due to the exponential growth of the Hilbert space with system size. This exponential scaling renders many accurate quantum chemistry calculations intractable for classical computers, particularly for systems exhibiting strong electron correlations. Since molecular systems are fundamentally governed by quantum mechanics, quantum computing presents a natural and highly promising approach for their study [2].

Although fault-tolerant quantum computers are not yet available, researchers are actively exploring algorithms that could demonstrate a quantum advantage using today's Noisy Intermediate-Scale Quantum (NISQ) devices [11]. Variational Quantum Algorithms (VQAs), such as the Variational Quantum Eigensolver (VQE) [10], have become prominent in this context. VQEs employ a hybrid quantum-classical approach: a parametrized quantum circuit, known as the ansatz, prepares a trial wavefunction on a quantum processor. The energy of

this state is then estimated on a quantum processor, while a classical optimizer updates the circuit parameters to minimize the energy.

However, conventional VQE methods can suffer from several limitations, including inefficient parameter optimization and the presence of numerous local minima in the energy landscape, especially for larger systems. To address these issues, the Adaptive Derivative-Assembled Pseudo-Trotter ansatz Variational Quantum Eigensolver (ADAPT-VQE) [3] was introduced. ADAPT-VQE dynamically builds a compact, problem-specific ansatz by iteratively adding operators from a predefined pool, selecting them based on their contribution to lowering the energy gradient. Extensions of this approach involve adding multiple operators per iteration to further improve efficiency. As an example, TETRIS-ADAPT-VQE [1] selects multiple operators with disjoint qubit supports to build denser circuits that are significantly shallower, a key consideration for hardware with limited coherence times.

Interestingly, this philosophy of adaptively building a compact and efficient wavefunction is not new to computational chemistry. In classical electronic structure theory, Selected Configuration Interaction (SCI) methods were developed to address the challenge of Hilbert space explosion in Full Configuration Interaction (FCI) calculations. Rather than considering all possible electron configurations, SCI algorithms iteratively select only the most energetically important determinants. Different selection criteria have led to several SCI variants, such as Configuration Interaction using a Perturbative Selection made Iteratively (CIPSI) [5], Heat-Bath Configuration Interaction (HCI) [4], and Adaptive Sampling CI [15]. Like ADAPT-VQE, these methods share a common adaptive, iterative philosophy: the wavefunction is built progressively, guided by quantitative measures of importance, such as energy contributions or Hamiltonian couplings.

In this work, we propose several improvements to the ADAPT-VQE algorithm aimed at enhancing its efficiency without compromising accuracy. Our modifications include an optimized subroutine that reduces the operator pool by pruning unnecessary operators and an improved selection process to further minimize the number of addresses to the quantum function. We call our selection process K-ADAPT-VQE. The key difference from the standard procedure is the simultaneous addition of K operators to the ansatz at each iteration. This "chunking" strategy significantly reduces the total number of quantum function evaluations required to achieve chemical accuracy.

Our benchmark focuses on gradient-free optimizers, such as COBYLA and CMA-ES, which are particularly interesting for NISQ-era hardware where gradient evaluations are costly. We assess the performance of our enhanced ADAPT-VQE implementation across a diverse set of small molecules at various bond lengths, demonstrating consistent improvements in convergence, accuracy, and overall computational resource requirements.

2 Method

This section outlines the methodology employed in our K-ADAPT-VQE approach. We begin by describing the preparation of the molecular Hamiltonian and the construction of the operator pool. We then detail the operator selection mechanism, highlighting the K-ADAPT-VQE criterion, and conclude with the full algorithmic procedure.

2.1 Hamiltonian preparation

The starting point of the K-ADAPT-VQE algorithm is the formulation of the chemical problem, given by the electronic structure Hamiltonian for a molecular system, typically expressed in the second quantization formalism. The general form of this Hamiltonian is:

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \tag{1}$$

where $\hat{a}_p^\dagger$ and $\hat{a}_q$ are fermionic creation and annihilation operators acting on spin-orbitals, and the coefficients h_{pq} and h_{pqrs} are the one- and two-electron integrals arising from the molecular orbital basis [7]. These interactions account for the kinetic energy of the electrons and the Coulomb interactions between them.

To obtain these integrals, we first use PySCF [14], an open-source quantum chemistry package, to perform initial Hartree-Fock (HF) calculations and define the molecular orbitals. The resulting integrals are then processed using OpenFermion [8], a library tailored for quantum simulations of chemistry, to construct the second-quantized Hamiltonian.

We then map this fermionic Hamiltonian into a qubit Hamiltonian suitable for quantum computation using the Jordan-Wigner (JW) transformation, a standard method that preserves the antisymmetric nature of fermionic wavefunctions. We choose the JW mapping for its conceptual and implementation simplicity, although we note that for other physically (chemically) inspired fermion-to-qubit mappings [9] the ADAPT-VQE pipeline would remain unchanged aside from the mapping step.

For small to moderate bond lengths, in the molecular systems studied in this work (e.g. BeH_2, LiH, N_2) a good approximation to the ground-state can be obtained from restricted HF methods. Conversely, for systems near-degeneracies or stretched bond lengths unrestricted HF may offer better approximations.

2.2 Operator pool

The ADAPT-VQE framework relies on an operator pool from which the VQE ansatz is dynamically constructed. In many classical quantum chemistry approaches, coupled-cluster (CC) theory is used to capture electron correlation beyond mean-field theory. In quantum computing, this leads to the Unitary

Coupled-Cluster (UCC) ansatz [13], which is compatible with VQE frameworks. In this work, we follow the UCCSD approach, which includes single and double excitations. This can, in principle, be extended to include triple and higher excitations, although the computational cost increases significantly.

The operator pool for the K-ADAPT-VQE used in this work is constructed from UCCSD excitation operators, including all symmetry-allowed double fermionic excitations, which are then mapped to qubit operators and used iteratively in the adaptive ansatz construction.

The UCC ansatz introduces a unitary form of the CC operator by symmetrizing it $|\Psi_{\mathrm{UCC}}\rangle = e^{\hat{T}-\hat{T}^\dagger}|\Psi_{\mathrm{HF}}\rangle$, where the CC operator $\hat{T} = \sum_k \hat{T}_k$ is composed of excitation operators, and $\hat{T} - \hat{T}^\dagger$ is anti-Hermitian, ensuring that the exponential is unitary and preserves normalization. Here, $\hat{T}$ is truncated to include only single ($\hat{T}_1 = \sum_{i,j} \theta_{ij}\hat{a}_i^\dagger \hat{a}_j$) and double ($\hat{T}_2 = \sum_{i,j,k,l} \theta_{ij}^{kl}\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l$) excitations. The exponential is then expanded using the Lie-Trotter formula into a product of operators.

In the JW transformation, these operators have a simple form. Let us consider, for example, double excitation for spin-orbitals p, q, r, and s. For compactness, we define the anti-Hermitian two-body generator $\hat{G}_{pqrs} \equiv \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s - \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_p \hat{a}_q$. Under Jordan–Wigner, $\hat{G}^{pqrs}$ maps to a sum of eight Pauli strings with appropriate Z-chains:

$$\begin{aligned}\hat{G}_{pqrs} &= \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s - \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_p \hat{a}_q = \\ &= 2i(X_pX_qX_rY_s + X_pX_qY_rX_s - X_pY_qX_rX_s - Y_pX_qX_rX_s - \\ &- Y_pY_qY_rX_s - Y_pY_qX_rY_s + Y_pX_qY_rY_s + X_pY_qY_rY_s) \bigotimes_{k=q+1}^{p-1} \hat{Z}_k \bigotimes_{l=s+1}^{r-1} \hat{Z}_l.\end{aligned} \tag{2}$$

We note that everything here commutes, so the expansion of $e^{\theta \hat{G}_{pqrs}} = e^{\sum_\alpha \pm 2i\theta \hat{P}_\alpha}$ to a product of exponents $\prod_\alpha e^{\pm 2i\theta \hat{P}_\alpha}$ is exact. Further in the paper we will be referring to an arbitrary sum of Pauli strings as $\hat{P}_{\mathrm{sum}} := \sum_\alpha c_\alpha \hat{P}_\alpha$. We also note that the double excitations do not act on any qubits between q and r.

To enhance computational efficiency, and focus the ansatz construction, when starting from the HF state we exclude from the operator pool: all single excitations (as HF is already optimized for those), all operators that do not conserve the total z-projection of spin (as the state we are searching for here is in the symmetry sector of zero spin projection), and we only allow hopping from two occupied in HF orbitals to two unoccupied as the influence of the rest is small on HF. A similar methodology for operator pool construction is used in [12], although in our case we include $\hat{a}_{\uparrow i}^\dagger \hat{a}_{\uparrow j}^\dagger \hat{a}_{\uparrow k} \hat{a}_{\uparrow l}$ and $\hat{a}_{\downarrow i}^\dagger \hat{a}_{\downarrow j}^\dagger \hat{a}_{\downarrow k} \hat{a}_{\downarrow l}$ operators in the pool, but exclude the operators representing electron jumps into already occupied spin-orbitals or from already vacant spin-orbitals.

2.3 Selecting the operators

A core component of ADAPT-VQE is the iterative selection of operators based on the largest contribution to the energy gradient. The energy gradient result-

ing from adding an operator $\hat{P}_{\text{sum}}$ to the ansatz ($|\psi\rangle$) has a convenient form: $\frac{\partial E}{\partial \theta} \mid_{\theta=0} = \frac{\partial}{\partial \theta} \langle \Psi | e^{-i\theta \hat{P}_{\text{sum}}} H e^{i\theta \hat{P}_{\text{sum}}} | \Psi \rangle \mid_{\theta=0} =$
$-i \langle \Psi | e^{-i\theta \hat{P}_{\text{sum}}} [\hat{P}_{\text{sum}}, H] e^{i\theta \hat{P}_{\text{sum}}} | \Psi \rangle \mid_{\theta=0} = -i \langle \Psi | [\hat{P}_{\text{sum}}, H] | \Psi \rangle$.

To optimize computational cost we precompute the analytical form of the commutators and, at each step of the adaptive procedure, the expectation values of all operator pool commutators. We note that the expectation values for all of the needed Pauli strings could also be precomputed at each step and then combined into the commutator expectations.

The main addition of the K-ADAPT-VQE to the ADAPT-VQE procedure is in the way operators from the pool are chosen and added to the ansatz. Instead of adding only the operator with the largest energy gradient, we choose the K operators with the highest absolute values of the commutator expectation and add them to the ansatz from the highest gradient to the lowest with the initial value of the parameters equal to zeros. The choice of $K = 5$ was guided by preliminary tests aimed at balancing efficiency and accuracy. For very small values of K (e.g., $K = 1$, corresponding to the original ADAPT-VQE), convergence is reliable but often requires many iterations and quantum function evaluations. Conversely, very large values of K risk overshooting by introducing too many operators simultaneously, which can increase ansatz redundancy and degrade optimization stability. In practice, $K = 5$ provided a consistent reduction in the number of VQE iterations and quantum function calls across all benchmarked systems, while maintaining chemical accuracy. Although this value worked robustly for the molecules considered here, the optimal choice of K may depend on the system size, electronic structure, and operator pool, and a systematic exploration of this dependence remains an interesting avenue for future work.

2.4 K-ADAPT-VQE Algorithm

1. **Prepare the system:** First, we compute the molecular integrals using a classical electronic structure package (e.g. PySCF), then we construct the second-quantized fermionic Hamiltonian which is mapped to qubit operators using one of the fermion-to-qubit transformations (e.g. JW)
2. **Initialize the ansatz:** The first step is to set the initial quantum state $|\Psi_0\rangle$ to the HF state. We then start with an empty ansatz.
3. **Construct the operator pool:** The operator pool includes all symmetry-allowed spin-conserving double excitation operators that correspond to electron hopping from two HF-occupied orbitals to two virtual orbitals. We proceed by mapping all allowed fermionic operators to qubit operators.
4. **Repeat until convergence:**
   - **a.** Compute the energy gradient for each operator $\hat{P}_i$ in the pool.
   - **b.** Select the top K operators with the largest absolute gradient magnitudes.
   - **c.** Append these K operators to the ansatz (in order of decreasing gradient).
   - **d.** Optimize the full ansatz parameters including previously added ones using a classical optimizer (e.g. COBYLA), allowing a maximum number of VQE iterations per ADAPT step (e.g. 200)

e. Check convergence: if all gradients are below a threshold ϵ, or if the energy difference between the last computed and the newly computed energy is below a threshold ϵ', stop. Otherwise, return to step 4a.

5. **Output:** We get the final optimized ansatz, parameter set, as well as the ground state energy estimate.

3 Results

We implemented the K-ADAPT algorithm using the MIMIQ simulator [6] to compute the ground state energies for several molecules, with systems up to 20 qubits. This section presents a comparative analysis of our K-operator chunking strategy (specifically with $K = 5$, termed 5-ADAPT) against the standard one-operator-at-a-time ADAPT-VQE (termed 1-ADAPT). All calculations presented utilize the STO-3G basis set and the COBYLA optimizer with a tolerance of 10^{-3} Hartree for the VQE subroutine. The gradient threshold ϵ was set to 10^{-5} and the energy difference threshold ϵ' was set to 10^{-15}. The HF and Full Configuration Interaction (FCI) energies, obtained from PySCF, serve as benchmarks.

Figure 1 showcases the performance benefits of the K-ADAPT strategy across various molecules and bond lengths.

For the linear BeH_2 molecule Fig. 1(a) shows the computed ground state energies as a function of the Be-H bond length. For both the 5- and 1-ADAPT procedures, the total number of VQE iterations is 1000 (200 iterations for 5-ADAPT chunks, 40 iterations for each 1-ADAPT operator), and the resulting ansatz consists of 25 operators. The 5-ADAPT approach consistently yields energies closer to FCI, achieving chemical accuracy (error $< 10^{-3}$ Ha) for most of the presented bond lengths. In contrast, the 1-ADAPT procedure generally results in energies with errors approximately an order of magnitude larger. The improved convergence of the K-ADAPT method is further detailed in Fig. 1(b), which shows the energy error relative to FCI as a function of the total number of VQE iterations for BeH_2 at the fixed bond length of 1.3 Å.

To reach chemical accuracy, the 5-ADAPT strategy required significantly fewer calls to the quantum function. Assuming three evaluations per COBYLA iteration, the 5-ADAPT approach (5 chunks of 5 operators, 200 VQE iterations per chunk for the first 800 iterations, then adapting) reached chemical accuracy with approximately 3300 total quantum function evaluations ($3\times800+180\times\frac{25}{5}$). The 1-ADAPT strategy (25 operators added one-by-one, 40 VQE iterations per operator) required roughly 14100 evaluations ($3\times3200+180\times25$). This signifies a computation cost reduction of ~ 4.3 times.

We observe that, in the ADAPT-VQE procedure, the parameters associated with the most recently added operators undergo the most significant changes during each VQE optimization. However, allowing previously added parameters to be fine-tuned in subsequent steps is crucial in finding the global energy minima and contributes to the overall efficiency of the optimization.

Figure 1(c, d) further demonstrates the advantages of the 5-ADAPT approach for other molecules. Fig 1(c) shows results for the LiH molecule (12

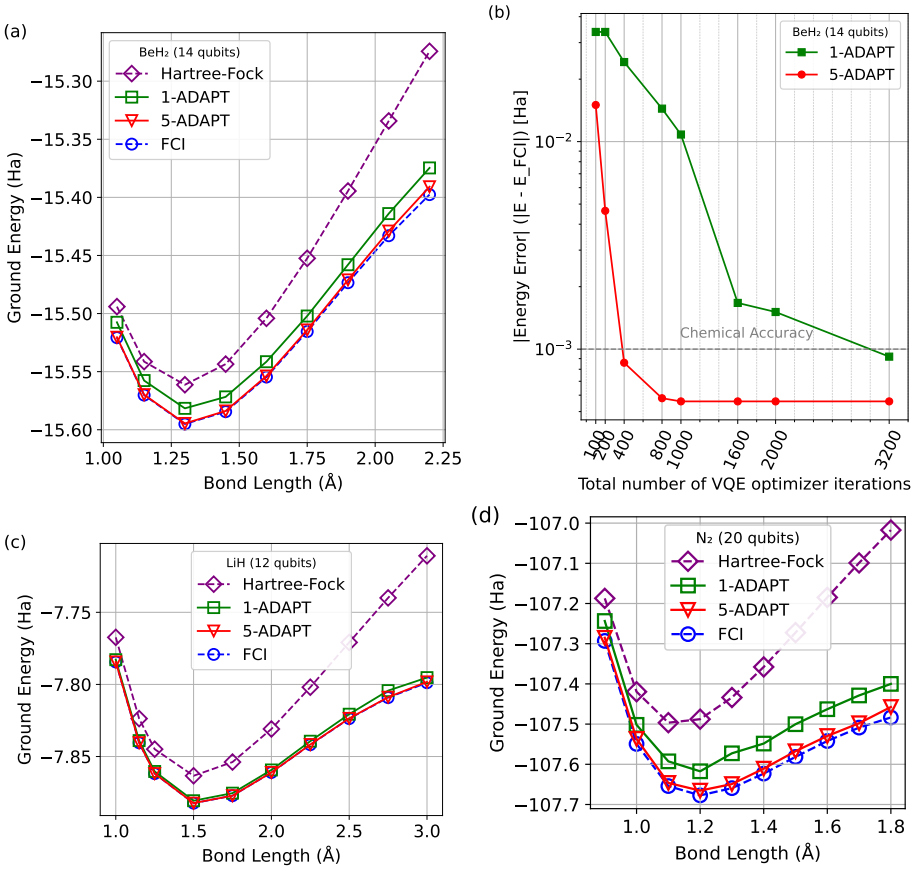


Fig. 1: Performance of 5-ADAPT and 1-ADAPT. All calculations employ the STO-3G basis, the HF and FCI energies serve as benchmarks. In figures a, c and d, the ground state energy (in Hartree) was computed as a function of the bond length for 3 different molecules using ADAPT-VQE with COBYLA optimizer (tolerance 1e-3) and a maximum of 200 VQE iterations per 5-ADAPT step, 40 VQE iterations per 1-ADAPT step. (a) BeH_2 molecule (14 qubits): with 25 operators added. (b) Convergence behavior at a fixed bond length for BeH_2 of 1.3 Å, showing the energy error (difference from FCI) versus the number of VQE iterations. (c) LiH molecule (12 qubits): with 25 operators added. (d) N_2 molecule (20 qubits): with 50 operators added.

qubits), where 25 operators were added in total. Fig 1(c) shows results for N_2 (20 qubits), with 50 operators added. In both cases, the 5-ADAPT method demonstrates robust performance across different bond lengths.

In Fig. 2 we show the resulting ansatz structure for the BeH_2 molecule at a bond length of 1.3 Å, generated by the 5-ADAPT algorithm. We observe that operators within the same chunk frequently act on intersecting sets of qubits.

This is a consequence of K-ADAPT which selects the top K operators based purely on gradient magnitude, without imposing a disjoint support criterion, distinguishing it from methods like TETRIS-ADAPT-VQE [1], which adds several operators at a time but only to non-overlapping qubits and allows mostly the reduction in depth of the circuit and not in the total number of gates needed. A direct numerical comparison of these two strategies would be a valuable direction for future work. We hypothesize that this way of focusing only on the physically relevant operators, even if overlapping, contributes to a more efficient search of the optimal parameters, without increasing the ansatz complexity with less impactful, yet parallelizable operators.

Furthermore, we note that the presence of spin-conserving excitations within the same spin type, in the fourth chunk of the ansatz, for example (Fig. 2), indicates their meaningful contribution to the ansatz, and hence supports our choice of including them in the operator pool.

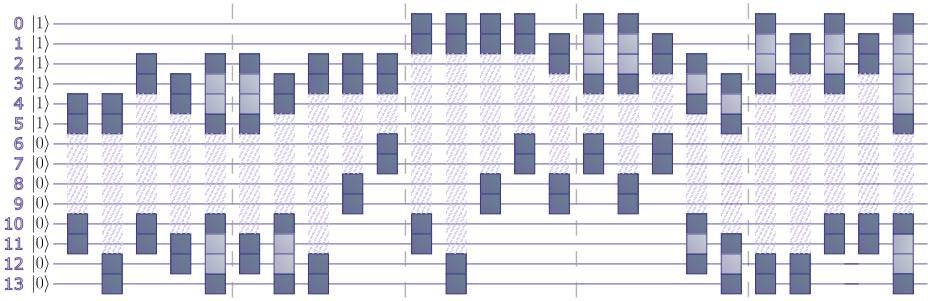


Fig. 2: The resulting structure of the ansatz for BeH_2 molecule with 1.3Å bond length in STO-3G basis. Here each gate denotes operator $e^{\theta \hat{G}_{pqrs}}$ (see Eq. 2), where each operator is expanded into eight sequential Pauli rotations $e^{\pm 2i\theta \hat{P}}$. Here θ is the variable coefficient, it is independent for each gate but same for each Pauli within one gate, p, q, r, s correspond to the positions of dark-gray squares, in the Pauli strings $\hat{P}$ there are X and on these positions, on the positions corresponding to light-gray squares there are Z, the rest are I.

4 Conclusion

In this work, we introduced K-ADAPT-VQE, a strategy that enhances the efficiency of the ADAPT-VQE algorithm by adding operators in chunks of K (empirically set to $K = 5$ in our demonstrations). This approach significantly reduces overall computational costs, most notably by decreasing the total number of VQE optimization iterations and, consequently, the calls to the quantum function required to reach chemical accuracy. For instance, our 5-ADAPT variant achieved a roughly 4.3-fold reduction in quantum function calls for the BeH_2

molecule compared to the standard one-operator-at-a-time ADAPT-VQE while attaining comparable accuracy.

The K-ADAPT-VQE method is effective even when operators within a single chunk act on overlapping sets of qubits, a practical scenario that distinguishes it from approaches mandating disjoint supports. This focus on adding the K most gradient-impactful operators, irrespective of their qubit intersection, prioritizes the inclusion of physically relevant correlations, thereby streamlining the ground-state search. These findings suggest that chunking operators is a beneficial strategy for molecular simulations, especially in resource-limited computational environments.

While we demonstrated that $K = 5$ provided a robust balance between accuracy and efficiency across BeH_2, LiH, and N_2, the optimal choice of K may depend on system size and operator pool characteristics, and thus represents an interesting direction for further investigation.

Acknowledgments. We thank Rosario Roberto Riso for insightful discussions on quantum chemistry. This research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 955479 (MOQS - Molecular Quantum Simulations).

Disclosure of Interests. Guido Masella is a shareholder of QPerfect.

References

1. Anastasiou, P.G., Chen, Y., Mayhall, N.J., Barnes, E., Economou, S.E.: Tetris-adapt-vqe: An adaptive algorithm that yields shallower, denser circuit ansätze. Phys. Rev. Res. **6**, 013254 (Mar 2024). https://doi.org/10.1103/PhysRevResearch.6.013254
2. Feynman, R.P.: Simulating physics with computers. International Journal of Theoretical Physics **21**(6–7), 467–488 (1982). https://doi.org/10.1007/BF02650179
3. Grimsley, H.R., Economou, S.E., Barnes, E., Mayhall, N.J.: An adaptive variational algorithm for exact molecular simulations on a quantum computer. Nature Communications **10**(1) (Jul 2019). https://doi.org/10.1038/s41467-019-10988-2
4. Holmes, A.A., Tubman, N.M., Umrigar, C.: Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling. Journal of Chemical Theory and Computation **12**(8), 3674–3680 (2016)
5. Huron, B., Malrieu, J., Rancurel, P.: Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. The Journal of Chemical Physics **58**(12), 5745–5759 (1973)
6. Leonteva, A., Masella, G., Outteryck, M., Orioli, A.P., Whitlock, S.: Comparative benchmarking of utility-scale quantum emulators. arXiv preprint arXiv:2504.14027 (2025). https://doi.org/10.48550/arXiv.2504.14027
7. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S.C., Yuan, X.: Quantum computational chemistry. Reviews of Modern Physics **92**(1), 015003 (2020)
8. McClean, J.R., et al.: Openfermion: the electronic structure package for quantum computers. Quantum Science and Technology **5**(3), 034014 (jun 2020). https://doi.org/10.1088/2058-9565/ab8ebc

9. Parella-Dilmé, T., Kottmann, K., Zambrano, L., Mortimer, L., Kottmann, J.S., Acín, A.: Reducing entanglement with physically inspired fermion-to-qubit mappings. PRX Quantum **5**, 030333 (Aug 2024). https://doi.org/10.1103/PRXQuantum.5.030333
10. Peruzzo, A., et al.: A variational eigenvalue solver on a photonic quantum processor. Nature Communications **5**(1) (Jul 2014). https://doi.org/10.1038/ncomms5213
11. Preskill, J.: Quantum computing in the nisq era and beyond. Quantum **2**, 79 (Aug 2018). https://doi.org/10.22331/q-2018-08-06-79
12. Ramôa, M., Anastasiou, P.G., Santos, L.P., Mayhall, N.J., Barnes, E., Economou, S.E.: Reducing the resources required by adapt-vqe using coupled exchange operators and improved subroutines. npj Quantum Information **11**(1), 1–19 (2025). https://doi.org/10.48550/arXiv.2407.08696
13. Romero, J., Babbush, R., McClean, J.R., Hempel, C., Love, P., Aspuru-Guzik, A.: Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz (2018), https://arxiv.org/abs/1701.02691
14. Sun, Q., et al.: Pyscf: the python-based simulations of chemistry framework. WIREs Computational Molecular Science **8**(1), e1340 (2018). https://doi.org/10.1002/wcms.1340
15. Tubman, N.M., Lee, J., Takeshita, T.Y., Head-Gordon, M., Whaley, K.B.: A deterministic alternative to the full configuration interaction quantum monte carlo method. The Journal of Chemical Physics **145**(4), 044112 (2016)

Neural Network Assisted Fermionic Compression Encoding: A Lossy-QSCI Framework for Scalable Quantum Chemistry Simulations

YuCheng Chen[1], Ronin Wu[2][0000−0002−3736−5391], M. H. Cheng[3,4], and MinHsiu Hsieh[1]

[1] Hon Hai Research Institute, Taipei, Taiwan
kesson.yc.chen@foxconn.com
[2] QunaSys Europe, Copenhagen, Denmark
ronin@qunasys.com
[3] Department of Physics, Blackett Laboratory, Imperial College London, SW7 2AZ, United Kingdom
[4] Fraunhofer Institute for Industrial Mathematics, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

Abstract. Quantum-chemistry calculations on noisy hardware are bottlenecked by qubit count and measurement overhead. We introduce **Lossy-QSCI**—a compact form of Quantum Selected CI that (i) uses a chemistry-aware lossy Random Linear Encoder (Chemical-RLE) to compress an M-orbital, N-electron Hamiltonian to $O(N \log M)$ qubits, and (ii) restores observables via a lightweight neural–network Fermionic Expectation Decoder (NN-FED). Applied to C_2 and LiH, Lossy-QSCI attains chemical accuracy with roughly half the qubits and determinants required by standard QSCI, pointing to a practical route for accurate quantum-chemistry on NISQ and early fault-tolerant devices.

No Title Given

No Author Given

No Institute Given

1 Introduction

Quantum computers promise chemically exact solutions of many-electron problems that overwhelm classical methods, from small molecules to catalytic clusters (1; 3). Near-term devices, however, are constrained by few qubits, noise, and the heavy sampling overhead of hybrid algorithms such as VQE and Quantum-Selected CI (QSCI) (8; 5).

VQE stalls on barren plateaus and requires millions of measurements, while standard QSCI avoids optimisation but still stores the full Fock space, exhausting qubits and amplifying post-processing cost (9).

We tackle both hurdles with **Lossy-QSCI**: a chemistry-aware, lossy Random Linear Encoder that compresses an M-orbital, N-electron problem to $\mathcal{O}(N \log M)$ qubits, and a neural-network Fermionic Expectation Decoder that restores observables with microsecond latency (2). Iterative sampling plus fast decoding yields compact CI subspaces that keep measurement counts low.

On C_2 and LiH we reach chemical accuracy using roughly half the qubits and 25% fewer determinants than qubit-uncompressed QSCI, pointing to a practical route for accurate quantum chemistry on NISQ and early fault-tolerant hardware.

Related work. Early mappings such as Jordan–Wigner encode the entire Fock space, consuming one qubit per spin-orbital(4). Recent number-conserving schemes shrink this to $\mathcal{O}(N \log M)$ qubits, but at the cost of either an exponential measurement count (non-linear variants) or a hard decoding problem, only partially resolved by the Fermionic-Expectation-Decoder (FED)(11; 2).[1] On the algorithmic side, Quantum-Selected CI (QSCI) samples a trial wavefunction, keeps the most frequent determinants, and classically diagonalises the Hamiltonian(5). Its 77-qubit hardware demo added self-consistent recovery and a LUCJ ansatz to fight noise(10); ADAPT-QSCI grows the circuit operator-by-operator(7); TE-QSCI sidesteps optimisation by time-evolving a simple reference state(6). All still store the full JW register, rely on deep or problem-specific ansätze, and require heavy sampling(9).

Our contribution. Lossy-QSCI closes this gap: a chemistry-biased lossy Random Linear Encoder (Chemical-RLE) compresses directly to the $\mathcal{O}(N \log M)$ register, while a neural-network FED (NN-FED) decodes observables in microseconds. The resulting workflow attains chemical accuracy on C_2 and LiH using roughly half the qubits and fewer determinants than uncompressed QSCI,

[1] Full derivations, scaling proofs and decoder benchmarks are given in of the Supplement.

Method	QSCI (5)	Scaled-up QSCI(10)	TE-QSCI(6)	Lossy-QSCI [This work]
Qubit Efficiency	O(M)	O(M)	O(M)	O(Nlog(M))
Bit-flip Error Mitigation	Post-Selection	Configuration Recovery	Post-Selection	Compression+Post-Selection
Chemical Prior	Customizable	LUCJ Ansatz	Time-Evolved-HF/UCCSD	Chemical-RLE

Table 1. Comparison of QSCI variants.

offering a practical path toward resource-frugal quantum chemistry on near-term and early fault-tolerant hardware.

2 Numerical Result of Lossy-QSCI

Benchmark protocol. We evaluate Lossy-QSCI on three systems, each defined by an active space $\left(N_{\text{Spin Orb}}, N_e\right)$:

- C_2**.** Vary the lossy–compression rate to reveal the trade-off between qubit count and energy accuracy.
- LiH**.** Combine a hardware–efficient ansatz (HEA) with different QSCI basis sizes R to probe basis–set convergence.
- H_2**.** Compare Lossy-QSCI and vanilla QSCI under a Pauli-X bit-flip channel to gauge noise robustness.

For every molecule we iteratively sample the R most-probable determinants from successive trial states, thereby refining a compact CI subspace that retains the dominant weight of the exact ground state.

2.1 C_2 Molecule

We begin by classically diagonalising the full `6-31G` Hamiltonian for C_2 at a bond length of 2.5 Å. The resulting eigenvector is taken as the exact ground state, $|\psi_g\rangle$. Figure 1 illustrates the performance of Lossy-QSCI as a function of the number of qubits. We compare our results with three other reference energies: (i) ED in the active space of size (22,4) (marked as "Chemical Accuracy (22, 4)"); (ii) ED in the active space of size (26, 4) (marked as "Chemical Accuracy (26, 4)"); (iii) uncompressed QSCI with R values of 65 at 26 qubits. The dashed blue and black lines respectively represent the values for Chemical Accuracy (22,4) and (26,4), respectively. The dashed green line depicts the uncompressed QSCI results for $R = 65$ at 26 qubits. The solid red line represents the Lossy-QSCI method with $R = 50$.

Lossy-QSCI with the number of qubits below 14 evaluates to a final energy that is higher than the (22, 4) Chemical Accuracy, but it quickly converges to similar values as the exact QSCI results with R values of 65 at 26 qubits, when the number of qubits increases to beyond 16.

The results demonstrate that Lossy-QSCI, which maps a larger active space to a smaller qubit representation, can outperform exact diagonalization in a smaller active space with the same number of qubits. Specifically, the Lossy-QSCI method with $R = 50$ achieves an energy closer to the exact QSCI results as the qubit count increases, surpassing the ED (22,4) benchmark. This suggests that Lossy-QSCI offers a promising approach for achieving more accurate quantum simulations with constrained qubit resources and circuit depth, effectively balancing computational efficiency and accuracy.

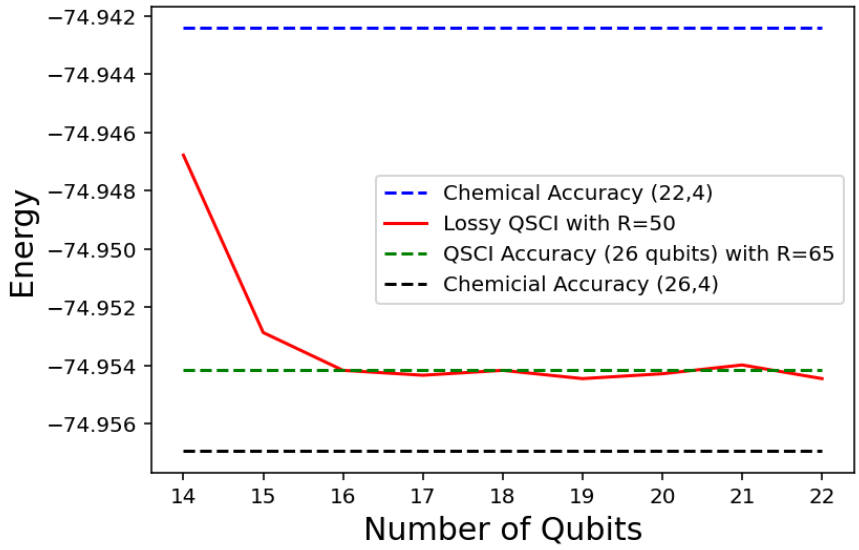


Fig. 1. The C_2 molecule Lossy-QSCI results with approximate ground state prepared by adding random noise to the exact ground state. The x-axis represents the number of qubits, while the y-axis shows the corresponding energy. The greater number of qubits will result in a less lossy rate of injectivity.

2.2 *LiH* Molecule

In the study of LiH molecule at a bond length of 2.5 Å with the STO-3G basis set, the candidate configuration sets, essential for a single round Chemical-RLE, is constructed from the 5 most frequently sampled configurations of the approximate ground states prepared via VQE. Figure 2 illustrates the result of the Lossy-QSCI method encoded in five qubits. The two dashed lines – colored in black and blue – provide references for the chemical accuracy in the (6, 2) and (10, 2) active space respectively. The red line depicts the energy convergence of the Lossy-QSCI method as we increase R, starting at slightly above the chemical accuracy of (6,2) active space for $R = 4$, converges quickly to the chemical accuracy of (10,2) active space and stabilizes as $R \geq 12$. The green line represents the collective optimized energy of VQE on randomly compressed Hamiltonians for five qubits.

The results demonstrate that iterative sampling of CI states from different compressed subspaces can significantly enhance the performance of Lossy-QSCI. Specifically, the Lossy-QSCI method achieves the (10,2) chemical accuracy with as few as 12 basis states, outperforming all collective VQE results that lies above -7.8. Improved convergence indicates a potential of Lossy-QSCI to deliver more accurate ground state energies with fewer basis states. This approach achieves the required accuracy with fewer qubit counts compared to conventional QSCI approaches through iterative VQE, thereby improving re-

source efficiency in quantum simulations, and boosts the simulation power of noisy hardwares.

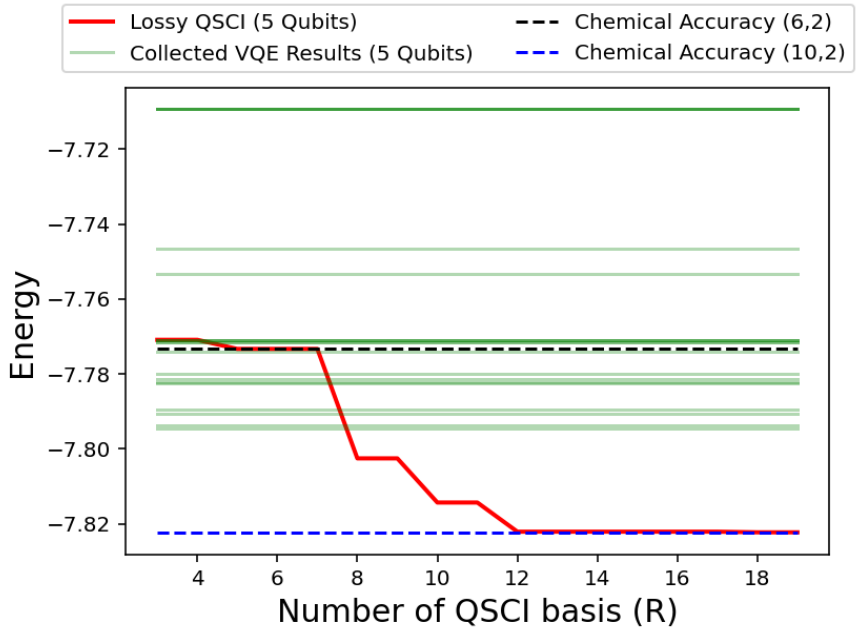


Fig. 2. The LiH molecule Lossy-QSCI Results with VQE. The x-axis represents the number of QSCI basis states (R) and the y-axis shows the computed energy in Hartree (Ha).

2.3 H_2 Molecule

We investigate the performance of noisy VQE optimization on the H_2 molecule with the 6-31G basis set at bond length 4Å. Figure 3 illustrates the performance of the Lossy-QSCI and QSCI under the influence of bit-flip error. For the Lossy-QSCI method, the candidate configuration set was constructed by sampling the 10 most frequent configurations from 20 subspaces of ground states (due to random compression) prepared by VQE. On the other hand, the QSCI method's candidate configuration set was constructed by sampling the 10 most frequent configurations from 50 subspaces of ground states prepared by VQE. Since Lossy-QSCI with Chemical-RLE is only optimized and sampled on the number-conserving subspace, we post-process the QSCI candidate configuration set only containing the number-conserving state to improve the construction of the effective Hamiltonian.

Both results demonstrate that enforcing number-conserving constraints can provide error robustness for both QSCI and noisy QSCI and enable them to recover from noisy VQE results, as depicted by the red and blue lines. Meanwhile, the Lossy-QSCI method (blue) outperforms QSCI (red), achieving energy below chemical accuracy with as few as 12 basis states compared to the 15 basis

states in QSCI. The total sampled configurations from Lossy-QSCI are 200 samples of 4-qubit configurations compared to the 500 samples of 8-qubit configurations for QSCI. This approach highlights the potential of Lossy-QSCI to deliver better searching efficiency. Robustness against bit-flip error shows the potential for making it far more suitable than QSCI in near-term devices.

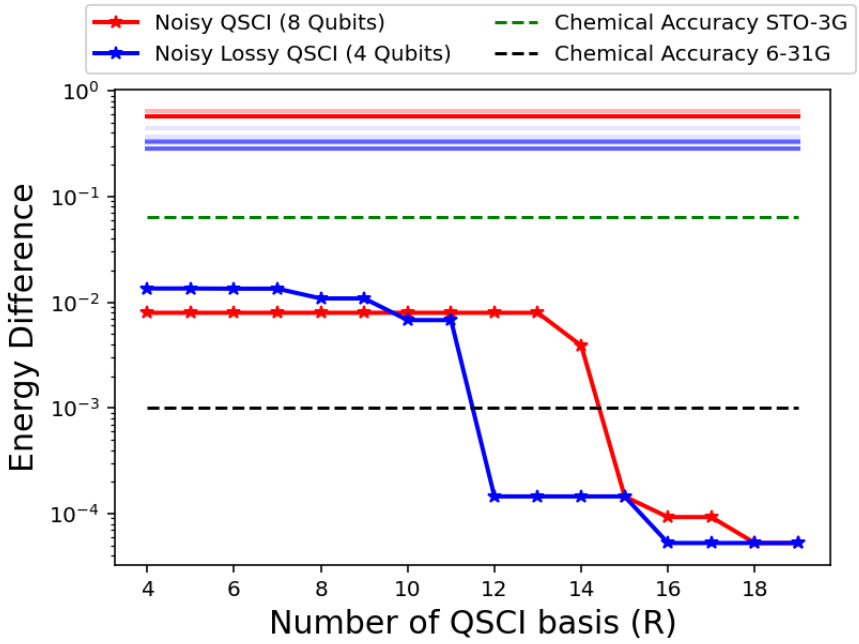


Fig. 3. This figure illustrates the H_2 molecule noisy simulation. A lossy compressed QSCI simulation not only significantly reduce the qubit cost, but also enhances the convergence w.r.t. increasing QSCI basis.

3 Conclusion

Lossy-QSCI marries a chemistry-aware lossy Random Linear Encoder with a neural-network Fermionic Expectation Decoder, compressing the register to $\mathcal{O}(N \log M)$ qubits while retaining all chemically relevant determinants.
The trained decoder recovers full-space observables in $\sim 10^{-3}$ s per shot, eliminating the exponential lookup table of earlier FED schemes.
Benchmarks on C_2, LiH, and H_2 achieve chemical accuracy with $\approx \frac{1}{2}$ the qubits and $\sim 25\%$ fewer CI states than uncompressed QSCI, and show built-in robustness to bit-flip noise.
These resource savings yield shallower circuits and lower sampling budgets, opening a practicable route to selected-CI on NISQ hardware and early fault-tolerant platforms.
Ongoing work will chart qubit–gate trade-offs for deeper encodings and extend the encoder–decoder paradigm to excited states and real-time dynamics.

Bibliography

[1] Aspuru-Guzik, A., Dutoi, A.D., Love, P.J., Head-Gordon, M.: Simulated Quantum Computation of Molecular Energies. Science **309**(5741), 1704–1707 (Sep 2005). https://doi.org/10.1126/science.1113479, https://www.science.org/doi/10.1126/science.1113479

[2] Cheng, M.H., Chen, Y.C., Wang, Q., Bartsch, V., Kim, M.S., Hu, A., Hsieh, M.H.: Optimal Particle-Conserved Linear Encoding for Practical Fermionic Simulation (Aug 2024). https://doi.org/10.48550/arXiv.2309.09370, http://arxiv.org/abs/2309.09370, arXiv:2309.09370 [quant-ph]

[3] Feynman, R.P.: Simulating physics with computers. International Journal of Theoretical Physics **21**(6-7), 467–488 (Jun 1982). https://doi.org/10.1007/BF02650179, http://link.springer.com/10.1007/BF02650179

[4] Jordan, P., Wigner, E.: ber das Paulische quivalenzverbot. Zeitschrift fr Physik **47**(9-10), 631–651 (Sep 1928). https://doi.org/10.1007/BF01331938, http://link.springer.com/10.1007/BF01331938

[5] Kanno, K., Kohda, M., Imai, R., Koh, S., Mitarai, K., Mizukami, W., Nakagawa, Y.O.: Quantum-Selected Configuration Interaction: classical diagonalization of Hamiltonians in subspaces selected by quantum computers (Feb 2023). https://doi.org/10.48550/arXiv.2302.11320, http://arxiv.org/abs/2302.11320, issue: arXiv:2302.11320 arXiv:2302.11320 [quant-ph]

[6] Mikkelsen, M., Nakagawa, Y.O.: Quantum-selected configuration interaction with time-evolved state (Feb 2025). https://doi.org/10.48550/arXiv.2412.13839, http://arxiv.org/abs/2412.13839, arXiv:2412.13839 [quant-ph]

[7] Nakagawa, Y.O., Kamoshita, M., Mizukami, W., Sudo, S., Ohnishi, Y.y.: Adapt-qsci: Adaptive construction of an input state for quantum-selected configuration interaction. Journal of Chemical Theory and Computation **20**(24), 10817–10825 (2024). https://doi.org/10.1021/acs.jctc.4c00846, pMID: 39642269

[8] Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.H., Zhou, X.Q., Love, P.J., Aspuru-Guzik, A., O'Brien, J.L.: A variational eigenvalue solver on a photonic quantum processor. Nature Communications **5**(1), 4213 (Jul 2014). https://doi.org/10.1038/ncomms5213, https://www.nature.com/articles/ncomms5213

[9] Reinholdt, P., Ziems, K.M., Kjellgren, E.R., Coriani, S., Sauer, S.P.A., Kongsted, J.: Fundamental Limitations in Sample-Based Quantum Diagonalization Methods (May 2025). https://doi.org/10.48550/arXiv.2501.07231, http://arxiv.org/abs/2501.07231, arXiv:2501.07231 [physics]

[10] Robledo-Moreno, J., Motta, M., Haas, H., Javadi-Abhari, A., Jurcevic, P., Kirby, W., Martiel, S., Sharma, K., Sharma, S., Shirakawa, T., Sitdikov, I., Sun, R.Y., Sung, K.J., Takita, M., Tran, M.C., Yunoki, S., Mezzacapo,

A.: Chemistry Beyond Exact Solutions on a Quantum-Centric Supercomputer (Nov 2024). https://doi.org/10.48550/arXiv.2405.05068, http://arxiv.org/abs/2405.05068, arXiv:2405.05068 [quant-ph]

[11] Shee, Y., Tsai, P.K., Hong, C.L., Cheng, H.C., Goan, H.S.: Qubit-efficient encoding scheme for quantum simulations of electronic structure. Physical Review Research **4**(2), 023154 (May 2022). https://doi.org/10.1103/PhysRevResearch.4.023154, http://arxiv.org/abs/2110.04112, arXiv:2110.04112 [cond-mat, physics:physics, physics:quant-ph]

Towards a Quantum Generative Graph-Based Clustering for Molecule Discovery

Julien Rauch[1 [0009−0009−4778−7692]], Damien Rontani[2,3 [0000−0002−8549−4040]], and Stéphane Vialle[4 [0000−0001−6336−2269]]

[1] LISN UMR-9015, Quantum-Saclay, CNRS, Université Paris-Saclay, Orsay, France
[2] LMOPS EA-4423, CentraleSupélec, Université de Lorraine, Metz, France
[3] Chair in Photonics, CentraleSupélec, Metz, France
[4] LISN UMR-9015, CentraleSupélec, CNRS, Université Paris-Saclay, Orsay, France

Abstract. In the field of drug discovery, it is necessary to speed up the exploration and analysis of very large datasets of molecules. A possible approach is the clustering of molecules with similar structures to limit the exploration of the chemical-space. This computationally-intensive operation can be mapped to the quantum domain offering inherent parallelism and providing possible higher-quality solutions.
We propose in this article a novel quantum-based clustering pipeline based on the use of quantum version of Generative Adversarial Network (QuGAN) to perform the clustering of open-access molecular data. Our contribution capitalizes on a previously-developped hybrid quantum data clustering algorithm that incorporates a molecule-specific quantum generative model for small molecular graphs (*e.g.*: QuMolGAN [11]).
We present preliminary results based on noiseless simulations of a quantum processing unit with 4 qubits. It shows gains in clustering quality after training our quantum generative pipeline on graph-based data from a quantum mechanical dataset (QM9) of small organic molecules. This motivates further investigation, in particular the fine-tuning hyperparameter optimization of the generative model, to achieve more distinct cluster separation and increased intra-cluster homogeneity.

Keywords: Generative Approach · Graph-based Clustering · Molecule Generation · Quantum Machine Learning

1 Motivation and Related Works

1.1 Molecule Graph Clustering

Molecule clustering is a fundamental cheminformatics technique to group molecules based on their structural and/or chemical similarities. The idea is that molecules belonging to the same cluster share multiple features whereas they differ significantly from molecules in other clusters. This technique is particularly interesting to accelerate the exploration of large-scale chemical compounds data sets [6, 2].

Molecular Data can be represented using a graph-theoretic approach called *molecular graph*, where vertices and edges represents the atoms and the chemical bonds, respectively. Grouping molecular graph data according to maximum

common subgraphs enables more efficient chemical space exploration. For example, by identifying families of similar graphs, one can dramatically speed up the search for novel molecules [17].

There are two main ways of grouping molecular graphs: directly using the graphs or using their fingerprints [19]. A molecule's fingerprint encodes its structural features in the form of a vector [8], which enables us to apply classical clustering algorithms (on these vectors). Fingerprints are widely used in machine learning, but we cannot use them to deduce the corresponding molecules in return. They therefore become ineffective in the context of a clustering algorithm based on a generative model that aims at learning how to directly generate candidate molecules.

Graph clustering refers to the process of classifying the nodes of a graph and is widely used in social networks, for example [14]. In our case, we perform graph-level clustering. This means that we group graphs together [10].

1.2 Quantum Generative Clustering

Many quantum clustering algorithms already exist, such as q-means [13], quantum spectral clustering [12], and others. However, these algorithms assume the availability of *Fault-Tolerant Quantum Computers* (FTQC), which are not yet a reality. To take advantage of current quantum hardware and get initial feedback, we use hybrid clustering algorithms in which certain subroutines only are implemented as quantum algorithms. The most suitable approaches for today's *Noisy Intermediate-Scale Quantum* (NISQ) devices are *Variational Quantum Algorithms* (VQAs), as they can be adapted to specific hardware constraints. For instance, the *Variational Quantum Eigensolver* (VQE) [5, 22] can be employed as a quantum subroutine.

In a previous work [16], we developed a novel clustering algorithm where quantum subroutines are discrete *Quantum Circuit Born Machine* (QCBM), trained and controlled by an *Expectation-Maximization* (EM) algorithm [4, 3]. A discrete QCBM is a variational, trainable quantum circuit, which output probability distribution matches that of our data [18]. In the context of data clustering, we use a mixture (*i.e.* a set) of QCBMs, where each of them is trained to reproduce the distribution of the data located in one given cluster. See Fig. 1 and Section 2 for a description of our ML hybrid pipeline tailored for Molecule clustering.

1.3 Quantum Generative Molecule Models

MolGAN is a generative model designed to produce molecular graphs [7], based on the framework of generative adversarial networks (GANs). The generator is a classical neural network that outputs two matrices. The first is the adjacency matrix of the molecular graph, with dimensions $V \times V \times L$, where V is the maximum number of atoms in a molecule, and L is the number of chemical bond types (represented using one-hot encoding). The second matrix encodes

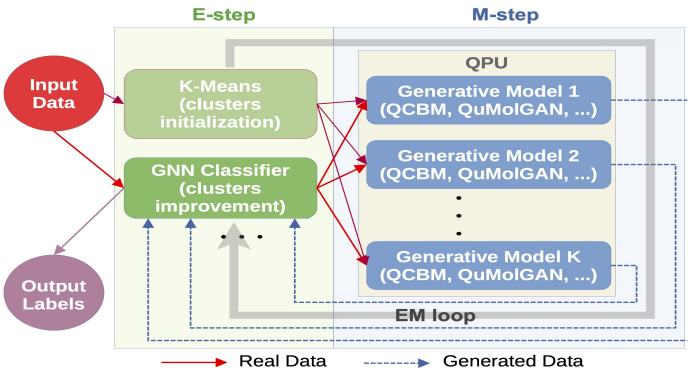


Fig. 1. Hybrid ML pipeline for data clustering

the types of atoms in the molecule, with dimensions $V \times M$, where M is the number of distinct atom types (also using one-hot encoding).

In the literature, several quantum variants of MolGAN have been proposed. These approaches adapt the GAN generator to a quantum or hybrid quantum-classical architecture, such as Hybrid Quantum MolGAN (HQ-MolGAN) [9], HQ-Cycle-MolGAN [1], and MolGAN-QC [11]. The latter also introduces three different versions: MolGAN-CQ with a quantum discriminator, MolGAN-QC with a quantum generator and QuMolGAN, which exploits quantum noise. The Gaussian noise usually used as input to the GAN generator is thus replaced by noise generated by a quantum circuit.

QuMolGAN has shown some quantum advantage in the generation of small molecules. Although it is more challenging to generate molecules with QuMolGAN compared to the classical MolGAN, the molecules it produces generally shows higher *Quantitative Estimate of Drug-likeness* (QED) scores. This mean that they possess better drug properties [11].

2 Development of our Hybrid Molecule Clustering

2.1 Hybrid Clustering Algorithm

We propose to adapt our generative clustering pipeline for molecules clustering and generation:

- For the first speed clustering we use the k-means clustering applicated on fingerprints of molecules.
- For the classifiers we took a Graph convolutional network (GNN). This ensures that molecule graphs sharing similar structures are located in the same cluster.
- For the generative model we took a Quantum MolGAN (QuMolGAN) exploiting quantum noise.

Our ML hybrid algorithm (see Fig. 1) starts by splitting the data in clusters with a naive and fast clustering algorithm (*e.g.*: k-means), then distribute the

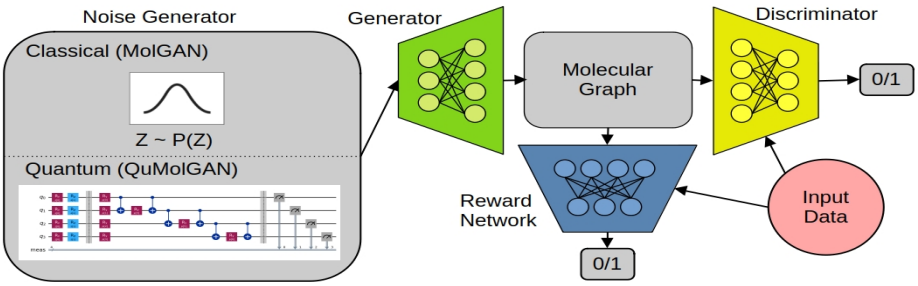


Fig. 2. Schema of QuMolGAN (inspired from [7])

data to our mixture of generative models. During a first M step each generative model learns its cluster attributed by the fast clustering algorithm. During a first E-step, new data clusters are generated by our generative models and we train a GNN classifier to match these new clusters while adjusting the input data clustering. Then, we enter a EM loop: The GNN classifier redistributes our data to the mixture of generative models (M-step). If data samples have a high probability to be in two clusters, then the GNN classifier can duplicate and send them to two distinct generative models. This mechanism guarantees an efficient exploration of the data space during the mixture evolution. The generative models generate new groups of data (E-step) and the classifier learns these new groups, adjusts the grouping of the input data one last time and returns the most probable label for each input data. Note that the GNN performs the calculation of distance, graph similarity in our algorithm.

As mentioned earlier, we use QuMolGAN as our generative model. The quantum circuit is located in the noise generator (see Fig. 2). In the original MolGAN, noise is sampled from a normal distribution and fed into the generator. In QuMolGAN, however, this noise is generated by a quantum parametric circuit. In our case, we use the same circuit architecture as the one proposed in [11]. In the MolGAN implementation we are using, the noise generator outputs a vector of dimension 4. To produce a comparable output, our quantum generator must also return a 4-dimensional vector. This requires a circuit of 4 qubits, where each component of the vector corresponds to the expectation value of the Pauli-Z observable on each qubit.

The reward network learns to assign a differentiable score to each molecule generated in order to drive the reinforcement learning process and enable MolGan to generate molecules with interesting chemical properties. The reward network is assisted in this task by the RDKit library, which provides it with non-differentiable scores (as QED score) for reference purposes.

2.2 Hybrid Software Stack

Figure 3 introduces the software stack of our python-based hybrid application. It includes several libraries (all with at less a Python interface): DeepChem [20]

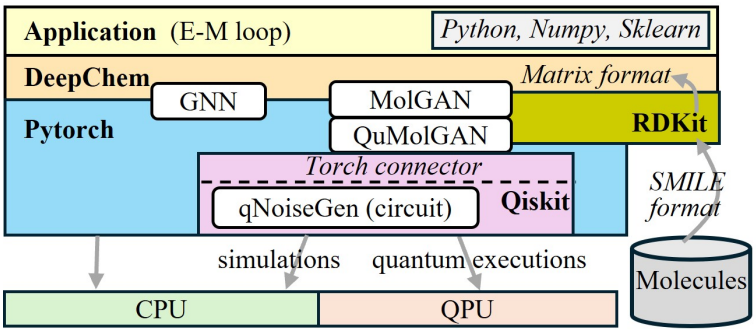


Fig. 3. Software stack for our hybrid application development

and RDKit[5] are chemical libraries, Pytorch is used to easily implement and control ML models, and Qiskit allows to implement quantum circuit and to run quantum simulations or to use real quantum hardware. The implementation of a consistent software stack was a lengthy process, requiring the identification of a set of compatible versions of all the libraries used. Finally, we managed to obtain a consistent stack based on Python 3.11.11, and from there, developing our application was easy.

We were therefore able to easily reuse our variational algorithm and its implementation in Python and Qiskit [16] by integrating MolGAN or qMolGAN objects, a GNN object, the qNoiseGen quantum circuit, and calls to RDKit. The latter library was mainly used to perform transformations of molecules from SMILE format to matrix format and to evaluate the chemical properties of the generated molecules.

To date, all of our executions and benchmarks have been performed using noise-free simulations. However, our software stack is ready to leverage real QPUs with Qiskit.

3 First experiments

3.1 Quantum Mechanical Dataset QM9

The QM9 Dataset is composed of 133,885 small organic molecules [21]. It has been used for machine learning tasks and the identification of structure-property relationships, making it adequate to benchmark our algorithm.

In this first phase of development of our work, we restrict ourselfs to a subset of 3,000 molecules from the initial data set mainly due to computational cost of the perfect simulation of noiseless quantum processing units (QPU) on multi-core GPUs (*e.g.* The simulation of one run with 200 iterations of the EM steps of our pipeline may take up to two days).

[5] RDKit: Open-source cheminformatics. https://www.rdkit.org

Metrics	3 clusters before training	3 clusters after training	5 clusters before training	5 clusters after training	15 clusters before training	15 clusters after training
Silhouette	-0.0465	-0.0409	-0.1571	-0.0008	-0.1890	-0.0696
Davies-Bouldin Index	6.0249	5.5015	4.2534	5.5445	3.786	4.3931
Calinski-Harabasz Index	22.794	33.017	21.128	47.4092	16.1305	39.4427

Table 1. Average scores over 5 runs of our hybrid algorithm for different clustering metrics on molecules of QM9 dataset (underlined scores indicate improved metrics).

3.2 Clustering Metrics

To evaluate the quality of our clustering we use several well-established *clustering metrics*[15]. We did not use the Normalized Mutual Information (NMI) or Adjusted Rand Index (ARI), since we do not have access to a base-line of properly clusterized molecular-graph data.

- The *Silhouette Score*: This metric assesses how similar an object is to its own cluster compared to other clusters. The score ranges from -1 to 1, where a value close to 1 indicates that the object is well-matched to its own cluster and poorly matched to neighboring clusters. A value close to -1 indicates the opposite.
- The *Davies-Bouldin Index* (DBI): This index evaluates the average similarity between each cluster and the one most similar to it, based on the ratio between intra-cluster distances and inter-cluster distances. It ranges in $[0; +\infty[$, and lower values indicate better clustering.
- The *Calinski-Harabasz Index*: Also known as the variance ratio criterion, this indicator assesses the ratio between inter-cluster dispersion and intra-cluster dispersion. It ranges in $[0; +\infty[$, and higher values indicate better-defined clusters.

However, it is difficult to apply these metrics directly to molecular graphs, as this involves defining a distance between two graphs. We therefore prefer to apply these metrics to the molecular fingerprint vectors established by RDKit, by calculating simple distances between two vectors. In our implementation all these metrics and distances are calculated by scikit-learn.

3.3 Preliminary Results

With the QM9 dataset, the exact number of clusters is unknown. As a result, we start with three clusters and gradually increase it to five. Our first results are introduced in Table 1. They show some improvements in clustering metrics between the start of the algorithm and the $200-$th iteration of the EM learning loop. Unfortunately, these first results are still quite low despite measurable effect associated with the evolution of mixture models during the learning phase. Thee underlying reasons for this limited performance level needs to be explored but could stem from sub-optimal selection of hyperparameters or improper choice in the number of clusters.

4 Conclusion and Next Steps

We have developed an hybrid quantum-classical algorithm and software architecture. They are executable and aim at generating candidate molecule structures directly (once trained) with similar properties to other existing molecules belonging to the same cluster rather than just generating fingerprints or features that would not allow the structures of the molecules to be retrieved.

Following this development phase and initial simulation experiments, we plan to continue this ongoing work in two main stages:

Short term developments: hybrid algorithm improvement

- Identify adapted graph-based clustering metrics to drive graph-based learning algorithm (instead of fingerprint-based metrics).
- Conduct an experimentation campaign to optimize the hyper-parameters (QuMolGAN hyper-parameters, GNN hyper-parameters, number of clusters, ...)
- Replace the fingerprint-based fast clustering method during algorithm initialization with a graph-based method.
- Compare our quantum-generative approach to classical graph-based clustering techniques.

Medium term work: experiments on real quantum architectures

- Switch to noisy quantum simulations and evaluate robustness of our hybrid algorithms.
- Test our hybrid algorithm on real (noisy) quantum computers.
- Use generated molecules scores as global quality metric to compare classic and hybrid approaches.
- Use larger molecule dataset (not limited to 3000 molecules).

Acknowledgments. This work was supported by the French government under the France 2030 program (QuanTEdu-France) under reference ANR-22-CMAS-0001. The authors acknowledge also the financial support of the Chair in Photonics and the Région Grand-Est, France.

References

1. Anoshin, M., Sagingalieva, A., Mansell, C., Zhiganov, D., Shete, V., Pflitsch, M., Melnikov, A.: Hybrid quantum cycle generative adversarial network for small molecule generation. IEEE Transactions on Quantum Engineering **5**, 1–14 (2024). https://doi.org/10.1109/tqe.2024.3414264
2. Backman, T.W.H., Cao, Y., Girke, T.: Chemmine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Research **39**(2), W486–W491 (2011). https://doi.org/10.1093/nar/gkr320
3. Benedetti, M., Garcia-Pintos, D., Perdomo, O., Leyton-Ortega, V., Nam, Y., Perdomo-Ortiz, A.: A generative modeling approach for benchmarking and training shallow quantum circuits. Npj Quantum Information **5**(1), 45 (2019). https://doi.org/10.1038/s41534-019-0157-8

4. Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M.: Parameterized quantum circuits as machine learning models. Quantum Science and Technology **4**, 043001 (2019). https://doi.org/10.1088/2058-9565/ab4eb5
5. Bermejo, P., Orus, R.: Variational quantum and quantum-inspired clustering (2024), https://arxiv.org/abs/2206.09893
6. Butina, D.: Unsupervised data base clustering based on daylight's fingerprint and tanimoto similarity: A fast and automated way to cluster small and large data sets. Journal of Chemical Information and Computer Sciences **39**(4), 747–750 (07 1999)
7. Cao, N.D., Kipf, T.: Molgan: An implicit generative model for small molecular graphs (2022), https://arxiv.org/abs/1805.11973
8. Capecchi, A., Probst, D., Reymond, J.L.: One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. Journal of Cheminformatics **12**(1) (Jun 2020). https://doi.org/10.1186/s13321-020-00445-4
9. Jain, P., Ganguly, S.: Hybrid quantum generative adversarial networks for molecular simulation and drug discovery (2022), https://arxiv.org/abs/2212.07826
10. Ju, W., Gu, Y., Chen, B., Sun, G., Qin, Y., Liu, X., Luo, X., Zhang, M.: Glcc: A general framework for graph-level clustering (2023), https://arxiv.org/abs/2210.11879
11. Kao, P.Y., Yang, Y.C., Chiang, W.Y., Hsiao, J.Y., Cao, Y., Aliper, A., Ren, F., Aspuru-Guzik, A., Zhavoronkov, A., Hsieh, M.H., Lin, Y.C.: Exploring the advantages of quantum generative adversarial networks in generative chemistry. Journal of Chemical Information and Modeling **63**(11), 3307–3318 (2023). https://doi.org/10.1021/acs.jcim.3c00562, pMID: 37171372
12. Kerenidis, I., Landman, J.: Quantum spectral clustering. Physical Review A **103**(4) (Apr 2021). https://doi.org/10.1103/physreva.103.042415
13. Kerenidis, I., Landman, J., Luongo, A., Prakash, A.: q-means: A quantum algorithm for unsupervised machine learning (2018), https://arxiv.org/abs/1812.03584
14. Liu, R., Feng, S., Shi, R., Guo, W.: Weighted graph clustering for community detection of large social networks. Procedia Computer Science **31**, 85–94 (2014). https://doi.org/https://doi.org/10.1016/j.procs.2014.05.248, https://www.sciencedirect.com/science/article/pii/S1877050914004256, 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
16. Rauch, J., Rontani, D., Vialle, S.: Data clustering on hybrid classical-quantum nisq architecture with generative-based variational and parallel algorithms. Journal of Systems Architecture p. 103431 (2025). https://doi.org/10.1016/j.sysarc.2025.103431
17. Raymond, J.W., Blankley, C., Willett, P.: Comparison of chemical clustering methods using graph- and fingerprint-based similarity measures. Journal of Molecular Graphics and Modelling **21**(5), 421–433 (2003). https://doi.org/10.1016/S1093-3263(02)00188-2
18. Riofrio, C.A., Mitevski, O., Jones, C., Krellner, F., Vuckovic, A., Doetsch, J., Klepsch, J., Ehmer, T., Luckow, A.: A characterization of quantum generative models. ACM Transactions on Quantum Computing **5**(2), 1–34 (Jun 2024). https://doi.org/10.1145/3655027
19. Seeland, M., Johannes, A.K., Kramer, S.: Structural clustering of millions of molecular graphs. In: Proceedings of the 29th Annual ACM Symposium on Applied Com-

puting. p. 121–128. SAC '14, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2554850.2555063
20. Shreyas, V., Siguenza, J., Bania, K., Ramsundar, B.: Open-source molecular processing pipeline for generating molecules (2024), https://arxiv.org/abs/2408.06261
21. Valdés, J.J., Tchagang, A.B.: Understanding the structure of qm7b and qm9 quantum mechanical datasets using unsupervised learning (2023), https://arxiv.org/abs/2309.15130
22. Yung, C., Usman, M.: Clustering by contour coreset and variational quantum eigensolver. Advanced Quantum Technologies **7**(8) (Jun 2024). https://doi.org/10.1002/qute.202300450

Quantum algorithm for anisotropic diffusion and convection equations with vector norm scaling

Julien Zylberman[1], Thibault Fredon[2], Nuno F.Loureiro[2], and Fabrice Debbasch[1]

[1] Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LUX, F-75005 Paris, France
[2] Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Abstract. In this work, we tackle the resolution of partial differential equations (PDEs) on digital quantum computers. Two fundamental PDEs are addressed: the anisotropic diffusion equation and the anisotropic convection equation. We present a quantum numerical scheme consisting of three steps: quantum state preparation, evolution with diagonal operators, and measurement of observables of interest. The evolution step relies on a high-order centered finite difference and a product formula approximation, also known as Trotterization. We provide novel vector-norm analysis to bound the different sources of error. We prove that the number of time-steps required in the evolution can be reduced by a factor $\Theta(16^n)$ for the diffusion equation, and $\Theta(4^n)$ for the convection equation, where n is the number of qubits per dimension, an exponential reduction compared to the previously established operator-norm analysis.

Keywords: quantum algorithms · partial differential equations · quantum computing · quantum numerical scheme · convection equation · diffusion equation

1 Introduction

The development of quantum numerical schemes for solving linear and non-linear partial differential equations is an active field of research, with open question about the computational efficiency, the reachable accuracies and the potential speedups when compared to classical methods. While the simulation of quantum systems by quantum computers has been extensively studied [1, 2, 3, 4, 5, 6], the simulation of classical (non-quantum) systems requires the development of novel quantum numerical schemes. Those need to take into account non-quantum evolution, i.e., non-unitary or non-linear evolution, that are not directly implementable on quantum computers.

In the following, we address the problem of solving anisotropic convection and anisotropic diffusion equations on quantum computers. First, we introduce a quantum numerical scheme based on high-order central finite difference and product formula approximation, also known as Trotter-splitting. Then, we present a numerical analysis based on vector norm instead of generic operator

norm analysis[3]. We show that the vector norm analysis predicts a reduction of the number of time-steps by an exponential factor $\theta(4^n)$ for the anisotropic convection equation, and by a factor $\theta(16^n)$ for the anisotropic diffusion equation, where n is the number of qubits per dimension.

2 Problem and approach

The first considered PDE is the d-dimensional anisotropic diffusion equation:

$$\partial_t \phi = \boldsymbol{\nabla} \cdot (K(\boldsymbol{x}, t) \boldsymbol{\nabla} \phi), \tag{1}$$

where ϕ is the unknown, $\boldsymbol{\nabla} = (\partial_{x_1}, \ldots, \partial_{x_d})$ and $K(\boldsymbol{x}, t) = \mathrm{diag}(\kappa_1(\boldsymbol{x}, t), \ldots, \kappa_d(\boldsymbol{x}, t))$ is a space- and time-dependent thermal conductivity. The second PDE of interest is the d-dimensional anisotropic convection equation:

$$\partial_t f + \boldsymbol{c}(\boldsymbol{x}, t).\boldsymbol{\nabla} f = 0, \tag{2}$$

where $\boldsymbol{c}(\boldsymbol{x}, t) = (c_1(\boldsymbol{x}, t), \ldots, c_d(\boldsymbol{x}, t))$ is a space- and time-dependent velocity field. Both equations are defined on the d-dimensional torus $\mathbb{T}^d$, i.e., the box $[0, 1]^d$ with periodic boundary conditions, with a time $t \in [0, T]$ for some $T > 0$, and with initial conditions $\phi(\boldsymbol{x}, t = 0) = \phi_0(\boldsymbol{x})$ and $f(\boldsymbol{x}, t = 0) = f_0(\boldsymbol{x})$ respectively. Additionally, we assume that the c_j and the κ_j functions are $2p+1$ differentiable, for some integer p, and do not depend on the j-th variables: $\forall j \in \{1, \ldots, d\}$, $\partial_{x_j} c_j = 0$, $\partial_{x_j} \kappa_j = 0$[4]. This last assumption is typical for different examples, such as the collisionless Boltzmann equation with a Lorentz force, the Liouville equation associated with non-linear Hamiltonian systems, or for diffusion processes in some functionally graded materials.

Real-space encoding. In order to numerically solve these equations on quantum computers, we defined the real space encoding of the solution function into a $n = n_1 + \ldots + n_d$ qubit state as: $|f\rangle_t = \sum_{\boldsymbol{x}} f(\boldsymbol{x}, t) |\boldsymbol{x}\rangle \in \mathcal{H}$, where $\mathcal{H} = \mathcal{H}_{x_1} \otimes \ldots \otimes \mathcal{H}_{x_d}$ is the Hilbert space associated with the n qubits and each $\mathcal{H}_{x_j}$ is a 2^{n_j}-dimensional Hilbert space. The ket vector $|\boldsymbol{x}\rangle$ is defined as the tensor product of the ket vector of each space $|\boldsymbol{x}\rangle = |x_1\rangle \otimes \ldots \otimes |x_d\rangle$. Each ket vector $|x_j\rangle$ is defined as $|x_j\rangle = |q_0\rangle \otimes \ldots \otimes |q_{n_j - 1}\rangle$, with $x_j = \sum_{k=0}^{n_j - 1} q_k / 2^{k+1} \in \{0, 1/2^{n_j}, \ldots, (2^{n_j} - 1)/2^{n_j}\}$ being the dyadic expansion of x and $\forall k, q_k \in \{0, 1\}$.

[3] In the literature, quantum algorithms for simulating Hamiltonian systems, or solving differential equations, often have a complexity given in terms of the spectral norm of the associated operator, such as the Hamiltonian or the operator $\hat{A}$ when solving a linear ordinary differential equation $df/dt = \hat{A} f$.

[4] This assumption is necessary to derive an efficient quantum circuit made only of quantum Fourier transforms and diagonal operators for the evolution step.

Quantum state preparation. The quantum numerical scheme is based on three steps. The first one is a quantum state preparation routine that encodes the initial condition into an n-qubit state: $\phi_0 \to |\phi_0\rangle = \frac{1}{\mathcal{N}} \sum_{\boldsymbol{x}} \phi_0(\boldsymbol{x})|\boldsymbol{x}\rangle$, $f_0 \to |f_0\rangle = \frac{1}{\mathcal{N}'} \sum_{\boldsymbol{x}} f_0(\boldsymbol{x})|\boldsymbol{x}\rangle$, where $\mathcal{N}, \mathcal{N}'$ are normalization constants. Many quantum state preparation routines have been published in the literature: exact protocols generically require a number of quantum gates proportionate to the number of components of the target quantum states. In other words, optimal circuit size[5] for preparing arbitrary n-qubit state is $O(2^n)$ [7, 8, 9]. Fortunately, efficient methods[6] have been developed for qubit states depending on differentiable functions, which leverage the structure of the state by using Walsh, Fourier or polynomial series [10, 11, 12, 13].

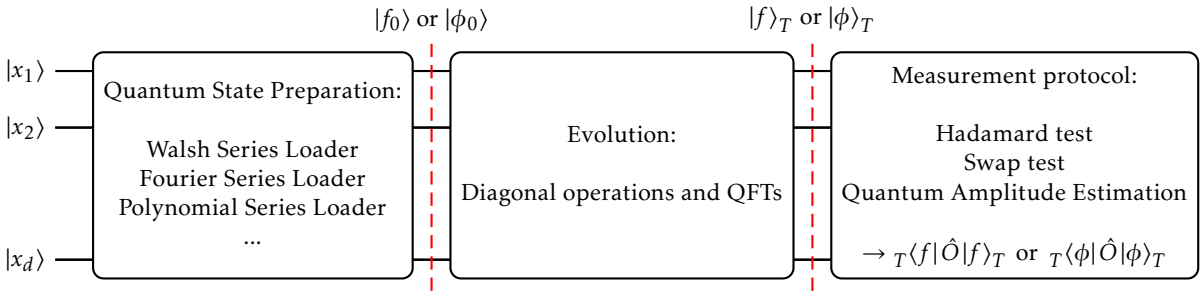


Fig. 1: Quantum numerical scheme associated with the resolution of the d-dimensional convection or diffusion equations with position- and time- dependent coefficients.

Evolution. The second step reproduces the PDE evolution using primitive quantum gates, evolving the initial qubit state $|\phi_0\rangle$ (resp. $|f_0\rangle$) toward a qubit state that encodes the PDE's solution at a target time $T > 0$: $|\phi_0\rangle \to |\phi\rangle_T = \frac{1}{\mathcal{N}''} \sum_{\boldsymbol{x}} \phi(\boldsymbol{x}, T)|\boldsymbol{x}\rangle$ (resp. $|f_0\rangle \to |f\rangle_T = \frac{1}{\mathcal{N}''} \sum_{\boldsymbol{x}} f(\boldsymbol{x}, T)|\boldsymbol{x}\rangle$). The evolution step is based on a high-order centered finite difference $\partial_{x_i}\phi(\boldsymbol{x}, t) \to (1/\Delta x_i) \sum_{k=-p}^{p} a_k \phi(\boldsymbol{x} + k\boldsymbol{\Delta x_i}, t) + O(\Delta x_i^{2p})$, where $a_0 = 0$ and, for $k \in \{-p, -p+1, ...p-1, p\}\backslash\{0\}$, $a_k = (-1)^{k+1}(p!)^2/(k(p-k)!(p+k)!)$[7]. By defining the discrete derivative operator associated with the j–th axis as $\hat{D}_j = \hat{I}_1 \otimes \cdots \otimes \hat{I}_{j-1} \otimes -i \frac{\sum_{k=-p}^{p} a_k (\hat{S}_j)^k}{\Delta x_j} \otimes \hat{I}_{j+1} \otimes \cdots \otimes \hat{I}_d$, where $\hat{S}_j = \sum_{x_j} |(x_j + \Delta x_j) \mathrm{mod} 1\rangle\langle x_j|$ is the shift operator over the j-th axis, one can recast the PDE problems as ordi-

[5] The size of a quantum circuit is the number of primitive quantum gates.

[6] A quantum circuit is said efficient if its circuit complexity expressed in terms of size, depth and ancilla qubits scale polynomially with n and $1/\epsilon$.

[7] The a_k coefficients are the finite difference coefficients, solution of the system of equations $\sum_{k=-p}^{p} a_k k^j = \delta_{j,1}$, where $j \in \{0, \ldots, 2p\}$.

nary differential equation (ODE) problems:

$$\frac{d|\tilde{f}\rangle_t}{dt} = -i(\sum_{j=1}^{d} \hat{c}_j(t)\hat{D}_j)|\tilde{f}\rangle_t, \tag{3}$$

$$\frac{d|\tilde{\phi}\rangle_t}{dt} = -(\sum_{j=1}^{d} \hat{\kappa}_j(t)\hat{D}_j^2)|\tilde{\phi}\rangle_t, \tag{4}$$

with the initial conditions $|\tilde{\phi}\rangle_{t=0} = |\phi_0\rangle$, $|\tilde{f}\rangle_{t=0} = |f_0\rangle$ and where $\hat{c}_j(t) = c_j(\hat{\boldsymbol{x}}, t)$, $\hat{\kappa}_j(t) = \kappa_j(\hat{\boldsymbol{x}}, t)$ with $\hat{\boldsymbol{x}} = (\hat{x}_1, ...\hat{x}_d)$, $\hat{x}_j = \hat{I}_1 \otimes \cdots \otimes \hat{I}_{j-1} \otimes \sum_{x_j} x_j |x_j\rangle\langle x_j| \otimes \hat{I}_{j+1} \otimes \cdots \otimes \hat{I}_d$. To be explicit, we distinguish $|\tilde{\phi}\rangle_t$ and $|\tilde{f}\rangle_t$, the solutions of the ODE problems, to the real space encoding $|\phi\rangle_t$ and $|f\rangle_t$ of the solution ϕ and f of the diffusion equation Eq.1 and convection equation Eq.2. Furthermore, the "hat" notation only refers to operators.

The solutions of the ODEs are given by time-ordered evolution operators as: $|\tilde{f}\rangle_t = \mathcal{T} e^{-i\int_0^t \sum_{j=1}^d \hat{c}_j(s)\hat{D}_j ds}|f_0\rangle$ and $|\tilde{\phi}\rangle_t = \mathcal{T} e^{-\int_0^t \sum_{j=1}^d \hat{\kappa}_j(s)\hat{D}_j^2 ds}|\phi_0\rangle$, where $\mathcal{T}$ is the time-ordering operator. In general, it is not possible to implement directly time-ordered exponentials on computers, whether classical or quantum. To overcome this issue, we consider first-order product formula to decompose the time-ordered evolution operators into efficiently implementable unitary operators. The standard product formula denoted by the index s and the generalized one denoted by g are defined as [14]:

$$\begin{aligned}
\hat{U}_s(t+h,t) &= \overleftarrow{\prod_{j=1}^{d}} \exp\Big(-ih\hat{c}_j(t+h)\hat{D}_j\Big), \quad \hat{U}_g(t+h,t) = \overleftarrow{\prod_{j=1}^{d}} \exp\left(-i\int_t^{t+h} \hat{c}_j(s)ds\hat{D}_j\right), \\
\hat{V}_s(t+h,t) &= \overleftarrow{\prod_{j=1}^{d}} \exp\Big(-h\hat{\kappa}_j(t+h)\hat{D}_j^2\Big), \quad \hat{V}_g(t+h,t) = \overleftarrow{\prod_{j=1}^{d}} \exp\left(-\int_t^{t+h} \hat{\kappa}_j(s)ds\hat{D}_j^2\right),
\end{aligned} \tag{5}$$

where $\overleftarrow{\prod_{j=1}^d} A_j = A_d ... A_1$ is a product with decreasing indexes for non-commuting operators. The product formula approximation is as accurate as h is small. Therefore, we discretize the time span $[0, T]$ into L equal segments and we define $|\tilde{\phi}_\alpha\rangle_t$ and $|\tilde{f}_\alpha\rangle_t$, with $\alpha = s, g$, the qubit states given by L trotterized-time-steps as:

$$\begin{aligned}
|\tilde{\phi}_\alpha\rangle_T &= \prod_{l=1}^{L} \hat{V}_\alpha(\frac{lT}{L}, \frac{(l-1)T}{L})|\phi_0\rangle, \\
|\tilde{f}_\alpha\rangle_T &= \prod_{l=1}^{L} \hat{U}_\alpha(\frac{lT}{L}, \frac{(l-1)T}{L})|f_0\rangle.
\end{aligned} \tag{6}$$

The implementation of the operators $\hat{V}_\alpha$ and $\hat{U}_\alpha$ is performed using quantum Fourier transforms and diagonal operators. The quantum Fourier transform

(QFT) diagonalizes the discrete derivative operators $\hat{D}_j$ as $\hat{D}_j = \widehat{QFT}_j^{-1}(\sum_{x_j} d(x_j) |x_j\rangle\langle x_j|)\widehat{QFT}_j$, where $\widehat{QFT}_j$ is the QFT acting on the j-th register and $d(x_j) = \frac{2}{\Delta x_j}\sum_{q=0}^{p} a_q \sin\left(2\pi q x_j\right)$ is one of the eigenvalues of the operator $\hat{D}_j$. Consequently, the operators $\hat{D}_j^2$ and each of the exponential defined in Equation 5 are also diagonalized by the QFT. The associated quantum circuit is well established for the QFT [15]. The diagonal operators depend on differentiable functions and can be efficiently implemented using (sparse) Walsh or Fourier series approximations (see references [11, 16] for the detailed quantum circuit constructions).

Measurement protocol. The last step of the numerical scheme is the extraction of information from the final qubit state $|\phi\rangle_T$ (resp. $|f\rangle_T$) using measurement protocols such as quantum amplitude estimation, the Hadamard test or the Swap test [15, 17, 18, 19, 20, 21]. These protocols enable the measurement of averaged values of observables $|\phi\rangle_T \rightarrow {}_T\langle\phi|\hat{O}|\phi\rangle_T$, where $\hat{O}$ is designed to correspond to relevant information, such as the value of the solution at a given position, the average position of the solution, its standard deviation, or higher-order moments[8].

3 Error analysis

Space discretization error. The quantum numerical scheme introduced two approximations: the space-discretization error and the product formula error. The first one is quantified by the difference between $|f\rangle_t$, the normalized vector encoding the solution of the convection equation Eq.(2) at time t, and $|\tilde{f}\rangle_t$, the solution of the ODE Eq.(3). One can bound their difference by writing the following equation for $|f\rangle_t$: $\partial_t|f\rangle_t = -i\sum_{j=1}^{d}\hat{c}_j(t)\hat{D}_j|f\rangle_t + |r(t)\rangle$ with $|r(t)\rangle = \sum_{j=1}^{d}\hat{c}_j(t)(i\hat{D}_j|f\rangle_t - |\partial_{x_j}f\rangle_t)$. The solution of this equation is given by variation of parameter formula as $|f\rangle_t = \hat{U}(t,0)|f_0\rangle + \int_0^t \hat{U}(t,s)|r(s)\rangle ds$, where $\hat{U}$ is the time-ordered evolution operator associated with the ODE Eq.(3). The variation of parameter formula makes it possible to bound the difference as $\||f\rangle_t - |\tilde{f}\rangle_t\|_{2,N} \leq t\max_{s\in[0,t]}\||r(s)\rangle\|_{2,N}$, where $\|.\|_{2,N}$ is simply the vector 2-norm of a vector of N-components. Then, using Taylor inequality, one can show $i\hat{D}_j|f\rangle = |\partial_{x_j}f\rangle + O((\Delta x_j)^{2p})$. Finally, the difference is bounded as $\||\tilde{f}\rangle_T - |f\rangle_T\|_{2,N} \leq tK\sum_{j=1}^{d}\|c_j\|_\infty(\Delta x_j)^{2p}$, where K is a constant depending on p and the maximum value over $j \in \{0,...,d\}$ and time of the $(2p+1)$ derivative of f_0 with respect to axis j. Since $\Delta x = 1/2^n$, one can choose a number of qubits associated

[8] Remark that, for the diffusion equation, the overall computational cost to estimate an observable of interest is dependent on the norm of the solution: either the evolution includes some projective measurement to ensure that $\||\tilde{\phi}_\alpha\rangle_T\|_{2,N} = 1$ with a probability of success proportional to $\||\tilde{\phi}_\alpha\rangle_T\|_{2,N}$, either one perform the measurement protocol with $\||\tilde{\phi}_\alpha\rangle_T\|_{2,N} \leq 1$ and the cost factor $\||\tilde{\phi}_\alpha\rangle_T\|_{2,N}$ is passed on the probabilistic cost of the measurement protocol.

with the j-th axis as $n_j = \lceil \frac{1}{2p} \log_2(TK\|c_j\|_\infty/\epsilon) \rceil$ to ensure an ϵ approximation level.

One can proceed analogously to bound the discretization error associated with the diffusion equation. Since the diffusion equation has a non-unitary evolution[9], one needs to consider the difference between the normalized vectors as: $\| \frac{|\tilde{\phi}\rangle_t}{\||\tilde{\phi}\rangle_t\|_{2,N}} - \frac{|\phi\rangle_t}{\||\phi\rangle_t\|_{2,N}} \|_{2,N} \leq TK \sum_{j=1}^{d} \|\kappa_j\|_\infty (\Delta x_j)^{2p}$, where K is independent of N and depends on the the $2p+1$ derivatives of ϕ.

Product formula approximation. A direct expansion of the Trotter formula gives $\| \left(e^{i(\hat{H}_1+\hat{H}_2)\Delta t} - e^{i\hat{H}_1\Delta t} e^{i\hat{H}_2\Delta t} \right) |\psi\rangle \|_{2,N} = \frac{\Delta t^2}{2} \| [\hat{H}_2, \hat{H}_1] |\psi\rangle \|_{2,N} + O(\Delta t^3)$. In the literature on quantum algorithms, most of the scaling uses the operator norm inequality by bounding the commutator as $\| [\hat{H}_2, \hat{H}_1] |\psi\rangle \|_{2,N} \leq \| [\hat{H}_2, \hat{H}_1] \|_2 \leq 2\|\hat{H}_1\|_2 \|\hat{H}_2\|_2$ [10]. However, in the case of the convection or the diffusion equation, the summand $\hat{H}_j$ depends on the discrete derivative operator such as $\|\hat{H}_j\|_2 = \|c_j \hat{D}_j\|_2 = O(1/\Delta x_j) = O(2^{n_j})$ or $\|\hat{H}_j\|_2 = \|\kappa_j \hat{D}_j^2\|_2 = O(1/\Delta x_j^2) = O(4^{n_j})$, resulting in exponential-with-n error bounds. In the following, we overcome these exponential scaling by considering the action of the operator $\hat{D}_j$ on the qubit state encoding the solution $|f\rangle_t$ (or $|\phi\rangle_t$), deriving a novel bound in vector norm with terms of the form $\| [\hat{H}_2, \hat{H}_1] |f\rangle_t \|_{2,N} = O(1)$. Precisely, one can prove that for the convection equation and $\alpha = s, g$[11] :

$$\| |\tilde{f}_\alpha\rangle_T - |\tilde{f}\rangle_T \|_{2,N} \leq a_\alpha \frac{T^2}{L} + r_\alpha, \tag{7}$$

where r_α contains higher order terms that are asymptotically negligible in the large $L \to +\infty$ limit and

[9] The evolution operator $\hat{U}$ associated with the discretized convection equation Eq.(3) is unitary thanks to the central finite difference and the assumption that $\forall j, \partial_{x_j} c_j = 0$, while the evolution operator associated with the discretized diffusion equation Eq.(4) is non-unitary, but has a spectral norm $\|\hat{V}_\alpha\| = 1$, since 0 is an eigenvalue of the discrete derivative operators: $\forall j, d_j(0) = 0$, which expresses the conservation over time of the average value of the solution function.

[10] The first reference introducing vector norm analysis for quantum algorithm is [14], which address the problem of solving the Schrödinger equation with a bounded potential, i.e., in the case where $\hat{H}_1$ is asymptotically unbounded and $\hat{H}_2$ is bounded. Our results generalize their approach since all the summands are composed of discrete derivative operators that are asymptotically unbounded for the convection and diffusion equations.

[11] Proof for the convection equation can be found in [22], and its generalization to the diffusion equation is straightforward under the assumptions of this work.

$$a_g = \frac{1}{2}\sum_{j=1}^{d-1}\sum_{m=j+1}^{d}\left(\|c_j\|_\infty\|\partial_{x_j}c_m\|_\infty \max_{t\in[0,T]}\frac{\|\partial_{x_m}f_t\|_{2,N}}{\|f_0\|_{2,N}} + \|c_m\|_\infty\|\partial_{x_m}c_j\|_\infty \max_{t\in[0,T]}\frac{\|\partial_{x_j}f_t\|_{2,N}}{\|f_0\|_{2,N}}\right),$$

$$a_s = a_g + \frac{1}{2}\sum_{j=1}^{d}\|\partial_t c_j\|_\infty \max_{t\in[0,T]}\frac{\|\partial_{x_j}f_t\|_{2,N}}{\|f_0\|_{2,N}} \leq a_g + \frac{d}{2}\max_{1\leq j\leq d, t\in[0,T]}\|\partial_t c_j\|_\infty \frac{\|\partial_{x_j}f_t\|_{2,N}}{\|f_0\|_{2,N}}.$$

Remark that $\|\partial_{x_m}f_t\|_{2,N}/\|f_0\|_{2,N}$ converges toward $\|\partial_{x_m}f_t\|_{L^2}/\|f_0\|_{L^2} = \|\partial_{x_m}f_0\|_{L^2}/\|f_0\|_{L^2}$ in the large N limits, which depends only on f_0[12].

Similarly, for the diffusion equation, one can show:

$$\left\|\frac{|\tilde{\phi}_\alpha\rangle_T}{\||\tilde{\phi}_\alpha\rangle_T\|_{2,N}} - \frac{|\tilde{\phi}\rangle_T}{\||\tilde{\phi}\rangle_T\|_{2,N}}\right\|_{2,N} \leq a'_\alpha\frac{T^2}{L} + r'_\alpha, \tag{8}$$

where r'_α contains higher order terms that are asymptotically negligible and

$$a'_g = \frac{1}{2}\sum_{j=1}^{d-1}\sum_{m=j+1}^{d}\left(\|\kappa_j\|_\infty\|\partial^2_{x_j}\kappa_m\|_\infty \max_{t\in[0,T]}\frac{\|\partial^2_{x_m}\phi_t\|_{2,N}}{\|\tilde{\phi}_T\|_{2,N}} + \|\kappa_m\|_\infty\|\partial^2_{x_m}\kappa_j\|_\infty \max_{t\in[0,T]}\frac{\|\partial^2_{x_j}\phi_t\|_{2,N}}{\|\tilde{\phi}_T\|_{2,N}}\right),$$

$$a'_s = a'_g + \frac{1}{2}\sum_{j=1}^{d}\|\partial_t\kappa_j\|_\infty \max_{t\in[0,T]}\frac{\|\partial^2_{x_j}\phi_t\|_{2,N}}{\|\tilde{\phi}_T\|_{2,N}} \leq a_g + \frac{d}{2}\max_{1\leq j\leq d, t\in[0,T]}\|\partial_t\kappa_j\|_\infty\frac{\|\partial^2_{x_j}\phi_t\|_{2,N}}{\|\tilde{\phi}_T\|_{2,N}},$$

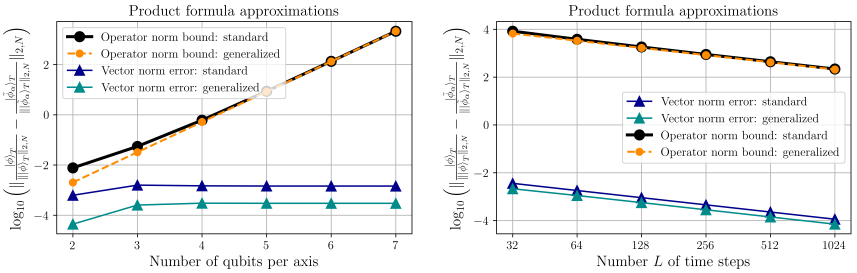
where $\|\tilde{\phi}_T\|_{2,N} = \|\widehat{QFT}\,\tilde{\phi}_T\|_{2,N} \geq |\langle 0|^{\otimes n}\,\widehat{QFT}\,|\phi\rangle_T| = |\sum_{\boldsymbol{X}}\tilde{\phi}(\boldsymbol{X},T)|/\sqrt{N} = |\sum_{\boldsymbol{X}}\phi_0(\boldsymbol{X})|/\sqrt{N}$, where the average value of $\tilde{\phi}$ is conserved over time and equal to the average value of ϕ_0. Then, by assuming that the average value of ϕ_0 is non-zero, one deduce that $1/\|\tilde{\phi}_T\|_{2,N} \leq \sqrt{N}/(|\sum_{\boldsymbol{X}}\phi_0(\boldsymbol{X})|) \xrightarrow{N\to+\infty} 1/(\sqrt{N}\langle\phi_0\rangle_{\mathbb{T}^d})$ with $\langle\phi_0\rangle_{\mathbb{T}^d} = \int_{\mathbb{T}^d}\phi_0(\boldsymbol{x})d\boldsymbol{x}$ and $\|\partial^2_{x_m}\phi_t\|_{2,N}/\|\tilde{\phi}_T\|_{2,N} \leq \sqrt{N}\|\partial^2_{x_m}\phi_t\|_{2,N}/|\sum_{\boldsymbol{X}}\phi_0(\boldsymbol{X})| \xrightarrow{N\to+\infty} \|\partial_{x_m}\phi_t\|_{L^2}/\langle\phi_0\rangle_{\mathbb{T}^d}$.

In both case, the product formula approximation scales as $O(T^2/L)$, with a prefactor independent of the discretization step. Consequently, the number of time steps L to reach an arbitrary precision $\epsilon > 0$ should scale as $L = O(T^2/\epsilon)$. It represents an exponential-with-n reduction compared to operator-norm analysis that predicts a number of time-step scaling with the operator norms $\|\hat{D}_j^2\|_2 = \Theta(4^{n_j})$ for the convection equation, and $\|\hat{D}_j^4\|_2 = \Theta(16^{n_j})$ for the diffusion equation.

These scalings are validated through simulations of a two-dimensional diffusion equation with space- and time-dependent thermal conductivity, as shown in Figure 2. In Figure 2a, the vector-norm error and operator-norm bounds are plotted against the number of qubits per axis, confirming that the vector-norm

[12] The L^2 norm of any derivative of f is constant over time, equal to its initial value at time $t = 0$.

error is asymptotically independent of n. Figure 2b then shows the same quantities as a function of the number of time steps, exhibiting the expected linear decrease with L. Notably, while the operator-norm bounds suggest that achieving an error of 10^{-3} would require approximately $L \simeq 200000$ time-steps, the simulations demonstrate that the same accuracy is effectively attained with only $L \simeq 64$ time-steps, proving the relevance of vector-norm scalings. Additionally, simulations of anisotropic convection equations that illustrate the vector-norm scalings can be found in reference [22].



(a) (b)

Fig. 2: Product formula errors for the generalized product formula $\alpha = g$ and the standard product formula $\alpha = s$ as a function of the number of qubits per axis with $L = 256$ (left) and, as a function of the number of time steps L with $n_1 = n_2 = 5$ (right). The simulations are performed for the diffusion equation of dimension $d = 2$ with $\kappa_1(x_2, t) = \cos(\omega t - kx_2) + 1, \kappa_2(x_1, t) = \cos(\omega t - kx_1) + 1$ and parameters $\omega = 128\pi, k = 2\pi$. The numerical solution $|\tilde{\phi}_\alpha\rangle_T$ is compared to the solution of the diffusion equation computed via a Runge-Kunta classical solver of order 5. Time is $T = 0.01$ for the left plot and $T = 0.1$ for the right plot. The initial condition is $f_0(x_1, x_2) \propto e^{-((x_1-\mu_1)^2+(x_2-\mu_2)^2)/(2\sigma^2)}$ with $\mu_1 = 0.5, \mu_2 = 0.5$ and $\sigma = 1/(10\sqrt{2})$.

4 Conclusion

In this work, we introduced quantum numerical schemes for an anisotropic convection equation and for an anisotropic diffusion equation. The numerical schemes are composed of three steps: a quantum state preparation to load the initial condition into a qubit state; an evolution step made of diagonal operators and quantum Fourier transforms; and a measurement protocol to estimate quantities of interest from the numerical solution. We presented an error analysis of the space discretization and product formula approximations. In particular, we introduced novel computations based on vector-norm analysis to prove that the required number of time-steps needed to reach a given accuracy $\epsilon > 0$ can be

reduced by an exponential-with-n factor, compared to operator norm analysis. These results pave the way for the development of efficient quantum numerical schemes for other PDE problems such as the Vlasov equation or the Fokker-Planck equation, and for the development of novel numerical analysis for other types of quantum numerical schemes. In particular, the vector-norm scaling is expected to extend to non-periodic PDEs, although a different quantum numerical scheme would be required, since simple QFT-based diagonalization may no longer be applicable.

References

[1] Ivan Kassal et al. "Polynomial-time quantum algorithm for the simulation of chemical dynamics". In: *Proceedings of the National Academy of Sciences* 105.48 (2008), pp. 18681–18686.

[2] Ian D Kivlichan et al. "Bounding the costs of quantum simulation of many-body physics in real space". In: *Journal of Physics A: Mathematical and Theoretical* 50.30 (2017), p. 305301.

[3] Ryan Babbush et al. "Exponentially more precise quantum simulation of fermions in the configuration interaction representation". In: *Quantum Science and Technology* 3.1 (2017), p. 015006.

[4] Guang Hao Low and Isaac L Chuang. "Optimal Hamiltonian simulation by quantum signal processing". In: *Physical review letters* 118.1 (2017), p. 010501.

[5] Guang Hao Low and Isaac L Chuang. "Hamiltonian simulation by qubitization". In: *Quantum* 3 (2019), p. 163.

[6] Ryan Babbush et al. "Quantum simulation of chemistry with sublinear scaling in basis size". In: *npj Quantum Information* 5.1 (2019), p. 92.

[7] Xiao-Ming Zhang, Tongyang Li, and Xiao Yuan. "Quantum state preparation with optimal circuit depth: Implementations and applications". In: *Physical Review Letters* 129.23 (2022), p. 230504.

[8] Pei Yuan and Shengyu Zhang. "Optimal (controlled) quantum state preparation and improved unitary synthesis by quantum circuits with any number of ancillary qubits". In: *Quantum* 7 (2023), p. 956.

[9] Xiaoming Sun et al. "Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 42.10 (2023), pp. 3301–3314.

[10] Julien Zylberman and Fabrice Debbasch. "Efficient quantum state preparation with walsh series". In: *Physical Review A* 109.4 (2024), p. 042401.

[11] Julien Zylberman et al. "Efficient quantum circuits for non-unitary and unitary diagonal operators with space-time-accuracy trade-offs". In: *ACM Transactions on Quantum Computing* (2024).

[12] Mudassir Moosa et al. "Linear-depth quantum circuits for loading fourier approximations of arbitrary functions". In: *Quantum Science and Technology* 9.1 (2023), p. 015002.

[13] Matthias Rosenkranz et al. “Quantum state preparation for multivariate functions”. In: *Quantum* 9 (2025), p. 1703.
[14] Dong An, Di Fang, and Lin Lin. “Time-dependent unbounded Hamiltonian simulation with vector norm scaling”. In: *Quantum* 5 (May 2021), p. 459. ISSN: 2521-327X. URL: http://dx.doi.org/10.22331/q-2021-05-26-459.
[15] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.
[16] Jonathan Welch et al. “Efficient quantum circuits for diagonal unitaries without ancillas”. In: *New Journal of Physics* 16.3 (2014), p. 033040.
[17] Gilles Brassard et al. “Quantum amplitude amplification and estimation”. In: *arXiv preprint quant-ph/0005055* (2000).
[18] Dmitry Grinko et al. “Iterative quantum amplitude estimation”. In: *npj Quantum Information* 7.1 (2021), p. 52.
[19] Patrick Rall and Bryce Fuller. “Amplitude estimation from quantum signal processing”. In: *Quantum* 7 (2023), p. 937.
[20] Tudor Giurgica-Tiron et al. “Low depth algorithms for quantum amplitude estimation”. In: *Quantum* 6 (2022), p. 745.
[21] Zhong-Xia Shang and Qi Zhao. “Estimating quantum amplitudes can be exponentially improved”. In: *arXiv preprint arXiv:2408.13721* (2024).
[22] Julien Zylberman et al. “Trotter-based quantum algorithm for solving transport equations with exponentially fewer time-steps”. In: *arXiv preprint arXiv:2508.15691* (2025).

Fully Quantum Algorithm for the 1-dimensional linear Lattice Boltzmann Method

Mohammed Bediche [1,2,3] (✉)[0009-0001-0255-3894], Matthijs van Waveren [1][0000-0001-7280-9114], Denis Ricot[2], and Pierre Sagaut [3][0000-0002-3785-120X]

[1] CS Research Lab, 3, boulevard Thomas Gobert, 91120 Palaiseau, France
mohammed.bediche@cs-soprasteria.com,
matthijs.vanwaveren@compilaflows.com
[2] CS GROUP, 22, avenue Galilée, 92350 Le Plessis-Robinson, France
denis.ricot@cs-soprasteria.com
[3] Aix Marseille Univ, Centrale Med, CNRS, M2P2 Laboratory, 13013 Marseille, France
pierre.sagaut@univ-amu.fr

Abstract. A fully quantum algorithm for solving the one-dimensional linear advection-diffusion equation using the Lattice Boltzmann method as a numerical procedure is presented in this work. We start by presenting a state of the art of the current usage of quantum algorithms for solving ordinary and partial differential equations. We then describe two algorithms for the one-dimensional Lattice Boltzmann method with two degrees of freedom. The first one is an existing hybrid quantum-classical algorithm with measurements at each time step, and the second one is our improved version, viz. a fully quantum algorithm where only one measurement is needed at the end of the algorithm. The fully quantum algorithm is first executed on a quantum simulator and then compared with a classical approach. Subsequently, the fully quantum algorithm is run on a quantum system with 133 qubits to investigate the effect of noise and the depth of the circuit on the output state. We find fluctuations in the final result due to the decoherence noise of the qubits.

Keywords: Quantum computing, Quantum algorithms, Fluid dynamics, Lattice Boltzmann Method, Linear partial differential equation

1 Introduction

Differential equations constitute the foundational framework for the mathematical modeling of physical systems. However, executing numerical simulations of nonlinear ordinary and partial differential equations (ODEs/PDEs) on classical computers remains a formidable challenge, primarily due to the extensive computational time and data processing resources required. In response to these limitations, quantum programming has emerged as a promising alternative, garnering increasing attention over recent decades. Certain computational problems that are intractable for classical machines can be addressed theoretically using quantum computers, for example, Shor's factoring algorithm, which underpins the concept of quantum supremacy. A key advantage of

quantum computing lies in its potential to exponentially expand the dimensionality of the state vector—hence the number of grid points in a simulation—as the number of qubits increases, thereby enabling the manipulation of high-dimensional state spaces. Quantum algorithms, owing to their inherently linear structure, are particularly well-suited for solving linear differential equations. Several such algorithms offer a speed-up over their classical counterparts [1,2]. Nevertheless, progress in developing quantum algorithms for nonlinear differential equations has been comparatively limited, and these approaches have yet to outperform established classical techniques. Broadly speaking, the numerical solution of PDEs can be approached via two principal methods. The first involves discretizing the system using integration techniques such as finite difference or finite element methods, implemented with either a HHL-based quantum algorithm [1,3] or a VQA algorithm [4]. The second strategy consists in transforming the problem into a Schrödinger-type equation, which can then be addressed using Hamiltonian simulation techniques [5].

In the field of fluid mechanics and the simulation of turbulent flows, the partial differential equations (PDEs) that are generally considered are the Navier–Stokes equations. For industrial flow configurations, simplified turbulence models must be employed within the system of equations, which leads to non-negligible discrepancies with respect to the actual physics. The potential capability of quantum simulation to address very large-scale models while maintaining the shortest possible computation times is of great interest for a wide range of physical domains, including electromagnetic computations, structural analysis, acoustics, materials, combustion, and many others. For fluid mechanics simulations, an alternative to solving the Navier–Stokes equations is the use of the Lattice Boltzmann Method (LBM) [6]. This numerical method enables the simulation of various convection–diffusion phenomena, both linear and nonlinear. Several academic and commercial software packages based on LBM are available, including ProLB, developed by CS GROUP within the framework of an industrial–academic partnership.

In this work, we present a quantum algorithm for the 1-dimensional linear advection-diffusion equation solved with the Lattice Boltzmann Method. We start by following the approach of L. Budinski [7], who wrote a hybrid quantum-classical algorithm that resembles the classical algorithm of the Lattice-Boltzmann Method. One of the drawbacks of this approach is the need of repeated measurements and reinitialization at each time step. This increases the complexity of the algorithm and hinders any potential for quantum advantage in the long run. We improve on the hybrid classical-quantum algorithm by proposing a fully quantum algorithm for the one-dimensional Lattice Boltzmann formulation D1Q2. We simulate the fully quantum algorithm on the Qiskit quantum simulator, and we execute it on the 133-qubit IBM quantum computer `ibm_torino`. The results from the computers have the same evolution of the flow as the classical results and the results from the quantum simulator. The results from the quantum computer however exhibit fluctuations due to the decoherence noise in the qubits.

2 Lattice Boltzmann Method

The advection-diffusion equation solved with the Lattice Boltzmann Method is the subject of this paper.

$$\frac{\partial \phi}{\partial t} + \frac{\partial (u_i \phi)}{\partial x_i} = \frac{\partial}{\partial x_i}\left(D \frac{\partial \phi}{\partial x_i}\right) \tag{1}$$

where ϕ is the depended variable (mass, momentum, energy, species, etc.), t is time, x_i is a Cartesian coordinate, u_i is the fluid velocity in the i-direction and D is the diffusion coefficient. The above equation is valid for general advection and diffusion phenomena, including both steady and unsteady situations.

In the Lattice Boltzmann Method, instead of solving the nonlinear equation, we solve an equivalent problem with a higher dimension. However, this formulation still contains a non-linear collision term that is not easily implemented on a quantum computer. Previous works have either neglected the non-linear part of this operator [7,8] or proceeded by linearizing it using Carleman linearization [9]. The non-linear part is zero if the advection velocity is independent of the concentration and can be neglected if the mass transport is dominated by diffusion.

In the Lattice Boltzmann method, a particle density function $f(x, c, t)$ is used to describe the probability of finding one distribution of particles within the elemental cube (x,x+dx) with a velocity comprised within the interval (c,c+dc) at the time t. The discretized Boltzmann equation under the Bhatnagar-Gross-Krook (BGK) approximation in the phase space [10] is given by:

$$\frac{\partial f_\alpha(\vec{x}, t)}{\partial t} + c_{\alpha,i} \frac{\partial f_\alpha(\vec{x}, t)}{\partial x_i} = -\frac{f_\alpha(\vec{x}, t) - {f^{eq}}_\alpha(\vec{x}, t)}{\tau} \tag{2}$$

The index i refers to the discretization in space and the index α defines the discretization in the velocity components. These distribution functions are linked to the macroscopic variables by:

$$\rho(\vec{x}, t) = \sum_\alpha f_\alpha(\vec{x}, t) \; ; \; \rho.\overrightarrow{u}(\vec{x}, t) = \sum_\alpha \overrightarrow{c_\alpha}.f_\alpha(\vec{x}, t) \tag{3}$$

The algorithm of the Lattice Boltzmann Method is a four-step process: i) initialization, ii) collision, iii) propagation and iv) macro-states calculation [11]. We develop these steps in the next section.

2.1 Fully Quantum Algorithm for the Linear D1Q2

We describe a fully quantum algorithm that uses the Lattice Boltzmann Method to solve the linear advection-diffusion equation in a one-dimensional space. This algorithm builds upon the work of Budinski [7] where he presented a hybrid classical-quantum algorithm for the linear D1Q2 model with a relaxation time equals to 1. In the D1Q2 model, the index α in (2) and (3) has values 1 and 2 with only two velocity

components, $e_1 = -e_2 = \frac{\Delta x}{\Delta t}$. Additionally, we consider $\Delta x = \Delta t$. The Boltzmann equation becomes:

$$f_\alpha(x + e_\alpha \Delta t, t + \Delta t) = f^{eq}{}_\alpha \tag{4}$$

The linear equilibrium distribution functions are given by (5) with $\omega_\alpha = 0.5$:

$$f_\alpha{}^{eq}(\vec{x}, t) = \omega_\alpha \rho(\vec{x}, t)\left(1 + e_{\alpha,i} u_i\right) + o(u_i) = A_\alpha \, \rho(\vec{x}, t) \tag{5}$$

A_α are square diagonal matrices of size $N \times N$, with N the number of lattice points. We consider the lattice points $x_0, x_1, \ldots, x_{N-1}$ each describing a position of the distribution function on this lattice. To encode the model into our quantum algorithm, $log_2(N)$ qubits are needed with one additional qubit used to encode information about the two degrees of freedom. One ancilla qubit is added to aid calculations.

We encode the density at t=0 in the normalized state vector of eq (6). We applied the reverse iterative procedure proposed by Shende et al. [15], which is implemented in the built-in function `initialize` in Qiskit.

$$|\psi_0\rangle = |0\rangle_a \sum_{i=0}^{2N-1} \frac{\rho_i}{\|\psi\|} |i\rangle \tag{6}$$

The following circuit in Fig.1 shows the original hybrid quantum circuit published by Budinski [7]. It represents the 4 steps needed in the quantum Lattice Boltzmann algorithm. The dashed vertical lines separate the four sections of the algorithm. We first describe this circuit before describing our improvement to convert it to a fully quantum algorithm.

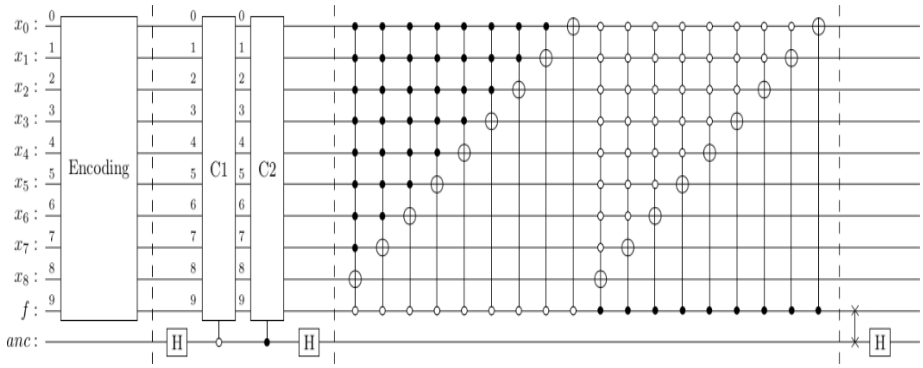


Fig. 1. The quantum circuit for one timestep. The leftmost section is the encoding step. The section with C1 and C2 is the collision step. The section to the right of it is the propagation step. The final rightmost section is the macro-states calculation step.

Collision. For the collision step, the Linear Combination of Unitaries approach [11] is used by applying the following set of operators:

$$|\psi_1\rangle = \left(H_a{}^{+}\otimes I_q\right)(|0\rangle\langle 0|_a C_1 + |1\rangle\langle 1|_a C_2)\left(H_a\otimes I_q\right)|\psi_0\rangle \tag{7}$$

where C_1 and C_2 are unitary diagonal operators constructed from the matrices A_1 and A_2 and are described in [7]. The state vector post-collision is:

$$|\psi_1\rangle = \frac{1}{\|\psi\|}\left(|0\rangle_a \sum_{i=0}^{2N-1} f_{\alpha,i}|i\rangle + i|1\rangle_a \sqrt{I - A^2} \sum_{i=0}^{2N-1} \psi_{\alpha,i}|i\rangle\right) \tag{8}$$

Propagation. The propagation step is carried out using the quantum walk procedure, taking advantage of the fact that the quantum state is already in superposition. This is done using a series of multiple-controlled NOT, a C-NOT gate and a NOT gate. The difference between the two directions of propagation is the state of the control qubit. These operators can be seen in Fig 1. The state vector $|\psi_2\rangle$ after propagation is:

$$|\psi_2\rangle = \hat{L}\hat{R}|\psi_1\rangle \tag{9}$$

Macro-states calculations. To calculate the particle density and the velocity, we first followed Budinski's algorithm and applied a SWAP gate between the ancilla qubit and the N-1'th qubit followed by a Hadamard. The final state is:

$$|\psi_3\rangle = \frac{1}{\sqrt{(2)}\|\psi\|}\left(|0\rangle_a|0\rangle_{N-1}\sum_{i=0}^{N-1}(f_1+f_2)|i\rangle + |1\rangle_a|0\rangle_{N-1}\sum_{i=0}^{N-1}(f_1-f_2)|i\rangle + |0\rangle_a|1\rangle_{N-1}(C1-C2)\sum_{i=0}^{N-1}\psi_{\alpha,i}|i\rangle\right) \tag{10}$$

In the hybrid scheme, the resulting state vector is measured, and classical post-processing is performed to extract the density and the velocity. These are then input in the calculation of $|\psi_0\rangle$ of the next timestep. Each timestep requires thus executing a quantum circuit, measuring the resulting state vector, and performing classical post-processing to extract the density and the velocity. This procedure incurs significant overhead. For bigger systems where the Hilbert space has a large dimension, the algorithm behaves worse than a classical algorithm. A fully quantum algorithm is needed to exploit the full potential of quantum machines.

Fully quantum algorithm. To build a quantum algorithm that avoids repeated measurements, we start by the quantum state $|\psi_3\rangle$ calculated previously. The state vector shows that the information about $f_1 + f_2$ and $f_1 - f_2$ is already stored in the final state vector just before the measurements. We now develop the expression of the equilibrium distribution function and show that it contains these two quantities. It is easy to show that:

$$\begin{aligned}\rho(\vec{x}, t+\Delta t) &= f_1(\vec{x}, t+\Delta t) + f_2(\vec{x}, t+\Delta t)\,; \\ \rho.\vec{u}(\vec{x}, t+\Delta t) &= f_1(\vec{x}, t+\Delta t) \;-\; f_2(\vec{x}, t+\Delta t)\end{aligned} \tag{11}$$

From (5) and (11), we derive the following expressions:

$$\begin{aligned}{f^{eq}}_1(x,t) &= \frac{1}{2}\big(\rho(\vec{x},t) + \rho.\vec{u}(\vec{x},t)\big); \\ {f^{eq}}_2(x,t) &= \frac{1}{2}\big(\rho(\vec{x},t) - \rho.\vec{u}(\vec{x},t)\big)\end{aligned} \tag{12}$$

This means that it is as if we had the result of the collision at the time step *t+1* encoded in the state obtained after the calculations at the instant *t*. We just need to rearrange the states again to separate these two values. To do that, we modify the circuit of Fig. 1 by adding a SWAP gate between the sixth qubit in the *q*-register and the ancilla qubits and apply a Hadamard gate on qubit *N-1*. We obtain the state in equation (13). $|\lambda\rangle_q$ is an additional state that holds no useful information to us.

$$|\psi_3\rangle = \frac{1}{\||\psi\|\|}\Bigg(|0\rangle_a|0\rangle_{N-1}\sum_{i=0}^{N-1}{f^{eq}}_{1,i}|i\rangle + |0\rangle_a|1\rangle_{N-1}\sum_{i=0}^{N-1}{f^{eq}}_{2,i}|i\rangle + |1\rangle_a|\lambda\rangle_q\Bigg) \tag{13}$$

The state $|1\rangle_a|\lambda\rangle_q$ is for the other amplitudes obtained that has no exploitable information in this algorithm which we didn't write for simplicity. We see that we do not need to switch to a classical algorithm to prepare the system for future time steps, and that we can stay in the quantum space. Consequently, we can directly follow up by the propagation and the macro-states calculations, which drastically optimizes the time needed for the algorithm. The overall architecture of the algorithm is given in Fig 2 where $|\psi_4\rangle$ is the state vector obtained after the quantum post-processing steps which allows us to bypass the intermediate measurement steps and classical processing in the hybrid algorithm to execute the next run.

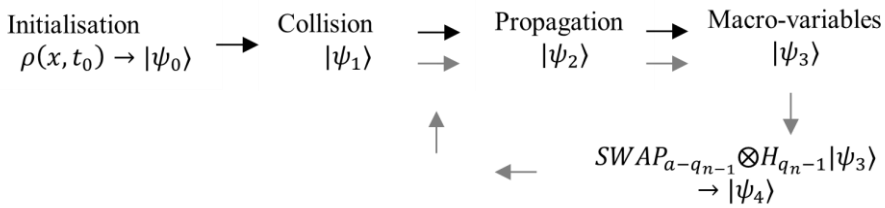


Fig. 2. The architecture of the quantum algorithm. The black arrows represent the first run of the algorithm and the grey ones the additional post treatment and the processes repeated in the *m* time loops afterwards for a total time of $t=m\Delta t$.

Our method eliminates the need for intermediate classical feedback. Instead, the required updates are achieved entirely within the quantum circuit by appending only two additional quantum gates at the end of each run. This results in a more efficient computation.

Complexity To calculate the complexity of the quantum algorithm in Fig. 2, we first divide it into two parts: (a) a first forward part with the four major steps: encoding, collision, propagation and macro-states calculation and (b) a second loop part containing the quantum post-processing, the propagation and the calculation of the macroscopic variables. This second part is the one which is looped over once for each timestep.

In the first forward part, encoding the input variables into the state vector requires $2 \times 4^n - (2n+3) \times 2^n + 2n$ CNOT gates with n being the number of qubits according to Table 1 in [15]. For the collision step, we require $O(\log_2 m)$ ancilla qubits for m directions of propagation. The LCU scales as $O(1)$ in n, since the number of unitaries does not grow with the system size. The multi-controlled NOT gates used in the propagation step scale as $O(n^2)$ in the number of CNOT gates [16]. The macro-states calculations require 3 CNOT gates and one Hadamard gate so its complexity is negligible.

For the second loop part of our algorithm, we require $O(t(n)^2)$ CNOT gates with t being the number of timesteps. We see that the depth of the second part of the circuit in terms of CNOTs is linear in t and quadratic in n. Note that the number of qubits $n = \log_2 N + \log_2 m$, with N the number of grid points and m the number of directions of propagation.

3 Results

3.1 Execution on a Quantum Simulator

We ran the simulations using a centered Gaussian flow as input with periodic boundary conditions. The domain is discretized using 512 regular mesh points with 1000-time steps. The encoding of the initial values of the concentration was done using the instruction `initialize` available on Qiskit, which encodes an arbitrary quantum gate into the register of n qubits. The results of the fully quantum algorithm matched the ones obtained using the classical one, as shown in Fig 3. Two time-instances, t=200 and t=600, were considered to show the evolution of the flow.

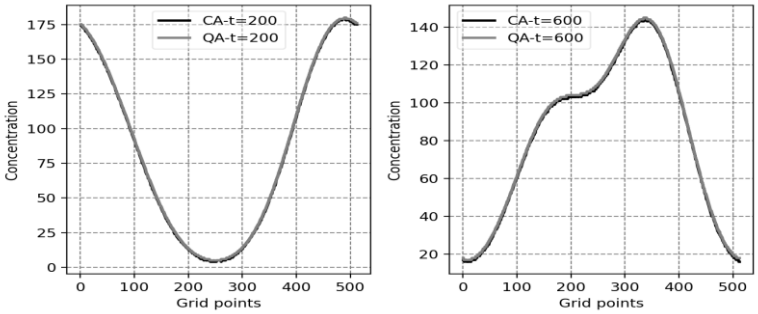


Fig. 3. Results of the fully quantum algorithm on a quantum simulator. CA corresponds to the Classical Algorithm, and QA to the Quantum Algorithm.

The original hybrid version of the algorithm was also compared with the classical one giving matching results. However, since the hybrid algorithm requires several measurement operations for each time step, it requires a longer time to execute.

3.2 Performance of the Fully Quantum Algorithm on a Quantum Computer

The fully quantum algorithm was then tested on the 133-qubit quantum machine `ibm_torino`. For every gate present in the quantum algorithm, a decomposition step is necessary to implement the gate on the quantum machine. For `ibm_torino`, there are 7 elementary gates which are the CZ, Id, RX, RZ, RZZ, SX and X. Every customized gate in the algorithm needs to be decomposed into these elementary gates before the algorithm is executed on the machine.

The quantum computer returns a histogram, which is the probability distribution obtained following the measurement operations at the end of the algorithm. Hence, we cannot directly access the quantum state. However, since the useful information of the quantum state only has a real part, it is possible to reconstruct the profile of the density as a function of the mesh points from the probability distribution. Due to the presence of noise on the quantum hardware, it was deemed necessary to reduce the size of the systems to 8 mesh points in order to decrease the depth of the algorithm. Multiple timesteps were also computed but as expected the output became noisier the longer the simulation.

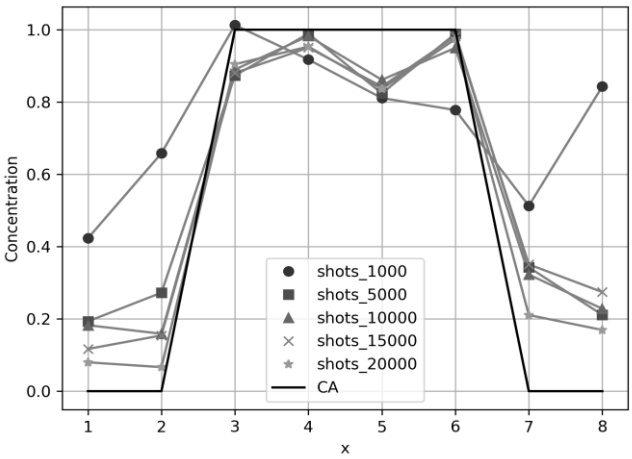


Fig. 4. Concentration results obtained from `ibm_torino`. The horizontal axis has the meshpoints.

Fig.4 represents the result obtained from `ibm_torino`, and it shows that we were successful in obtaining the same evolution of the flow as the results on the classical algorithm. We also note that increasing the number of shots from 1000 to 20000 led to

better results. The values of the concentration from the run on the quantum machine fluctuate and differ from the values of the classical ones, but they have the same qualitative evolution of the flow. Our hypothesis is that this difference is due to the decoherence noise in the qubits, because larger systems and longer simulations lead to more noise. The noise in the transmon qubits arises mainly from the effects of dielectric loss and its manifestation through the interaction with two-level systems (TLS) [13]. To overcome the problem of noise in the qubits, quantum error mitigation and quantum error correction techniques are utilized. The basic idea of quantum error correction is the introduction of redundancy in the qubits, which means that one logical error-free qubit is represented by dozens of physical noisy qubits. Brayyi et al [14] for example present an end-to-end quantum error correction protocol that implements fault-tolerant memory on the basis of a family of low-density parity-check codes. We expect that once the error mitigation and correction techniques have been implemented in production machines, we will be able to simulate larger systems.

4 Discussion

The goal of this work was to show that we can build a fully quantum algorithm for the one-dimensional linear advection-diffusion equation using the lattice Boltzmann method as a numerical procedure. The focus was more on avoiding measurements as much as possible and not as much on optimizing the steps and the gates in the algorithm. We were inspired by the hybrid quantum-classical algorithm of L. Budinski [7]. This hybrid quantum-classical algorithm is limited in its scope for quantum speedup due to the classical measurement bottleneck. We developed a fully quantum version of the algorithm by replacing the measurement step with a state vector rearrangement step. We tested this algorithm both on the Qiskit quantum simulator and on the 133-qubit IBM quantum computer `ibm_torino`. The results from the quantum computer have the same evolution of the flow as the results from the quantum simulator. The results from the quantum computer exhibit fluctuations due to the noise in the qubits. Our future work is to reduce these fluctuations by decreasing the depth of the circuits by optimizing the encoding, collision and propagation steps. We are also working on extending this algorithm to 2D systems and adding non-linearity to the model.

Appendix

The quantum algorithm was implemented using a Jupyter notebook. The version of Python that was used is 3.12.7. The API Qiskit version 1.2.4 was used with:

- Qiskit-aer: version 0.15.1
- Qiskit-ibm-runtime: version 0.30.0

The initial density is a centered Gaussian distribution, and the initial uniform velocity field is 0.2. The code is not open source, thus we cannot publish it on GitHub.

References

1. Montanaro, A., Pallister, S. Quantum Algorithms and the Finite Element Method, Physical Review A 93(2), 032324 (2016). https://doi.org/10.1103/PhysRevA.93.032324
2. Childs, A.M., Liu, J.-L., Ostrander, A. High-Precision Quantum Algorithms for Partial Differential Equations. Quantum 5, 574 (2021). https://doi.org/10.22331/q-2021-11-10-574
3. Liu, J.-P., Kolden, H.Ø., Krovi, H.K., Loureiro N.F., Trivisa, K., Childs, A.M. Efficient Quantum Algorithm for Dissipative Nonlinear Differential Equations. In: Proceedings of the National Academy of Sciences, 118 (35), 1-6 (2021). https://doi.org/10.1073/pnas.2026805118.
4. Lubasch, M., Joo, J., Moinier, P., Kiffner, M., Jaksch, D. Variational quantum algorithms for nonlinear problems. Physical Review A. 10.1103/physreva.101.010301. https://doi.org/10.1103/PhysRevA.101.010301
5. Jin S., Liu N., Yu U. Quantum simulation of partial differential equations via Schrödingerization. Phys. Rev. Lett 133, 230602 (2024). https://doi.org/10.1103/physrevlett.133.230602
6. Chen, S. and Doolen, G.D., Lattice Boltzmann Method for fluid flows, Annual Review of Fluid Mechanics, 30 (1998). https://doi.org/10.1146/annurev.fluid.30.1.329
7. Budinski, L., Quantum Algorithm for the Advection–Diffusion Equation Simulated with the Lattice Boltzmann method. Quantum Information Processing 20(2), 1-17 (2021). https://doi.org/10.1007/s11128-021-02996-3
8. Kocherla S. et al. Fully Quantum Algorithm for Lattice Boltzmann Methods with Application to Partial Differential Equations. AVS Quantum Science, 6(3), 033806 (2024). https://doi.org/10.1116/5.0217675
9. Itani W., Succi S. Analysis of Carleman Linearization of Lattice Boltzmann. Fluids, 7(1), 24 (2022). https://10.3390/fluids7010024
10. Bhatnagar, P.L., Gross, E.P., Krook, M.A. Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems. Physical Review 94(3), 511-525 (1954). https://10.1103/PhysRev.94.511
11. Succi, S.: Lattice Boltzmann Method for Fluid Dynamics and Beyond. Oxford University Press, Oxford (2001).
12. Low, G.H., Chuang L.L. Hamiltonian Simulation by Qubitization. Quantum 3, 163 (2019). https://doi.org/10.22331/q-2019-07-12-163.
13. Murray, C.E., Material matters in superconducting qubits, Material Science and Engineering: R: Reports, 146 (2021) 100646 ISSN 0977-796X https://doi.org/10.1016/j.mser.2021.100646.
14. Brayyi S et al, High-threshold and low-overhead fault-tolerant quantum memory, Nature 627 (2024) 778-782. https://doi.org/10.1038/s41586-024-07107-7
15. Shende V.V., Bullock S.S. and Markov I.L., Synthesis of quantum-logic circuits, IEEE Trans Comput Aided Design, 25(6), pp. 1000-1010 (2006) https://doi.org/10.1109/TCAD.2005.855930
16. Barenco A. et al., Elementary gates for quantum computation, Phys Rev A **52**, 3457 (1995). https://doi.org/10.1103/PhysRevA.52.3457

Acknowledgements. This study was partly funded by the French state through CIFRE.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Tucker iterative quantum state preparation

Carsten Blank[1 [0000−0003−3450−0823]] and Israel F. Araujo[1 [0000−0002−0308−8701]]

data cybernetics ssc GmbH, Landsberg am Lech 86899, Germany
blank@data-cybernetics.com
https://data-cybernetics.com

Abstract. Quantum state preparation is a fundamental component of quantum algorithms, particularly in quantum machine learning and data processing, where classical data must be encoded efficiently into quantum states. Existing amplitude encoding techniques often rely on recursive bipartitions or tensor decompositions, which either lead to deep circuits or lack practical guidance for circuit construction. In this work, we introduce Tucker Iterative Quantum State Preparation (Q-Tucker), a novel method that adaptively constructs shallow, deterministic quantum circuits by exploiting the global entanglement structure of target states. Building upon the Tucker decomposition, our method factors the target quantum state into a core tensor and mode-specific operators, enabling direct decompositions across multiple subsystems.

Keywords: Quantum state preparation · Entanglement · Circuit decomposition · Tensor decomposition · Tucker decomposition.

1 Introduction

Quantum computing has emerged as a rapidly advancing field with significant recent progress in both theoretical and experimental domains. Current quantum hardware resides in the so-called Noisy Intermediate-Scale Quantum (NISQ) era, characterized by devices that operate with a limited number of qubits and are susceptible to noise and decoherence. These limitations pose substantial challenges in the engineering, control, and deployment of quantum systems. In particular, NISQ devices lack full fault tolerance and require sophisticated quantum error mitigation or correction strategies. Additional complications arise in the design and implementation of quantum circuits, especially concerning the reliable execution of quantum gates. A critical bottleneck in practical quantum computation is the task of encoding classical data into quantum states – a process referred to as *quantum state preparation*, or more generally, data encoding and loading.

Quantum state preparation is the process of encoding classical data into quantum states, which serve as inputs for quantum algorithms in applications such as quantum machine learning and data processing tasks [5,16,21]. Multiple approaches have been developed to address the quantum state preparation problem. Among the most widely used are *Amplitude Encoding* [1–4], *Qubit Encoding* (also referred to as *Angle Encoding*) [9], and *Hamiltonian Encoding* [6,21].

In this work, we introduce a novel amplitude encoding method named *Tucker Iterative Quantum State Preparation* (Q-Tucker). This technique aims to provide a deterministic approach to state preparation by generating quantum circuits whose depth is adaptively determined by the entanglement structure of the target quantum state [1].

Entanglement has long been recognized as a fundamental resource in quantum computing, enabling capabilities that are unattainable in classical computation. Nevertheless, quantum features such as entanglement and superposition can, to a certain extent, be simulated on classical hardware. In practice, the preparation of an n-qubit quantum state – represented by a complex vector of $N = 2^n$ amplitudes – for execution on current NISQ-era quantum processing units (QPUs) often depends on classical pre-processing.

Delegating these preparatory computations to classical resources becomes increasingly challenging as the dimensionality of the quantum state grows, a scenario commonly encountered in quantum machine learning and quantum data processing tasks. A well-established strategy for mitigating this complexity involves decomposing the target state into approximately low-entangled subsystems. Even approximate decompositions can offer substantial advantages, such as reduced quantum circuit depth and, in many cases, improved fidelity when compared to exact state initialization. As a result, efficiently characterizing and exploiting the entanglement structure of quantum states is a critical objective in the design of scalable quantum algorithms and state preparation protocols.

1.1 Related work

Several previous studies have explored low-rank quantum state preparation by leveraging hierarchical decompositions of multipartite quantum systems. Araujo et al. introduced an iterative scheme based on successive applications of the Schmidt decomposition to identify and disentangle bipartitions within a quantum system [1]. In their method, at each iteration, a suitable bipartition is selected, and the entanglement is reduced until no further disentanglement is possible, yielding a compact, low-rank representation of the original quantum state.

Regarding quantum state encoding, a variety of approaches have been proposed to map the N complex-valued amplitudes of an arbitrary quantum state onto a physical quantum circuit. One of the earliest deterministic techniques was presented by Grover et al., who proposed a method to construct quantum superpositions corresponding to efficiently integrable probability distributions [10]. While conceptually straightforward, this method requires $\mathcal{O}(N)$ quantum operations, rendering it impractical for large-scale quantum systems.

An alternative family of methods exploits tensor network structures, particularly the matrix product state (MPS) representation, to facilitate efficient state encoding. Schön et al. proposed a sequential scheme to construct MPS representations via one- and two-qubit quantum gates, enabling scalable circuit generation for a broad class of quantum states [20]. Owing to their strong approximation capabilities, MPS methods have become widely adopted in quantum many-body

physics. Building upon this foundation, subsequent works have integrated parameterized tensor-network models to reduce quantum circuit depth and to support hybrid quantum-classical training schemes [19]. In some instances, iterative application of so-called *disentangler* operators to MPS representations has been proposed to further improve state compression and circuit efficiency.

A conceptually distinct approach based on the Tucker decomposition has been proposed by Protasov et al. [22]. In contrast to methods relying on the Schmidt decomposition or singular value decomposition (SVD), the Tucker decomposition generalizes these techniques by enabling one-versus-all decompositions across multiple modes. While the authors advocate for the use of the Tucker decomposition in quantum state preparation [22], their approach does not specify a selection criterion for determining the appropriate decomposition among the many non-unique possibilities. Finally, the work provides no concrete procedure for mapping the decomposed representation into a quantum circuit, leaving a critical gap between theoretical formulation and practical implementation.

1.2 Objectives and organization

The primary objective of this work is to address the limitations of existing state preparation techniques by introducing a method based on tensor decomposition that is naturally suited for multipartite quantum systems. Unlike prior approaches that rely on recursive bipartitions – often resulting in complex hierarchical structures – our method circumvents these constraints by directly exploiting the global structure of the quantum state.

The remainder of the paper is organized as follows. Section 2 introduces the Tucker Iterative Process, detailing its decomposition strategy, fidelity control mechanisms, and the circuit synthesis procedure. Section 3 presents an analysis of the method's convergence. Section 4 reports the experimental results. Finally, Section 5 concludes the paper with a summary of the findings and future research directions.

2 Tucker iterative quantum state preparation

This section introduces a hardware-efficient method for encoding classical data into a quantum device using an iterative tensor decomposition strategy. Using a multipartite-specific decomposition based on the Tucker tensor representation, the method addresses the limitations of previous methods [1] and eliminates the need for complex hierarchical bipartitions. The quantum circuit is constructed from the Tucker representation, with a performance metric guiding the iterative refinement process (see Fig. 1)

Any vector $v \in \mathbb{C}^N$ can be represented as a m-tensor $T \in \mathbb{C}^{d_1} \otimes \ldots \otimes \mathbb{C}^{d_m}$, for an arbitrary $m > 1$, with the constraint $\prod_i d_i = N$. Given such a tensor T, a representation or decomposition can be determined. One such representation is the Tucker decomposition:

$$T = G \times_1 W^{(1)} \times_2 W^{(2)} \times_3 \ldots \times_m W^{(m)},$$

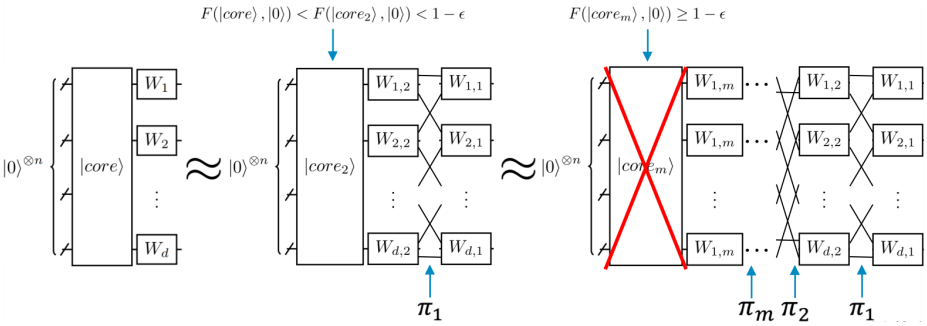


Fig. 1. Quantum circuit representation of the Tucker Iterative Process for circuit compression. The initial quantum state is decomposed into a core vector $|\text{core}\rangle$ and local hardware-efficient unitary (or isometry) factors W_i. At each iteration, the fidelity $F(|\text{core}_j\rangle, |0\rangle)$ is evaluated to determine whether the core can be approximated by the ground state. If the fidelity exceeds the threshold ($F \geq 1-\epsilon$), the core is removed (indicated by the red cross). Between iterations, logical qubit swaps π_j are applied to optimize the placement of the factors. These swaps are guided by minimizing the bond dimension between qubit partition blocks – each corresponding to a factor W_i – which reflects the entanglement between subsystems.

where $\times_i$ denotes the mode-i product, $G \in \mathbb{C}^{p_1} \otimes \ldots \otimes \mathbb{C}^{p_m}$ is the core tensor (with $p_i \leq d_i$), and the factor matrices $W^{(i)} \in \mathbb{C}^{d_i \times p_i}$. This decomposition is naturally suited for quantum computing since the tensor decomposition of T directly corresponds to a quantum state decomposition:

$$|\psi\rangle = (W^{(1)} \otimes \ldots \otimes W^{(m)})|G\rangle,$$

where we denote the operators $W^{(i)}$ using the same symbol as the factor matrices, even if their dimensions do not match ($p_i \neq d_i$), in which case columns are padded with basis vectors. The core tensor G is interpreted as a quantum state $|G\rangle$ via vectorization.

A key feature is that the factor matrices can be interpreted as unitary operators acting on separate subspaces (hence, the Tucker format is also known as the *subspace format*). From a quantum computing perspective, these unitary operations can be executed in parallel – if hardware permits – making the depth of any resulting quantum circuit determined by the longest individual factor circuit $W^{(i)}$. The geometric entanglement of the core tensor is never greater than that of the original state and can, in many cases, be significantly lower. This property enables a recursive or iterative approach that terminates. If recursion halts after r steps, the resulting decomposition is:

$$|\psi\rangle = W_1 \ldots W_r|0\rangle.$$

Each iteration j of the process yields a factorization $W_j = W_{(j,1)} \otimes \ldots \otimes W_{(j,m)}$. If the resulting core tensor – or quantum state $|G_r\rangle$ – can be expressed as a product state (i.e., composed solely of operations that can be absorbed

into the adjacent factors), or if the fidelity between the core state and the true ground state exceeds a predefined threshold, then the core state is adopted as the ground state. In this case, it requires no additional quantum operations. Each factor $W_{(j,i)}$ is implemented as a unitary or isometry operator $W^{(i)} \in \mathbb{C}^{d_i \times p_i}$. If a full unitary operator is not required ($p_i < d_i$), then isometries can be used – an important detail that enables circuit simplification and motivates the method described below.

It is well known that the tensorization of a vector $v \in \mathbb{C}$ into a tensor $T \in \mathbb{C}^{d_1} \otimes \ldots \otimes \mathbb{C}^{d_m}$ is not unique. Prior work has shown that different sequences of logical swap operations can lead to different entanglement behaviors [1]. This result extends to tensorization itself, resulting in a search space defined by the number of modes m, the dimension tuple $\mathbf{d} = (d_{i_1}, \ldots, d_{i_m})$, and a permutation $\pi \in \mathrm{Sym}([m])$ that encodes the swap pattern. Each *tensorization configuration* is thus characterized by a tuple $(m, \mathbf{d}, \pi)$. Since the optimal configuration $(m, \mathbf{d}, \pi)$ is not unique, it is possible to seek those that balance approximation quality with implementation complexity.

In general, there exists a configuration $(m, \mathbf{d}, \pi)$ such that the entanglement in the core tensor is smaller than in the original tensor T. This allows the method to be applied recursively until a product state is reached, which can be trivially initialized. Alternatively, methods such as that of Rudolph & Grover [10], or their concrete implementations [4,14], can be employed.

The search can be simplified by taking into account the coupling map limitations of the target device. On currently available quantum hardware, efficient two-qubit operations are typically restricted to neighboring qubits. Under this constraint, we can fix $m = n/2$, leading to $d_i = 4$. Since the dimension of these subsystems is fixed, the cost of each SVD is constant. This eliminates the scalability bottleneck often associated with tensor decompositions such as Tucker, where SVD costs scale cubically with the input dimension.

Even with such simplifications, the search process still depends on logical qubit swaps, which are costly. Identifying the partition with the minimal bond dimension – that is, the one producing the core vector with the least entanglement – requires a combinatorial search. The complexity of this search is determined by the size of the partition blocks, $n_i = \log_2 d_i$, and grows according to the binomial coefficient $\binom{n}{n_i}$. In the worst-case scenario, where all blocks have equal size, the complexity becomes $\mathcal{O}(2^n)$. For large systems, the computational overhead of evaluating the bond dimension at each step becomes prohibitive, making entanglement quantification infeasible.

To address this, heuristic strategies and physically motivated constraints are essential to guide the search process. These include limiting the partition search space to locality-preserving cuts, prioritizing low-overlap subsystems, or leveraging prior knowledge about the structure of the target state (e.g., symmetries or sparsity). Such strategies drastically reduce the search space and enable approximate yet effective decomposition, balancing computational tractability with circuit efficiency.

3 Convergence guarantee of Q-Tucker

Here we formalize the correctness of the iterative procedure described in Section 2. Throughout, an *oracle* provides a partition $\mathcal{P} = \{B_1, \ldots, B_m\}$ (a qubit set per block) that maximizes "inner" entanglement within blocks (equivalently, minimizes inter-block entanglement or bond dimension). Given this $\mathcal{P}$, we run a Tucker step and then *fix the gauge* ("monotone gauge") by choosing the first column of each block operator to maximize the block-product overlap with the current state. Tucker/HOSVD and the non-uniqueness of factors up to rotations inside singular subspaces are classical facts [7, 12].

Notation and per-step objective. Let $|G_{j-1}\rangle$ be the core state before iteration j and $|0\rangle = \bigotimes_i |0\rangle_{B_i}$ the block-wise ground state. For a partition $\mathcal{P}$, define the multipartite product-overlap

$$\alpha_{\mathcal{P}}(|\phi\rangle) \;:=\; \max_{\|u_i\|=1} \big|\langle u_1 \otimes \cdots \otimes u_m \mid \phi\rangle\big|,$$

which is called the *entanglement eigenvalue*, and corresponds to the closest separable state [24]. The *monotone gauge* sets the first column of each block unitary so that $W^{(i)}|0\rangle_{B_i} = u_i^\star$ where $(u_i^\star)$ attains the maximum above.

Lemma 1 (monotone-gauge identity). *For the (oracle-provided) partition $\mathcal{P}_j$ used at iteration j,*

$$\langle 0 \mid G_j\rangle \;=\; \alpha_{\mathcal{P}_j}(|G_{j-1}\rangle), \qquad F_j \;=\; \alpha_{\mathcal{P}_j}(|G_{j-1}\rangle)^2.$$

Proof. By construction, $(\otimes_i W_j^{(i)\dagger})|G_{j-1}\rangle = |G_j\rangle$ and $W_j^{(i)}|0\rangle_{B_i} = u_i^\star$; hence $\langle 0 \mid G_j\rangle = \langle \otimes_i u_i^\star \mid G_{j-1}\rangle$, which is the maximizing value.

Corollary 1 (per-step monotonicity for any partition). *For any partition $\mathcal{P}$,*

$$\alpha_{\mathcal{P}}(|G_{j-1}\rangle) \;\geq\; \big|\langle 0 \mid G_{j-1}\rangle\big|,$$

since $|0\rangle$ is a feasible block-product vector. Therefore, with the monotone gauge on $\mathcal{P}_j$,

$$F_j \;=\; \alpha_{\mathcal{P}_j}(|G_{j-1}\rangle)^2 \;\geq\; \big|\langle 0 \mid G_{j-1}\rangle\big|^2 \;=\; F_{j-1}.$$

This mirrors the monotonicity of ALS/HOOI-style block updates for Tucker models [8, 12, 23].

Role of the oracle. The oracle's partition, which maximizes inner entanglement, is used to *accelerate* progress: it typically enlarges the cut-based ceiling for block-product overlaps (see below). However, monotonicity $F_j \geq F_{j-1}$ holds *for any partition* when the monotone gauge is applied.

Cut ceiling and strict improvement. For a partition $\mathcal{P}$, let $\beta(\mathcal{P}) := \min_C \lambda_{\max}(C)$ be the minimum, over block cuts C, of the largest Schmidt coefficient of $|G_{j-1}\rangle$ across C. For any bipartition, the largest Schmidt coefficient equals the maximal overlap with a product state across that cut [15,17]; therefore any block-product is bounded by

$$F_j \;=\; \alpha_{\mathcal{P}_j}(|G_{j-1}\rangle)^2 \;\le\; \beta(\mathcal{P}_j).$$

Partitions maximizing inner entanglement tend to increase $\beta(\mathcal{P}_j)$, raising a tight upper bound on F_j. A *strict* increase $F_j > F_{j-1}$ is guaranteed whenever the "worst" cut for $\mathcal{P}_j$ is unique and its top Schmidt vectors factor across blocks, in which case $\alpha^2_{\mathcal{P}_j} = \beta(\mathcal{P}_j)$ and the larger ceiling is attained.

Stalls and block-size expansion. Fix a maximum block size k (so each block unitary acts on $\le k$ qubits). If $F_j = F_{j-1}$ persists, we say the process *stalls at level* k. This occurs because, with fixed k, each step (via the monotone gauge) already attains the *best possible* block-product overlap within the current model class, i.e., $F_j = \alpha_{\mathcal{P}_j}(|G_{j-1}\rangle)^2$. When the state's overlap is capped by the *cut ceiling* for the chosen partition ($\alpha_{\mathcal{P}_j} = \beta(\mathcal{P}_j)$), or when all admissible partitions yield the same ceiling (or the worst-cut top Schmidt vectors fail to factor across blocks), no strict improvement is possible. Increasing k strictly enlarges the feasible set and (weakly) raises the ceilings β, so the limiting fidelity cannot decrease; allowing k to reach n (a single block) makes $F = 1$ attainable.

Theorem 1 (global convergence, stall-and-grow). *At each iteration j, (i) receive only the partition $\mathcal{P}_j$ from the oracle; (ii) run a Tucker step on $\mathcal{P}_j$ and apply the monotone gauge to form $|G_j\rangle$; (iii) if stalled at block size k, increase k. Then (F_j) is non-decreasing and converges for fixed k. Moreover, under stall-and-grow with k allowed to reach n (one block), the process attains $F = 1$ (exact preparation) in finitely many expansions.*

Proof. Monotonicity and boundedness ($0 \le F_j \le 1$) yield convergence for fixed k. Enlarging k cannot reduce the achievable limit; with $k = n$ any pure state can be mapped to $|0\rangle$ by a single block unitary [15], whence $F = 1$.

Summary. Q-Tucker's per-iteration fidelity *never decreases* when the monotone gauge is enforced after the oracle's partition. Oracle partitions that maximize inner entanglement typically improve the *rate* of progress by enlarging the cut ceiling. If the process stalls at a given block size, raising the block size restores progress; allowing it to reach the full register guarantees exact convergence.

4 Numerical results

To evaluate the utility of the method, we apply it to the MNIST 784 dataset [13], compressing the circuit with a maximum factor size of two. Each digit is embedded into 10 qubits, and each iteration introduces five two-qubit general unitaries. These decompose into a circuit of approximate depth 13–14 with respect to the

gate set $\{R_x, R_y, R_z, \mathrm{CX}\}$. Consequently, the circuit depth scales linearly with the number of iterations. For example, after six iterations the depth is approximately 78, while after 260 iterations it reaches about 3500, without applying any circuit optimization.

For the digit "zero" (index 0 of [13]), a clear fidelity-depth trade-off is observed: increasing the number of iterations reduces the fidelity loss, but simultaneously increases the circuit depth. This relationship is illustrated in Fig. 2. The standard Qiskit [18] initializer, which employs the isometries method [11], is used as a baseline. Any circuit depth exceeding this baseline is marked as *No Use*, as it provides no practical advantage. Approximations below this threshold – even those associated with relatively high fidelity loss – may still offer sufficient quality for certain applications.

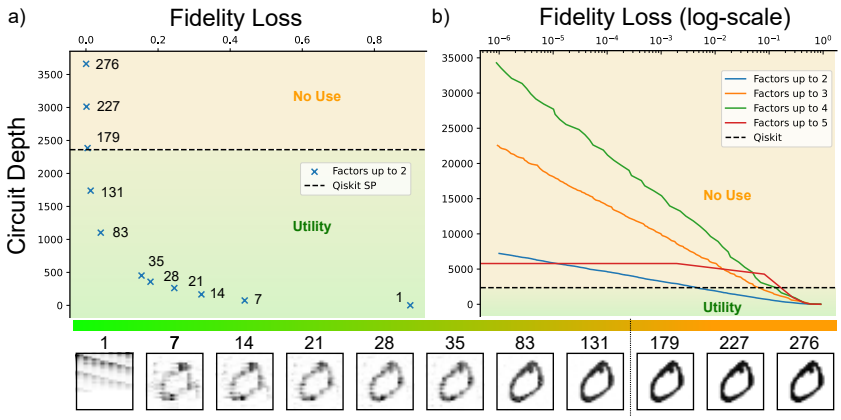


Fig. 2. This study examines the relationship between iteration count, circuit depth, and fidelity loss in the Q-Tucker method. As illustrated by the scatter plot, the method achieves rapid convergence – requiring fewer iterations – when an acceptable fidelity loss is targeted, highlighting its potential utility.

To examine the effect of the factor size, values of 2, 3, 4, and 5 were compared on the same digit (cf. Fig. 2). Since the synthesis of general three-, four-, and five-qubit unitaries is expensive, the method does not scale efficiently in this regime, highlighting the need for improved multi-qubit unitary synthesis techniques. Nevertheless, larger factor sizes exhibit a clear advantage in terms of convergence speed (not shown in the figure). For the digit "zero", convergence below a precision of 10^{-6} is achieved after 553 iterations with factor size 2, 163 with factor size 3, 51 with factor size 4, and only 8 with factor size 5. These results indicate that advances in unitary synthesis could yield substantial improvements for the iterative Tucker decomposition.

5 Comments and Conclusion

Empirical tests on real-world data show that even when Q-Tucker fails to fully converge – an infrequent scenario – it still produces high-quality approximations, often matching the best achievable results within a given classical computational budget. In our implementation, the number of iterations is capped at the square of the number of qubits. This parameter, which directly influences circuit depth, can be tuned to match the constraints of the target device and application. Within this limit, the method consistently delivers efficient decompositions compatible with available classical resources.

A key strength of the approach lies in its reliance on SVDs over two-qubit subsystems. Since these subsystems have fixed dimensionality, each SVD incurs a constant cost. This avoids the scalability bottleneck typical of tensor decompositions like Tucker, where SVD costs scale cubically with input size. Additionally, the heuristic used to search for a good configuration tuple has linear complexity with respect to the number of state amplitudes. As a result, the overall asymptotic cost of the method is linear in the number of amplitudes – a significant improvement over prior approaches [1].

The method also includes a convergence detection mechanism. When convergence is deemed unlikely, it falls back on the Schmidt method [1], incorporating partial disentanglement. This concept of partial disentanglement enables compressions that standard methods often overlook. However, the fallback remains comparatively slow, even when combined with greedy heuristics, highlighting the need for faster alternatives.

In summary, Q-Tucker achieves the best possible approximation allowed by the classical resources at hand, offering two key guarantees: (1) The approximation improves proportionally with increased classical resources. (2) The algorithm always returns a result – even for large, complex states where only a coarse approximation may be feasible.

Acknowledgments. We acknowledge funding of the Federal Ministry of Research, Technology and Space Project Number 13N17157 (Q-ROM).

Disclosure of Interests. Carsten Blank is the co-founder and CEO of data cybernetics. Furthermore, a patent application covering the methods described herein has been filed and is pending; details are under confidentiality restrictions.

References

1. Araujo, I.F., Blank, C., Araújo, I.C.S., da Silva, A.J.: Low-rank quantum state preparation. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **43**(1), 161–170 (2024)
2. Araujo, I.F., Park, D.K., Ludermir, T.B., Oliveira, W.R., Petruccione, F., Da Silva, A.J.: Configurable sublinear circuits for quantum state preparation. Quantum Information Processing **22**(2), 123 (2023)
3. Araujo, I.F., Park, D.K., Petruccione, F., da Silva, A.J.: A divide-and-conquer algorithm for quantum state preparation. Scientific reports **11**(1), 6329 (2021)

4. Bergholm, V., Vartiainen, J.J., Möttönen, M., Salomaa, M.M.: Quantum circuits with uniformly controlled one-qubit gates. Phys. Rev. A **71**, 052330 (2005)
5. Blank, C., Park, D.K., Rhee, J.K.K., Petruccione, F.: Quantum classifier with tailored quantum kernel. npj Quantum Information **6**(1), 1–7 (2020)
6. Cade, C., Mineh, L., Montanaro, A., Stanisic, S.: Strategies for solving the fermi-hubbard model on near-term quantum computers. Phys. Rev. B **102**, 235122 (2020)
7. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM Journal on Matrix Analysis and Applications **21**(4), 1253–1278 (2000)
8. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank-(r1 ,r2 ,. . .,rn) approximation of higher-order tensors. SIAM Journal on Matrix Analysis and Applications **21**(4), 1324–1342 (2000)
9. Grant, E., Benedetti, M., Cao, S., Hallam, A., Lockhart, J., Stojevic, V., Green, A.G., Severini, S.: Hierarchical quantum classifiers. npj Quantum Information **4**(1), 65 (2018)
10. Grover, L., Rudolph, T.: Creating superpositions that correspond to efficiently integrable probability distributions. arXiv (2002)
11. Iten, R., Colbeck, R., Kukuljan, I., Home, J., Christandl, M.: Quantum circuits for isometries. Physical Review A **93**(3), 032318 (2016)
12. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review **51**(3), 455–500 (2009)
13. LeCun, Y., Cortes, C., Burges, C.: Mnist handwritten digit database. ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist **2** (2010)
14. Möttönen, M., Vartiainen, J.J., Bergholm, V., Salomaa, M.M.: Quantum circuits for general multiqubit gates. Physical review letters **93**(13), 130502 (2004)
15. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press (2010)
16. Park, D.K., Petruccione, F., Rhee, J.K.K.: Circuit-based quantum random access memory for classical data. Scientific Reports **9**(1), 3949 (2019)
17. Plbnio, M.B., Virmani, S.: An introduction to entanglement measures. Quantum Info. Comput. **7**(1), 1–51 (2007)
18. Qiskit contributors: Qiskit: An open-source framework for quantum computing (2023). https://doi.org/10.5281/zenodo.2573505
19. Ran, S.J.: Encoding of matrix product states into quantum circuits of one- and two-qubit gates. Phys. Rev. A **101**, 032310 (2020)
20. Schön, C., Hammerer, K., Wolf, M.M., Cirac, J.I., Solano, E.: Sequential generation of matrix-product states in cavity qed. Phys. Rev. A **75**, 032311 (2007)
21. Schuld, M., Petruccione, F.: Supervised Learning with Quantum Computers. Quantum Science and Technology, Springer International Publishing (2018)
22. Stanislav, P., Marina, L.: Faster quantum state decomposition with Tucker tensor approximation. Quantum Machine Intelligence **5**(2), 25 (2023)
23. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. J. Optim. Theory Appl. **109**(3), 475–494 (2001)
24. Wei, T.C., Goldbart, P.M.: Geometric measure of entanglement and applications to bipartite and multipartite quantum states. Phys. Rev. A **68**(4) (2003)

Quantum Approaches to the Minimum Edge Multiway Cut Problem

Ali Abbassi[1,2], Yann Dujardin[1], Eric Gourdin[1], Philippe Lacomme[3], and Caroline Prodhon[2]

[1] Orange Research, Chatillon, France
[2] LIST3N, University of Technology of Troyes, Troyes, France
[3] LIMOS - UMR CNRS 6158, ISIMA - Université Clermont Auvergne, France
ali.abbassi@utt.fr,orange.com, yann.dujardin@orange.com, eric.gourdin@orange.com, philippe.lacomme@isima.fr, caroline.prodhon@utt.fr

Abstract. We investigate the minimum edge multiway cut problem, a fundamental task in evaluating the resilience of telecommunication networks. This study benchmarks the problem across three quantum computing paradigms: quantum annealing on a D-Wave quantum processing unit, photonic variational quantum circuits simulated on Quandela's Perceval platform, and IBM's gate-based Quantum Approximate Optimization Algorithm (QAOA). We assess the comparative feasibility of these approaches for early-stage quantum optimization, highlighting trade-offs in circuit constraints, encoding overhead, and scalability. Our findings suggest that quantum annealing currently offers the most scalable performance for this class of problems, while photonic and gate-based approaches remain limited by hardware and simulation depth. These results provide actionable insights for designing quantum workflows targeting combinatorial optimization in telecom security and resilience analysis.

Keywords: Quantum annealing · Multiway cut · Variational quantum circuits · QAOA · Photonic computing

1 Introduction

The *Minimum Edge Multiway Cut* (MEMC) problem arises in tasks where specific terminal nodes must be separated to satisfy specific resilience criteria for a network or unauthorized communication [7]. Given an undirected graph and a set of k terminals, the objective is to identify a minimum-weight set of edges whose removal ensures that no two terminals lie in the same connected component. For $k = 2$, the problem reduces to a classical min-cut, solvable in polynomial time via max-flow algorithms [14]. For $k \geq 3$, the problem is NP-hard [9].We evaluate the MEMC problem using three quantum computing paradigms: (i) quantum annealing on D-Wave hardware, (ii) the Quantum Approximate Optimization Algorithm (QAOA) implemented with IBM Qiskit, and (iii) variational photonic

quantum circuits simulated using Quandela's Perceval platform. Each paradigm operates on a shared QUBO formulation of MEMC, enabling consistent problem modeling across solvers [8].

The QUBO model is mapped to an Ising formulation for execution on the D-Wave platform and is employed as the cost Hamiltonian for QAOA and photonic variational circuits. D-Wave serves as a natural baseline due to its native support for QUBO-based quantum annealing and available hybrid solvers [10]. We report numerical results on small graph instances, including annealing behavior, QAOA convergence trends, and bitstring sampling accuracy in photonic circuits. This study aims to characterize the behavior of different quantum optimization approaches on a unified problem encoding, highlighting trade-offs in scalability, model expressiveness, and hardware compatibility [2]. The paper is structured as follows: Section 2 defines the MEMC problem and recalls key complexity results. Section 3 presents the QUBO formulation used across all quantum paradigms. Section 4 introduces the three quantum optimization approaches: quantum annealing, QAOA, and photonic variational circuits. Section 5 reports numerical results obtained on representative instances. Section 6 concludes with a discussion of current limitations and directions for future work.

2 Problem Statement

We consider the *Minimum Edge Multiway Cut* (MEMC) problem, a classical NP-hard graph partitioning problem with applications in network interdiction, secure communications, and containment strategies [20].

Input. An undirected connected graph $G = (V, E)$, a non-negative cost function $C : E \to \mathbb{R}_{\geq 0}$, and a set of $k \geq 2$ terminals $T = \{t_1, \ldots, t_k\} \subseteq V$.

Goal. Find a minimum-cost set of edges $F \subseteq E$ such that in the graph $G' = (V, E \setminus F)$, no two terminals lie in the same connected component.

The problem generalizes the classic *minimum s–t cut* (when $k = 2$), solvable in polynomial time via max-flow algorithms [14]. For $k \geq 3$, the problem becomes NP-hard [9], and remains hard even under restrictions such as bounded degree or treewidth [6][22]. While fixed-parameter tractable algorithms exist for certain parameters [4][21], the decision variant—testing whether a cut of cost at most K exists—is NP-complete in general. Greedy heuristics offer a $2 - \frac{2}{k}$ approximation ratio [9], but exact solutions for larger instances often require exponential-time methods or relaxation techniques. We consider the optimization variant and reformulate it as a Quadratic Unconstrained Binary Optimization (QUBO), which serves as a common ground for deploying quantum solvers.

3 QUBO formulation

We adopt the QUBO formulation for the MEMC problem from [18], designed for efficient embedding on quantum hardware. Given a graph $G = (V, E)$, each

vertex $u \in V$ is assigned to one of the terminals $t \in T$ via binary variables $x_{u,t} \in \{0,1\}$, where $x_{u,t} = 1$ if node u is assigned to terminal t. This defines a partition of the graph vertices. We hence consider the following Hamiltonian:

$$H(x) = \alpha \left(\sum_{u \in V} \left(1 - \sum_{t \in T} x_{u,t} \right)^2 + \sum_{\substack{t,t' \in T \\ t \neq t'}} x_{t,t'} \right) + \sum_{\{u,v\} \in E} \sum_{t \in T} \sum_{\substack{t' \in T \\ t \neq t'}} C(\{u,v\}) x_{u,t} x_{v,t'} \tag{1}$$

The first term ensures that each node is assigned to exactly one terminal, and the second term prevents conflicting terminal overlaps (no terminal is assigned to another terminal). The final term accumulates the cost of cutting edges that connect vertices assigned to different terminals. The scalar α regulates the strength of the constraint terms relative to the objective.

4 Quantum Optimization

We describe the different quantum optimization workflows considered in this study and briefly recall their theoretical soundness and encoding strategies [1][2].

4.1 Quantum Annealing

Quantum annealing (QA) is a continuous-time optimization method based on the quantum adiabatic theorem [11]. The system evolves under a time-dependent Hamiltonian

$$H(t) = A(t)H_0 + B(t)H_p, \quad t \in [0,T], \tag{2}$$

where H_0 is a transverse-field driver and H_p encodes the cost function. Provided $A(t), B(t)$ vary slowly enough and the spectral gap remains non-negligible, the system ends near the ground state of H_p.

Our QUBO formulation is mapped to an Ising energy function via $x_i = (1 - z_i)/2$ with $z_i \in \{-1, 1\}$ [15]:

$$E(z) = \sum_{i<j} J_{ij} z_i z_j + \sum_i h_i z_i, \tag{3}$$

where $J_{ij} = \frac{1}{4} Q_{ij}$ and $h_i = -\frac{1}{4} \sum_j (Q_{ij} + Q_{ji})$. This defines an Ising spin glass whose ground state solves the original problem.

The Ising cost function is promoted to a diagonal quantum Hamiltonian:

$$H_p = \sum_{i<j} J_{ij} Z_i Z_j + \sum_i h_i Z_i, \tag{4}$$

where Z_i is the Pauli-Z operator on qubit i. On D-Wave systems, this Hamiltonian is implemented using programmable couplings and local fields over a sparse hardware graph [10]. Minor embedding is used to satisfy hardware connectivity constraints by linking physical qubits into chains, enabling annealing toward a minimum-energy configuration that corresponds to low cost solution.

4.2 Quantum Approximate Optimization Algorithm

The Quantum Approximate Optimization Algorithm (QAOA) [12] is a variational quantum algorithm for solving discrete optimization problems using alternating Hamiltonian dynamics on shallow circuits. It operates on a cost Hamiltonian H_C, derived from the QUBO model, and a mixer Hamiltonian H_M. The QAOA$_p$ ansatz prepares the state

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \prod_{\ell=1}^{p} e^{-i\beta_\ell H_M} e^{-i\gamma_\ell H_C} |\psi_0\rangle,$$

where $\boldsymbol{\gamma}, \boldsymbol{\beta} \in \mathbb{R}^p$ are tunable parameters and $|\psi_0\rangle = |+\rangle^{\otimes n}$ is the initial uniform superposition. In practice, H_C is implemented using single-qubit R_Z gates; the standard mixer $H_M = \sum_i X_i$ corresponds to global R_X rotations.

The parameters are optimized by minimizing the energy expectation value $\langle\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})|H_C|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle$ using classical optimizers such as COBYLA. The QAOA structure can be interpreted as a Trotterized approximation of evolution under $H = H_C + H_M$, with a "bang-bang" control profile [5].

While standard QAOA explores the full Hilbert space, constrained extensions like the Quantum Alternating Operator Ansatz (QAOA-A) [17] use custom mixers to enforce problem constraints. These were not used in our experiments; instead, constraints are embedded directly in the QUBO energy landscape.

4.3 Variational Circuits on Photonic Quantum Computers

Photonic variational quantum circuits offer an alternative paradigm for solving combinatorial optimization problems using linear optical components [19] [3]. These circuits operate in the discrete-variable regime, where Fock states—photon number states—form the computational basis. Unitary transformations are implemented via passive interferometers built from beam splitters and phase shifters. Classical optimization is used to adjust circuit parameters to minimize a QUBO-derived cost function.

Each Fock state $|\boldsymbol{n}\rangle = |n_0, n_1, \ldots, n_{M-1}\rangle$ encodes the number of photons in M modes, with $\sum_j n_j = P$ photons in total. For binary optimization, encodings such as dual-rail and parity are used. In dual-rail encoding, a logical qubit spans two modes:

$$|0\rangle = |1\rangle_0 \otimes |0\rangle_1, \quad |1\rangle = |0\rangle_0 \otimes |1\rangle_1.$$

We use a parity-based mapping [24], where each bit is computed as $x_i = n_i \bmod 2$, enabling direct evaluation of QUBO objectives: $H(x) = x^\top Q x$.

The circuit applies a unitary transformation $U(\boldsymbol{\theta}, \boldsymbol{\phi})$ to the input Fock state $|\boldsymbol{n}_{\text{in}}\rangle$. This unitary is decomposed into parameterized beam splitters and phase shifters:

$$\text{BS}(\theta, \phi) = \begin{pmatrix} \cos(\theta) & -e^{i\phi}\sin(\theta) \\ e^{-i\phi}\sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \text{PS}(\varphi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{pmatrix}.$$

These gates evolve the quantum state to:

$$|\psi_{\text{out}}\rangle = U(\boldsymbol{\theta}, \boldsymbol{\phi})|\boldsymbol{n}_{\text{in}}\rangle.$$

Measurements in the Fock basis yield output configurations $\boldsymbol{n}_{\text{out}}$, which are mapped to bitstrings $x \in \{0,1\}^N$ via parity. The QUBO energy is estimated as:

$$\mathbb{E}_{x\sim p(\boldsymbol{\theta},\boldsymbol{\phi})}[H(x)] = \sum_x p(x) \cdot x^\top Q x.$$

Minimizing this expected value over $(\boldsymbol{\theta}, \boldsymbol{\phi})$ forms the variational optimization loop. We simulate small MEMC instances using `Perceval` [19] under idealized conditions and evaluate the circuit's ability to sample near-optimal bitstrings.

5 Numerical Analysis

Preliminary results confirm that *small-scale instances* of the MEMC problem can be solved to optimality across all three quantum workflows—quantum annealing (QA), gate-based QAOA, and photonic variational circuits (VQC). In cases with $k = 2$ and $|V| \leq 10$, all platforms return the optimal solution, demonstrating that constrained combinatorial problems are tractable in low-dimensional quantum settings.

Table 1. Numerical summary

Backend	Small instances ($k = 2$, $\lvert V\rvert \leq 10$)	Larger instances ($k = 2$, $10 < \lvert V\rvert < 70$)	Convergence / notes
D-Wave QA	Optimal (100%)	Gap $\approx$ 20–40% (topology-dependent)	Chain breaks observed, avg. chain length $\sim$ 8; hybrid solver recommended for stability.
D-Wave Hybrid	Optimal (100%)	Near-optimal; typical gap $\lesssim$ 10–15%	Longer runtime; classical post-processing mitigates embedding effects.
QAOA (sim.)	Optimal bitstring sampled with $\sim$ 40% probability at $p = 1$; no improvement for larger iterations.	Gap $\sim$ 15–30% at fixed p; no systematic improvement for larger p (barren plateau)	Converges in $\sim$10–20 iterations; decreasing probability of sampling best bitstring at $p = 1$, effect of BP for medium to large instances.
Photonic (sim.)	Optimal sampled with $>$ 90% probability.	Gap $\sim$ 20–40%; infeasible to simulate larger systems with SLOS backend	$\sim$300–1500 parameter updates; Nelder–Mead more stable than COBYLA.

Figure 1 compares cut costs obtained by simulated annealing, quantum annealing, and D-Wave's hybrid solver across increasing graph sizes. For small

graphs, all solvers yield the same optimal values. However, quantum annealing performance deteriorates sharply starting from the $(10, 70)$ instance, where QA incurs significantly higher cut costs. This degradation is likely due to *embedding overheads, chain break errors, or insufficient annealing schedule expressivity* in larger QUBO encodings. Figure 2 shows runtime scaling on D-Wave hardware. QA remains efficient (under 1 second) for small graphs but exhibits a two-order-of-magnitude spike on the $(10, 70)$ instance, likely reflecting increased overhead for embedding and retries. In contrast, the hybrid solver maintains stable runtime (10–20 seconds) across all sizes, while consistently achieving better solution quality. These observations highlight a key trade-off between computational time and output accuracy, with hybrid workflows offering robustness as instance size increases.

Figure 3 shows the QAOA circuit structure for a depth-$p = 1$ instance on a small MEMC graph. The ansatz alternates between entangling ZZ-cost layers and parametrized single-qubit mixers. While the structure is amenable to gate-based simulation, circuit width and entanglement depth increase quadratically with instance size, limiting simulation feasibility to $n \lesssim 14$ qubits.

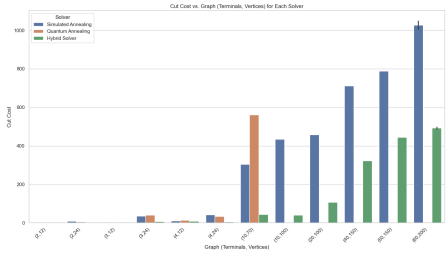


Fig. 1. Hybrid vs Quantum Annealing Cut Cost

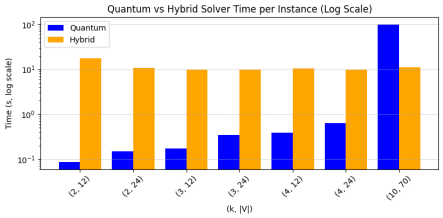


Fig. 2. D-Wave Runtime vs Instance Size

Figure 4 reports the optimization trajectory of the expected cost across 25 classical steps Despite operating in a noise-free environment, the energy landscape remains rugged, leading to unstable convergence and local minima. This reflects known challenges such as barren plateaus in variational algorithms. While convergence is eventually reached, QAOA failed to scale to larger MEMC instances ($k \geq 3$ and $|V| \geq 15$) in our setting. Nevertheless, it remains more expressive than photonic workflows in terms of circuit design and model capacity.

Photonic simulations were conducted using Perceval's `GenericInterferometer` and the SLOS backend, combined with classical optimizers. For each parameter configuration, $N = 10000$ measurement shots estimated the QUBO cost. In Figure 6, a toy instance ($|V| = 4$, $k = 2$) demonstrates strong sampling concentration on the optimal bitstring. Figure 5 shows convergence of QUBO energy using multiple optimizers, with Nelder–Mead yielding greater stability.

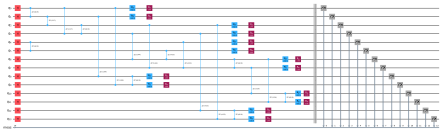


Fig. 3. QAOA Circuit Structure ($p = 1$)

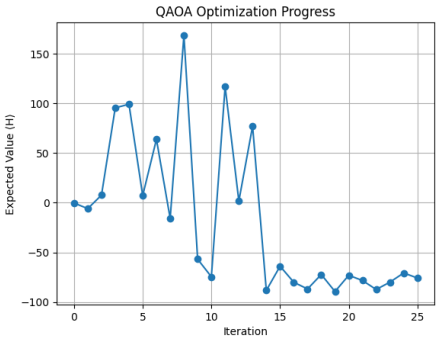


Fig. 4. QAOA Training Convergence

Nonetheless, scalability of the photonic workflow remains limited: performance degrades significantly for problems with more than 10 variables or $k \geq 3$. Embedding effects, photon loss, and decoherence are not modeled here. Future directions include testing with qudit encodings or hybrid analog encodings to address mode limitations.

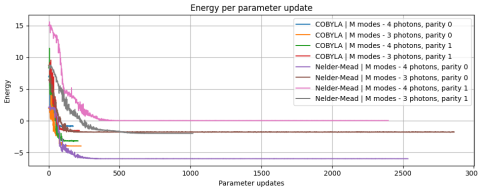


Fig. 5. Photonic Energy Convergence

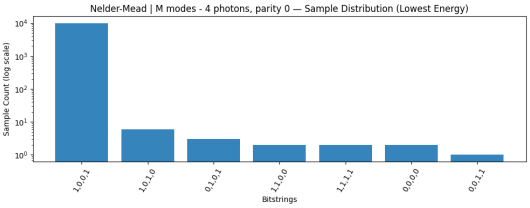


Fig. 6. Photonic Bitstring Frequencies

6 Conclusion

We benchmarked the MEMC problem across three quantum paradigms under the same QUBO encoding. For small instances ($|V| \leq 10$, $k \leq 2$), all methods—quantum annealing, QAOA, and photonic variational circuits—recovered optimal solutions, confirming baseline compatibility. Among these, D-Wave showed the highest scalability, solving larger graphs with reasonable solution quality despite embedding overhead. The QUBO formulation, designed with annealing in mind, favored sparse, pairwise interactions, reinforcing D-Wave's suitability. QAOA was limited by circuit depth and barren plateaus. Despite convergence in simulation, performance did not extend to medium-scale problems. More advanced designs (e.g., constraint-preserving mixers [17], parameter concentration [5]) remain to be tested. Photonic simulations reliably captured optimal

bitstrings on toy instances but degrade rapidly beyond ∼10 variables. Current parity encodings and the exponential state space are major constraints. Hardware-oriented enhancements (e.g., qudit encodings) are promising.

Overall, Quantum annealing remains the most mature option at scale. Future work will target more expressive encodings, noise-resilient QAOA circuits, and physical deployment of photonic models.

7 Supplementary Material

Table 2. System properties and algorithmic parameters across the three backends.

D-Wave Advantage (eu-central-1)	
QPU	Advantage_system5.4, Pegasus topology (5614 working qubits)
Temperature	16.4 mK
Supported Problems	QUBO, Ising
Annealing Time	[0.5, 2000] μs (default 20 μs)
Embedding	Automatic minor-embedding, avg. chain length
Coupling Ranges	$J \in [-2, 1]$, $h \in [-18, 15]$
Reads / Postproc	Multiple reads, majority-vote unembedding
Quandela Perceval (Photonic VQC, simulated)	
Backend	SLOS shot-based simulator
Ansatz	Generic interferometer (beam splitters + phase shifters)
Encoding	Parity-mapped Fock states ($j = 0, 1$), full/reduced occupancy
Parameters	Phases $\{\theta_i, \phi_{tr,i}, \phi_i\}$ uniform in $[0, 2\pi]$
Optimizers	Powell, COBYLA, Nelder–Mead (max. 1000 it.)
Loss Function	Variational energy $\langle\psi(\boldsymbol{\theta})\|H\|\psi(\boldsymbol{\theta})\rangle$
Samples	10 000 per evaluation
Postproc	Parity conversion, histogramming of distributions
IBM Qiskit QAOA (gate-based, simulated)	
Backend	Qiskit AerSimulator (statevector + 4000 shots)
Ansatz	QAOAAnsatz, depth $p = 1$ (U_C, U_M)
Initialization	$\beta_0 = \pi/2$, $\gamma_0 = \pi/4$
Optimizer	COBYLA, bounds $[0, \pi]$
Objective	$\langle\psi(\boldsymbol{\beta}, \boldsymbol{\gamma})\|H\|\psi(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle$
Compilation	Transpilation, preset pass manager (opt. level 0)
Sampling	All qubits measured, normalized distributions

Reproducibility and scope. All implementations use public SDKs (Ocean, Qiskit, Perceval), with parameter ranges listed in Table 2. Classical baselines are not included here; they remain essential for competitiveness analysis but outside the scope of this feasibility study.

Contributions and scope. This work provides one of the first cross-paradigm studies of the *Minimum Edge Multiway Cut* (MEMC) using a single QUBO formulation across three modalities: D-Wave annealing, gate-based QAOA, and photonic variational circuits. The main contributions are: (i) a unified encoding enabling consistent benchmarking of embedding overheads and workflow constraints; (ii) a qualitative analysis of scalability bottlenecks (iii) an application-driven link between MEMC and telecom-network resilience;

Unlike prior work focused on native device problems (e.g., Max-Cut on hardware graphs), we benchmark a combinatorial problem with direct industrial motivation. The target is feasibility and methodology, not quantum advantage.

Telecom resilience Finding minimum sets of "critical" edges (such that a simultaneous breakdown of these edges would imply a shutdown of a service) is essential for designing "resilient" telecom networks with low probability of failure. This is a hard task since there is a trivial reduction from the NP-hard problem "min edge multiway cut" to this problem.

References

1. Abbas, A., et al. (2024). Challenges and opportunities in quantum optimization. *Nature Reviews Physics*.
2. Bochkarev, A., Heese, R., Jäger, S., Schiewe, P., Schöbel, A. (2024). Quantum computing for discrete optimization: A highlight of three technologies. arXiv:2409.01373.
3. Bombardelli, M., Fleury, G., Lacomme, P., Vulpescu, B. (2025). Foundations of photonic quantum computation. *arXiv preprint* arXiv:2509.04266.
4. Bousquet, N., Daligault, J., Thomassé, S. (2011). Multicut is FPT. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pp. 459–468.
5. Brady, L. T., Van Heck, B., Keating, C. T. (2021). Bang-bang control and quantum approximate optimization. *Quantum*, 5, 500.
6. Calinescu, G., Karloff, H., Rabani, Y. (2003). Multicuts in unweighted graphs and digraphs with bounded degree and bounded tree-width. *Journal of Algorithms*, 48(2), 333–359.
7. Costa, M.-C., Coudert, D., Jaumard, B. (2005). Minimal multicut and maximal integer multiflow: A survey. *European Journal of Operational Research*, 162(1), 55–69.
8. W. Cruz-Santos, C. Salas, and J. C. López-Sánchez, "A QUBO formulation of minimum multicut problem instances in trees for D-Wave quantum annealers," *Sci. Rep.*, vol. 9, p. 1721, 2019.
9. Dahlhaus, E., Johnson, D. S., Papadimitriou, C. H., Seymour, P. D., Yannakakis, M. (1994). The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23(4), 864–894.
10. D-Wave Systems Inc. (2024). Hybrid workflow solvers. Technical documentation.
11. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M. (2000). Quantum computation by adiabatic evolution. arXiv:quant-ph/0001106.
12. Farhi, E., Goldstone, J., Gutmann, S. (2014). A quantum approximate optimization algorithm. arXiv:1411.4028.

13. Flamini, F., Spagnolo, N., Sciarrino, F. (2018). Photonic quantum information processing: A review. *Reports on Progress in Physics*, 82(1), 016001.
14. Garg, N., Vazirani, V. V., Yannakakis, M. (1996). Approximate max-flow min-(multi)cut theorems and their applications. *SIAM Journal on Computing*, 25(2), 235–251.
15. Glover, F., Kochenberger, G., Du, Y. (2019). A tutorial on formulating and using QUBO models. arXiv:1811.11538.
16. Guo, J., Niedermeier, R. (2008). Complexity and exact algorithms for vertex multicut in interval and bounded treewidth graphs. *European Journal of Operational Research*, 186(2), 542–553.
17. Hadfield, S., Wang, Z., O'Gorman, B., Rieffel, E., Venturelli, D., Biswas, R. (2019). From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2), 34.
18. Heidari, S., Dinneen, M. J., Delmas, P. (2022). An equivalent QUBO model to the minimum multi-way cut problem. CDMTCS Research Reports, Tech. Rep. CDMTCS-565.
19. Heurtel, C., Arrazola, J. M., Delgado, A., Thiel, V., Wright, L. G., Orieux, A. (2023). Perceval: A software platform for discrete-variable photonic quantum computing. *Quantum Science and Technology*, 8, 014006.
20. Magnouche, Y., Mahjoub, A. R., Martin, S. (2020). The multi-terminal vertex separator problem: Branch-and-cut-and-price. *Discrete Applied Mathematics*, 286, 168–189.
21. Marx, D., Razgon, I. (2011). Fixed-parameter tractability of multicut parameterized by the size of the cutset. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pp. 469–478.
22. Papadopoulos, C. (2012). Restricted vertex multicut on permutation graphs. *Discrete Applied Mathematics*, 160(12), 1791–1797.
23. Venegas-Andraca, S. E., Barrera, J. F., Boyer, M. L., Carrasco, A. E. (2018). A cross-disciplinary introduction to quantum annealing-based algorithms. *Contemporary Physics*, 59(2), 174–196.
24. Brádler, K., Wallner, H.: Certain properties and applications of shallow bosonic circuits. arXiv:2112.09766 [quant-ph] (2021). https://doi.org/10.48550/arXiv.2112.09766

Systematic improvement of the quantum approximate optimisation ansatz for combinatorial optimisation using quantum subspace expansion

Yann Beaujeault-Taudière[1][0000−0002−7317−1287]

Independent researcher, France
yann.beaujeault-t@protonmail.com

Abstract. The quantum approximate optimisation ansatz (QAOA) is one of the flagship algorithms used to tackle combinatorial optimisation on graphs problems using a quantum computer, and is considered a strong candidate for early fault-tolerant advantage. In this work, I study the enhancement of the QAOA with a generator coordinates method (GCM), and achieve systematic performances improvements in the approximation ratio and fidelity for the maximal independent set on Erdös-Rényi graphs. The cost-to-solution of the present method and the QAOA are compared by analysing the number of logical CNOT and T gates required for either algorithm. Extrapolating on the numerical results obtained, it is estimated that for this specific problem and setup, the approach surpasses QAOA for graphs of size greater than 75 using as little as eight trial states.

Keywords: Quantum approximate optimisation ansatz · quantum subspace expansion · combinatorial optimisation

1 Introduction

Among the myriad fundamental and applied use cases that quantum computing is expected to disrupt, combinatorial optimisation on graphs stands a peculiar role, finding applications across several fields. The possibility to achieve quantum advantage for solving NP-hard instances has been the focus of extensive experimental and theoretical research in the recent years, within the scope of both analogue and digital approaches. The quantum approximate optimisation ansatz (QAOA) [1,2] has emerged as one of the most promising variational algorithms to tackle graph problems. Several variations have been put forth, aiming for instance at reducing the circuit depth by using more sophisticated mixers, or the number of evaluations of the cost function during the optimisation (see [2] for a review). The existing flavours of QAOA all add refinements of varying complexity to the original ansatz. In this work, an alternative and complementary route is followed, combining the QAOA ansatz with a generator coordinate method

(GCM), which is a standard tool for approximating the solution to eigenvalues problems on classical processors.

This paper is organised as follows. Section 2, 3, 4 respectively present the general aspects of the MIS problem, of the QAOA, and of the GCM. Section 5 defines a simplified metric that can be used to determine whether applying the QSE protocol on top of a circuit optimised by QAOA leads to an improvement in the time-to-solution. The evaluation of the so-called Hamiltonian kernels, which represent the most challenging quantity to evaluate on a quantum processing unit (QPU), is discussed in section 6 ; three different procedures are briefly analysed. Section 7 presents numerical results.

2 Maximal independent set

The maximum independent set is a classic problem of graph theory [3]. Given a graph $G = (V, E)$, where V is the set of vertices and E is the set of edges, the objective is to find the largest set of independent vertices, that is, the largest set where no two vertices are linked by an edge. The MIS problem is generally NP-hard, which means that there is no known classical algorithm that can find the optimal solution in polynomial time. There exist however efficient heuristic algorithms that can get close to the optimal solution, provided the MIS in question belongs to the appropriate complexity class [4]. The MIS problem can be formalised as finding the minimum of the cost Hamiltonian

$$\hat{H}_C = -\sum_{i \in V} \hat{n}_i + \sum_{(i,j) \in E} \hat{n}_i \hat{n}_j, \tag{1}$$

where the number operator $\hat{n}_i = (1 - Z_i)/2$ corresponds to rejecting ($\langle \hat{n}_i \rangle = 0$) or including ($\langle \hat{n}_i \rangle = 1$) the node i in the solution. Expanding (1) in terms of the Pauli matrices Z_i puts the Hamiltonian in a form suited for quantum simulation.

3 Quantum approximate optimisation ansatz

The quantum approximate optimisation ansatz (QAOA) [1,2] is one of the leading algorithms for solving combinatorial optimisation problems [2]. In its original version, QAOA makes use of two non-commuting operators $\hat{H}_C$ and $\hat{H}_M$, called the cost and mixer Hamiltonians, respectively. Together, they define the alternating parametric unitary of depth L

$$U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{\ell=1}^{L} e^{-i\beta_\ell \hat{H}_M} e^{-i\gamma_\ell \hat{H}_C}, \tag{2}$$

and one seeks to minimise the following expectation value:

$$\begin{aligned} C(\boldsymbol{\gamma}, \boldsymbol{\beta}) &= \langle \phi | U^\dagger(\boldsymbol{\gamma}, \boldsymbol{\beta}) \hat{H}_C U(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \phi \rangle \\ &\equiv \langle \Phi_0 | \hat{H}_C | \Phi_0 \rangle, \end{aligned} \tag{3}$$

In order to lower the energy, the mixer Hamiltonian should not commute with $\hat{H}_C$. It should furthermore allow reaching the ground state of the cost Hamiltonian, which can be formalised as the requirement that the nested commutators of the cost and mixer produce an algebra whose exponential map contains at least one path transforming $|\phi\rangle$ into the proper ground state. These two conditions can be fulfilled regardless of the initial state by the simple yet efficient choice of a coherent rotation mixer:

$$\hat{H}_M = -\sum_{i=0}^{N-1} X_i, \tag{4}$$

where X_i is the Pauli X operator acting on qubit i.

4 Generator coordinate method

The generator coordinate method (GCM) [5,6] is a general method to obtain ground and excited states properties of a Hamiltonian. The GCM is based upon expanding the eigenstates of the cost Hamiltonian on a set of (usually non-orthogonal) generating states $\{|\chi_k(\boldsymbol{q}_k)\rangle\}_{k=0,\ldots,K-1}$ according to

$$|\Psi_j\rangle = \sum_{k=0}^{K-1} f_{k,j}\ |\chi_k(\boldsymbol{q}_k)\rangle, \tag{5}$$

the parameters $\boldsymbol{q}_k$ being referred to as the generator coordinates. One has full freedom in choosing the generating states; for instance, this formulation encompasses real-time quantum subspace expansion (QSE) [12] as the special case $\boldsymbol{q}_k = t_k$ and $|\chi_k(t_k)\rangle = e^{-i\hat{H}_C t_k}\ |\Phi_0\rangle$. The GCM approach has been recently applied on quantum computers in order to access properties of strongly coupled fermionic systems, such as the one-body Green's functions [16,17], ground and excited states building on a unitary coupled clusters ansatz [18,19] or coherent one-body rotations in the case of Hamiltonians possessing permutation invariance [20].

Given the expansion (5), approximate eigenstates of $\hat{H}_C$ can be obtained by applying the Rayleigh-Ritz variational principle to the cost function

$$C^{(j)}_{\mathrm{GCM}}(\{f_{k,0}\}) = \sum_{kk'} f^*_{k,j} f_{k',j} \langle\chi_k|\hat{H}_C|\chi_{k'}\rangle, \tag{6}$$

with the normalisation

$$\sum_{kk'} f^*_{k,j} f_{k',j} \langle\chi_k|\chi_{k'}\rangle = 1, \tag{7}$$

leading to a generalised eigenvalue equation

$$\mathcal{H}\ |\Psi\rangle = E\mathcal{S}\ |\Psi\rangle, \tag{8}$$

where $\mathcal{H}$ and $\mathcal{S}$ are the so-called Hamiltonian and overlap kernels, whose elements read

$$\mathcal{H}_{kk'} = \langle \chi_k | \hat{H}_C | \chi_{k'} \rangle, \tag{9}$$
$$\mathcal{S}_{kk'} = \langle \chi_k | \chi_{k'} \rangle. \tag{10}$$

In practice, the kernels (9)-(10) estimated on the QPU, and (8) is solved classically.

Due to the non-orthogonality of the generating states, the overlap matrix may contain small eigenvalues which spoil the numerical resolution of Eq.(8), even when the matrix elements are computed exactly [13,14]. This instability is exacerbated in the presence of noise. A simple prescription is to project out the states that are quasi-degenerate with the rest of the basis, by removing the eigenvalues that are smaller than a given threshold ε_{cut}. This has been used successfully in the nuclear physics context since over three decades [13,14], delivering robust and reliable results for virtually any reasonable choice of the threshold. The convergence and stability properties of this approach were studied in detail in [12].

5 When is GCM a good FTQC algorithm?

Whether an initial state algorithm should be employed rather than another is measured by whether it reduces the overall computational resources, that is, runtime, number of gates, and/or number of logical qubits required. Among these metrics, the number of gates plays the most critical role, as it essentially dictates the balance between space and time tradeoffs during the execution on the QPU. Focusing on this quantity alone is thus a good first first-order estimate of an algorithm's performances, and in the following discussion, I use the term "cost" to refer to the number of gates required to run the algorithms.

The fidelity between the ground state(s) and the state prepared by a given algorithm "X" is written $F^{(X)}$. The cost-to-fidelity ratio of the GCM is favourable when the fractional increase in the fidelity outweighs the fractional increase in computational cost. This is achieved if the following condition is met:

$$\frac{F^{(\text{GCM})}}{F^{(\text{QAOA})}} \gtrsim 4(K-1)\left(1+\frac{1}{L'}\right) \times f, \tag{11}$$

where f is a scaling factor depending on which of the subcircuits introduced in Sec. 6 is used to extract the matrix elements of the kernels (see [21]), and $L' \leq L$ is the number of layers retained after the optimisation of the QAOA angles.

From (11), it can readily be argued that under some assumptions, there will always exist some critical graph size beyond which the GCM is always favourable over the plain QAOA, in the case of a fixed depth. Indeed, while there are no rigorous results on the practical scaling of the QAOA fidelity with the number of nodes, a simple empirical fit of the form $F^{(X)}(N) = \beta_X(1+e^{N\alpha_X})^{-1}$ captures the most salient expected features of the fidelity, namely its exponential decay

from one to zero as N grows, and its plateau effect for small N. As the left-hand side of (11) grows exponentially, while the right-hand side scales polynomiallly, the GCM will deliver a better cost-to-fidelity ratio than QAOA for larger graphs. In Sec. 7, this simple metric is used to extrapolate the QAOA and GCM results in the special case of Erdös-Rényi graphs with fixed average density.

6 Computation of the kernels

The matrix elements $\mathcal{A}_{kk'} \equiv \langle\Psi_k|\hat{A}|\Psi_{k'}\rangle$ can be evaluated using two Hadamard tests, as shown in Fig. 1. While evaluating $\mathcal{S}_{kk'}$ poses little challenge, estimating

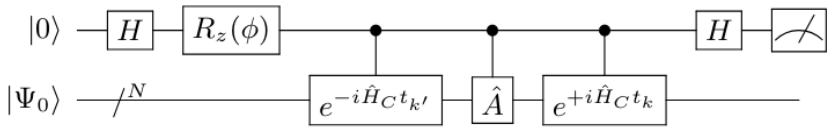


Fig. 1. Circuit for the evaluation of $\mathcal{A}_{kk'}$ with Hadamard tests. The real part of $\mathcal{A}_{kk'}$ is obtained by setting ϕ to zero. The imaginary part is obtained using $\phi = -\pi/2$, which corresponds to the modified Hadamard test. Both parts are recovered as the difference between the probabilities of measuring the ancillary qubit in the state zero and one.

the kernels $\mathcal{H}_{kk'}$ requires more care, as the cost Hamiltonian is not a unitary operator. This can be addressed in several ways (see [21] for a detailed discussion and derivation):

- By estimating the expectation values of the Pauli strings $\{Z_n, Z_n Z_{n'}\}$ directly. With this approach, $N/2$ two-qubits expectation values can be estimated simultaneously by grouping the operators into commuting sets. This approach furthermore avoids the use of extra CNOT and T gates beyond the ones required to encode the states. Because $\mathcal{O}(\sqrt{\rho}N)$ rounds of measurements are performed, each with approximately $\sqrt{\rho}N$ terms evaluated, the total number of shots scales as $(\sqrt{\rho}N)^3$.
- By using real time evolution and forming linear combinations. In that case, one estimates $\langle\Psi_k|e^{-i\hat{H}_C t_j}|\Psi_{k'}\rangle = \mathcal{S}_{kk'} - it_j\mathcal{H}_{kk'} + \mathcal{O}(t_j^2)$ for several propagation times $\{t_j\}_{j=1,\cdots,p}$. The kernels $\mathcal{S}_{kk'}, \mathcal{H}_{kk'}$ can then be recovered to precision $\mathcal{O}(t^p)$ by forming linear combinations of the p expectation values. The principal disadvantage of this approach is that it requires large values of p in order to estimate matrix elements to high precision, which in turns increases the final variance.
- By encoding $\hat{H}_C$ into a larger space using the linear combination of unitaries (LCU) approach [15]. This approach involves $\mathcal{O}(\log_2(N))$ ancillary qubits. The LCU is a probabilistic algorithm, and close to the optimal solution, $p_{\text{success}} \sim (\log_2(N)/N)^4$.

Which approach is optimal depends on several parameters, including the availability of ancillary qubits, the required precision to which the unitaries must be realised, which algorithm is used to decompose them on the elementary gate set, and the scaling of the off-diagonal moments $\langle \Psi_k | \hat{H}_C^p | \Psi_{k'} \rangle$. The three approaches are analysed and compared in more depth in [21].

7 Numerical results

7.1 Parameter setting

For all the numerical applications shown below, the number of QAOA layers is set to $L = 20$, and only one layer is optimised at a time in order to keep the runtime reasonable. That is, once the ℓth layer's parameters $(\gamma_\ell, \beta_\ell)$ have been optimised, they are fixed before moving on to the optimisation of the $(\ell+1)^{\text{th}}$ layer. The angles are optimised using Scipy's implementation of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) gradient descent algorithm [8,9]. Finally, after the L layers, the optimal depth $L' \leq L$ is chosen simply by selecting the depth that leads to the lowest cost function. All simulations are carried out using PennyLane library [22], using a noiseless statevector simulator.

For the QSE stage, the evolution times are taken equally spaced in the $[-\pi(1-1/K); \pi(1-1/K)]$ interval. This gives both $\mathcal{H}$ and $\mathcal{S}$ a Toeplitz structure, such that the number of matrix element to estimate is linear in K. When diagonalising the overlap kernel, states with eigenvalue below $\varepsilon_{\text{cut}} = 10^{-3}$ are discarded. These parameters are chosen somewhat arbitrarily, and several problem-specific improvements are possible. As the QSE is guaranteed to deliver improved solution with respect to the QAOA, the simple choice for the values of the generator coordinates nonetheless leads to substantially better solutions.

7.2 Erdös-Rényi graphs

To showcase and benchmark the capabilities of the QSE more systematically, I apply the machinery on Erdös-Rényi (ER) graphs of average density $\rho = 0.5$. These graphs are NP-hard in general [4]. In particular, they cannot be addressed using analogue quantum processors when the number of nodes becomes large, as the probability that a random ER graph with fixed non-zero density is a unit-ball graph approaches zero in the large N limit [10]. Although for small graph size, the probability of generating a disconnected graph is non-negligible, this probability decreases exponentially fast as N grows and ρ remains fixed and strictly nonzero. For ER graphs, there is therefore no practical need to identify whether a graph is connected or not [11,10].

The QAOA-plus-QSE workflow is applied to graphs of size $N = 2$ to 10. In addition to the approximation ratio and the fidelity, the error in the particle number (i.e. in the Hamming weight) and parity are also shown on Fig. 2.

The QAOA achieves rather poor approximation ratios, reaching 0.5 for $N = 10$. Naturally, the fidelity degrades to even lower values, namely 0.15 at $N = 10$.

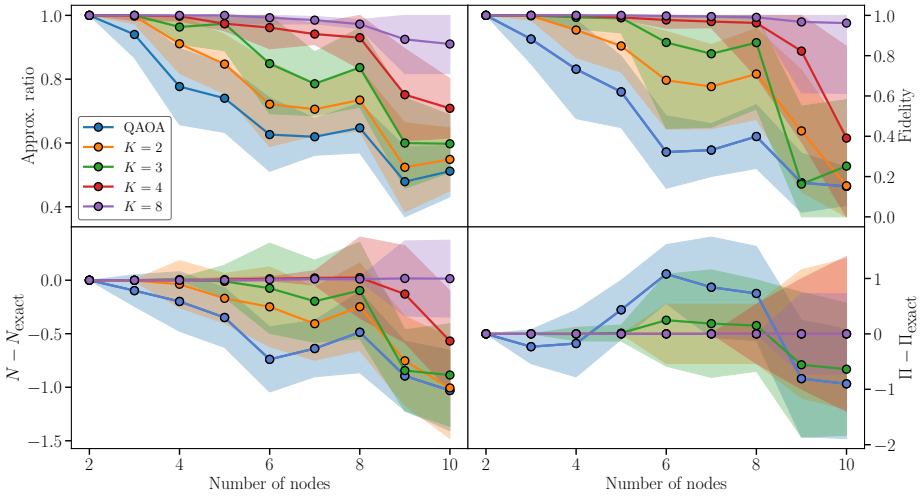


Fig. 2. Approximation ratio (top left), fidelity (top right), error in the Hamming weight/particle number (bottom left), and error in the parity (bottom right). On all subplots, the blue curve corresponds to the QAOA results, while the other curves represent the results obtained when adding a QSE on top, with $K = 2, 3, 4, 8$ trial states. For each number of nodes, 14 random Erdös-Rényi graphs are generated. The shaded areas correpond to one standard deviation.

Furthermore, the error in both the Hamming weight and the parity of the QAOA solutions is of order one, indicating that QAOA fails to evolve the initial state towards the correct symmetry subspace.

Adding the QSE step resolves all the issues of the QAOA optimisation, and increasing the number of trial states systematically improves all the relevant performance metrics. Notably, for $N = 10$, the approximation ratio is almost double. The performance of the QSE is even more striking for the fidelity, which remains significantly higher that the QAOA solutions, with a value of 0.96 for $N = 10$ nodes. As a consequence of the high fidelity, the Hamming weight (and therefore the parity) of the QSE solution quickly converge to the exact values.

Fitting the fidelities of the QAOA and the QSE to Fermi-Dirac distributions, one can extract the critical graph size N_* beyond which the QSE presents a favourable time-to-solution as the smallest N for which

$$\frac{\beta_{\text{QAOA}}}{1 + e^{N\alpha_{\text{QAOA}}}} \frac{1 + e^{N\alpha_{\text{QSE}}(K)}}{\beta_{\text{QSE}}(K)} > 2K(\sqrt{\rho}N)^3, \tag{12}$$

assuming the Hadamard plus Pauli method is used. In that specific setting and for $K = 8$, the QAOA-plus-QSE surpasses QAOA for $N > 75$, and increasing K should further reduce N_*. It should be stressed that this number implicitly depends on the choice of the QAOA initial state and mixer, on the number of layers, on the optimisation algorithm and strategy, on the type of graphs,

and on the optimisation problem itself. As such, the precise value obtained here bears little meaning, and the only merit of such extrapolation is to illustrate the advantages of the QSE over QAOA for graphs with size of practical interest.

Acknowledgments. I thank Denis Lacroix and Jing Zhang for useful discussions. I also wish to warmly thank E. Beaujeault-Taudière for enthusiastic comments at all stages of this work.

Disclosure of Interests. I declare no competing interests.

References

1. Farhi, E., Goldstone, J. & Gutmann, S. A Quantum Approximate Optimization Algorithm. (arXiv,2014,11), http://arxiv.org/abs/1411.4028, arXiv:1411.4028 [quant-ph]
2. Blekos, K., Brand, D., Ceschini, A., Chou, C., Li, R., Pandya, K. & Summer, A. A review on Quantum Approximate Optimization Algorithm and its variants. *Physics Reports*. **1068** pp. 1-66 (2024,6), http://dx.doi.org/10.1016/j.physrep.2024.03.002
3. West, D. Introduction to Graph Theory (2nd Edition). (Prentice Hall,2000,8)
4. Dalyac, C., Henry, L., Kim, M., Ahn, J. & Henriet, L. Exploring the impact of graph locality for the resolution of MIS with neutral atom devices. (arXiv,2023,6), http://arxiv.org/abs/2306.13373, arXiv:2306.13373 [quant-ph]
5. Hill, D. & Wheeler, J. Nuclear Constitution and the Interpretation of Fission Phenomena. *Physical Review*. **89**, 1102-1145 (1953,3), https://link.aps.org/doi/10.1103/PhysRev.89.1102
6. Ring, P. & Schuck, P. The Nuclear Many-Body Problem. (Springer Berlin, Heidelberg,1980,1)
7. Parrish, R. & McMahon, P. Quantum Filter Diagonalization: Quantum Eigendecomposition without Full Quantum Phase Estimation. (arXiv,2019,9), http://arxiv.org/abs/1909.08925, arXiv:1909.08925 [quant-ph]
8. Byrd, R., Lu, P., Nocedal, J. & Zhu, C. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal On Scientific Computing*. **16**, 1190-1208 (1995), https://doi.org/10.1137/0916069
9. Zhu, C., Byrd, R., Lu, P. & Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*. **23**, 550-560 (1997,12), https://doi.org/10.1145/279232.279236
10. Cazals, P., François, A., Henriet, L., Leclerc, L., Marin, M., Naghmouchi, Y., Coelho, W., Sikora, F., Vitale, V., Watrigant, R., Garzillo, M. & Dalyac, C. Identifying hard native instances for the maximum independent set problem on neutral atoms quantum processors. (arXiv,2025,2), http://arxiv.org/abs/2502.04291, arXiv:2502.04291 [quant-ph]
11. Kim, K., Kim, M., Park, J., Byun, A. & Ahn, J. Quantum Computing Dataset of Maximum Independent Set Problem on King's Lattice of over Hundred Rydberg Atoms. *Scientific Data*. **11**, 111 (2024,1), http://arxiv.org/abs/2311.13803, arXiv:2311.13803 [quant-ph]
12. Epperly, E., Lin, L. & Nakatsukasa, Y. A theory of quantum subspace diagonalization. (arXiv,2023,6), http://arxiv.org/abs/2110.07492, arXiv:2110.07492

13. Bonche, P., Dobaczewski, J., Flocard, H., Heenen, P. & Meyer, J. Analysis of the generator coordinate method in a study of shape isomerism in 194Hg. *Nuclear Physics A.* **510**, 466-502 (1990,4), https://www.sciencedirect.com/science/article/pii/037594749090062Q
14. Marević, P. Towards a unified description of quantum liquid and cluster states in atomic nuclei within the relativistic energy density functional framework. (Université Paris-Saclay,2018), http://www.theses.fr/2018SACLS358/document, Thèse de doctorat dirigée par Khan, Elias Structure et réactions nucléaires Université Paris-Saclay (ComUE) 2018
15. Childs, A. & Wiebe, N. Hamiltonian Simulation Using Linear Combinations of Unitary Operations. *Quantum Information And Computation.* **12** (2012,2), http://arxiv.org/abs/1202.5822, arXiv:1202.5822 [quant-ph]
16. Jamet, F., Agarwal, A. & Rungger, I. Quantum subspace expansion algorithm for Green's functions. (arXiv,2022,12), http://arxiv.org/abs/2205.00094, arXiv:2205.00094
17. Umeano, C., Jamet, F., Lindoy, L., Rungger, I. & Kyriienko, O. Quantum subspace expansion approach for simulating dynamical response functions of Kitaev spin liquids. (2024), https://arxiv.org/abs/2407.04205
18. Zheng, M., Peng, B., Wiebe, N., Li, A., Yang, X. & Kowalski, K. Quantum algorithms for generator coordinate methods. (arXiv,2022,12), http://arxiv.org/abs/2212.09205, arXiv:2212.09205
19. Zheng, M., Peng, B., Li, A., Yang, X. & Kowalski, K. Unleashed from Constrained Optimization: Quantum Computing for Quantum Chemistry Employing Generator Coordinate Method. (arXiv,2024,8), http://arxiv.org/abs/2312.07691, arXiv:2312.07691
20. Beaujeault-Taudiere, Y. & Lacroix, D. Solving the Lipkin model using quantum computers with two qubits only with a hybrid quantum-classical technique based on the Generator Coordinate Method. (arXiv,2023,12), http://arxiv.org/abs/2312.04703, arXiv:2312.04703
21. Beaujeault-Taudière, Y. Systematic improvement of the quantum approximate optimisation ansatz for combinatorial optimisation using quantum subspace expansion. (2025), https://arxiv.org/abs/2506.18594
22. Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M., Alonso-Linaje, G., AkashNarayanan, B., Asadi, A., Arrazola, J., Azad, U., Banning, S., Blank, C., Bromley, T., Cordier, B., Ceroni, J., Delgado, A., Matteo, O., Dusko, A., Garg, T., Guala, D., Hayes, A., Hill, R., Ijaz, A., Isacsson, T., Ittah, D., Jahangiri, S., Jain, P., Jiang, E., Khandelwal, A., Kottmann, K., Lang, R., Lee, C., Loke, T., Lowe, A., McKiernan, K., Meyer, J., Montañez-Barrera, J., Moyard, R., Niu, Z., O'Riordan, L., Oud, S., Panigrahi, A., Park, C., Polatajko, D., Quesada, N., Roberts, C., Sá, N., Schoch, I., Shi, B., Shu, S., Sim, S., Singh, A., Strandberg, I., Soni, J., Száva, A., Thabet, S., Vargas-Hernández, R., Vincent, T., Vitucci, N., Weber, M., Wierichs, D., Wiersema, R., Willmann, M., Wong, V., Zhang, S. & Killoran, N. PennyLane: Automatic differentiation of hybrid quantum-classical computations. (2022), https://arxiv.org/abs/1811.04968

The Current Landscape of Quantum Hardware Development - An Overview

Siddharth R. Chander[1 [0000−0001−9391−807X]]

California Institute of Technology, Pasadena CA 91125, USA
schander@caltech.edu

Abstract. Quantum computing has developed significantly since the 1980s, with substantial progress in both theoretical foundations and practical implementations. This paper provides a comprehensive analysis of the current state of quantum hardware development, categorizing platforms by commercial maturity and research focus. We examine superconducting quantum processors (currently leading commercial deployment), trapped ion systems (notable for high fidelity), photonic approaches (offering room-temperature operation), and neutral atom arrays (showing recent rapid advancement). Additionally, we explore emerging platforms including semiconductor spin qubits. For each approach, we discuss operating principles, current performance metrics, scalability challenges, and potential applications. The analysis reveals that while significant challenges remain in decoherence, error correction, and scalability, quantum hardware continues to advance rapidly toward practical implementation across multiple technological pathways.

Keywords: Quantum Computing · Quantum Hardware · Qubit Technologies · Quantum Processors

1 Introduction

Quantum computing represents a paradigm shift from classical computation, leveraging quantum mechanical phenomena like superposition and entanglement to solve problems intractable for classical systems. Since Peter Shor's 1994 algorithm demonstrated quantum computing's potential to break widely-used encryption, the field has accelerated dramatically.

The development of robust quantum hardware forms the critical foundation for realizing quantum computing's transformative potential. This paper examines the current landscape of quantum processor technologies, analyzing their relative strengths, limitations, and application domains. Understanding these hardware platforms is essential for assessing the field's progress and future trajectory.

2 Superconducting Quantum Processors

Superconducting quantum processors utilize Josephson junctions to create artificial atoms serving as qubits, operating at cryogenic temperatures (10-15

mK) where superconductivity eliminates electrical resistance. Qubit states correspond to different energy levels manipulated via microwave pulses, leveraging the unique property of superconductors where resistivity drops to near zero at optimal temperatures [3].

2.1 Operating Principles

The system is quantized from classical to quantum mechanics using amplitude complexity, modifying wave phases through superconductors. Processor initialization involves fine-tuning electrical elements by classically adjusting capacitance and inductance, then modifying the total energy following quantum mechanical laws including Kirchoff's circuit laws and quantum Hamiltonian properties [4].

Fabrication employs lithography to etch metals and pattern materials for precise junctions and components. These processors uniquely use Josephson Junctions where two thin superconductive wires separated by a few-atom-thick insulator create superconducting currents that activate qubits [5].

2.2 Current State and Challenges

Operational control involves sending specific-frequency microwave pulses to each qubit for designated times to apply unitary operators. Current systems face significant challenges including decoherence from environmental interactions requiring cooling to near absolute zero, crosstalk between qubits and control lines introducing noise, fabrication difficulties in creating high-quality low-loss circuits, control complexity in signal generation and error correction, and interfacing issues with classical systems requiring efficient signal conversion between microwave, optical, and electrical domains [7, **?**,**?**].

Note: Schematic of an IBM Superconducting Quantum Processor. Note that the processor is only the bottommost component of this figure, with the rest being cooling and interactive equipment [6].

Despite these challenges, superconducting processors represent the most mature quantum technology with commercial systems available through cloud platforms. Notable implementations include Google's Sycamore demonstrating quantum advantage, IBM's Q System One, Rigetti's Aspen, and Intel's Horse Ridge. These systems show significant potential in cryptography, optimization, machine learning, and physics applications. Various superconducting qubit types exist including fixed-frequency and tunable qubits, with the latter allowing individual qubit frequency adjustment [10].

3 Trapped Ion Quantum Processors

Trapped ion quantum processors confine atomic ions (typically Yb+ or Ca+) using electromagnetic fields, utilizing their electronic states as qubits. These systems currently lead in gate fidelities and coherence times, making them strong contenders for fault-tolerant quantum computation.

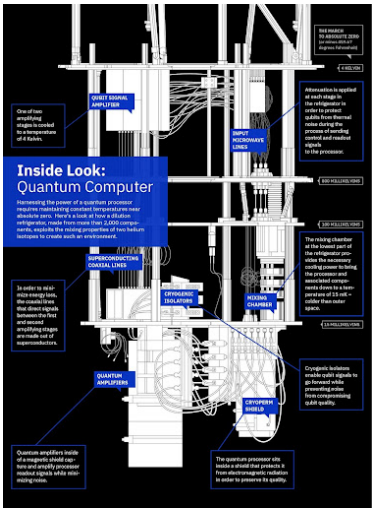


Fig. 1. Inside Look of Superconducting Quantum Processor

3.1 System Operation

The process begins with trapping and cooling ions using electromagnetic fields to create confining potential wells, typically arranged linearly. Laser cooling techniques reduce kinetic energy, bringing ions to their ground state for enhanced qubit stability against external noise [11, **?**].

State initialization prepares qubits in well-defined states, typically the ground state, accomplished through laser pulses or microwave radiation that excite or de-excite ions to specific energy levels. Quantum operations employ carefully engineered laser beams coupling different internal energy levels, with precise control achieved by manipulating laser intensity, frequency, and duration.

3.2 Entanglement and Error Correction

Entanglement generation leverages strong inter-ion interactions through designed laser pulses that create correlated systems where individual ion states cannot be described independently, enabling complex quantum computations [12].

Measurement extracts information by mapping quantum states to classical states through separate laser beams that ionize atoms and produce detectable electrical currents. Quantum error correction combats decoherence through techniques like encoding multiple qubits into single logical qubits and implementing error-detecting codes, enabling fault-tolerant computing.

Recent progress includes IonQ's 2021 commercial trapped ion computer availability, newer systems employing surface code error protection, and Tsinghua University's 2023 programmable quantum phononic processor combining trapped ion and photonic approaches that offers improved scalability for complex problems [14].

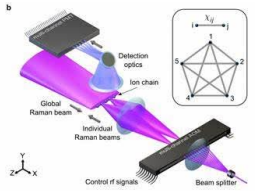


Fig. 2. Diagram of how a generic trapped ion quantum processor works, not considering how one interacts and implements programs on the processor [13].

4 Photonic Quantum Processors

Photonic quantum processors utilize photons as quantum information carriers, encoding qubits in optical properties such as polarization, path encoding, or time bins. These systems leverage quantum interference and measurement-induced nonlinearities to perform quantum computations.

4.1 Quantum Principles and Advantages

Photonic processors work by initializing photons in superposition states then controlling them using optical elements that flip states or create entanglement between photons. Calculations employ quantum algorithms operating on photons, with results read out through optical measurement elements. Since photons cannot interact in vacuum, additional media facilitate photon interactions within the processor architecture [15].

Key advantages include light-speed operation enabling faster calculations than classical computers, inherent scalability requiring less effort for multi-qubit setups compared to other platforms, room-temperature operation eliminating cryogenic requirements, natural networking capability through fiber optic compatibility, and low decoherence from weak environmental interactions [16].

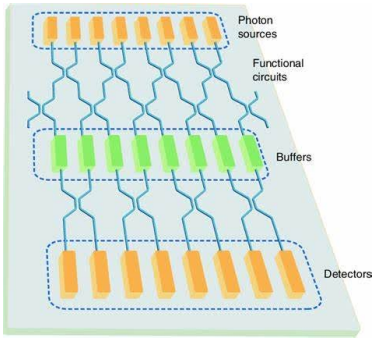


Fig. 3. Overview of a Photonic Quantum Processor

Note: Simplified diagram of how a photonic quantum processor works. Note the photon emitters, circuits, and the detectors [15].

4.2 Limitations and Recent Progress

Significant limitations include susceptibility to noise causing quantum property loss and calculation errors, extreme manufacturing complexity preventing mass production, and high fabrication costs. Despite these challenges, photonic processors have advanced beyond pure research to commercial applications.

Recent progress includes Xanadu's Borealis processor demonstrating quantum advantage using 216 squeezed-light qubits and Gaussian boson sampling, development of new optical elements enabling more complex calculations, improved manufacturing methods creating more reliable and affordable systems, and Chinese research teams demonstrating photonic systems solving specialized problems $10^14 times faster than classical supercomputers$[16].

Primary applications include quantum simulation of molecular vibrations and bosonic systems, Gaussian boson sampling for graph problems and machine learning, quantum networking and quantum key distribution, and specialized optimization problems.

5 Neutral Atom Quantum Processors

Neutral atom quantum processors utilize arrays of individual atoms (typically rubidium or cesium) trapped in optical tweezers and cooled to microkelvin temperatures. These systems encode qubits in atomic energy levels and leverage Rydberg interactions to create entanglement between atoms, offering unique advantages in scalability and connectivity [17].

5.1 Technical Implementation

The first step involves laser cooling atoms to microkelvin temperatures using repeated photon absorption and emission to reduce kinetic energy. Once cooled, atoms are trapped using optical tweezers or magnetic fields that create confining potential wells at standing wave nodes where potential energy is minimized [17].

Qubits are typically encoded in the spin states of atoms, where up and down states represent classical 0 and 1 values, with lasers manipulating these states through precise rotations. Multiple atoms can be arranged into reconfigurable arrays to represent multiple qubits. Quantum operations include rotations for state manipulation, entanglement through Rydberg blockade effects that create correlated quantum states, and measurement operations that collapse quantum states to classical values [17].

Note: Schematic of how a generic neutral atom quantum processor works [18].

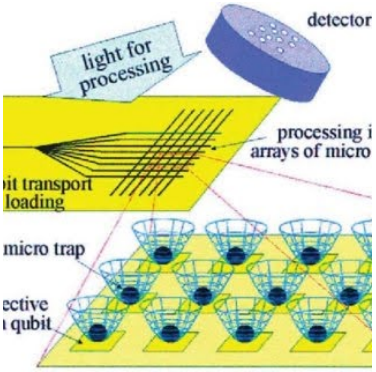


Fig. 4. Neutral Atom Processor Diagram

5.2 Recent Advances and Challenges

Recent advances include QuEra's 256-qubit Aquila processor (2022) and 280-qubit system (2023), PASQAL's 200-qubit quantum processing unit, and Atom Computing's 100-qubit system with nuclear spin qubits. These systems have demonstrated quantum error correction, high-fidelity two-qubit gates (99.5%) using Rydberg interactions, and dynamic qubit rearrangement during computation [19].

Despite progress, significant challenges remain. Decoherence presents particular difficulties as neutral atoms are highly sensitive to environmental interactions, making it challenging to maintain coherent states for extended periods [19]. Precise control of individual atoms remains difficult, and scaling to larger qubit arrays is constrained by decreasing coupling strength with distance. Additionally, these systems require specialized equipment and facilities, making them expensive to build and operate [20].

The future outlook for neutral atom quantum processors remains promising due to their exceptional scalability, long coherence times, and flexible qubit arrangements. With continued research addressing control precision, error reduction, and system integration, neutral atom systems are progressing toward 1000+ qubit demonstrations.

6 Semiconductor Spin Qubits

Semiconductor spin qubits leverage the quantum properties of electron or hole spins confined in quantum dots within semiconductor structures, primarily silicon. These qubits benefit from compatibility with existing semiconductor manufacturing infrastructure, offering a promising path toward scalable quantum computing [21].

6.1 Technical Progress

Significant progress has been made with researchers achieving 99.9% single-qubit gate fidelities in isotopically enriched silicon-28, 99.8% two-qubit gate fidelities demonstrating entanglement, coherence times exceeding 10 seconds through dynamical decoupling techniques, and operation temperatures up to 1.5 K, reducing cooling requirements. Intel's 2023 demonstration of a 12-qubit array with all-to-all connectivity represents the largest silicon spin qubit system to date, with their Horse Ridge cryogenic controller enabling scalable control of qubit arrays [21].

Spin qubits are implemented using electron spins in quantum dots defined by electrostatic gates, hole spins in germanium-silicon heterostructures, or donor atoms (phosphorus) implanted in silicon substrates. Qubit control is achieved through ESR techniques for microwave manipulation of electron spins, electric dipole spin resonance for electric field control of hole spins, and exchange coupling for mediating two-qubit operations through controlled interactions [21].

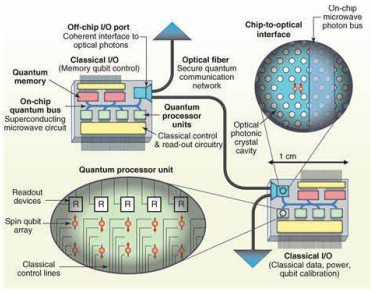


Fig. 5. User Interface with Silicon Spin Quantum Processor

Note: How one can work and use a silicon spin quantum processor [22].

6.2 Advantages and Future Outlook

The primary advantage lies in CMOS fabrication compatibility, enabling potential integration with classical control electronics. Recent developments include 3D integration of qubits and control circuitry, cryogenic CMOS controllers operating at 4 K, and quantum dot arrays with pitch sizes compatible with advanced lithography [23].

While spin qubits show exceptional coherence properties and fabrication compatibility, challenges remain in maintaining uniformity across large qubit arrays, scaling control electronics for thousands of qubits, improving readout fidelity and speed, and implementing error correction in dense arrays. The field is progressing rapidly toward 50-100 qubit demonstrations within the next 2-3 years, with several companies and research institutions pursuing different architectural approaches [23].

7 Conclusion

Quantum computing represents a rapidly developing field with transformative potential across numerous industries, with quantum hardware development serving as the critical foundation for realizing this potential. This review has comprehensively examined the current landscape of quantum processors, including superconducting, trapped ion, photonic, neutral atom, and semiconductor spin systems.

Each platform demonstrates unique advantages and limitations: superconducting processors represent the most mature technology with demonstrated quantum advantage; trapped ion systems offer exceptional coherence times and high fidelities; photonic approaches provide rapid operation and scalability but face noise susceptibility; neutral atom arrays show promising scalability and long coherence times; semiconductor spin qubits benefit from classical electronics compatibility and mass-production potential. Other platforms such as topological systems promise inherent noise resistance though remain in early development; NMR processors offer operational simplicity despite scalability constraints; and diamond-based systems provide remarkable stability with room-temperature operation despite fabrication challenges.

The future of quantum computing appears increasingly promising as technology continues advancing toward more powerful and efficient quantum processors capable of solving problems intractable for classical computers. Several critical challenges must be addressed to achieve full potential: decoherence remains a fundamental obstacle as qubits maintain fragility to environmental interactions across all platforms; error correction presents complex implementation hurdles despite being essential for scalable quantum computation; scalability demands substantial advancement to achieve practical qubit counts for meaningful applications; and cost reduction is necessary to overcome current economic barriers to widespread adoption.

Despite these challenges, considerable optimism persists within the research community, with continued development expected to yield solutions in the coming years. Quantum computing's potential to revolutionize numerous fields makes this an exceptionally promising and exciting area of technological development.

Acknowledgments. This paper was done individually by the author, Siddharth Chander, without any guidance or association with anyone else.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Shor, P. W.: Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th Annual Symposium on Foundations of Computer Science, pp. 124–134. IEEE (1994)

2. Arute, F. et al.: Quantum supremacy using a programmable superconducting processor. Nature 574(7779), 505–510 (2019)
3. Krantz, P. et al.: A quantum engineer's guide to superconducting qubits. Applied Physics Reviews 6(2), 021318 (2019)
4. Oliver, W. D., Welander, P. B.: Materials in superconducting quantum bits. MRS Bulletin 38(10), 816–825 (2013)
5. Clarke, J., Wilhelm, F. K.: Superconducting quantum bits. Nature 453(7198), 1031–1042 (2008)
6. IBM Quantum: IBM Quantum System One. Retrieved from https://www.ibm.com/quantum/system-one (2023)
7. Devoret, M. H., Schoelkopf, R. J.: Superconducting circuits for quantum information: an outlook. Science 339(6124), 1169–1174 (2013)
8. Gambetta, J. M. et al.: Building logical qubits in a superconducting quantum computing system. npj Quantum Information 3(1), 1–7 (2017)
9. Wendin, G.: Quantum information processing with superconducting circuits: a review. Reports on Progress in Physics 80(10), 106001 (2017)
10. Koch, J. et al.: Charge-insensitive qubit design derived from the Cooper pair box. Physical Review A 76(4), 042319 (2007)
11. Cirac, J. I., Zoller, P.: Quantum computations with cold trapped ions. Physical Review Letters 74(20), 4091 (1995)
12. Monroe, C., Kim, J.: Scaling the ion trap quantum processor. Science 339(6124), 1164–1169 (2013)
13. IonQ: How IonQ Trapped Ion Quantum Computers Work. Retrieved from https://ionq.com/technology (2023)
14. Zhang, J. et al.: A programmable quantum phononic processor based on trapped ions. Nature Physics 19(3), 345–350 (2023)
15. Xanadu: Photonic Quantum Computing. Retrieved from https://www.xanadu.ai/technology (2023)
16. Madsen, L. S. et al.: Quantum computational advantage with a programmable photonic processor. Nature 606(7912), 75–81 (2022)
17. Saffman, M. et al.: Quantum information with Rydberg atoms. Reviews of Modern Physics 82(3), 2313 (2010)
18. QuEra Computing: Neutral Atom Quantum Computers. Retrieved from https://www.quera.com/technology (2023)
19. Browaeys, A., Lahaye, T.: Many-body physics with individually controlled Rydberg atoms. Nature Physics 16(2), 132–142 (2020)
20. Bernien, H. et al.: Probing many-body dynamics on a 51-atom quantum simulator. Nature 551(7682), 579–584 (2017)
21. Veldhorst, M. et al.: An addressable quantum dot qubit with fault-tolerant control fidelity. Nature Nanotechnology 10(7), 589–592 (2015)
22. Silicon Quantum Computing: Our Technology. Retrieved from https://www.siliconquantum.com.au/our-technology (2023)
23. Zwanenburg, F. A. et al.: Silicon quantum electronics. Reviews of Modern Physics 85(3), 961–1019 (2013)
24. Nayak, C. et al.: Non-Abelian anyons and topological quantum computation. Reviews of Modern Physics 80(3), 1083–1159 (2008)
25. Kitaev, A. Y.: Fault-tolerant quantum computation by anyons. Annals of Physics 303(1), 2–30 (2003)
26. Google AI Quantum Team: Observation of Majorana fermions in a superconducting InAs nanowire. Nature 556(7700), 74–79 (2018)

27. Satzinger, K. J. et al.: Realizing topologically protected qubits in a superconducting circuit. Nature 595(7868), 508–513 (2021)
28. Microsoft Quantum: Topological Qubits. Retrieved from https://www.microsoft.com/en-us/quantum/topological-qubits (2023)
29. Vandersypen, L. M., Chuang, I. L.: NMR techniques for quantum control and computation. Reviews of Modern Physics 76(4), 1037–1069 (2005)
30. Jones, J. A.: Quantum computing with NMR. Progress in Nuclear Magnetic Resonance Spectroscopy 59(2), 91–120 (2011)
31. Vandersypen, L. M. et al.: Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. Nature 414(6866), 883–887 (2001)
32. Laflamme, R. et al.: NMR quantum information processing. Quantum Information & Computation 2(3), 166–176 (2002)
33. Biamonte, J. et al.: Quantum machine learning. Nature 549(7671), 195–202 (2018)
34. Doherty, M. W. et al.: The nitrogen-vacancy colour centre in diamond. Physics Reports 528(1), 1–45 (2013)
35. Quantum Diamond Technologies: NV Center Diamond Quantum Sensors. Retrieved from https://www.quantumdiamondtech.com/technology (2023)

Coherence Development at IQM Quantum Computers

Matthew Steggles, Arthur Rebello, Ivan Takmakov, Andrew Guthrie, Lihuang Zhu, Azad Karis, Alejandro Gómez-Frieiro, Tuomas Mylläri, Eelis Takala, Leonid Abdurakhimov, and Juha Hassel

IQM Quantum Computers, Keilaranta 19, 02150 Espoo, Finland
matthew.steggles@meetiqm.com

Abstract. We showcase recent developments in improving coherence times of superconducting qubits at IQM Quantum Computers. We demonstrate how optimized fabrication and design result in increases in coherence times, enabling enhanced performance when integrated into quantum processing units (QPUs). We also report current efforts to mitigate the detrimental effects on coherence of remaining two-level-system (TLS) defects. Our results showcase state-of-the-art fabrication capabilities at IQM Quantum Computers, exhibiting T_1 and T_2 echo times up to the millisecond level in test devices. These improvements, when applied to QPU-compliant designs and combined with IQM's quantum computer control software stack, yield consistently improved performance, enabling increased gate fidelities and paving the way to building fault-tolerant quantum computers.

Keywords: Superconducting Quantum Computing · Fabrication · Design · Coherence · TLS defects

1 Introduction

Coherence times are one of the most foundational performance indicators of quantum processing units (QPUs). Long energy relaxation and coherence times (hereafter 'coherences') enable high fidelities of single-qubit gates, two-qubit gates, and readout [1]. Nevertheless, there are numerous decay channels that substantially limit the relaxation times that must be dealt with to increase QPU performance. The coherences of superconducting qubit platforms are frequently limited by dielectric losses via so-called two-level-system (TLS) defects, a broad class of microscopic defect states that couple to the qubit electric field via their dipole moment [2–4].

While defect states are always present in any nanofabrication process, there are several strategies which can be employed to decrease their effect. By improving fabrication, design, and parking strategies we present here significant improvements in the lifetimes of transmon qubits, showing the benefits of an integrated approach to coherence development.

2 Improvements from fabrication and design

Here we discuss how qubit coherence can be improved by decreasing the impact of the TLS bath. In the section 2.1 we demonstrate high coherence by improving fabrication quality, which we attribute to decreasing the density of TLSs. Then, in the section 2.2 we optimize the qubit design to be less sensitive to TLSs, which we attribute to shifting the distribution of coupling strengths towards weaker qubit-TLS coupling.

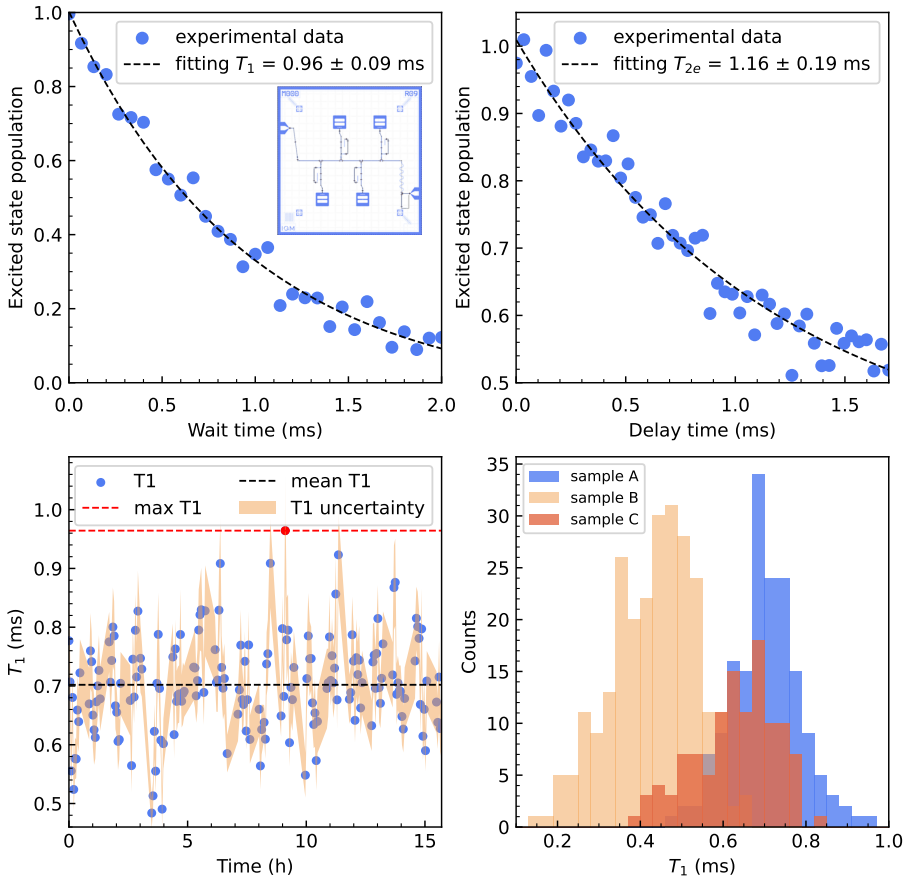


Fig. 1. Performance of isolated double-pad test qubits. (a) Best qubit T_1 trace. (b) Best qubit T_2 echo trace. (c) T_1 temporal fluctuations. (d) Histograms of T_1 distributions for best qubits from three different samples.

2.1 Fabrication improvements

The most straightforward way to improve the relaxation times of a qubit is to improve the quality of the fabrication. We optimized multiple steps in the fabrication process for niobium-on-silicon isolated double-pad qubits, including silicon substrate cleaning, base-layer Nb film deposition, and Al/AlOx/Al Josephson junction double-angle evaporation. Fabrication was performed in IQM's internal clean room facilities.

The double-pad qubit is used as a reference design in order to compare more directly with the literature, as there is a significant wealth of papers showing high coherence with double-pad designs [7–10]. We note that millisecond coherence times have been obtained with different materials, such as tantalum [10], but keep the materials used constant to more effectively benchmark the fabrication process.

In figure 1 we show results from 3 test devices with the highest maximum T_1, highlighting the maximum T_1 time of 0.96 ± 0.09 ms and maximum T_2 echo time of 1.16 ± 0.19 ms. The fluctuation data presented also show mean performance $T_{1,\,\mathrm{mean}} > 0.7$ ms. Our maximum T_1 compares favourably with the literature: table 2.1 compares this work to various high-coherence transmon qubits, all fabricated with the same materials.

	Max T_1	Median T_1	Qubit Frequency	Max Q-factor
This work, IQM [6]	0.96 ms	0.70 ms	3.6 GHz	22 M
R.T. Gordon et al [7]	0.6 ms	0.40 ms	3.7 GHz	14 M
S. Kono et al [8]	0.45 ms	N/A	4.8 GHz	14 M
M. Tuokkola et al [9]	0.67 ms	0.50 ms	2.9 GHz	12 M

Table 1. Comparison of highest niobium-on silicon transmon T_1 data published in academic literature. Maximum and median T_1 values are included where available.

We close this section by noting that our maximum Q-factor of 22M is higher than for any niobium-on-silicon double-pad transmon published in the academic literature. We acknowledge that some transmons made of other materials have demonstrated Q-factors up to 25M [10] with tantalum-on-silicon. In that work, they conclude that they are limited by substrate losses. The conclusion that our maximum Q-factor is also near the substrate limit is supported by energy participation ratio simulations.

We are also aware of reports of higher T_1 [11] but due to the materials platforms and fabrication details not being disclosed in those sources we cannot make direct comparison.

2.2 Design improvements

Another avenue for improving coherence is through optimizing the distribution of the qubit electric field through design. The qubit and its immediate environment

consist of several components made of materials with differing loss factors [5]. By optimizing the energy participation ratio (EPR) of the qubit electric field in each of these regions we reduce the coupling of the qubit to the TLSs at lossy interfaces. The size of the interaction between a TLS and the qubit scales like $\boldsymbol{p}_{\mathrm{TLS}} \cdot \boldsymbol{E}_{\mathrm{qb}}$ [3,4], where $\boldsymbol{p}_{\mathrm{TLS}}$ is the TLS dipole moment and $\boldsymbol{E}_{\mathrm{qb}}$ the qubit electric field. Hence, reducing the field strength in lossy regions where the TLS density is high reduces the total coupling to the TLS bath. We have already seen this at work in the double-pad qubits; there the electric field is distributed largely away from the lossy interfaces, explaining the high coherence. However, there are challenges associated with scaling qubit counts with double pad qubits due to their large size, so instead we optimize the design of an IQM QPU-compliant qubit.

In figure 2 we show data comparing the T_1 performances of isolated QPU-compliant qubits before and after EPR optimization. The data indicate an improvement in mean T_1 of 19% following the geometry changes.

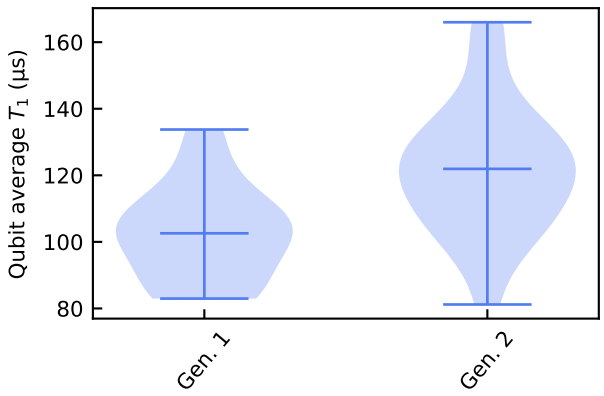


Fig. 2. QPU-compliant test device performances, demonstrating improved average performance. (a) A violin plot of mean T_1 values across 13 first generation qubits. (b) A violin plot of mean T_1 values across 22 second generation qubits.

The main benefits of this enhanced coherence are seen readily once this qubit design is incorporated into a QPU-compliant test device. When combined with advanced gate calibration techniques this qubit design has allowed for two qubit gate fidelities of over 99.9% fidelity [6, 15].

3 Mitigating the effects of TLS defects

Even with improvements to fabrication quality and design, one cannot completely remove the effects of TLSs. With this in mind, we have studied different

strategies to dynamically reduce the impact of near-resonant TLSs on the performance of the qubits.

3.1 Electric field tuning of TLS defects

A promising approach to mitigating the impact of TLSs on superconducting qubits involves directly manipulating the defect frequencies using an external electric field. This methodology has been explored in the literature [11, 13, 14]. However, this technique requires careful consideration of the chip architecture to ensure compatibility as several implementations require dedicated control lines to manipulate the TLS frequency.

We demonstrate the feasibility of tuning TLS defects by introducing a bias tee following reference [13] and applying DC voltage to the qubit drive line. By performing AC-Stark TLS spectroscopy [12], we observed a TLS defect frequency shifting under applied voltage, illustrated in figure 3.

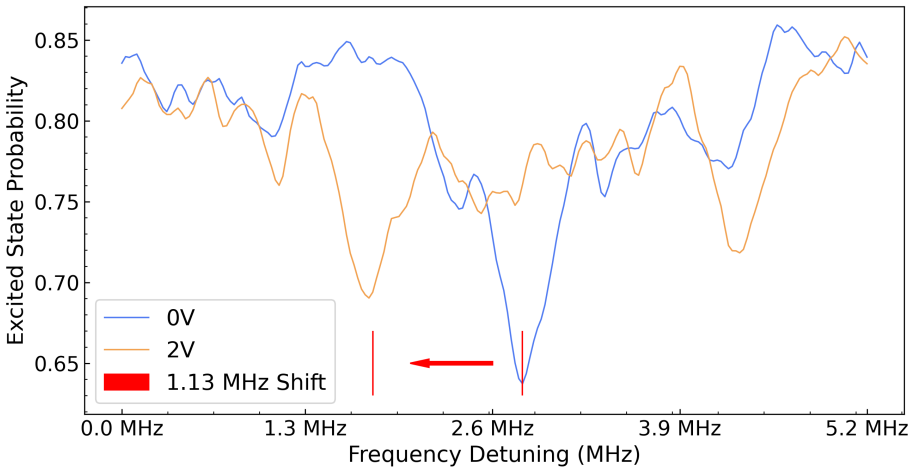


Fig. 3. Results of AC-Stark TLS spectroscopy at two different biasing voltages, showing the frequency shift of a TLS defect by over 1 MHz at 2V.

With a biasing voltage of 2V, the TLS is seen to have been tuned by 1.13 MHz. However, the majority of TLSs are not tuned. An electromagnetics simulation demonstrates that the qubit geometry shields the surroundings, so only a small minority of TLSs are exposed to the biasing field. This highlights that trade-offs must be made; optimizing the design for resilience to TLS and for high coupling means TLS tuning is not viable at present in our devices, although other groups with different designs have employed it to good effect [13].

3.2 TLS defect avoidance

Alternatively, a different approach that does not require hardware modifications is based on qubit frequency parking optimization. Possible parking strategies based on particular types of optimizers were reported in [16, 17]. We are investigating other optimization frameworks, and, in figure 4, we share results of optimizing coherence times with such an approach, improving T_1 by 14% on a 45 qubit square topology QPU. By utilizing advanced pulse shaping methodologies [18] and crosstalk avoidance [19], we can achieve fast high-fidelity gates on QPUs which are mostly coherence-limited, so optimizing coherence within this framework can directly improve QPU performance.

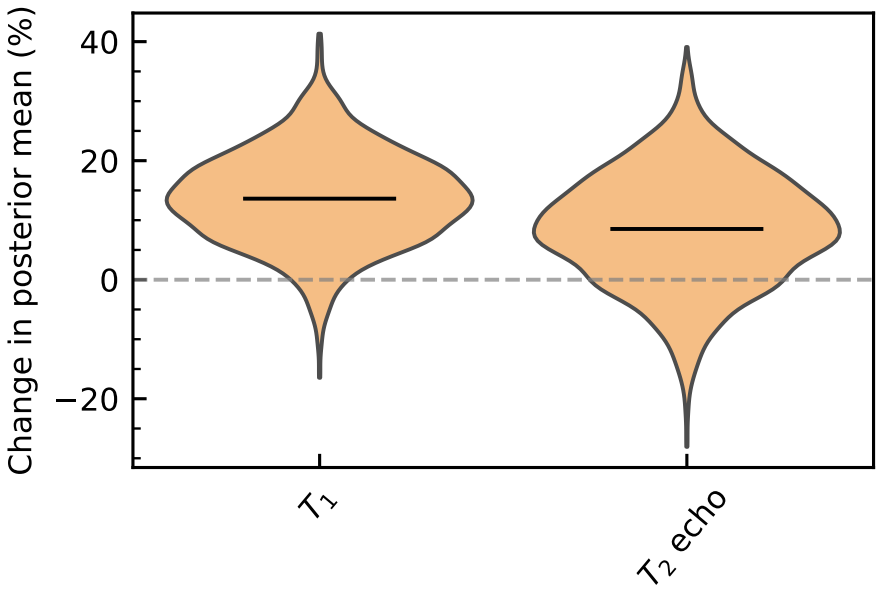


Fig. 4. Posterior distributions for the change in mean T_1 and T_2 echo on a 45 qubit QPU. The maximum *a posteriori* (MAP) estimate for the increase in coherence is 14%.

4 Outlook

In this paper we have summarized several strategies for improving coherence times at IQM Quantum Computers, demonstrating the role state-of-the-art fabrication and integrated approaches to design and software play in improving QPU performance. We are currently implementing these changes in larger scale QPUs to reap similar benefits and adapting what we have learned from TLS avoidance into our calibration methodologies.

Acknowledgments. We would like to acknowledge Marco Marin Suarez, Sampo Saarinen, Snigdha Kumar for measurement support; Lan-Husan (Lance) Lee, Seung-Goo Kim for sample fabrication; Tianyi Li, Liuqi Yu, Jeroen Verjauw, Vladimir Milchakov, Jean-Luc Orgiazzi for technical discussions.

This research has been partially supported by the Horizon Europe program through project OpenSuperQ+ of the EU Flagship on Quantum Technologies.

Disclosure of Interests. The authors are employed by IQM Quantum Computers.

References

1. Abad, T., Fernández-Pendás, J., Frisk Kockum, A., Johansson, G.; Universal Fidelity Reduction of Quantum Operations from Weak Dissipation, Phys. Rev. Lett. 129, 150504 (2022)
2. Simmonds, R. W., Lang, K. M., Hite, D. A., Nam, S., Pappas, D. P., Martinis, J. M.; Decoherence in Josephson phase qubits from junction resonators, Phys. Rev. Lett. 93, 077003 (2004).
3. Martinis, J. M., Cooper, K. B., McDermott, R., Steffen, M., Ansmann, M., Osborn, K. D., Cicak, K., Oh, S., Pappas, D. P., Simmonds, R. W., Yu, C. C.; Decoherence in Josephson qubits from dielectric loss, Phys. Rev. Lett. 95, 210503 (2005).
4. Müller, C., Cole, J. H., Lisenfeld, J.; Towards understanding two-level-systems in amorphous solids: insights from quantum circuits, Reports on Progress in Physics 82, 124501 (2019).
5. Wang, C., Axline, C., Gao, Y. Y., Brecht, T., Chu, Y., Frunzio, L., Devoret, M. H., Schoelkopf, R. J.; Surface participation and dielectric loss in superconducting qubits, Appl. Phys. Lett. 107, 162601 (2015).
6. IQM Quantum Computers Press Release, https://meetiqm.com/press-releases/iqm-quantum-computers-achieves-new-technology-milestones-with-99-9-2-qubit-gate-fidelity-and-1-millisecond-coherence-time/, (Published on 2024-07-15, URL Date Accessed 2025-08-25).
7. Gordon, R. T., Murray, C. E., Kurter, C., Sandberg, M., Hall, S.A., Balakrishnan, K., Shelby, R., Wacaser, B., Stabile, A. A., Sleight, J. W., Brink, M., Rothwell, M. B., Rodbell, K. P., Dial, O., Steffen, M.; Environmental radiation impact on lifetimes and quasiparticle tunneling rates of fixed-frequency transmon qubits, Appl. Phys. Lett. 120, 074002 (2022).
8. Kono, S., Pan, J., Chegnizadeh, M., Wang, X., Youssefi, A., Scigliuzzo, M., Kippenberg, T.J.; Mechanically induced correlated errors on superconducting qubits with relaxation times exceeding 0.4 ms, Nat Commun 15, 3950 (2024).
9. Tuokkola, M., Sunada, Y., Kivijärvi, H., Albanese, J., Grönberg, L., Kaikkonen, J.-P., Vesterinen, V., Govenius, J., Möttönen, M.; Methods to achieve near-millisecond energy relaxation and dephasing times for a superconducting transmon qubit, Nat Commun 16, 5421 (2025).
10. Bland, M. P., Bahrami, F., Martinez, J. G. C., Prestegaard, P. H., Smitham, B. M., Joshi, A., Hedrick, E., Pakpour-Tabrizi, A., Kumar, S., Jindal, A., Chang, R. D., Yang, A., Cheng, G., Yao, N., Cava, R. J., de Leon, N. P., Houck, A. A.; 2D transmons with lifetimes and coherence times exceeding 1 millisecond, arXiv:2503.14798 (2025).
11. Dane, A., Balakrishnan, K., Wacaser, B., Hung, L.-W., Mamin, H. J., Rugar, D., Shelby, R. M., Murray, C., Rodbell, K., Sleight, J.; Performance Stabilization of High-Coherence Superconducting Qubits, arXiv:2503.12514 (2025).

12. Carroll, M., Rosenblatt, S., Jurcevic, P., Lauer, I., Kandala, A.; Dynamics of superconducting qubit relaxation times, npj Quantum Information 8, 132 (2022).
13. Chen, L., Lee, K.-H., Liu, C.-H., Marinelli, B., Naik, R. K., Kang, Z., Goss, N., Kim, H., Santiago, D. I., Siddiqi, I.; Scalable and Site-Specific Frequency Tuning of Two-Level System Defects in Superconducting Qubit Arrays, arXiv:2503.04702 (2025).
14. Lisenfeld, J., Bilmes, A., Megrant, A., Barends, R., Kelly, J., Klimov, P., Weiss, G., Martinis, J. M., Ustinov, A. V.; Electric field spectroscopy of material defects in transmon qubits, npj Quantum Information 5, 105 (2019).
15. Marxer, F. et al; Above 99.9% Fidelity Single-Qubit Gates, Two-Qubit Gates, and Readout in a Single Superconducting Quantum Device, arXiv:2508.16437 (2025).
16. Google Quantum AI and Collaborators; Quantum error correction below the surface code threshold, Nature 638, 920 (2025).
17. Jiang, T. et al; Generation of 95-qubit genuine entanglement and verification of symmetry-protected topological phases, arXiv:2505.01978 (2025).
18. Hyyppä, E., Vepsäläinen, A., Papić, M., Chan, C. F., Inel, S., Landra, A., Liu, W., Luus, J., Marxer, F., Ockeloen-Korppi, C., Orbell, S., Tarasinski, B., Heinsoo, J.; Reducing Leakage of Single-Qubit Gates for Superconducting Quantum Processors Using Analytical Control Pulse Envelopes, PRX Quantum **5**, 030353 (2024).
19. Wesdorp, J. et al; In preparation (2025).

ST Microelectronics/ Quobly path for Spin based large scale quantum core

Tristan Meunier[1], Nicolas Daval[1], Maud Vinet[1], Franck Arnaud[2], Jean-Charles Barbé[1,3]

[1] Quobly, Grenoble, France
[2] STMicroelectronics, Crolles, France
[3] CEA-Leti, Grenoble, France
tristan.meunier@quobly.io

Abstract.
Recent advancements in qubit manipulation in quantum dot arrays, as well as in the classical/quantum co-integration of FD-SOI spin-based quantum circuits shed light on an industrialization path for large-scale quantum computing. We present this path to designing and engineering good qubits using technology that is as close as possible to the most advanced industrial FD-SOI nodes. We then investigate qubit design accounting for the constraints arising from the established industrial fabrication process. More precisely, we repurpose the W vias and, in a single contact patterning step, define the gates that enable to define the electrochemical potential of quantum dots (QDs), as well as the W vias that control the coupling barriers between adjacent QDs. We present simulation-based and experimental results on the individual coupling control of QDs in arrays fabricated on the industrial 28nm FD-SOI technology. We present detailed wafer-level transfer characteristics of each barrier implemented on a 1x3 linear array at room temperature and at 2K. These results demonstrate that the vias behave like MOSFET gates, providing effective electrostatic control over the silicon channel. This validates the compatibility of the 28 nm FD-SOI industrial route with essential requirements for demonstrating a two-qubit gate.

Keywords: Spin qubit, semiconductor, Industrialization

1 Introduction

Silicon-based qubits are considered the most promising experimental system for scaling quantum computing. Due to the intrinsic compatibility with semiconductor industry, engineering quantum processors at large scale would benefit from the well-established (mature) semiconductor manufacturing. In particular, the ability to co-integrate qubits and transistors would be an asset for controlling large scale quantum circuits and keep their size comparable to classical processors. Among the various Si-based quantum technology platforms, FinFeET- and FD-SOI-based quantum technologies are the closest to advanced commercial CMOS technologies [1].

Demonstrating both hole and electron spin qubits [2], charge array control with tunnel coupling compatible with spin manipulation [3], and cryoelectronic circuits for qubit

control and read-out [4] has been key to validating the potential of the FD-SOI spin qubit technology. In particular, the last demonstration showed that the level of control in quantum dot arrays was sufficient to reach the last remaining milestone of the technology: the two-qubit gate in the form pulsed induced coherent exchange oscillations [5].

Precise control of the electrical parameters of a quantum dot array is essential for well controlled manipulation of qubits and, consequently, for reliable quantum computation [2]. Controlled exchange interactions can be achieved by independently tuning the tunnel coupling between adjacent QDs, which is a key requirement for implementing two qubit gates in spin-based quantum processors. Previous approaches in commercial FD-SOI linear arrays have relied on local back biasing [6] (Fig. 1 – top) or an extra front-gate [7] (Fig. 1 – Middle) to tune QD tunnel coupling. While these approaches are functional, they have the respective inconveniences of not being scalable, even in linear arrays, and of sacrificing one front-gate from two, which leads to a larger effective pitch for neighboring 2-qubit cells to have full control.

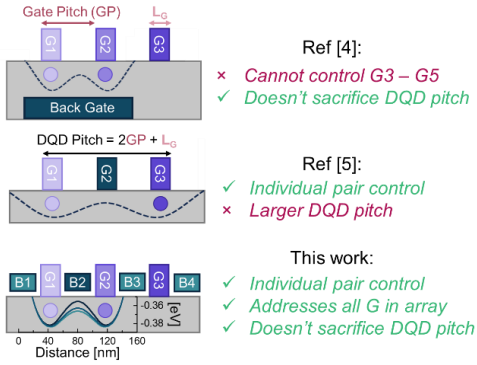


Figure 1. Cross-section schematics of our design (bottom), depicting a 1x3 front gate array (G1 – G3), leveraging W vias as barrier gates, B1 – B4, as compared to previous work on foundry fabricated devices using the 22FDX® platform. In each case, the gate used for coupling control is shown in dark blue. Simulations of our design show the barrier height control by the bias applied on B2, between G1 and G2.

Here, we review our effort to design QD arrays that can be manufactured using known industrial fabrication methods, i.e. design rules compatible. The coupling control between adjacent QDs fabricated on 28nm FD-SOI [8] by using vias to tune the barrier height is discussed. In comparison to previous designs, this integration approach allows in-situ tunnel coupling control over individual pairs of QDs in linear arrays while respecting a tighter pitch and reserving the FD-SOI back gate for tuning other important aspects of the device performance, such as mobility [9] and charge noise [10].

2 Device geometry and characterization

The devices measured were fabricated on the industrial 28nm FD-SOI technology platform. Nanowires of various widths were defined by STI etching on an SOI wafer

and were covered by a series of high-κ metal gates using standard gate-first processes with a 90 nm pitch. These front gates (G1 – G3) are used to trap charges (electrons or holes) in the Si nanowire to form QDs (Fig. 1 - Bottom). Both the nanowire and the front gates were contacted by standard W vias, but additional vias were engineered to serve as barrier gates (B1 – B4) to tune the tunnel coupling between adjacent dots, inspired by our previous work [9] based on non-commercial FD-SOI technology.

Figure 2a shows a 2D map of $I_D(V_{barrier},V_{GT})$ for a typical 1x3 qubit array (Fig. 1 – Bottom) at room temperature. The four barrier gates (vias) are swept together ($V_{barrier}$) while the front gates (G1-G3) are biased at different V_G to achieve the same V_{GT} ($V_{GT} = V_G – V_{TH,G}$). Figures 2b and 2c show how the I-V curves change with the voltages of the barrier and front gates, while Figure 2d shows how the barrier's threshold voltage depends on the back gate voltage. Figure 2e-h show the impact of capacitive coupling of the barrier gate with the front and back gates at room temperature and at 2K.

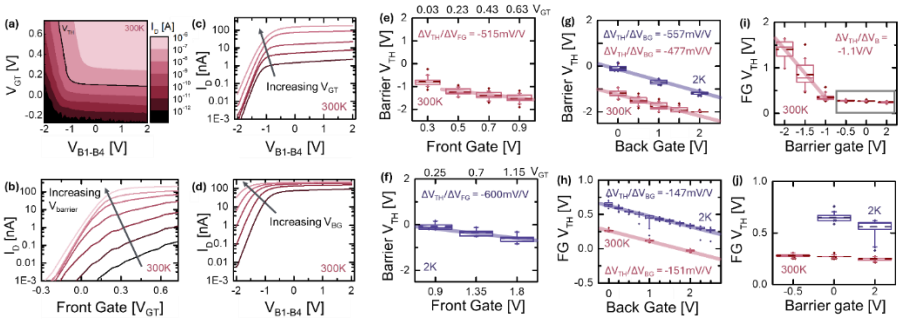


Figure 2. **(a)** Typical $I_D(V_{Br},V_{GT})$ sweeping all 4 barrier vias (B1–B4) in a 1x3 array with different front gate (G1–G3) V_{GT} values. **(b-c)** Sample I_D-V_{gate} with back gate = 0V for (b) different $V_{barrier}$ while sweeping V_{GT} and (c) different V_{GT} while sweeping $V_{barrier}$. **(d)** $I_D(V_{barrier})$ for different back gate voltages (V_{GT} = 230mV). **(e-h)** Full wafer statistics showing the impact of capacitive coupling between gates: (e-f) barrier's threshold voltage variation with V_{GT} (V_{BG} = 0V) at (e) 300K and (f) 2K; Barrier's (g, V_{GT} = 0V) and Front gate (h, $V_{barrier}$ = 0V) threshold voltage variation with V_{BG} at 300K (red) and 2K (blue). **(i-j)** front gate threshold voltage variation with $V_{barrier}$ (V_{BG} = 0V) at 300K (red) and 2K (blue). V_{DS} = 50mV for all panels. V_{TH} calculated at I_D = 1nA. Lines overlayed to guide the eye for linear extrapolation.

As barrier gate's threshold voltage is strongly affected by the operating V_{GT} (Fig. 2 e-f), to enable comparison of the individual barriers, the front gates were biased at near-threshold V_{GT} while the remaining barrier were biased at 0V. A MOS-like I_D-V_{Bi} of each barrier gate B_i was measured at room temperature on a typical die (see Fig. 3a). This was also collected at wafer level and at 300K [2K] displaying a mean threshold voltage over the 4 barrier gate $V_{TH,Bi}$ = –1.2V [–0.54V] (Fig. 3b), mean I_{ON} = 92.85 nA, and a mean subthreshold swing (SS) of 233 mV/dec (Fig. 4d), with good matching between barrier gates. At 2K, the SS was visibly smaller but, due to current oscillations induced by the formation of sub-bands at cryogenic temperatures [11], an accurate SS extraction was not possible. We observe a ~250mV interquartile range for $V_{TH,Bi}$ and 1V min-max range at both temperatures. While large, the EOT of the barrier gate is

~10-15nm, leading to an electric field strength that is ~3-12x as weak as a typical high-κ front gate (EOT~1.25-3.5nm); the range we observe is therefore comparable to the ~300-400mV range typically observed for a standard commercial front gate in a qubit arrays [8,12].

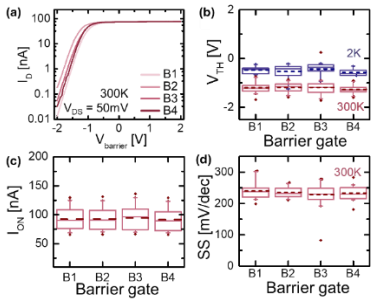


Figure 3. **(a)** Typical I_DV_B for each of the 4 barrier gate at 300K and V_{GT} = 127mV showing MOSFET-like control over the Si channel **(b-d)** Boxplots of (b) $V_{TH,Bi}$ (c) I_{ON} and (d) SS_{Bi} for each of the 4 barrier vias, at V_{GT} = 127mV [40mV] at 300K [2K] showing individual barrier gate matching at 300K (26 dies) and 2K (10 dies). V_{TH} calculated using linear extrapolation due to low voltages.

3 Quantum dot coupling control

To test the effect of the barrier on the tunnel coupling control, we first performed wafer-level screening of various array designs (Design A-C). To do so, we collect I_DV_G for each front gate at low V_{DS} = 5mV and target regular patterns of Coulomb peaks (SET behavior) which have been predicted to be evenly spaced by 10-20 mV by simulation. As shown in Fig. 4, we observe an improvement in SET behavior by modifying the layout and characteristic size of the qubit array. In Design A we observe a purely MOSFET-like behavior, which improves slightly but shows poor Coulomb blockade in Design B and finally, good electrostatic definition of the QDs in Design C. The observation of some Coulomb peaks for Design B shows an improvement in quantum confinement when compared to Design A due to a 60% reduction of the front gate size. The undesired current in Design B, (MOSFET-like superposed to the peaks,) indicates that the structure is not operating properly in the Coulomb blockade regime and that uncontrolled carriers are flowing through the device. This behavior is improved for Design C, where the Coulomb peak heights are more even and the mean peak spacing is 15mV which is in good agreement with simulations.

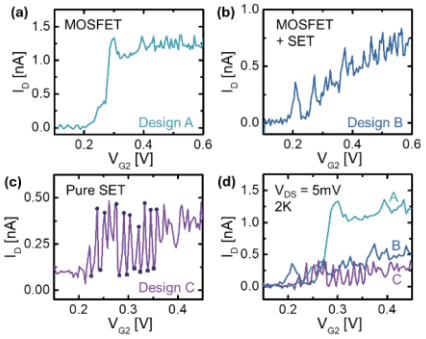


Figure 4. Typical $I_D V_G$ at V_{DS} = 5 mV for three different qubit array designs measured during our screening protocol on the same die displaying **(a)** pure MOSFET **(b)** mixed MOSFET + SET and **(c)** pure SET characteristics. **(d)** all three $I_D V_G$ overlaid to show comparison. In design C, Coulomb peak heights and valleys are highlighted to calculate average peak distance.

For Design C, we show 2D maps of $I_D(V_{G2},V_{G3})$ in Fig. 5. The existence of different sets of regular Coulomb peaks controlled by each of the two gates confirm that electrostatic QDs are defined by G2 and G3. The dots are found to be coupled with each other since honeycomb patterns can be distinguished [14]. For this design, we see over an 80% yield of the dies tested (20/24) spread across the center, middle and edge radii of the wafer. The yield indicates that a double QD system was identified through qualitative analysis of the 2D maps (*i.e.* visual inspection of patterns).

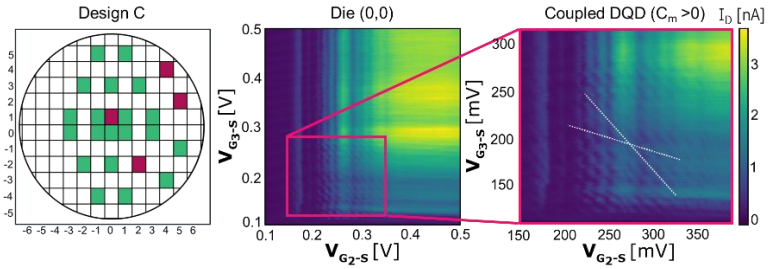


Figure 5. Wafer-level 2D stability diagrams are collected at 2K for all neighboring front gate combinations to confirm SET behavior, collect information on their positions as well as verify the existence of coupling without the application of a barrier via bias. 80% of Design C devices displayed a DQD. Example of a coupled DQD shown for Die (0,0) defined by front gates G2 and G3.

In Fig. 6 we show simulations and experimental results for different regimes of coupling between two QDs, formed underneath front gates G1 and G2, for the same wafer, die and device as shown in Fig. 5. Simulations suggest, and experiment confirms tunnel coupling control between two neighboring QDs in our linear array design. The stability diagrams show $I_D(V_{G1},V_{G2})$ collected for V_{B2} = –0.5V, +0.75V and +1.5V, where the

tunnel coupling varies from weak (observation of bias triangles) to strong. In the moderate coupling regime, honeycomb patterns are clearly observed, indicative of capacitive coupling.

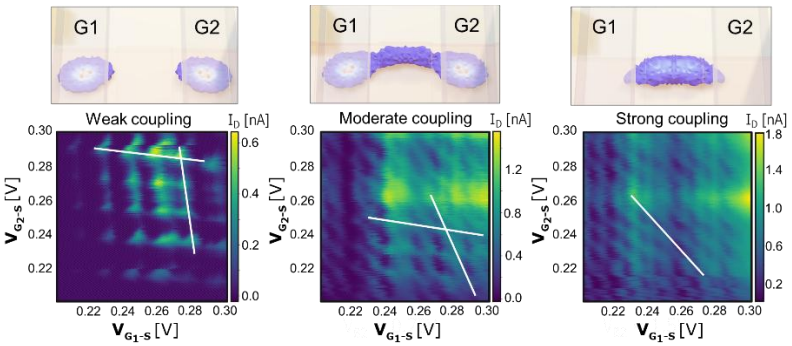


Figure 6. **(top)** Simulation of envelope probability density, showing the effect of applied bias on the barrier gate via, B2. **(bottom)** Stability diagrams of G1 and G2 from Design C, with coupling controlled by the barrier gate B2, collected at 2K. White solid lines are added to guide the eyes.

4 Conclusions

We summarize our effort to design and operate quantum circuits using industrial 28nm FD-SOI technology platform. We proposed the use of W vias as barrier gates in qubit devices and demonstrate the ability to control the tunnel coupling between two adjacent QDs operating in the many-electron regime. The vias act as proper electrostatic gates and control both conduction within the Si channel as well as allow for coupling control in the Coulomb blockade regime. Our design fulfills the requirement of individual coupling control needed for implementing quantum gates in spin-based quantum processors while maintaining an aggressive pitch. Our work demonstrates the first step towards an industrialization of spin qubit processors based on the existing semiconductor ecosystem.

Acknowledgments

We thank the Quobly characterization and engineering team for providing the data.

References

1. Meunier, T., Daval, N., Perruchot, F., & Vinet, M. Silicon spin qubits: a viable path towards industrial manufacturing of large-scale quantum processors. *The European Physical Journal A*, *61*(3), 1-11 (2025)

2. B.C. Paz *et al.*, 'FDSOI platform for quantum computing', *2024 IEEE International Electron Devices Meeting (IEDM)*, pp 1-4, Dec. 2024. doi: 10.1109/IEDM50854.2024.10873521
3. Hamonic, Pierre, et al. "Combining multiplexed gate-based readout and isolated CMOS quantum dot arrays." *Nature Communications* 16.1 (2025): 6323.
4. Q. Schmidt *et al.*, 'Compact frequency multiplexed readout of silicon quantum dots in monolithic FDSOI 28nm technology', in *2024 IEEE European Solid-State Electronics Research Conference (ESSERC)*, Sep. 2024, pp. 157–160. doi: 10.1109/ESSERC62670.2024.10719580.
5. P. Hamonic et al, Coherent exchange oscillation manuscript (submitted)
6. S. Bonen, S. P. Tripathi, J. McIntosh, T. Jager, and S. P. Voinigescu, 'Investigation of p- and n-Type Quantum Dot Arrays Manufactured in 22-nm FDSOI CMOS at 2–4 K and 300 K', *IEEE Electron Device Lett.*, vol. 45, no. 10, pp. 2025–2028, Oct. 2024, doi: 10.1109/LED.2024.3435380.
7. S. V. Amitonov *et al.*, 'Commercial CMOS Process for Quantum Computing: Quantum Dots and Charge Sensing in a 22 nm Fully Depleted Silicon-on-Insulator Process', Dec. 20, 2024, *arXiv*: arXiv:2412.08422. doi: 10.48550/arXiv.2412.08422.
8. N. Planes *et al.*, '28nm FDSOI technology platform for high-speed low-voltage digital applications', in *2012 Symposium on VLSI Technology (VLSIT)*, Jun. 2012, pp. 133–134. doi: 10.1109/VLSIT.2012.6242497.
9. G. E. Elbaz *et al.*, 'Transport characterization and quantum dot coupling in commercial 22FDX', Jan. 17, 2025, *arXiv*: arXiv:2501.10146. doi: 10.48550/arXiv:2501.10146.
10. C. Spence *et al.*, "Probing Low-Frequency Charge Noise in Few-Electron CMOS Quantum Dots", *Phys. Rev. Appl.*, vol. 19, pp. 044010, Apr. 2023; DOI: 10.1103/PhysRevApplied.19.044010
11. T. Bédécarrats *et al.*, 'A new FDSOI spin qubit platform with 40nm effective control pitch', in *2021 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2021, pp. 1–4. doi: 10.1109/IEDM19574.2021.9720497.
12. [10] J.-P. Colinge, 'Quantum-wire effects in trigate SOI MOSFETs', *Solid-State Electron.*, vol. 51, no. 9, pp. 1153–1160, Sep. 2007, doi: 10.1016/j.sse.2007.07.019.
13. S. Neyens *et al.*, 'Probing single electrons across 300-mm spin qubit wafers', *Nature*, vol. 629, no. 8010, pp. 80–85, May 2024, doi: 10.1038/s41586-024-07275-6.
14. W. G. van der Wiel *et al.*, "Electron transport through double quantum dots", *Rev. Mod. Phys.*, vol. 75, No. 1, Dec. 2002; doi: 0.1103/RevModPhys.75.1

A Hybrid Quantum-Classical Approach for Urban Drone Path Planning – Insights from Quantum Computer Implementation

Hian Lee Kwa, Bing Hong Teh, Teck Yoong Chai, and Yung Sze Gan

CortAIx Singapore, Thales Solutions Asia, Singapore
{hianlee.kwa, binghong.teh, teckyoong.chai}@thalesgroup.com
yungsze.gan@asia.thalesgroup.com

Abstract. As drone traffic and air taxi services expand globally, regulators face increasing pressure to manage complex airspace safely, while operators aim to optimise energy-efficient routes. Traditional computing may not scale to meet future path planning demands, making quantum computing a promising alternative. However, the variety of qubit modalities across quantum processing units (QPUs) raises questions about which platforms are most suitable for such applications. This paper presents a hybrid quantum-classical drone path planning approach that accounts for urban obstacles and its implementation on two QPUs: D-Wave's quantum annealer and Pasqal's neutral atom processor. We compare solution quality, time-to-solution, and hardware-specific trade-offs against a classical solver, providing insights into the scalability, embedding challenges, and the practical feasibility of different quantum computing platforms for urban air mobility.

Keywords: Drone Path Planning · Hybrid Quantum-Classical Algorithm · Neutral Atom Quantum Computers · Quantum Annealing · Quantum Computing · Urban Air Mobility

1 Introduction

The use of unmanned aerial vehicles (UAVs) in urban environments has recently started to grow rapidly, with applications ranging from building inspection and surveillance to parcel delivery. As the uptake of urban air mobility increases, regulators face mounting challenges in ensuring the safe and efficient management of airspace. At the same time, drone operators seek to minimise energy usage by optimising flight paths. A centralised planner that accounts for drone requirements and environmental constraints could address both needs. However, traditional computing approaches may struggle with scalability, a challenge that quantum computing could overcome.

Quantum computing in such applications can be achieved through hybrid quantum-classical frameworks that combine classical techniques with quantum algorithms. A common method involves converting real-world problems into Quadratic Unconstrained Binary Optimisation (QUBO) form [3], which can be

solved using quantum annealers, such as those provided by D-Wave, or gate-based processors using variational quantum algorithms (VQAs) [9].

In this paper, the contributions are as follows: (1) extending the method proposed by James and Raheb [4] to one that has been adapted to account for environmental obstacles, (2) implementing this approach using a D-Wave quantum annealer and a simulated neutral atom QPU from Pasqal, as well as benchmarking the resulting solutions for a single drone case against those produced by a classical algorithm, and (3) evaluating the two QPUs in terms of solution quality and generation time, and discuss their implementation, scalability, and the challenges involved in real-world deployment.

2 Background and Motivation

In classical computing, path planning for autonomous vehicles typically relies on graph search algorithms, which aim to identify a sequence of edges connecting a start node to a destination [7]. In practice, classical routing problems often map nodes to points on a fine grid, allowing near-continuous movement across a space. These approaches can be adapted to quantum computing since graph search problems are readily reformulated as Quadratic Unconstrained Binary Optimisation (QUBO) problems and can subsequently be solved via Quantum Annealing or other Variational Quantum Algorithms (VQAs) [1].

However when formulating the problem with quantum computing, nodes are typically mapped to key landmarks, such as drop-off locations or intersections, due to the limited qubit count and connectivity in current quantum processing units (QPUs) [6,14]. As a result, drones are often constrained to predefined routes and are unable to take advantage of variations in their flight capabilities, such as the speed of fixed-wing drones or the agility of multi-rotor drones.

To better exploit UAV manoeuvrability, an iterative approach that does not rely on fixed graphs is required. This enables drones to carry out control decisions anytime during the flight instead of at pre-determined waypoints. Similar quantum-based methods have been explored in contexts like robot swarms [2] and airport landing path generation [4]. However, these do not fully account for obstacles such as buildings and no-fly zones. In what follows, we formulate a QUBO problem that allows the optimiser to route drones from their starting points to their destinations while avoiding collisions with both environmental obstacles and other drones. This is done without mapping waypoints to graph nodes, making it applicable to any area with known obstacles, and easily scalable to accommodate an arbitrary number of drones.

3 QUBO Problem Formulation

To guide drones toward their destinations while avoiding obstacles, we use a combination of attractive and repulsive potential fields. The quantum optimiser attempts to minimise the total potential field experienced by each UAV at each time-step through choosing a speed and turning rate instruction the UAV is

to carry out. This potential value is related to the distance of a drone to its destination and to other nearby drones. For the original formulation, the reader is directed to [4]. In what follows, we describe an extension of their QUBO formulation that accounts for the presence of buildings and other obstacles, which is done via classical computing methods.

The overall potential field guiding the drone is given by the equation:

$$P_i(x_i) = P_{\text{att},i}(x_i) + P_{\text{rep},i}(x_i). \tag{1}$$

Where, $P_{\text{att},i}$ and $P_{\text{rep},i}$ is the attractive field and repulsive field experienced by a Drone i respectively. This combined field is visualised in Fig. 1.

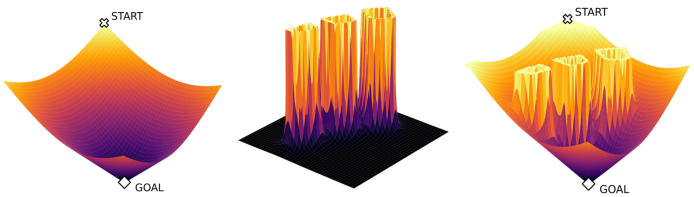


Fig. 1: Attractive potential field that guides a drone to its target (Left). Repulsive potential field generated by three towers (Centre). The overall potential field of the environment (Right). The potential fields were all generated at one altitude.

3.1 Attractive Potential Fields

In both the work presented here and the original, an attractive potential is used to guide drones towards their destinations. This is defined for Drone i at position x_i as a function of its distance to its destination:

$$P_{\text{att},i}(x_i) = k_{\text{att}}||\big(x_i - x_{\text{target},i}\big)^{\alpha}||, \tag{2}$$

where $x_{\text{target},i}$ is the target location, and tunable constants $k_{\text{att}} = 1 \times 10^3$ and $\alpha = 2$. This function ensures that the potential field experienced by the drone reduces the closer it is to its destination.

3.2 Repulsive Potential Fields

In addition to the attractive potential field, a repulsive potential field is used to push drones away from obstacles. A Drone i perceives a neighbouring Drone j as a peak in its repulsive field given by:

$$P_{\text{rep},i,j}(x_i) = k_{\text{rep}}\left(\frac{1}{D_{i,j}} - \frac{1}{R}\right)^{\beta} ||\big(x_i - x_{\text{target},i}\big)||, \tag{3}$$

where $D_{i,j}$ is the distance between Drones i and j, $R = 30$m is the Drone j's influence range, and the tunable constants are $k_{\text{rep}} = 1 \times 10^7$ and $\beta = 2$.

To extend this to obstacles, we use point clouds to delimit restricted areas around buildings. The repulsive fields from the areas are computed by summing the contributions from each point with Equation 3, where $D_{i,j}$ is the drone's distance to each point. Furthermore, buildings located close to each other are aggregated into one large restricted area to avoid the potential trapping of drones in potential wells generated within narrow corridors between buildings.

3.3 Control Decision Vectors

To optimise drone control at each time-step, the planner seeks a binary solution vector U that minimises the cost function:

$$C(U) = U^T Q(X) U, \tag{4}$$

where $Q(X)$ is the QUBO matrix derived from the potential field values of the controlled drones. The vector U represents all discretised control decisions across UAVs and can be split into smaller vectors u, each corresponding to a single UAV. For instance, a fixed-wing Drone i with a speed range of 10-15 m/s and a maximum turning rate of 25 deg/s may have two binary variables for speed and three for turning rate, giving:

$$u_i^T = [\text{Turning Rate: } -25, 0, 25 \text{ deg/s; Speed: } 10, 15 \text{ m/s}]. \tag{5}$$

This formulation essentially turns the path planning problem into a combinatorial optimisation problem, where the optimiser selects the combination of control actions that minimises the cost function. With this scheme, a drone flying at 15 m/s with a turning rate of 0 deg/s is associated with the binary control vector $u_i^T = [0, 1, 0, 1, 0]$. Fig. 2 illustrates this setup. While the current approach handles two-dimensional motion, it can be extended to three dimensions by a separate control group for climb and descent rates.

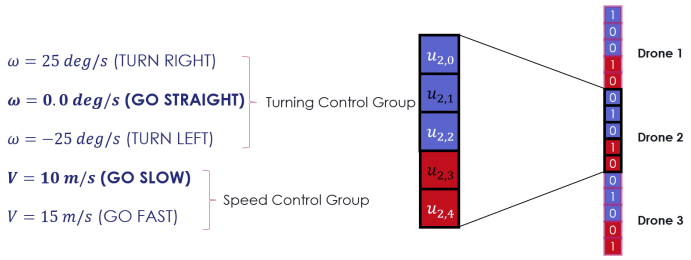


Fig. 2: The overall control vector, U, with the control vector, u_2, for Drone 2 highlighted. The drone is travelling at a speed of 10 m/s and a turning rate of 25 deg/s.

With this formulation, combinations of control decisions results in different positions for the drone at the subsequent time-step, which is used to calculate the future potential value of the drone. These serve as the coefficient values for the QUBO matrix. Formally, for an element q_{ij} in the QUBO matrix, $Q(x)$,

$$q_{ij} \coloneqq P(x(u_i, u_j)), \tag{6}$$

where $x(u_i, u_j)$ is the future position of the drone when control decisions u_i and u_j are selected. Consequently, the number of logical qubits required to encode the QUBO problem scales as $O(N \times M)$, where N is the number of drones and M is the number of controls per drone.

3.4 Mutual Exclusivity Groups

Some control combinations are invalid. E.g., $u_i = [1, 0, 1, 0, 0]$ instructs the drone to turn both left and right while omitting a speed command. To prevent such conflicts, mutual exclusivity constraints are applied using the following identity:

$$\sum_{g=1}^{G} \left(1 - \sum_{c=1}^{C} u_{gc}\right)^2 = 0, \tag{7}$$

Here, G is the number of exclusivity groups (e.g., speed and turning), and C is the number of control options in each group. In Fig. 2, there are two groups: turning (3 controls) and speed (2 controls), giving $G = 2$ and $C = 3, 2$ respectively. These constraints are incorporated into the optimiser via a weighted minimiser, which penalises the selection of invalid control combinations:

$$K \min_{u} \sum_{g=1}^{G} \left(1 - \sum_{c=1}^{C} u_{gc}\right)^2 = 0, \tag{8}$$

where K is a large penalty constant.

4 QPU Implementation

4.1 Quantum Annealer

Quantum annealing is an approach for solving combinatorial optimisation problems by exploiting the adiabatic theorem: if a quantum system evolves slowly enough from an easy-to-prepare initial Hamiltonian to a final problem Hamiltonian, it remains in its ground state, thereby yielding the optimal solution. In the QUBO-to-Ising mapping used here, the Hamiltonian evolves according to

$$H(t) = A(t)H_i + B(t)H_f, \tag{9}$$

where $H_i = \sum_i \sigma_i^x$ is the initial Hamiltonian, and $H_f = \sum_i h_i \sigma_i^z + \sum_{i,j} J_{i,j} \sigma_i^z \sigma_j^z$ encodes the QUBO problem through local fields h_i and couplings $J_{i,j}$. The schedule functions $A(t)$ and $B(t)$ interpolate between the two.

On D-Wave's Advantage 4.1 quantum annealer, QUBO problems are represented as dictionaries of linear and quadratic coefficients. These logical variables are then mapped to physical qubits arranged in a Pegasus topology. Because the Pegasus graph has sparse, fixed connectivity, a *minor embedding* process is required to map QUBO variables onto physical qubits. D-Wave's SDK provides a heuristic embedding algorithm that automates this step, after which the annealing schedule is executed to sample low-energy solutions.

4.2 Neutral Atom Processor

On Pasqal's neutral atom QPUs, the coefficients of the QUBO matrix, which define the binary variables' interactions, are encoded through the controlled Rydberg interaction between atom pairs [8,11]. The interaction strength is defined as $U_{ij} = C_6/r_{ij}^6$, where C_6 is a constant and r_{ij}^6 is the inter-atomic distance between atoms i and j. To encode a QUBO problem, atoms are arranged in a register to set the desired inter-atomic distances, which then determine the interaction strengths representing the off-diagonal coefficients. Formally, this can be formulated as a minimisation problem:

$$\underset{X}{\text{minimize}} ||U(X) - Q||_F, \tag{10}$$

where Q is the QUBO matrix, X is a vector of the atoms' placements to be optimised, and $U(X)$ is the $n \times n$ matrix with the Rydberg interaction levels between atom pairs, given as:

$$U_{ij} = \begin{cases} \frac{C_6}{||\mathbf{x_i} - \mathbf{x_j}||^6}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}, \tag{11}$$

where $\mathbf{x_i}$ and $\mathbf{x_j}$ are vectors containing the x and y positions of atoms i and j respectively. In our implementation, these were found using the Nelder-Mead method repeated 30 times with random atom initialisations. The configuration with the lowest cost was chosen as the final atomic placements. This qubit placement method effectively maps one QUBO variable to one physical qubit. While this is sufficient for our current implementation, one must consider alternative embedding strategies should large numbers of controls be considered due to the finite level of connectivity for each qubit, including the use of quantum wires to connect physically distant qubits [5,13].

After placing the atoms, we optimise the time evolution of the laser pulse's amplitude and detuning during the adiabatic process, subject to two conditions: (1) the amplitude must start and end at zero, and (2) the detuning must increase monotonically through zero. Using an interpolated waveform, we optimise the pulse at three equally spaced points (six parameter values), which is done based on the expected cost of the sampled bitstrings.

5 Flight Path Planning

Three different solvers were implemented to evaluate the flight planning method. First was D-Wave's Leap software package running on their Advantage-4.1 system, using the quantum annealing process described in Section 4.1. The second solver was Pasqal's neutral atom QPU that was emulated via their Pulser library and implemented as described in Section 4.2. Finally, a comparison was made between the two quantum approaches against a classical approach, simulated annealing, which was also available via D-Wave's SDK.

For this work, we considered flights taking place in the Marina Bay area of Singapore. As mentioned in Section 3, restricted areas were drawn around buildings to prevent drone collisions. In addition, buildings that were located very close to each other were grouped into a larger restricted zone to prevent drones from being trapped between them (see Fig. 3). In this work, for simplicity and to demonstrate its ability to avoid buildings, only fixed-winged drones with two speed controls (10 and 15 m/s) and three turning rate controls (-25, 0, 25 deg/s), as well as a time-step of $dt = 1$s were considered. We also restrict the drones to flying at a fixed altitude, similar to how a controller may assign a specific altitude for drone operations [12]. Despite this, we note that this path planning framework will work for drones with different available controls, including altitude controls, as demonstrated by James and Raheb in clear environments [4].

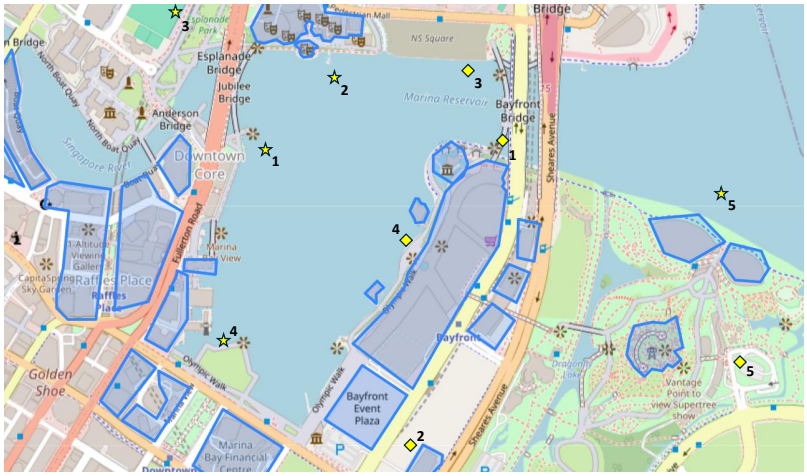


Fig. 3: Flight planning area over Marina Bay, Singapore. Restricted zones are demarcated in blue. Route start points (stars) and destinations (diamonds) are labelled with route numbers.

6 Results

6.1 Solution Quality

During testing, five flight scenarios were used, each with its own launch and destination point. One example route can be seen in Fig. 4. Table 1 presents the total distance covered by the fixed-wing drone for each test case, as planned by the different solvers. This table shows that the quantum methods produced solution qualities similar to, and in some cases better than, the classical simulated annealing method, with all results falling within 4% of each other.

At this stage, we note that the neutral atom-based solver used a single optimized laser pulse sequence to generate all the paths, including those for flying in both clear airspace and near obstacles (see Fig. 4). This is an advantage over the

Route No.	D-Wave (m)	Pasqal Simulated Annealing (m)	(m)
1	639	648	**633**
2	**1271**	1295	1290
3	915	900	**898**
4	**535**	555	556
5	**510**	528	520

Table 1: Total distance planned for a fixed-wing drone flying on different routes.

use of VQAs as although it is possible to run VQAs on neutral atom QPUs, doing so would require additional classical and quantum computational resources. This is because for VQAs, the pulse sequence needs to be optimized at each time-step, thereby requiring additional calls to the QPU and a classical optimiser.

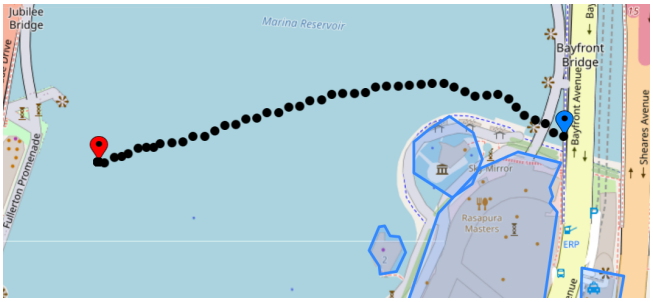


Fig. 4: Flight path for a fixed-wing drone on Route 1, as produced by the simulated Pasqal neutral atom solver.

6.2 Solution Time

Despite the relatively similar result quality, the amount of time required to derive these results from the two QPUs differs. For D-Wave annealers, the time required to return a solution is split into four broad stages[1]: (1) preparation, i.e., the mapping of variables to physical qubits, (2) QPU access time, which includes QPU programming and sampling, (3) post processing, and (4) worker and QPU queues, with steps 2 and 3 occurring in parallel. For our framework, the time required to obtain the solution to each submitted QUBO problem was $\sim$ 6s, with the QPU access time being measured at $\sim$ 550ms per result. This suggests that the majority of the end-to-end latency arises from queueing for both the QPU and the pre-processing worker and pre-processing overhead rather than from quantum computation itself.

In contrast, the preparation phase for Pasqal's QPU occurs locally using the process described in Section 4.2. For the 5-control vector single drone case used, this took $\sim$ 60ms on an Intel i5-1145G7 CPU. However, it should be noted that initialising the atomic register using a structured layout that reflects the connectivity of the QUBO problem, rather than a random placement, may reduce the preprocessing time associated with atom positioning and potentially improve

[1] D-Wave Operation and System Timing

solution quality by providing a better starting point for layout optimisation. Although simulating quantum processes to solve the QUBO problem took less than 100ms per instance, this step is expected to take significantly longer on a physical QPU. Due to the qubit preparation processes, these QPUs currently have a base repetition rate of $2 - 3$Hz [11]. As such, solving a QUBO problem for each time-step with 5000 samples on a physical QPU is expected to take around 30mins, while solving for a drone's entire trajectory over 100 time-steps is expected to take 50 hours.

7 Discussion

In this paper, we presented an extension to the method proposed by James and Raheb to account for fixed environmental obstacles such as buildings, and its implementation on two QPUs. Unlike waypoint-based systems, this approach permits free UAV flight without predefined routes, assuming obstacles have been pre-mapped. The required number of logical qubits scales with the number of drones and control variables rather than flight area size. Tests with a single fixed-wing UAV demonstrate the feasibility of hybrid quantum–classical frameworks for obstacle-aware urban path planning, with quantum solvers producing path lengths comparable to or better than a classical solver. Further validation and scalability analysis is required for multi-drone operations.

Despite the similar solution qualities, solution times varied significantly depending on the QPU used. D-Wave's quantum annealers returned each time-step's solution in ~ 6s while physical execution on Pasqal's QPU is estimated to take ~ 30mins per QUBO problem. This is mainly due to the time required for atomic register preparation. As such, D-Wave is more suitable for pre-flight or urgent replanning, while Pasqal's QPUs are better suited to generating routes well in advance of the flight. Fundamentally, implementing such a framework in live path planning requires a reduction in solving time for each submitted QUBO problem. This means reducing the base repetition rate for neutral atom QPUs[2] or the preparation and queue time for D-Wave's quantum cloud services.

Scaling the number of drones or control variables produces larger, denser QUBO graphs, increasing both embedding complexity and physical qubit requirements. For instance, moving from one drone with five controls to ten drones increases the QUBO size and logical qubit requirement by a factor of 10. This could quickly exceed the capacity of current annealing and neutral atom systems without qubit-efficient encodings or improved mapping strategies. While D-Wave offers mature minor embedding heuristics and gate-based QPUs support qubit efficient encodings [6,10], no comparable approach exists for neutral atom architectures. Future work should focus on geometry-aware heuristics or variational embedding strategies tailored to neutral atom QPUs to reduce preprocessing overhead and improve scalability.

[2] Pasqal's QPU development roadmap states their aim to achieve a base repetition of 100Hz in 2029.

Acknowledgements The authors would like to thank Prof. Dimitris Angelakis, Mr. Andreas Stratakis, and Mr. Gordon Ma from AngelQ for the insightful discussions on problem formulation, and Dr. Krisztian Benyo from Pasqal for the discussions on implementing quantum algorithms on their QPUs emulators.

References

1. Bauckhage, C., Brito, E., Cvejoski, K., Ojeda, C., Schücker, J., Sifa, R.: Towards shortest paths via adiabatic quantum computing. In: Proc. Mining Learn. Graphs (2018)
2. Chella, A., Gaglio, S., Mannone, M., Pilato, G., Seidita, V., Vella, F., Zammuto, S.: Quantum planning for swarm robotics. Robotics and Autonomous Systems **161**, 104362 (2023)
3. Glover, F., Kochenberger, G., Du, Y.: Applications and computational advances for solving the QUBO model. In: The Quadratic Unconstrained Binary Optimization Problem: Theory, Algorithms, and Applications, pp. 39–56. Springer (2022)
4. James, S.H., Raheb, R.N.: Path planning for critical ATM/UTM areas. In: 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC) (2019)
5. Kim, M., Kim, K., Hwang, J., Moon, E.G., Ahn, J.: Rydberg quantum wires for maximum independent set problems. Nature Physics **18**(7), 755–759 (2022)
6. Leonidas, I.D., Dukakis, A., Tan, B., Angelakis, D.G.: Qubit efficient quantum algorithms for the vehicle routing problem on noisy intermediate-scale quantum processors. Advanced Quantum Technologies **7**(5), 2300309 (2024)
7. Sanchez-Ibanez, J.R., Pérez-del Pulgar, C.J., García-Cerezo, A.: Path planning for autonomous mobile robots: A review. Sensors **21**(23), 7898 (2021)
8. Silvério, H., Grijalva, S., Dalyac, C., Leclerc, L., Karalekas, P.J., Shammah, N., Beji, M., Henry, L.P., Henriet, L.: Pulser: An open-source package for the design of pulse sequences in programmable neutral-atom arrays. Quantum **6**, 629 (2022)
9. Symons, B.C., Galvin, D., Sahin, E., Alexandrov, V., Mensa, S.: A practitioner's guide to quantum algorithms for optimisation problems. Journal of Physics A: Mathematical and Theoretical **56**(45), 453001 (2023)
10. Tan, B., Lemonde, M.A., Thanasilp, S., Tangpanitanon, J., Angelakis, D.G.: Qubit-efficient encoding schemes for binary optimisation problems. Quantum **5**, 454 (2021)
11. Tibaldi, S., Leclerc, L., Vodola, D., Tignone, E., Ercolessi, E.: Analog QAOA with bayesian optimisation on a neutral atom QPU (2025), https://arxiv.org/abs/2501.16229
12. Veytia, A.M., Badea, C.A., Ellerbroek, J., Hoekstra, J., Patrinopoulou, N., Daramouskas, I., Lappas, V., Kostopoulos, V., De Vries, V., van Ham, J., et al.: Metropolis II: Benefits of centralised separation management in high-density urban airspace. In: 12th SESAR Innovation Days (2022)
13. Wintersperger, K., Dommert, F., Ehmer, T., Hoursanov, A., Klepsch, J., Mauerer, W., Reuber, G., Strohm, T., Yin, M., Luber, S.: Neutral atom quantum computing hardware: performance and end-user perspective. EPJ Quantum Technology **10**(1), 32 (2023)
14. Yarkoni, S., Neukart, F., Tagle, E.M.G., Magiera, N., Mehta, B., Hire, K., Narkhede, S., Hofmann, M.: Quantum shuttle: traffic navigation with quantum computing. In: Proceedings of the 1st ACM SIGSOFT International Workshop on Architectures and Paradigms for Engineering Quantum Software. pp. 22–30 (2020)

Quantum-Backed Integrity for Industrial Blockchain Events: A Secure Logging Framework Using Qiskit and Ethereum

Benoît Prieur [1]

[1] Medialoco LLC, USA
benoit.prieur@protonmail.com

Abstract. In the era of Industry 4.0, where automation, traceability, and secure data flows are foundational to modern industrial ecosystems, blockchain technology has emerged as a powerful tool to ensure transparency, auditability, and resistance to tampering. By offering an immutable ledger of transactions or events, blockchains are increasingly adopted in manufacturing, logistics, healthcare, and energy sectors to track critical operations and ensure trust among stakeholders. However, the integrity of any blockchain-based system critically depends on the quality of the data it ingests. If the data recorded is predictable, manipulated, or inserted without sufficient entropy or verification, the immutability of the blockchain becomes moot. This paper introduces a proof-of-concept framework that integrates Quantum Random Number Generation (QRNG) and smart contract-based logging on Ethereum to strengthen the cryptographic robustness of industrial event recording. Each event is timestamped, tagged with a quantum-generated number, described via metadata, and hashed using SHA-256 to produce a unique fingerprint. The hash is then immutably stored on a local Ethereum blockchain. The system demonstrates a fusion of quantum physical entropy and decentralized blockchain trust.

Full source code: https://github.com/Medialoco/quantum-secure-blockchain.

Keywords: Quantum Random Number Generation (QRNG), Blockchain immutability, Ethereum smart contract, Industrial event logging, Cryptographic integrity.

1 Motivation and Threat Model

Modern industrial systems increasingly depend on the real-time logging of operational events to guarantee quality control, safety compliance, and system transparency. In domains such as energy, aerospace, pharmaceuticals, and defense, industrial logs serve not only as technical records but as legal and operational anchors for decision-making and accountability.

While blockchain offers immutability once data is recorded, it cannot guarantee the quality or unpredictability of the input itself. Attackers may intercept or forge log entries before they are committed to the chain. Classical sources of entropy, such as pseudorandom number generators (PRNGs), are vulnerable to manipulation if seeds or initial conditions are exposed. Predictable or replayable data allows adversaries to mimic legitimate events or tamper with historical records.

Quantum Random Number Generation (QRNG), based on the inherent unpredictability of quantum measurement, provides a remedy. By sourcing entropy from quantum states, one ensures the non-reproducibility of each event, enhancing trust at the data insertion layer. The hash of this event, constructed with SHA-256, becomes a tamper-evident commitment to the quantum measurement, timestamp, and metadata.

In our threat model, an attacker may attempt to replay events, inject forged data, or alter metadata before blockchain commitment. The proposed system thwarts such attempts by binding unpredictable quantum entropy with cryptographic sealing and immutable storage.

2 System Design

The proposed framework integrates four components: a quantum entropy source (QRNG with Qiskit), a cryptographic hashing layer (SHA-256), a smart contract for event recording (QSecure.sol), and a blockchain interaction layer (Node.js orchestration).

2.1 Quantum Entropy Extraction

To obtain physically unpredictable values, we construct an 8-qubit quantum circuit using Qiskit. Each qubit is placed in superposition via a Hadamard gate, producing a uniform distribution over all 256 possible bitstrings (see Fig. 1). Upon measurement, the circuit collapses to one classical outcome, interpreted as a random integer.

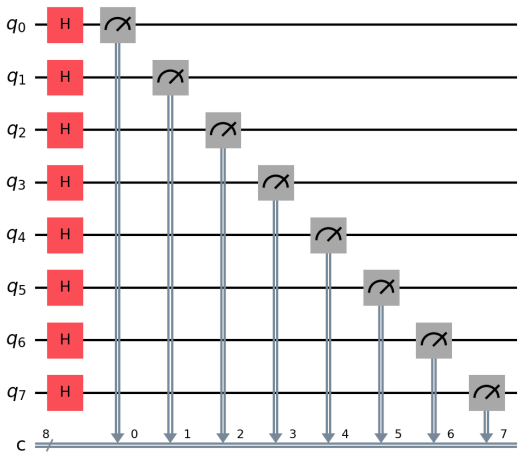


Fig. 1. Diagram of the quantum circuit.

The following Python method (file qrng.py) implements this QRNG pipeline and persists all artifacts, bitstring, integer value, timestamp, and ASCII diagram, in a JSON file for forensic inspection:

```
def generate_quantum_random_bits(n_bits=8,
save_path="quantum_proof.json"):
    circuit = QuantumCircuit(n_bits, n_bits)
    circuit.h(range(n_bits))
    circuit.measure(range(n_bits), range(n_bits))
    backend = Aer.get_backend("qasm_simulator")
    job = backend.run(circuit, shots=1)
    result = job.result()
    counts = result.get_counts()
    bitstring = list(counts.keys())[0]
    value = int(bitstring, 2)
    proof = {
        "timestamp": datetime.now(timezone.utc).isofor-
mat(),
        "bitstring": bitstring,
        "value": value,
        "n_bits": n_bits,
        "circuit_ascii": circuit.draw(out-
put='text').single_string()
    }
    with open(save_path, "w") as f:
```

```
            json.dump(proof, f, indent=2)
    return value
```

This quantum-generated integer serves as the entropy anchor for each logged event.

2.2 Event Fingerprint and Integrity Seal

The quantum value is combined with metadata (e.g., component label, checkpoint name) and a system timestamp to form a unique event fingerprint. Using Node.js's crypto module, this string is hashed with SHA-256 to produce a tamper-evident commitment.

```
const eventData = `${timestamp}-${randomNumber}-${meta-
data}`;
const eventHash = crypto.createHash('sha256').up-
date(eventData).digest('hex');
```

2.3 Smart Contract for Immutable Logging

The integrity seal, together with the quantum proof (the quantum-generated integer) and metadata, is transmitted to the QSecure smart contract, implemented in Solidity and deployed locally via Hardhat. The contract stores a timestamp, the quantum-generated integer, metadata, and the SHA-256 hash as an immutable record:

```
struct Event {
    uint256 timestamp;
    uint256 quantumProof;
    string metadata;
    bytes32 eventHash;
}

function recordEvent(
    uint256 _quantumProof,
    string calldata _metadata,
    bytes32 _eventHash
) external { ... }
```

Access control ensures only the deployer may append new events, while any participant can read back entries to verify provenance.

2.4 Orchestration Pipeline

The orchestration script (scripts/setQuantum.js) automates the full pipeline: it runs the Python QRNG (qrng.py) to obtain the quantum generated integer, computes the SHA-256 hash, invokes recordEvent on the deployed QSecure contract via Web3.js, retrieves

the most recent event (getEvent) to confirm successful commitment, and serializes all results to JSON for external archiving.

```
const eventData = `${timestamp}-${randomNumber}-${meta-
data}`;
const eventHash = crypto.createHash('sha256').up-
date(eventData).digest('hex');

await qsecure.methods
  .recordEvent(randomNumber, metadata, '0x' + eventHash)
  .send({ from: signer.address });
```

With this end-to-end pipeline design [qrng.py → setQuantum.js → QSecure.sol] each industrial event is cryptographically bound to unpredictable quantum entropy and immutably anchored on the blockchain.

2.5 Example Execution Output

The following output, captured from a local Hardhat testnet, demonstrates a full end-to-end run. After deploying the QSecure contract, a quantum-generated integer is produced via Qiskit, combined with metadata and timestamp, and then hashed using SHA-256. This hash, along with the raw data, is recorded on-chain. Finally, the event is retrieved from the blockchain to verify the consistency of the stored values.

```
QSecure deployed at:
0xB7f8BC63BbcaD18155201308C8f3540b07f84F5e
Quantum-generated number: 214
SHA-256 event hash:
d4f8665e0bc8a3ffb589eb1e0f48e889cdc2a3af893f0451c994a48f7
9d82ac2
Total events recorded: 1
Event details:
  Timestamp      : 1744230722
  Quantum Proof  : 214
  Metadata       : Component validated - Checkpoint A
  Stored SHA-256 :
0xd4f8665e0bc8a3ffb589eb1e0f48e889cdc2a3af893f0451c994a48
f79d82ac2
```

This console output confirms that all components of the pipeline, quantum entropy generation, cryptographic hashing, smart contract interaction, and blockchain storage, executed successfully. The SHA-256 hash retrieved from the blockchain exactly matches the one derived from the original quantum value, timestamp, and metadata. This confirms that the event was stored immutably and verifiably. Any attempt to replay or forge the event would inevitably produce a hash mismatch.

3 Classical vs Quantum Randomness and the Role of Qubit Count

Pseudorandom number generators (PRNGs) are algorithms whose unpredictability depends on how well a secret seed is protected. Physical noise sources, often used as true random number generators (TRNGs), improve unpredictability but can drift or be influenced unless carefully conditioned. A quantum random number generator (QRNG) is different: each bit comes from a quantum measurement whose outcome cannot be reconstructed from hidden device state, so unpredictability stems from superposition and collapse rather than secrecy. For logging, the key metric is extractable entropy, not raw bit rate. In practice, QRNGs combine shallow per-qubit measurements with vetted randomness extractors to remove bias and correlations and to deliver an auditable minimum-entropy budget.

Adding qubits mostly increases throughput under ideal conditions, where n qubits can yield up to n bits per shot. Quality is limited by noise, calibration drift, and crosstalk, so moving to 16 or 32 qubits raises subtle asymmetries that require the extractor to use larger windows and lower rates to match measured minimum entropy. The secure yield is therefore $\alpha \cdot n$ with α less than 1 and determined during operation. Beyond a few dozen qubits, bottlenecks often shift to input and output (I/O), post-processing, and anchoring costs, so the right target is verified entropy per second rather than raw counts. For secure logging, simple, parallel, well-calibrated measurements generally outperform deeper, entangling circuits.

4 Security Analysis

Security links three properties end to end: unpredictability at the source, binding by hash, and immutability at the ledger. An attacker might tamper with firmware, reorder network traffic, skew clocks, or attempt chain reorganizations. Quantum generation supplies non-derivable entropy at the edge, continuous health tests enforce fail-closed behavior, and periodic minimum-entropy estimates throttle extraction so commitments never overstate entropy. Each event is sealed with the Secure Hash Algorithm 256-bit (SHA-256) over quantum bits, metadata, and a timestamp. Even with Grover's quadratic speedup, margins remain comfortable, and the Secure Hash Algorithm 384-bit (SHA-384) or 512-bit (SHA-512) can be substituted if policy requires. Non-repudiation comes from signing before anchoring, with a migration path to post-quantum cryptography (PQC) for future resilience.

Time is treated as a security parameter by recording two timestamps, an attested local time and the block time at anchoring, which exposes back-dating and replay. Practical immutability is achieved through finality using confirmations on public chains or Byzantine Fault Tolerant (BFT) consensus in consortium settings. Costs and availability are managed by batching event hashes into a Merkle root on chain (the root hash of a Merkle tree) and keeping encrypted logs off chain under hardware security module

(HSM) protected keys, revealing only commitments and minimal metadata. Redundant entropy sources mixed cryptographically prevent single-device bias, while telemetry on entropy, health-test failures, anchoring latency, and clock drift ensures the system fails loudly rather than silently.

5 Conclusion

This framework turns operations into verifiable facts by combining quantum-origin entropy, cryptographic commitments, and an immutable ledger. Quantum generation removes reliance on seed secrecy, while disciplined extraction and monitoring convert raw outcomes into audited minimum entropy. Scaling to 16 to 32 qubits mainly boosts throughput; quality remains a calibration issue, so the right objective is secure entropy per second. Strong commitments, signatures with a path to post-quantum cryptography, dual timestamps, and batched anchoring provide an audit trail resilient to device compromise and network manipulation, suitable for industrial compliance at scale.

References

1. IBM Qiskit Documentation, https://qiskit.org/documentation, last accessed 2025/09/06.
2. Ethereum Smart Contract Best Practices, https://consensys.github.io/smart-contract-best-practices, last accessed 2025/09/06.
3. NIST Quantum Randomness Initiative, https://www.nist.gov/programs-projects/quantum-randomness, last accessed 2025/09/06.
4. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008).
5. Herrero-Collantes, M., Garcia-Escartin, J.-C. : Quantum Random Number Generators, Rev. Mod. Phys., 89, 015004 (2017).
6. Boneh, D., Shoup, V. : A Graduate Course in Applied Cryptography (2020).

QUBO Formulation for Fault Location on Power Grid with Sparse Measurements

Eloi Gravot[1,2][0009−0003−6162−7536], Beatriz Moya[1], Sergio Torregrosa[3], Nicolas Hascöet[1], Xavier Kestelyn[1,2], Francisco Chinesta[1,4], Fikri Hafid[5], and Paul-Henri Langlois[5]

[1] RTE Chair @ PIMM Lab, Arts et Métiers Institute of Technology, Paris, France
[2] L2EP Lab, Centrale Lille, Junia ISEN, Arts et Metiers, University of Lille, France
[3] PIMM Lab, Arts et Métiers Institute of Technology, Paris, France
[4] CNRS @ CREATE, CREATE Tower, Singapore
[5] RTE, Réseau de Transport d'Electricité, La Défense, France

Abstract. The paper proposes a Quadratic Unconstrained Binary Optimization (QUBO) formulation in a bipartite power grid fault location problem using sparse measurements. The sparse approximation problem consists of solving an underdetermined system of complex-values equations. Through enforcement of the grid-depending sparsity of the problem, the amount of quantum bits necessary to solve the system is reduced, opening possibilities for industrial, large grid applications with lowered computational cost.

Keywords: QUBO · Fault Location · Power Grid.

1 Introduction

Transmission operators often face electrical faults on their lines due to various external events leading to an isolation of the line through circuit breakers. Some faults are considered fugitive; the short circuit vanishes in few milliseconds after extinguishment of the arc by circuit breakers or end of the physical contact with the line. These faults are mostly due to lightning strikes (+90%) and do not require any sort of intervention by field teams. However, electric flash can cause serious dielectric and thermal damage to conductors and isolators, reducing their respective reliability and performance. Locating these faults has an added value for implementing preventive maintenance strategies. Permanent faults are the consequence of persistent contact with a conductor (vehicle, farming equipment, trees, severed line, etc...). The line is therefore isolated until the contact is removed by field teams. To ensure fast recovery of the line and ensure system stability, various methods exist to locate the fault. Two main fault location families arise: traveling wave based, and impedance based.

The traveling wave method determines the fault location by measuring the time delay between the disturbance's arrival at both terminals of the transmission line [1]. As the disturbance travels around the speed of light, the location estimation leads to good results. This method can also be adapted to a single

terminal using the disturbance echoes. The latter may prove less accurate, as it becomes difficult to distinguish echoes from noise. Despite good localization results, traveling wave methods require extremely precise measuring equipment.

Impedance-based methods rely on the steady-state phasors of the network and impedance data of the lines. Depending on the amount of data, there are many variations. One-ended method use phasor from one end of the faulted line and two-ended from both ends, synchronized or not. One-ended method, due to many unknowns, are based on various hypotheses and lack robustness. Two-ended methods are proven to be accurate as they do not depend on the hypothesis of the network. However, they require IEDs (Intelligent Electronic Device) at each bus/node of the system [2].

The last category of impedance methods is the so-called wide area measurement. Using synchronized PMUs (Phasor Measurement Unit) scattered around the network, an underdetermined system of linear complex-value equations with a sparse solution is solved. Different approaches can be found such as LARS (Least Angle Regression) to solve the Lasso problem [3, 4] or Compressive Sensing [5–7]. These methods bring a compromise of having accurate fault location while minimizing the number of recorders on the grid. To enforce the sparsity of the solution and convergence of the optimization problem, the mentioned methods use ℓ_1-norm minimization as followed : given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ with $m \ll n$, we seek a vector $x \in \mathbb{R}^n$ that minimizes :

$$\min_x \|Ax - b\|_2^2 + \lambda \|x\|_1. \tag{1}$$

However, depending on topology of the network in fault location, the number of k nonzero values in the solution can be known a priori. Yet the ℓ_1 norm relaxation does not strictly enforce this constraint and is sensitive to the choice of the regularization parameter λ, often requiring extensive tuning to approximate the desired sparsity. By contrast, formulating the problem with the ℓ_0 norm as a combinatorial minimization would directly enforce k nonzero values, addressing both the known sparsity and the sensitivity to regularization. Although, such a problem can not be handled by iterative approach on classical computers due to the discontinuity of the ℓ_0 norm.

In recent years, Quadratic Unconstrained Binary Optimization (QUBO) formulations have gained attention for their application in solving linear problems [8, 9] due to the growing availability of adiabatic quantum computer (AQC) or GPU-based Annealing Engines (AE). QUBO problems map naturally to quantum annealing computing frameworks, which are designed to handle binary quadratic formulations directly. This approach enables true l_0 minimization, directly seeking the sparsest solution unlike LARS or CS. However, one of the downsides of solving such system with QUBO is the need of converting real-valued variables to binary ones which can become extremely costly in terms of qubits, and therefore not suited for industrial application on AQCs, or even on AEs that allow higher ranked QUBO matrices. Nevertheless, recent studies have investigated the application of quantum annealing for solving industrial-

scale problems, such as fault localization in graph-based network structures like electrical distribution grids [10].

This paper proposes a QUBO formulation for fault location on a power bipartite grid with sparse measurements. By optimizing PMUs placement, the sparsity of the solution is enforced according to the topology of a bipartite graph/grid. This formulation offers scalability advantages by reducing the amount of qubits to solve the system.

2 Problem Formulation

Consider a N-bus power grid. The nodal equation in steady-state at each bus is the following :

$$ZI = V \tag{2}$$

With Z the impedance matrix of the system, I and V the vector of buses' respective current and voltage. When a fault occurs on a power system, voltage variations ΔV can be observed on the nodes of the system. The nodal equations under a steady-state fault becomes :

$$Z\Delta I = \Delta V \tag{3}$$

The resulting ΔI is referred as the sparse virtual fault current injection vector [3]. When studying the positive sequence of the system (for asymmetrical fault), ΔI is $N \times 1$ with two non-zero complex values I_g and I_h corresponding to the two buses g and h surrounding the faulted line. The fault location on the line (from g) can be computed as follows [7] :

$$\textit{Fault location} = \frac{I_h}{I_g + I_h} * 100\% \tag{4}$$

However, solving (3) requires having PMUs on all buses of the networks. When PMUs are spread out, some buses aren't measured. The number of these unmeasured buses is denoted by u, and it's usually larger than the number of measured ones. The whole system is therefore not observable, and the nodal equation becomes underdetermined. To apply sparsity recovery algorithms such as LARS and CS, the complex-value system $Ax = b$ is transformed into a real one:

$$A_r = \begin{bmatrix} real(A) & -imag(A) \\ imag(A) & real(A) \end{bmatrix} \tag{5}$$

$$x_r = \begin{bmatrix} real(x) \\ imag(x) \end{bmatrix} \tag{6}$$

$$b_r = \begin{bmatrix} real(b) \\ imag(b) \end{bmatrix} \tag{7}$$

$$A_r x_r = b_r \tag{8}$$

3 QUBO formulation

3.1 Ising Model

The Ising model is a mathematical model from statistical mechanics [11]. It represents a system of n discrete bipolar variables called spins. Each spin s_i can take values in $\{-1, +1\}$. Spins interact with their neighbors and with an external field. The energy of the system is defined by the Hamiltonian:

$$H(s) = -\sum_{i<j} J_{ij} s_i s_j - \sum_i h_i s_i \tag{9}$$

The coefficients J_{ij} define the interaction strength between spins i and j. The values h_i represent the effect of the local field. The objective is to find the configuration of spins that minimizes the energy. In quantum computing, optimization problems are often written in QUBO form. In this model, variables are binary: $x_i \in \{0, 1\}$. The QUBO objective function is:

$$Q(x) = \sum_{i<j} Q_{ij} x_i x_j + \sum_i Q_{ii} x_i \tag{10}$$

To map the Ising model to QUBO, we use a change of variables. Each spin s_i is rewritten as $s_i = 2x_i - 1$. This transformation converts Ising variables into binary variables. It preserves the structure of optimization. This equivalence allows solving Ising problems on quantum annealers. Quantum processors like D-Wave natively work with Ising or QUBO models, as well as AEs.

3.2 Solving Linear Least Squares

In QUBO formulation, variables are binary, which is not the case in equation (8). We approximate each real-valued variable x_j using binary variables $q_{j,l} \in \{0, 1\}$ as follows:

$$x_j \approx \sum_{l=0}^{m} 2^l q_{j,l} - 2^{m+1} q_{j,\emptyset}, \tag{11}$$

where $q_{j,\emptyset} \in \{0, 1\}$ acts as a sign bit (0 for non-negative values, 1 for negative values). Substituting this binary expansion into the original system leads to a reformulation of each equation. Specifically, each row i of the expression $A\mathbf{x}$ becomes:

$$\sum_{j=1}^{n} a_{ij} x_j \longrightarrow \sum_{j=1}^{n} a_{ij} \left(\sum_{l=0}^{m} 2^l q_{j,l} - 2^{m+1} q_{j,\emptyset} \right), \tag{12}$$

resulting in a quadratic objective in terms of binary variables, suitable for QUBO formulation. We minimize a least-squares objective function:

$$min \, \|Ax - b\|_2^2 \tag{13}$$

$$\|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b \tag{14}$$

where $b^T b$ is a constant. It can be rewritten as a QUBO formulation :

$$min\ y = x^T Q x \tag{15}$$

$$Q = A^T A + diag(-2b^T A) \tag{16}$$

The major limitation of such an approach is that, even if we enforce a sparsity constraint, there will be $m+2$ qubits to express each variable. As m increases for higher accuracy, the size of the QUBO matrix scales linearly with the number of unknown u and m.

3.3 Enforcing a graph-dependent sparsity

We consider a power grid with a bipartite graph topology: vertices can be divided into two distinct and independent sets U and V. The IEEE 9-bus satisfies this conditions (Fig 1) where $V = \{v_1, v_2, v_3\}$ and $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$.

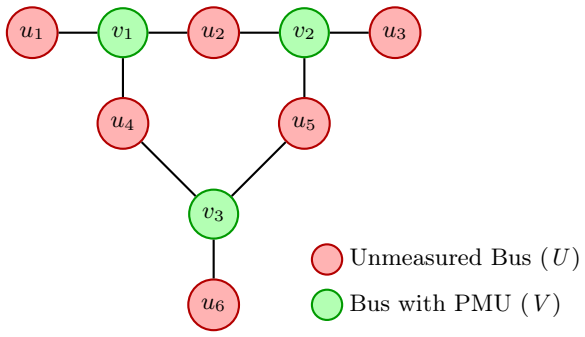


Fig. 1. IEEE 9-bus system consisting of 9 buses and 9 lines. By putting the PMUs on set V, every line has exactly one measured end, and one unmeasured.

If PMUs are placed on all the buses belonging to one set (V in Fig 1), strictly one bus of the unmeasured ones (U in Fig 1) will have a non-zero injected current value during a fault on the grid. Therefore, $real(x)$ and $imag(x)$ (6) both have a single non-zero entry at the same index. This implies that the two vectors share exactly the same support. To make the structure explicit, we introduce a additional qubits $q_{bus,i} \in \{0, 1\}$, one for each bus $i \in \{1, ..., |U|\}$, that takes 1 if the bus is at one end of the faulted line, 0 otherwise. Using this notation, the original expression of x_r (17) can be transformed to (18) where X_1 and X_2 are single real-valued approximations obtained from the underlying binary variables using (11). That is, only one approximation per vector is needed, and it is shared across all active entries, with the binary coefficients $q_{bus,i}$ determining which position is active. Moreover, to enforce sparsity, the sum of activation bus qubits is equal to one: The shunt fault happens on only one line (19).

$$x_r = \begin{bmatrix} \sum_{l=0}^{m} 2^l q_{1,l} - 2^{m+1} q_{1,\emptyset} \\ \sum_{l=0}^{m} 2^l q_{2,l} - 2^{m+1} q_{2,\emptyset} \\ \sum_{l=0}^{m} 2^l q_{3,l} - 2^{m+1} q_{3,\emptyset} \\ \dots \end{bmatrix} \tag{17}$$

$$\begin{cases} real(x) = \begin{bmatrix} q_{bus,1} * X_1 \\ q_{bus,2} * X_1 \\ q_{bus,3} * X_1 \\ \dots \end{bmatrix} & \text{with } X_1 = \sum_{l=0}^{m} 2^l q_{1,l} - 2^{m+1} q_{1,\emptyset} \\ imag(x) = \begin{bmatrix} q_{bus,1} * X_2 \\ q_{bus,2} * X_2 \\ q_{bus,3} * X_2 \\ \dots \end{bmatrix} & \text{with } X_2 = \sum_{l=0}^{m} 2^l q_{2,l} - 2^{m+1} q_{2,\emptyset} \end{cases} \tag{18}$$

$$\left(\sum_{i=1}^{|U|} q_{\text{bus},i} = 1 \right) \quad \Rightarrow \quad \forall i \neq j,\ q_{\text{bus},i} \cdot q_{\text{bus},j} = 0 \tag{19}$$

The variable change from (17) to (18) leads to, once (15) is expanded, some quadratic, cubic and quartic terms. The quartic terms necessary include bus activation qubits products such as in (19) and are therefore removed. To enforce sparsity of (19), a penalty is added to the expression (20, 21) [12].

$$P \left(\sum_{i=1}^{|U|} q_{\text{bus},i} - 1 \right)^2 \tag{20}$$

$$\Leftrightarrow P \left(2 \sum_{i<j} q_{\text{bus},i} \cdot q_{\text{bus},j} - \sum_{i=1}^{|U|} q_{\text{bus},i} + 1 \right) \tag{21}$$

The cubic terms are due to products of X_1 and X_2 qubits. These products are replaced by an additional binary variable y_i [12] and an added penalty as in the following example :

$$q_{\text{bus},i}\ q_{1,l}\ q_{2,l} \Rightarrow q_{\text{bus}}\ y_1 + P(q_{1,l}\ q_{2,l} - 2q_{1,l}\ y_1 - 2q_{2,l}\ y_1 + 3y_1) \tag{22}$$

By considering those added variables, the total amount of qubits necessary for fault location on a bipartite graph where PMUs are placed on set V and where set U contains u unmeasured bus is :

$$nb_{qubits} = \frac{(2m+4)(2m+3)}{2} + 2(m+2) + u \tag{23}$$

The first term corresponds to the added qubits due to the cubic terms of X_1 and X_2 that each have $m+2$ qubits (second term) plus the u unmeasured buses with their respective activation qubits. This formulation offers a significant scalability

advantage over the classic one (17), which requires $2(m+1)\times u$ qubits. While the proposed formulation introduces a quadratic term in m, the number of qubits needed grows linearly with the system size u, whereas formulation (16) scales as $\mathcal{O}(mu)$. It makes the approach much more efficient for large systems.

4 Example

The formulation is tested on complex-value system inspired by the IEEE 9-bus and solved on the Fixstars Amplify AE [13]. We consider the following system with A of size 3×6 (3 measured buses, 6 unknowns) :

$$\begin{cases} A = & \begin{bmatrix} 35-36j & 36-28j & -20+32j & -17-28j & -24-39j & -1+36j \\ 36-27j & -34+22j & -13+10j & -3+11j & 26+7j & -22-20j \\ -24+38j & -11-25j & -20-26j & -36-9j & 28+2j & -7+19j \end{bmatrix} \\ b = & \begin{bmatrix} -656+2960j \\ 1938+1504j \\ -840+1904j \end{bmatrix} \end{cases} \tag{24}$$

The sparse solution of the system is $-98+38j$ for x_5. Using the proposed method, the size of Q is 142×142 with accuracy set at $m = 6$. The output of the AE is displayed in Fig 2. The computation time, including the server connection, is 2.6 seconds. Once the binary values of X_1 and X_2 are transformed back to integers with (11), the computed solution is $-98 + 38j$, obtained on bus 5. Although based on synthetic data, this example demonstrates how the resulting complex output could be used to localize the fault along the line in a real-world electrical network using (4).

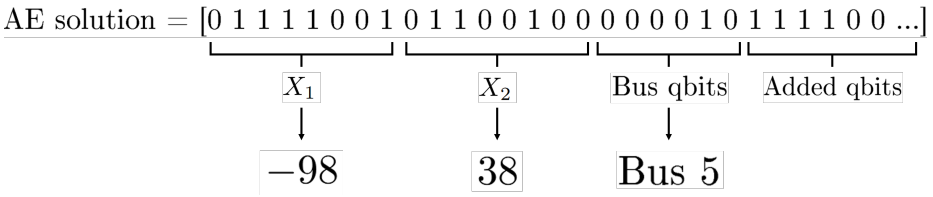


Fig. 2. First qubits of the 142 long AE output. First 8 corresponds to the real part of the solution, 8 following to the imaginary part, then the 6 bus qubits, followed by 120 added qubits due to cubic terms

5 Conclusion

This paper proposed a QUBO formulation of an underdetermined sparse complex-value linear system to solve a fault location problem on a bipartite power grid

with sparse measurements. Through enforcement of sparsity based on the topology of the graph, the formulation proposed demonstrates scalability advantages and opens doors to application on larger grids. Future work aims to adapt the method to tripartite graph that are more representative of transmission grids with also less PMUs: the proposed method relies on having a PMU at least one end of each line. Their optimal placement on the grid could be studied as a constrained graph coloring problem. Also, the data considered in this study is purely analytical, and future work will include simulations, incorporating tests with noise, to assess the performance of the proposed formulation under more realistic system conditions.

Acknowledgments. The authors acknowledge the support of the RTE research chair at Arts et Métiers Institute of Technology.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Lee, H., Mousa, A. M.: GPS Travelling Wave Fault Locator Systems: Investigation into the Anomalous Measurements Related to Lightning Strikes. *IEEE Transactions on Power Delivery* **11**(3), 1214–1223 (1996).
2. Das, S., Santoso, S., Gaikwad, A., Patel, M.: Impedance-Based Fault Location in Transmission Networks: Theory and Application. *IEEE Access* **2**, 537–557 (2014).
3. Feng, G., Abur, A.: Identification of Faults Using Sparse Optimization. In: *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1040–1045. IEEE, Urbana, IL (2014).
4. Mouco, A., Abur, A.: Improving the Wide-Area PMU-based Fault Location Method Using Ordinary Least Squares Estimation. *Electric Power Systems Research* 189, 106620 (2020).
5. Jia, K., Yang, B., Bi, T., Zheng, L.: An Improved Sparse-Measurement Based Fault Location Technology for Distribution Networks. *IEEE Transactions on Industrial Informatics* 16(1), 1–1 (2020).
6. Majidi, M., Etezadi-Amoli, M., Fadali, M. S.: A Sparse-Data-Driven Approach for Fault Location in Transmission Networks. *IEEE Transactions on Smart Grid* 8(2), 548–556 (2017).
7. Shan, H., Zhang, L., Wu, Q. H., Li, M.: Location of Asymmetric Ground Fault Using Virtual Injected Current Ratio and Two-stage Recovery Strategy in Distribution Networks. *CSEE Journal of Power and Energy Systems* 10(1), 151–161 (2024).
8. Borle, A., Lomonaco, S. J.: Analyzing the Quantum Annealing Approach for Solving Linear Least Squares Problems. In: *arXiv*, 1809.07649, pp. 1–5 (2018).
9. Jun, K.: QUBO Formulations for a System of Linear Equations. *Results in Control and Optimization* 14, 100380 (2024).
10. Bi, Z., Yang, X., Wang, B., Zhang, W., Dong, Z., Zhang, D.: Quantum annealing algorithm for fault section location in distribution networks. *Applied Soft Computing*, vol. 149, p. 110973 (2023)
11. Brush, S. G.: History of the Lenz-Ising Model. *Review of Modern Physics* 39(4), 1–10 (1967)

12. Glover, F., Kochenberger, G., Du, Y.: Quantum Bridge Analytics I: A Tutorial on Formulating and Using QUBO Models. *Annals of Operations Research*, (2018)
13. Fixstars: Fixstars Amplify. *Online documentation*, 2024. https://amplify.fixstars.com/en/docs/amplify/v1/index.html