
Operating practical quantum devices in the pre-threshold regime

By

ANTONIO ANDREAS GENTILE



Department of Physics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

OCTOBER 2019

Word count (approx): 55096

ABSTRACT

The last 10 years have seen an accelerating increase in the interest towards quantum technologies. Such interest has multifaceted motivations.

The miniaturisation of components and devices has achieved a level whereby quantum effects are not negligible, such that testing and measurement tasks require enhanced resolution and accuracies. In another context, we are observing a surge in the need for computational power, driven by the development of novel automated technologies. These are but two examples of areas where what can be achieved by mere incremental advances in classical protocols and toolsets might soon incur in physical as well as information theoretical limitations. Research efforts were thus driven towards the realm of *quantum sensing and computation*, respectively. However, such efforts are often frustrated by the important role played by noise against a successful deployment of quantum technologies, at a time when the scale of quantum devices does not satisfy the requirements for full error correction mechanisms (i.e. a *pre-threshold* regime).

This Thesis puts forward several methods, especially based on *machine learning* protocols, to mitigate the effects of noise in. We begin our work by reviewing fundamental concepts in quantum information, as well as in *integrated photonics* and *solid-state atomic defects*, here adopted as experimental platforms. Using quantum photonic chips, a novel proposal for a fundamental building block of quantum computers is here demonstrated. Moreover, we propose and demonstrate a quantum protocol for the efficient characterisation of untrusted quantum devices, automated via machine learning, and a variational algorithm for solving important instances of the eigenproblem, which is of fundamental importance to quantum chemistry. These novel protocols, and their efficiency, leverage upon the known principles of quantum simulation. We also apply machine learning methods to demonstrate efficient quantum sensing with a solid-state defect. Throughout these investigations, we deal with real-world pre-threshold devices, intended to be deployed in applications of crucial importance: this makes them *practical* regardless of achieving fundamentally superior performances over their classical counterparts.

Our approaches contribute towards a positive answer to a crucial research question: whether near-term quantum devices can deliver useful applications.

DEDICATION

- - - / - - - / - . . . - - - . - / - - . - - . - - - - / - - . - - . - - - - / - - - / - - - . - - - - - - / - - -
/ - - - / - - - - - - - - / - - - . - - - / - - - - - - - - / - - - - / - - - - / - - - - / - - - - / - - - - /
- - - - - - - - - - / - - - - - - / - - - - / - - - - / - - - - / - - - - - - - - / - - - - - - - - /

- - - / - - - / - - - - - - - - - - - - / - - - - - - - - - - / - - - - - - - - - - / - - - - - - - - / - - - - - - -
/ - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - - - /

- - - / - - - - - - - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - / - - - - - - - - - - - -
/ - - - - - - - - - - - - - - - - - -

ACKNOWLEDGEMENTS

A comprehensive acknowledgement to all the people who somehow supported (in)directly the research reported in this Thesis is... out of the scope of this Acknowledgement section! Yet, I believe at least giving it a good try is essential part of the duties for a newly “passed” Doctor. Hence, whoever is interested in reading this, bear with me a little.

First and foremost, the biggest acknowledgement goes to those, to whom this Thesis is dedicated. Words are too little to thank you, but you know how important you’ve been all along!

Concluding successfully a PhD is hardly possible, without starting one in a subject we are passionate about. After years in a (different) industry, and a highly non-linear parallel academic path, I must thank prof. Thompson who believed I could re-start in the quantum photonic realm. I might have been good in convincing you.. but still, it was you triggering this all, and giving me plenty of opportunities and freedom to explore my interests first, and present my results afterwards. The (very German) title of *DoktorVater* would belong to you.

The burden of supervising my steps in this novel (to me) area has been shared also by Dr. Barreto and Prof. Laing. I will gratefully remember Jorge’s continuous support, patience and -brutally honest where needed- advices with my projects. I admired his (apparent?) aplomb, whilst my research plans kept evolving in unpredictable ways (a sure mark of research working well, I came to learn). I thank Anthony for having believed in the projects related to machine learning applications from the very beginning, as well as including me in his very well organised (and big!!!) group of PhD students. Anthony’s genuine interest and challenging questions were a big push to refine our research outcomes.

A very special acknowledgement goes to the doctrine¹ of Doctors in Bristol’s *QETlabs*, that I am so honoured to have joined. A concentration of so many brilliant, knowledgeable people is a rare find, more so when they can freely interact to achieve a common goal, without feeling trapped in the quest. You all were a huge inspiration, and help in my everyday tasks.

Dr. Santagati is without doubt the first in this list. Raffaele’s passion towards the research he pursues played no little role in me choosing Bristol for my PhD. The time we spent discussing Science&science -or in general working together- is enormous, and contributed to shaping my interests and opinions about the field. I feel lucky I count him among my friends, not only former colleagues.

Dr. Wang and Knauer also played a special role. Jianwei and Raf helped me moving my first steps in the dark (quite literally!) of photonic labs, when almost everything was new to me. Sebastian(o) introduced me to the world of NV-centres, which have been so crucial to parts of this Thesis. His perfectionism ranging from laboratory practice to languages and cooking is monumental.

There is so many other staff people whose expert advices and hints I will sorely miss: Laurent, monsieur les plus recherchés dans les QETlabs; Döndü, our Lady of the Cleanroom; Josh, whose drivers make-the-photons-go-round; Graham, the problem-solver (ask him he’s a) doctor; Magnus,

¹ Yes, I waited years to use this collective noun in an official document, and the moment has come.

Andrew (Murray) & Pisu, irreplaceable PCB and/or chip-surgeons; Andy the polyglot engineer; Stefano (Valle) harvesting results from vibrations; the team of the Physics mech workshop ... and many others!

Research in Bristol wouldn't have been the same without the -outstanding- number of peer students who worked hard and enthusiastically towards a PhD, alongside with me. Those who contributed to specific results of this Thesis are mentioned at the end of the various Chapters. For all the others...

There might be a sprinkle of national pride here, but let me start with The BrIsTalians (in order of appearance?). Giacomo & Francesco were the very first to warmly welcome me in Bristol. With Stefano we shared a long stretch of the PhD path together, falling in love with the same topics: I have learnt loads from him, including considering Mathematica more than an outreach tool! Alberto III, the last so far of a glorious photonic lineage, with his memorable, fundamental doubts. Nicola, who could switch from mastering intricate math calculations to managing a basketball... yet I suspect he will go down in History for the first ones. Caterina, alongside with Giacomo one of the most artsy physicists I know (who can carry a tune too!). And finally Maurangelo & Massimo, highly productive meteors in the Bristol firmament.

For a scientist, outreach should always be a duty, as ultimately we are cherished² for by a community which might struggle to perceive the beauty (if not utility) of our work. I must hence thank the BOSS chapter for reminding me this principle, and I'll do this in the persons of the fellow committee members who served with me: Sam, Jeremy, Imad, Joseph & Alex, you have been great comrades of yet one more nice experience!

The cryo-team has been literally the coolest of QETlabs, and it's been great fun to be part of it! If the walls and ceilings of the Physics Labs in Bristol were not bursted away, it is also the merit of the expert hands of Gerardo, Nicola, & Ben (Slater). The pictures with the ICE-oxford cryostat being assembled will be forever the epitome of how Physics can be heavy-duty work...

The "wider Qsim group" of students under Anthony has been mentioned already, but not its many-of-a-breakfast, nor all its members, which if potentially infinite, are yet countable. I will hence thank Alex (Neville) for showing me that efficient algorithms have a life of their own, out of Computer Science departments; Patrick & Alexandre for their out-of-hours company, when the recommended daily dose of work was really not enough; Jake with his encyclopaedic knowledge of past, present (and possibly upcoming) Arxiv papers; Ross and his untiring efforts to take us numericals to the orderly elegance of theorematic approaches.

Also, it's been a special pleasure and a honour to follow closely and/or bootstrap the work of some younger PhD students. Here I am referring in particular to: Brian, the One to rule the many (computing) Cores & Hunter of bugs; Konstantina, the Lady protectress of the CU4 chip, and (sadly for only a very quick interlude at the end of my PhD) Rachel with her zoo of variables. If some projects in Bristol extended further than my physical presence there, it is your merit!

My life in Bristol has extended beyond the University buildings, and the same should happen for these acknowledgements. If my mental health is way better than it might have been, I owe this also to all the non-colleagues with whom I shared accommodation, meals, coffees, music, laughs, walks ... Anna and Eddy, my fantastic leftist landlords and their intriguing, never-ending circle of friends; Letizia, Stefano (Dal Canale) & Stella, some more great members of the wider BrIsTalians community; Bertrand, Waka & Martine as some of the best representatives of the BISCers; Flora, Michael & Samuel, who kept alive in me the sparkle of pottery and classical studies; Natasha, the very amusing find of a Welsh opera lover, along with Nuhaadh, Christina,

² I am clearly speaking of an ideal world here...

Freddie & Katrin a constant presence in artsy events; Akanksha, Mia and all the other committee members from the Enactus society...

Last but not least, the people who collaborated with me, proving distances can be easily overcome, when achieving something together is the priority. Dr. Wiebe & Dr. Granade (formerly) from *Microsoft Research*, for everything they taught me and for hinting me at how vast are Lands of Knowledge beyond my little Shire of Comfort... often without realising they were doing it! The group of Prof. Jelezko in Ulm, for sharing their world-class excellence in experiments with NV-centres. Dr. McClean and Prof. Tew for their useful help with Quantum Simulation topics. And the old time friends and board members of *MRS s.r.l.*, Salvatore & Gianluca, who kept the company alive and well in my (physical) absence...

These acknowledgements might well continue much further, hence I must finally write my apologies to all those who may read this and not find themselves mentioned: this does not mean you are forgotten! In the end, this section is possibly the most important to myself, a testament that will remind me of all the bricks that laid the foundations of my Self, when the memory of those years will fade.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE: 22/10/2019

TABLE OF CONTENTS

| | Page |
|--|-------------|
| List of Figures | xv |
| Introduction | 1 |
| 1 Methodology | 5 |
| 1.1 Quantum Computing and Information | 5 |
| 1.1.1 Ideal scenarios | 7 |
| 1.1.2 Noise in realistic quantum devices | 10 |
| 1.1.3 Error correction strategies for realistic implementations | 14 |
| 1.1.4 Potential and limitations of pre-threshold quantum devices | 15 |
| 1.2 Machine learning and quantum data processing | 17 |
| 1.3 Platforms for quantum technologies | 18 |
| 1.4 Integrated quantum photonics | 20 |
| 1.4.1 Quantum photonic gates | 22 |
| 1.4.2 Other photonic components | 27 |
| 1.4.3 A scheme for arbitrary photonic controlled-unitaries | 32 |
| 1.4.4 The CU(2) quantum photonic chip | 34 |
| 1.5 NV-centres in diamond | 40 |
| 1.6 <i>Author's contributions to the Chapter</i> | 46 |
| 2 Deploying noise-resilient building blocks of a quantum computer | 47 |
| 2.1 Phase Estimation (PE) algorithms | 47 |
| 2.1.1 Performance metrics | 48 |
| 2.1.2 Information theoretic PE | 49 |
| 2.1.3 Sequential PE Algorithm | 51 |
| 2.1.4 Kitaev's PE algorithm | 52 |
| 2.1.5 Iterative Phase Estimation Algorithm | 54 |
| 2.1.6 Rejection Filtering Phase Estimation (RFPE) | 56 |
| 2.1.7 Time-series PE | 58 |
| 2.1.8 Multi-round PE strategies | 60 |

TABLE OF CONTENTS

| | | |
|----------|--|------------|
| 2.2 | The impact of noise upon PE algorithms: a preliminary overview | 61 |
| 2.3 | Testing PE strategies on a quantum photonic reconfigurable device | 65 |
| 2.3.1 | Implementation of the IPEA protocol | 65 |
| 2.3.2 | Implementation of the RFPE protocol | 67 |
| 2.3.3 | Testing protocols' resilience against external noise | 68 |
| 2.4 | <i>Author's contributions to the Chapter</i> | 72 |
| 3 | Characterisation, Verification and Validation of quantum devices | 73 |
| 3.1 | Process tomography | 75 |
| 3.2 | Quantum Hamiltonian Learning via Likelihood Estimation | 77 |
| 3.2.1 | The prior representation problem and the SMC approximation | 80 |
| 3.2.2 | Efficient time heuristics | 84 |
| 3.3 | Efficient magnetometry with NV-centres | 86 |
| 3.3.1 | Demonstrating efficient magnetometry via CLE: Magnetic Field Learning | 88 |
| 3.3.2 | Resource efficient magnetometry with MFL and noise modelling | 95 |
| 3.3.3 | Multi-parameter estimation in MFL | 98 |
| 3.3.4 | Applications in magnetic field tracking | 99 |
| 3.4 | Characterising an (almost) blackbox quantum device: the model learning agent . | 103 |
| 3.4.1 | The QMLA protocol | 105 |
| 3.5 | Experimental CVV of NV-centres and their environment via Hahn-echo sequences | 111 |
| 3.5.1 | QMLA with a classical trusted simulator | 115 |
| 3.5.2 | Environment characterisation via hyperparametrised QMLA | 124 |
| 3.5.3 | QMLA with a photonic chip as trusted quantum simulators | 128 |
| 3.6 | Discussion and final remarks | 130 |
| 3.7 | <i>Author's contributions to the Chapter</i> | 133 |
| 4 | Quantum Chemistry with quantum simulators | 135 |
| 4.1 | Mapping quantum chemistry Hamiltonians onto a quantum computer | 136 |
| 4.1.1 | Spin-orbital configurations and qubit states | 137 |
| 4.1.2 | The mapping of Hamiltonians | 138 |
| 4.1.3 | Gadgetisation | 141 |
| 4.1.4 | Trotterisation | 141 |
| 4.2 | Quantum eigensolvers with PE-based strategies | 142 |
| 4.2.1 | Adiabatic state preparation | 142 |
| 4.3 | Variational approaches to Quantum eigensolvers | 143 |
| 4.3.1 | Energy estimation and the variational principle | 144 |
| 4.3.2 | Probe state parametrisation | 145 |
| 4.3.3 | VQE beyond ground-state estimation | 149 |
| 4.4 | A Witness-Assisted Variational EigenSolver | 150 |

TABLE OF CONTENTS

| | |
|--|---------|
| 4.4.1 The WAVES protocol | 150 |
| 4.4.2 Complexity analysis of the WAVES protocol | 153 |
| 4.5 Testing the WAVES protocol | 155 |
| 4.5.1 Molecular Hydrogen test case | 156 |
| 4.5.2 Excitonic Hamiltonians test case | 156 |
| 4.5.3 Simulated performances | 158 |
| 4.5.4 Experimental demonstration | 164 |
| 4.6 Noise resilience of the WAVES variational search | 167 |
| 4.6.1 Resilience to experimentally realistic noises | 169 |
| 4.6.2 Synthetically reduced ansätze | 170 |
| 4.7 Discussion and final remarks | 171 |
| 4.8 <i>Author's contributions to the Chapter</i> | 175 |
| Conclusive remarks and outlook | 177 |
| List of Abbreviations | 181 |
| A Appendix | 185 |
| A.1 The QFT gate and its complexity | 185 |
| A.2 Pseudo-algorithms for PE applications | 186 |
| A.3 Algorithms for Hamiltonian Learning | 187 |
| A.4 Algorithms for the Quantum Model Learning Agent | 189 |
| A.5 The Swarm Optimisation Algorithm for variational searches | 192 |
| B Appendix | 195 |
| B.1 Extracting binary outcomes from a likelihood signal | 195 |
| B.2 Qualitative reasoning supporting the Heisenberg-limit | 195 |
| B.3 Comparison of WAVES and the FS method | 196 |
| B.4 Finite-size effects of the nuclear spins bath in experimental QMLA | 197 |
| C Appendix | 203 |
| C.1 Purity as an eigenstate witness | 203 |
| C.2 Details about the excitonic unitary | 204 |
| C.2.1 Appropriate choice of evolution t for WAVES | 204 |
| C.3 Adiabatic State Preparation | 206 |
| C.4 Degeneracies in Hamiltonian models and dynamics | 206 |
| D Appendix | 209 |
| D.1 Simulation of an MZI | 209 |
| D.2 Qudit tomographies with the CU(4) chip | 210 |

TABLE OF CONTENTS

| | |
|--|------------|
| D.3 Low-power reconfigurability for cryogenic photonics | 212 |
| D.4 Characterising vibration levels in a cryogenic setup | 215 |
| Bibliography | 219 |

LIST OF FIGURES

| FIGURE | Page |
|---|-------------|
| 1.1 Bloch–sphere representation for a single qubit | 8 |
| 1.2 Photonic waveguides | 21 |
| 1.3 Physical components of a quantum photonic chip | 23 |
| 1.4 Additional physical components of a quantum photonic chip | 27 |
| 1.5 Schematic representation of the photonic chip. | 34 |
| 1.6 Schematic representation of the experimental set-up. | 36 |
| 1.7 Verification of $C\hat{U}$ chip thermal heaters | 37 |
| 1.8 Preliminary $C\hat{U}$ chip characterisation | 38 |
| 1.9 The NV–centre in diamond | 40 |
| 1.10 Coherent control of an NV–centre | 43 |
| 2.1 Quantum Circuits for PE algorithms | 50 |
| 2.2 The rejection filtering scheme | 55 |
| 2.3 General modelling of noise effects in PE schemes | 62 |
| 2.4 RFPE preliminary testing | 64 |
| 2.5 RFPE preliminary testing | 67 |
| 2.6 Effects of different experimental noises on PE strategies | 69 |
| 3.1 QHL quantum circuits | 81 |
| 3.2 The SMC approximation and LWR resampling, across different steps of the inference process | 82 |
| 3.3 Intuition behind the Particle Guess Heuristic | 84 |
| 3.4 Flowchart for the MFL protocol | 90 |
| 3.5 An epoch of the MFL algorithm including a Ramsey sequence and readout | 92 |
| 3.6 Experimental results for scaling of uncertainty and precision with MFL | 93 |
| 3.7 Analysis of the behaviour of the PGH in MFL | 94 |
| 3.8 Noise-compensation in MFL | 96 |
| 3.9 Experimental results for scaling of absolute precision | 98 |
| 3.10 Simultaneous learning of B and T_2^* | 100 |
| 3.11 Magnetic field tracking | 101 |

LIST OF FIGURES

| | |
|---|-----|
| 3.12 Schematic representation of a learning macrostep for QMLA | 104 |
| 3.13 Experimental QHL with NV–centres, inaccurate model behaviour | 112 |
| 3.14 Simulated testing of QMLA | 118 |
| 3.15 Experimental QMLA with NV–centres, success rates | 119 |
| 3.16 Experimental QMLA with NV–centres, overview | 121 |
| 3.17 Experimental QMLA with NV–centres, distribution of evolution times | 123 |
| 3.18 Average cDAG search in QMLA | 125 |
| 3.19 Hyperparametrised QMLA | 127 |
| 3.20 Quantum photonic CU(4) chip | 129 |
| 3.21 Experimental results of QMLA implemented on a photonic chip | 131 |
| | |
| 4.1 Summary of steps and techniques to map quantum chemistry problems onto quantum simulators | 137 |
| 4.2 The WAVES protocol | 151 |
| 4.3 Numerical simulations for molecular–Hydrogen Hamiltonians | 158 |
| 4.4 Convergence of the WAVES algorithm to a subspace of excited states for different Hydrogen systems | 159 |
| 4.5 Synopsis of numerical simulations of excited states searches for the molecules H ₂ and H ₃ ⁺ | 161 |
| 4.6 Behaviour comparison of the variational search in WAVES, and an equivalent implementation of the FS method | 162 |
| 4.7 Comparison between different ansätze adopted in the search for excited states in the H ₃ ⁺ system | 163 |
| 4.8 Experimental results for the variational search in WAVES | 166 |
| 4.9 Experimental results for the IPEA–based refinement in WAVES | 167 |
| 4.10 Robustness of WAVES against gate infidelities | 169 |
| 4.11 Numerical simulations of WAVES with a synthetically truncated PH ansatz | 171 |
| | |
| A.1 The circuit for the Quantum Fourier Transform | 185 |
| A.2 Flowchart for the MFL protocol | 187 |
| | |
| B.1 Absolute precision for other QPE protocols | 196 |
| B.2 Finite–size bath simulations | 199 |
| B.3 Bayes Factor stability - simulations | 200 |
| B.4 Bayes Factor stability - simulations | 201 |
| | |
| C.1 Choice of evolution t to determine the experimental \hat{U}_{exc} | 205 |
| C.2 Simulation of adiabatic ground state preparation | 206 |
| | |
| D.1 Simulation and experimental analysis of a typical MZI structure | 210 |

| | |
|---|-----|
| D.2 State tomographies on the CU(4) chip | 211 |
| D.3 Power consumption in the CU(2) chip | 213 |
| D.4 Characterisation of vibrations in a cryogenic setup | 217 |

INTRODUCTION

Recent years have witnessed an increasing interest in quantum computation and information technologies, both from a theoretical as well as a technological perspective. The motivations behind this surge are rather various, but for the scope of this Thesis we will mention but two: the augmenting miniaturisation and integration of technological devices, and the race towards computing facilities of better performances. We believe that in the first case, mainly technological challenges triggered theoretical speculations, whereas in the second scenario, the latter ones stimulated the investigation of technological possibilities. The establishing of a virtuous cycle between these research lines has possibly acted as the main drive behind the development of quantum technologies: it is hence worthy of a more detailed discussion.

Computational and simulation capacity installed worldwide relies mostly on the extraordinary development of digital, semiconductor-based technologies developed since mid 20-th century. Leveraging upon such technologies, enhancements in hardware computing performances have been mostly achieved by moving towards progressively higher integration of components in monolithic devices, thus allowing for higher speed as well as reduced consumption capabilities. This development has led to the famous *Moore's law*, an empirical statement that, formulated more than 50 years ago, has predicted accurately an exponential integration trend for decades [1]. Since the early 1990s, semiconductor manufacturing technologies have achieved miniaturisation levels whereby quantum effects cannot be neglected anymore³, and need to be explicitly modelled in the design of e.g. novel chip architectures and components. Such impressive achievements also testify the effort and investment in the fabrication of complementary metal-oxide semiconductor (CMOS) devices, estimated to total above 1 trillion USD along a few decades [5]. Nevertheless, very large scale integration is nowadays reaching intrinsic limits, with FinFET technologies achieving 5 nm in lateral dimension of gate features [6, 7]: this is approximately 20 times the distance between two Si atoms in the host lattice composing the silicon wafer.

Notwithstanding the approach of these technological challenges, it might be argued that the installed capacity is nowhere close to saturate the current needs. To quote but a few examples, historically important fields in the usage of supercomputing facilities have been climate science,

³ E.g. the behaviour of transistors at the single electron level [2], or when tunnelling currents are present [3, 4]

fluid engineering, high energy physics and atomic-level simulations⁴ of chemical systems and dynamics [8, 9]. For all these fields, an increase in computational power can at least provide higher accuracy, by making accessible better approximations to the physical systems available. Simultaneously, one can reasonably expect the rise of new fields demanding additional computational resources. This has been the case for recent developments in cryptocurrencies, and blockchain technologies more in general, which require substantial computational power for the validation of transactions [10]. An even more compelling case, as closely connected to one of the main topics in this Thesis, is provided by a renovated interest towards Artificial Intelligence (AI) or Machine Learning (ML) [11, 12]. The latest developments in this field were triggered by advances in memory availability and distributed computing that rendered feasible the exploration of new algorithms and applications for ML approaches, with the latter ones pushing in turn for further hardware developments [12].

Therefore, if further enhancements in the computational and simulation power of novel devices is required to proceed at the current rate, a substantial paradigm shift might soon be required. Along this Thesis, we will advocate for the adoption of quantum information technologies as a viable option to address the limitations of current approaches, at least in some instances of interest. We do so by presenting several examples where the performance of state-of-art quantum devices can be greatly enhanced, especially by employing appropriately classical co-processing power.

To this purpose, we provide in Chap. 1 a brief introduction of the main conceptual and technological tools adopted in the research chapters, and related to quantum computing, simulation and sensing: the areas of quantum technologies that are of interest for this work. More in detail, we will describe how information is stored and processed in quantum devices, discussing why noise -and its propagation into errors⁵- plays a fundamental role against quantum computing platforms. As it is likely to think that quantum devices will augment, but not replace, their classical counterparts, a dedicated section⁶ discusses a particular case of this interplay: machine learning methods and the processing of quantum data. In this way, we introduce a fundamental argument of this Thesis: how enhanced data processing techniques might substantially eradicate limitations that would be intricate, if not impossible, to solve via hardware. We conclude Chap. 1 with an overview of the different platforms for quantum technologies, discussing their (dis)advantages and focusing in particular upon the cases of photonic devices and atomic defects in solid state. Indeed, these two will be the implementations of our choice to perform all the experiments discussed in the research chapters.

Chap. 2 introduces a routine algorithm (QPE) of fundamental importance for a quantum

⁴ This specific case will be discussed at length in Chap. 4.

⁵ In this work, we shall often refer to a scenario described by the following epistemic nomenclature. *Experimental* stochastic and systematic *noises* inevitably affect any realistic device, and hence any measurement protocol. Such noise translates into *uncertainty* of measurement outcomes (data), and if this is not properly mitigated or handled, it ultimately leads to *errors* in the conclusions inferred by said outcomes, e.g. the result of a computation.

⁶Sect. 1.2.

computer, discussing and comparing alternative schemes suggested so far for its implementation. Such comparison is performed also experimentally for a small system, where we employed a photonic device that accomplishes for the first time a recently proposed QPE scheme (RFPE), based upon machine learning ideas. The exquisite control over the photonic chip also allowed us to experimentally investigate different noise regimes, demonstrating the superiority of RFPE in many instances of interest. Chap. 2 provides a first evidence of the advantageous performances enabled by optimal classical processing of quantum data (thus a hybrid approach), particularly in terms of noise robustness.

Chap. 3 moves from the analysis of standard building blocks of a quantum computer to the more generic problem of characterising, verifying and validating (CVV) an unknown quantum device. The problem is particularly relevant, as a trusted classical device is assumed impractical to provide the characterisation of a classical counterpart⁷. Ideas from both machine learning and quantum simulation are employed to propose a solution to the CVV problem. At first, we adopt a recently proposed method (QHL) to address instances where a relevant amount of aprioristic information is available. We test QHL experimentally using atomic defects in diamond (NV centres), demonstrating a quantum sensing protocol that achieves record sensitivities without the need for demanding setup characteristics, as further evidence of how expensive hardware resources can be sometimes efficiently replaced by software improvements. We then extend QHL ideas, introducing a novel scheme (QMLA) to approach characterisation problems with very limited aprioristic information. QMLA leverages again on hybrid quantum platforms, and we demonstrate it for the complete characterisation of an NV–centre and its environment, either with the only aid of classical simulators, or by employing a quantum photonic chip as an additional, potentially scalable quantum simulator.

Chap. 4 is dedicated to investigating the realm of applications, choosing as a study case the experimental study of quantum chemistry systems with a digital quantum simulator. Again, we observe how application instances, which would be beyond the capabilities of current quantum devices alone, become feasible when classical co-processing power is introduced. In particular, we expand on known variational algorithms to propose a protocol (WAVES) capable of finding not only the ground state of a quantum system (and its energy), but also its excited states. At the same time, WAVES enhances noise–robustness properties intrinsic of variational approaches, making it amenable for the implementation in near term quantum devices. WAVES is tested in this Thesis not only in simulations, but also in a photonic platform, reporting for the first time a successful retrieval of excited states in an experiment.

We believe that a widespread application of said approaches might focus the research attention towards those limitations that represent instead bottlenecks hard to address in post–processing. Within the space of this work, we thus systematically address and remove obstacles preventing the practicality of quantum devices and techniques, ranging from logical and physical components,

⁷As it will be discussed in the following.

INTRODUCTION

to device characterisation and application.



METHODOLOGY

1.1 Quantum Computing and Information

When considering the scenario mentioned in the Introduction, it is perhaps unsurprising that quantum computing represents one of the most pursued paradigm shifts in computing, along the last decades [5]. However, the reason behind this choice lies not only in technological developments approaching dimensions were the quantum nature of the objects can be exploited. A more compelling vision is rather provided by the conjecture that, because of their intrinsic features, quantum computers provide a broader computational model, compared to their classical counterparts. Famously,

Definition 1.1. The *extended Church–Turing thesis* states that any physically realisable device¹ can be efficiently simulated by a (probabilistic²) Turing machine [5, 13].

This statement is rather ambitious. Probabilistic Turing machines (PTM) naturally define the complexity class **BPP** [13], including all problems solvable in polynomial time by non-deterministic algorithms³, with a success rate strictly higher than 1/2. As the PTM ensemble admits a *universal*⁴ machine [14], then if the hypothesis in Def. 1.1 is true, it implies that a single universal probabilistic Turing machine provides the most general computational model that can be *efficiently* deployed.

¹ The original version of the statement targeted the simulation of an arbitrary algorithmic process, however we report here the version as in [5], which in turn reflects the approach later taken by Wolfram and Deutsch [13].

² Also here, the original formulation is adapted to take into account the challenge already posed by *randomised algorithms* [5].

³ We remind how there is increasing evidence that **BPP**=**P**, the class of problems solvable in polynomial time by a deterministic Turing machine, however this is yet to be proven [5, 14].

⁴ I.e. a Turing machine capable of simulating exactly all other machine instances of the same class.

Now much of the expectations for the computational model offered by a *universal quantum computer*⁵ (UQC), as formulated in [15], comes from the widely believed conjecture that **BPP** ⊂ **BQP**, the class of problems that can be solved in polynomial time, within a bounded error, by a quantum computer [5]. Well-known instances support this conjecture. For example, order-finding algorithms implemented on a quantum computer offer a $\mathcal{O}(n^2 \log n \log \log n)$ solution to the factorisation in prime numbers of a generic integer n [16]. This is an exponential speedup compared to the most efficient known classical algorithm, the number field sieve, which is (asymptotically) $\mathcal{O}(\exp(\log^{1/3} n \log^{2/3} \log n))$ [17]. Other examples do not even require a universal quantum computer, but more specialised devices known as *analogue quantum simulators*. These devices are not universal, but nevertheless they could offer solutions to problems that are classically intractable. One such example is the simulation of systems derived from quantum chemistry, a case discussed in detail in Chap. 4. Other examples are offered by even more special-purpose devices: it has been shown that sampling from the output of a network of linear-optical elements is unlikely⁶ to be efficient on a generic classical device [18], whereas recent claims for accomplished quantum supremacy⁷ involved sampling the output from randomly generated quantum circuits [19, 20].

Finally, a separate remark is required to cover the field of *quantum sensing*. In this case, computational and simulation performances are not the focus, but enhancements in sensitivity and/or noise-robustness of sensors and measurement protocols, for a wide variety of physical quantities. Classically, one expects to reduce the measurement uncertainty for an unknown parameter x , at a rate proportional to $1/\sqrt{N}$, when using N physical resources⁸. E.g. these might be N independent probes, and measurements thereafter, used to test the phenomenon depending upon x . This scaling is known as the *shot-noise limit* (SNS) [21].

The definition of a quantum sensor encompasses a broad range of properties that might be exploited in order to gain quadratically⁹ advantageous usage of measurement resources, thus overcoming the SNS. The speedup is thus intrinsically related to optimally extracting the quantum information provided by probe states interacting (and thus being manipulated) by the system or phenomenon under study [22]. According to [23], the properties identifying a quantum sensor might be (not mutually exclusive):

- i the availability of quantised stable energy levels, because the sensor is a quantum object;
- ii quantum coherence among the probe states employed for the sensing;

⁵ Note how a gate-based quantum computer is universal under a weaker definition than Turing's, as the requirement for *exact* simulation of all other machines in the same class is replaced by the arbitrary precision simulation condition, as in Thm. 1.1.

⁶ Here intended, that it implies a multiple-level collapse in computational classes, an occurrence that in the community is regarded as a *reductio ad absurdum* [13].

⁷ Refer to Def. 1.2 below.

⁸ This scaling can be formally deduced from the central limit theorem if x estimates the mean of a random distribution, as e.g. generated by stochastic noise affecting each measurement.

⁹ This will be detailed in Chap. 3.

iii entangled¹⁰ or squeezed¹¹ quantum probe states.

When the third case holds, the literature usually refers to *quantum metrology* [25]. In some cases, the availability and addressability of systems as in (i) is made possible by exquisite manufacturing and control at the nanoscale, as it will be shown in Sect. 1.5, so that extreme spatial resolution might be possible simultaneously with sensitivity beyond what classically achievable [23]. However, this need not be the case, with macroscopic setups providing flamboyant demonstrations of the possibilities offered by quantum metrology, e.g. in the detection of gravitational waves [26]. Therefore, quantum sensing is often regarded along with *quantum communication* as one of the most promising areas, where commercial applications will be made available via the so called “2nd quantum revolution” [27].

In conclusion, if fully developed and deployed, quantum devices would make tractable instances of problems, or offer new generations of sensors and sensing protocols, that are well-beyond mere technological improvements in their classical counterparts. In our opinion, this justifies the efforts in attempting the development of either a UQC, or specialised instances of quantum devices.

1.1.1 Ideal scenarios

It is well-known how *bits* are the fundamental units for digital classical information. *Qubits* are their counterparts in the realm of quantum information, and describe the state of a quantum system represented by a 2-dimensional Hilbert space \mathcal{H} . Therefore, each qubit is a vector in \mathcal{H} and thus admits a decomposition in terms of any orthonormal basis $\{|b_i\rangle\}$ of \mathcal{H} . A particularly convenient choice is known as the *computational basis*, given by the states $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. A generic *pure* qubit state $|\psi\rangle$ can be any superposition state:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle, \quad (1.1)$$

whose normalisation is ensured by $\cos^2(\theta/2) + e^{2i\phi}\sin^2(\theta/2) = 1$. The parameters describing the state are known as *azimuthal* (θ) and *polar* (ϕ) angles [5], because they can be immediately visualised as the angles of a unitary-radius sphere, the *Bloch sphere* (Fig. 1.1). It is crucial to stress that digital classical systems do not admit any superposition like Eq. 1.1, as the only states allowed are {0, 1}.

Quantum states need not be pure, as they might be an incoherent statistical mixture of pure states. A generic description for such *mixed states* is given by their *density matrix*:

$$\rho = \sum_i p_i |b_i\rangle\langle b_i| \quad (1.2)$$

¹⁰See Sect. 1.1.1.

¹¹A quantum state represented by an ellipse in the quadrature phase space, see [24] for details, as squeezing will not be used in the following of this Thesis.

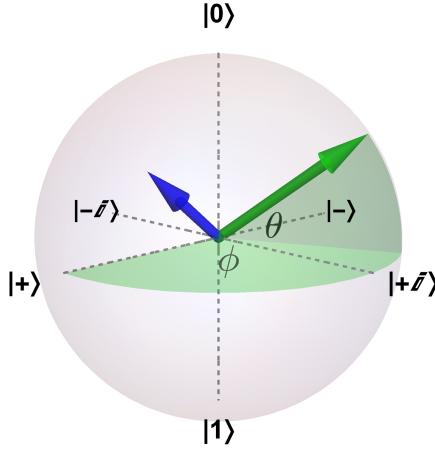


FIGURE 1.1. Bloch–sphere representation for a single qubit. Pure states have norm $|r| = 1$ thus lying on the surface of the Bloch sphere, whereas mixed states have norm $|r| < 1$. Angles θ and ϕ are reported for a generic pure state.

where the terms p_i are the probability of occupation for each basis state $|b_i\rangle$, satisfying $\sum_i p_i = 1$. ρ is thus a Hermitian, positive operator of trace 1. The norm of the state vector $|r|$ is used to convey information about the quantum state’s purity \mathcal{P} , as $(1 + |r|^2)/2 = \text{Tr}(\rho^2) \equiv \mathcal{P}$. Therefore, pure states have norm $|r| = 1$, whereas for mixed states $|r| < 1$ (see Fig. 1.1).

Quantum operations describe the process of evolution of a generic input state ρ injected in the quantum device. Their most general representation is given by trace–preserving, completely positive linear maps $\mathcal{U}(\rho)$ [5]. $\mathcal{U}(\rho)$ provides a valid description also for an open quantum system¹², whereby the input and output Hilbert spaces are not necessarily identical. Therefore, it is also an effective way to represent the evolution of states through a noisy quantum channel.

However, in several cases of interest the quantum device is approximated as an isolated system. In other words, the evolution of states injected in the system is similar to the evolution of bits through a noiseless classical channel. Within this approximation, the generic manipulation of a qubit state is given by a unitary operator \hat{U} of \mathcal{H} [5]. The evolution of a generic quantum state under \hat{U} :

$$|\psi\rangle \xrightarrow{\hat{U}} \hat{U}|\psi\rangle \quad (1.3)$$

$$\rho \xrightarrow{\hat{U}} \sum_i p_i \hat{U} |b_i\rangle\langle b_i| \hat{U}^\dagger = \hat{U} \rho \hat{U}^\dagger, \quad (1.4)$$

with Eq. 1.4 descending immediately from Eq. 1.3 and the definition 1.2. Within the context of quantum computing, a \hat{U} operating on a qubit state is also known as a *quantum gate*, being the equivalent of gates in the classical computing realm.

¹² The $\mathcal{U}(\rho)$ formalism holds under general assumptions for any open quantum system [28], starting from an initial factorised state, i.e. the same case assumed later in Eq. 3.26.

General purpose quantum computations intuitively require the arbitrary manipulation of qubits. Implementing a dedicated physical gate for any \hat{U} would evidently undermine the very feasibility of the approach. Fortunately, physical implementability is ensured by the:

Theorem 1.1. (Solovay–Kitaev) theorem. *A finite sequence of quantum gates, selected from a finite universal set of gates, is sufficient to approximate any other quantum gate operation¹³, with arbitrary precision.*

Several universal gate sets are known, with $\{\text{H}, \text{T}, \text{CNOT}\}$ often quoted as an exemplary case, where:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{pmatrix}, \quad \text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (1.5)$$

H is known as the *Hadamard* gate, whereas the T gate is also named the $\pi/8$ gate. The controlled-NOT (*CNOT*) gate is a two-qubit gate, and its name can be easily inferred from the action upon qubits prepared in the computational basis. Indeed, it can be described in words as a logical not operation on the second qubit, conditional upon the state of the first qubit (which remains unaffected by the operation). E.g. for two qubits: $|00\rangle \xrightarrow{\text{CNOT}} |00\rangle$, but $|10\rangle \xrightarrow{\text{CNOT}} |11\rangle$. More generally, with the first qubit in a superposition of the computational basis states:

$$(a|0\rangle + b|1\rangle) \otimes |0\rangle = a|00\rangle + b|10\rangle \xrightarrow{\text{CNOT}} a|00\rangle + b|11\rangle \quad (1.6)$$

Eq. 1.6 shows why controlled operations are rather central in quantum computation: we started from a *separable* state¹⁴ and evolved it into a non-separable one, i.e. an *entangled* state. The generalisation of a *CNOT* gate is formalised by the introduction of *controlled unitaries* (CU), where the *NOT* operation is replaced by an arbitrary unitary operation \hat{U} , acting upon all (*target*) qubits other than the first, that acts again as a *control* qubit.

In the simpler case of a single qubit, its arbitrary manipulation requires complete access to the states as described by the Bloch sphere in Fig. 1.1. This case can thus be mapped on the problem of controlling the two angles θ, ϕ describing a generic qubit state. Therefore, a complete set of 1-qubit gates is simply $\{R_z, R_y\}$:

$$R_z(\phi) = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}, \quad R_y(\theta) = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}. \quad (1.7)$$

It must be emphasised how the practicality in the implementation of quantum computing is not ensured by Thm. 1.1 alone. Indeed, a generic \hat{U} for an n -dimensional system involves

¹³ The original formulation of the Solovay–Kitaev theorem approximates any *single-qubit* gate operation, yet this result was readily extended to multiple qubit gates [29]. Therefore, we adopted here a comprehensive definition.

¹⁴ I.e. as emphasised at the start of Eq. 1.6, a state that can be expressed as a product state of the various qubit systems.

2^{2n} free parameters, so that the number of gates decomposing it, if finite, might still scale exponentially with the system's size [30]. A solution to the conundrum is offered by making additional hypotheses about the nature of \hat{U} . For example, a typical unitary of interest is offered by the equation of motion for an isolated quantum system, governed by its Hamiltonian \hat{H} : in this case, one is interested in the implementation of $\hat{U} = e^{-i\hat{H}\tau}$, with τ the evolution time. As we will detail in Chap. 4, for many systems of interest the global Hamiltonian can be written as $\hat{H} = \sum_i^M \hat{h}_i$, with M scaling sub-exponentially, and the primitives \hat{h}_i are two-body¹⁵ Hamiltonians [5]. In such cases, a crucial result is offered by the *Liu-Trotter decomposition* [32]:

$$\lim_{L \rightarrow \infty} \left(\prod_i e^{\frac{-i\hat{h}_i \tau}{L}} \right)^L = e^{-i \sum_i \hat{h}_i \tau} \quad (1.8)$$

where already low-order (in L) approximations provide reasonable results. In particular, it has been estimated that to encode the evolution of an N -orbitals system, less than $N^{8+o(1)}$ exponential operations are required [33]. In turn, approximating the latter ones using a universal gates set accrues an overhead that scales logarithmically with the targeted accuracy [34]. Hence Eq. 1.8 leads to an efficient protocol for decomposing a unitary evolution in a practical number of gates.

As a quantum device admits as input a generic superposition state, say $\sum_i b_i |x\rangle_i$, and quantum operations are linear, the device will also prepare a superposition of the outputs, e.g. $\sum_i b_i \hat{U} |x\rangle_i$, that encodes the evolution outcome for each input $|x\rangle_i$. This property is known as *quantum parallelism* [5].

Finally, after the intended evolution of the qubit has been performed, one needs to extract information from the outcome. This is typically achieved by *projective measurements*, i.e. once chosen a measurement basis $|b\rangle_i$, the generic output ρ from the quantum evolution can be processed by extracting the probabilities of occupation:

$$p_i(\rho) = \text{Tr}(\rho |b_i\rangle\langle b_i|) \quad (1.9)$$

for each basis state. Measurement is a crucial operation to achieve quantum advantages over classical counterparts. In fact, from Eq. 1.9 it is evident how trivial measurements of the computational output would invalidate any advantage provided by the aforementioned quantum parallelism, as the classical information extracted collapses the overall information carried by ρ . Therefore, to retain a quantum advantage, algorithms must be envisioned, capable of extracting information from a “superposed processing” efficiently [5]. In the next Chapters we will encounter and discuss some exemplary algorithms.

1.1.2 Noise in realistic quantum devices

In Sect. 1.1.1, we have provided a brief outline of how state preparation, evolution and measurement occur in a quantum device, with a particular focus dedicated to gate-based devices,

¹⁵ The fact that two-body contributions might suffice to render the dynamics generated by \hat{H} can be thought as loosely connected to the N -representability problem [31], checking where an N -body hamiltonian problem can be reduced to a corresponding two-body problem.

as they are to provide the basis for most of the results in this Thesis. However, insofar we implicitly assumed a perfect implementation of any component. Evidently, this is far from true, and imperfections driven by the physical limitation of realistic devices can influence greatly what can be achieved, or not. Quantum devices suffer from a wide variety of noise sources, which might be either generic or peculiar to a specific implementation. A non-exhaustive list helps to depict a broad scenario for mechanisms that alter ideal behaviour:

1. (in)fidelities in the state preparation and evolution;
2. (de)coherence phenomena;
3. losses;
4. imperfect measurement.

The noise leading to (1) describes the ultimate unavoidable inaccuracies of a realistic device, when implementing gate-based operations such as those described in Sect. 1.1.1. In order to validate the behaviour of a device, two effective metrics can be introduced. The first is the *state fidelity*:

$$F_{\psi^*} = |\langle \psi | \psi^* \rangle|^2 \quad (1.10)$$

$$F_{\rho^*} = \left[\text{Tr} \left(\sqrt{\sqrt{\rho^*} \rho \sqrt{\rho^*}} \right) \right], \quad (1.11)$$

respectively valid for pure and mixed states [35], where the ideally targeted state is labelled $|\psi^*\rangle$ (ρ^*), and the one actually prepared in the device $|\psi\rangle$ (ρ).

In order to quantify instead the fidelity for a *quantum operation* \mathcal{U}^* , the usual approach is not direct, as it would be to characterise e.g. the matrix describing the unitary evolution \hat{U} for the system, and then compute a distance against the targeted unitary \hat{U}^* . Instead, one usually introduces a representative set of input states $\{\rho_{\text{in}}\}$, that is used to define an average gate fidelity [36]:

$$\bar{F}_{\{\rho_{\text{in}}\}} = \sum_{\{\rho_{\text{in}}\}} F_{\rho_{\text{in}}} / \dim(\{\rho_{\text{in}}\}), \quad (1.12)$$

where each term $F_{\rho_{\text{in}}}$ is computed as in Eq. 1.11, with $\rho^* \equiv \mathcal{U}^*(\rho_{\text{in}})$, whilst $\rho \equiv \mathcal{U}(\rho_{\text{in}})$ is the state actually prepared by the real quantum channel \mathcal{U} . The reason behind this approach is pragmatic¹⁶, as a subset of input states might be of special interest, rendering the overhead of a complete process characterisation pointless¹⁷, whilst providing a simple metric also in the generic case of a quantum channel \mathcal{U} . However, it was pointed out that the approach in Eq. 1.12 does not highlight the readiness for fault-tolerant quantum computation [37], as this is bound to the error rate $\epsilon \in [0, 1]$, with which the correct output is extracted from the quantum device. However, also in [37] is claimed that a tight upper bound for the error rate can be inferred from \bar{F} :

$$\epsilon \leq d \sqrt{(1 + d^{-1})(1 - \bar{F})} \quad (1.13)$$

¹⁶ For example, it can be easily evaluated via the randomised-benchmarking mentioned in Sect. 3.1.

¹⁷ See also Sect. 3.1.

with d the dimension of the Hilbert space involved.

Decoherence is intrinsic to many realistic quantum devices, capturing the high sensitivity of most quantum states, to contributions from an external environment. It is an exclusively quantum effect, that has no parallel in the theory of classical noisy channels [5]. If we assume to start with a 1-qubit pure superposition state $|\psi\rangle_{\text{in}} = a|0\rangle + b|1\rangle$, decoherence can be modelled via a *phase damping* channel \mathcal{D} [5], whose output is:

$$|\psi\rangle_{\text{in}} \xrightarrow{\mathcal{D}} \begin{pmatrix} |a|^2 & ab^* e^{-\lambda} \\ a^* b e^{-\lambda} & |b|^2 \end{pmatrix} \quad (1.14)$$

where the crucial parameter is the *dephasing rate* λ , dictating the exponential decay in the off-diagonal elements of ρ , which in turn describe the relative phase between the system's eigenstates. In many quantum devices of interest (see below), $\lambda \propto \tau$, the evolution time, so that a (*de*)coherence time¹⁸ T_2 is introduced, defined by $e^{-t/T_2} \equiv (1 - \lambda)$. Decoherence phenomena can also be modelled via a series of discretised *phase-flip* processes [5].

Another typically quantum noise effect is the depolarisation of a qubit, i.e. its transformation, with probability $1 - p$, in the corresponding maximally mixed state, $\tilde{\rho} \equiv \sum_{b \in \{|0\rangle, |1\rangle\}} |b\rangle \langle b| / 2$. Depolarising channels can also be understood in terms of decoherence phenomena. E.g. taking $\lambda = +\infty$ in Eq. 1.14 and $|\psi\rangle_{\text{in}} = |+\rangle$, we obtain from \mathcal{D} a depolarising channel with $p = 0$.

Losses have an immediate parallelism with the theory of classical noisy channels, where they intuitively describe the loss of the physical carrier of information. Also losses in quantum devices require the interaction with an environment, being dissipative phenomena. Their role in altering quantum information can be interpreted via some of the famous Di Vincenzo's criteria for an ideal qubit, requiring that no energy levels other than the two defining the qubit are accessible, and the latter are stable for times much longer than those required by the computation [38]. As no real qubit can be thought as perfectly isolated from the environment, the system will naturally decay at a certain rate to its ground state. Now, this is often used to represent the $|0\rangle$ in the computational basis, so that the system will tend to collapse in it. This phenomenon is known as *amplitude damping*, and can be modelled as a map \mathcal{A} [5]:

$$\begin{pmatrix} a & c \\ c^* & b \end{pmatrix} \xrightarrow{\mathcal{A}} \begin{pmatrix} 1 - (1 - \gamma)(1 - a) & c\sqrt{1 - \gamma} \\ c^*\sqrt{1 - \gamma} & b(1 - \gamma) \end{pmatrix} \quad (1.15)$$

Eq. 1.15 describes an energy loss to a dissipative environment with rate $\gamma \in [0, 1]$, however the extreme case $\gamma = 1$ can also be used to build a circuit-model where the information carrier leaks to a level that is not embedded in the qubit description. For example, this could be the case for additional energy levels in a spin-qubit, or photonic qubits scattered away from the quantum channel by the optical medium, as we will detail later.

Amplitude damping is a purely dissipative process. However, some environments might interact reversibly with the system upon short timescales [28], so that preserving our simple

¹⁸ Also named the *dephasing time*.

example of $|0\rangle$ representing the ground state, environment-induced transitions $|0\rangle \rightarrow |1\rangle$ become also possible. This requires the introduction of *bit–flip* channels, defined via the map \mathcal{F} [5]:

$$\begin{pmatrix} a & c \\ c^* & b \end{pmatrix} \xrightarrow{\mathcal{F}} p \begin{pmatrix} a & c \\ c^* & b \end{pmatrix} + (1-p) \begin{pmatrix} b & c^* \\ c & a \end{pmatrix} \quad (1.16)$$

so that the probabilities of occupancy of the eigenstates are left untouched with probability p , or flipped with probability $1 - p$. Amplitude damping or, more in general, bit–flip provide a description for relaxation phenomena that occur with inelastic spin–flip [39]. We will return to this later on.

Even if the evolution of each qubit was ideal, errors might still occur (or be amplified) when extracting classical information from the evolved quantum state. Projective measurements are the most common way to extract information about a quantum output ρ , as it will be discussed in details throughout this Thesis. Restricting again for simplicity to a single qubit case, it is well-known how the variance of a projective measurement with the operator $\hat{P}_b \equiv |b\rangle\langle b|$ is $\sigma^2(\hat{P}_b) = \text{Pr}[b](1 - \text{Pr}[b])$, with $\text{Pr}[b]$ the probability of the output state to be projected onto $|b\rangle$ [40]. This *quantum projection noise*¹⁹ leads to binomially distributed outcomes, which play a fundamental contribution to the ultimate uncertainty that can be achieved via a single measurement. Such uncertainty is maximal when the output is a superposition state in the measurement basis chosen.

To this fundamental limit adds also the non–ideal behaviour of the measurement apparatus. Ultimately, measurements upon qubits for all the platforms described in Sect. 1.3 rely on recovering a signal from a noisy background, be it originally electronic or converted e.g. from an optical signal. Therefore, established techniques from signal recovery apply, and in particular the *receiver operating characteristics* for binary classification tasks²⁰ [41]. With this approach, a unit processing signals can be characterised by its false–positive and false–negative rates. These might be low, but not null for a realistic device²¹

The usual way to overcome the difficulties of probabilistically distributed outcomes from the final measurement is to repeat the complete computation and measurement, in order to accrue outcome statistics and thus estimate more accurately each $\text{Pr}[b]$ [5], by improving the signal-to-noise ratio (SNR). However, it must be born in mind how many repeated measurement imply overheads that might impact heavily upon the performances of the protocol²², but in contrast insufficient statistics might lead to ambiguous results. These will require appropriate post–processing, not to affect the final outcome.

¹⁹ We here retain the traditional *noise* nomenclature used in [40], even if here the noise affecting the measurement is readily modelled into an outcome uncertainty.

²⁰ Which are routinely performed in qubits–operated architecture, as projective measurements will e.g. occur upon the binary computational basis.

²¹ For example, the optical, electronically read–out detectors that are introduced in Sect. 1.4, trigger unwanted detection events because of thermal and electronic noise. At the same time, dissipative mechanisms or photon losses within the detection device might lead to undetected genuine events.

²² See in particular Sect. 3.3.2 for an exemplification regarding quantum sensing.

1.1.3 Error correction strategies for realistic implementations

Reducing the impact of noise has involved considerable research efforts already in the fields of communication and storage of classical information. The additional fragility of its quantum counterpart makes error correction strategies even more crucial. The basic idea of all error–correcting protocols is to *encode* quantum information in a redundant way²³, thus adding resilience against its degradation due to noise [5]. A final *decoding* allows to recover the original information, when needed.

An important logical step in the development of error–correcting codes for quantum computation is that a general modelling of the noise occurring in a realistic quantum device can be made, e.g. following Sect. 1.1.2. In this way, robustness can be achieved making but a few generic assumptions about the physical mechanisms introducing errors in the targeted evolution of the qubit states [5]. Several strategies have been developed to detect errors as they occur in the quantum circuit, and correcting them on–the–fly to achieve *fault–tolerant* quantum computation [5, 42, 43] and interesting small–scale demonstrations were already achieved [44, 45]. We will not list nor discuss error–correcting protocols, as they are no necessary background for the rest of this Thesis. However, the general idea of *concatenating* the encodings adopted to cope with noisy devices leads us to a central result [5]:

Theorem 1.2. (*Threshold for fault–tolerance*) *Given a quantum circuit of n qubits and G_n gates, using noisy hardware whose gates have all an upper–bounded error probability $\epsilon_g < \epsilon_{th}$,*

$$\mathcal{O}(\text{poly}[\log G_n/\epsilon]G_n) \quad (1.17)$$

gates suffice for the overall circuit error probability to be at most ϵ .

The scalable overhead provided by Eq. 1.17 is among the strongest arguments that advocate in favour of the feasibility of a UQC. Nevertheless, a favourable scaling is not enough to ensure optimism for near–term, large–scale fault tolerant quantum computations. Further analysis of Thm. 1.2 has unveiled important practical consequences. First, in order to avoid excessive resource overheads, ϵ_g should be well below 10^{-2} , possibly in the range of 10^{-4} [5, 42, 46, 47], which is already beyond the limits of many experimental platforms (Sect. 1.3). Moreover, these results often make the assumption of available parallelisation of the encoded states evolution, i.e. also overheads in the number of qubits must be imposed. It is currently estimated that $\sim 10^3$ physical qubits are required to obtain a generic, error–protected *logical qubit* in current platforms [48], even if proof–of–concept encodings might be much less demanding [43].

These results led to an interesting research question [48], that is at the core of Chaps. 2–4 in this Thesis. Once chosen a particular experimental platform, we are bound by its physical

²³ We remind that the parallelism with classical error correction codes, and the redundant encoding of information, must not be intended in a strict way, as the famous *no-cloning* theorem proves the impossibility to clone an arbitrary quantum state [5].

limitations to a certain maximum overhead in the number of gates, and this in turn dictates the necessary ϵ_{th} to achieve fault tolerance. Now, can we use a noisy quantum device with $\epsilon_g > \epsilon_{th}$ (i.e. *pre-threshold*²⁴), to perform useful quantum algorithms, and to what extent?

We will start answering this question in the following sections.

1.1.4 Potential and limitations of pre-threshold quantum devices

At the time of writing, it is still unclear whether the limitations outlined in Sect. 1.1.2 might represent a fundamental obstacle to the development of a fully error-corrected quantum computer. If confirmed, such a noise-related backstop would be similar to analogue classical computing (ACC) failing to provide a computational model, more general than PTM machines'. Indeed, the speculated capability of analogue computers to invalidate the extended Church–Turing thesis, as in Def. 1.1, was almost completely dismissed after analyses considering reasonable noise levels proved such computational advantage to be unfeasible [5, 49].

It must be noted that a major difference between quantum computers and classical analogue computers, is the development of error correction methods²⁵, ideally capable to counteracting noise sources, and not available in ACC architectures. Nevertheless, in recent years some works were published, criticising the very feasibility of handling the overheads, especially in terms of ancillary qubits to achieve threshold gates' fidelities, that error-corrected QC architectures require [50, 51]. Nevertheless, the conjectural nature of these works does not provide definitive evidence against the ideas in Sect. 1.1.3.

We will not address further this dispute, as a full error-corrected quantum computing platform is well beyond the scope of this Thesis, that instead exploits in different directions the capabilities of Noisy Intermediate-Scale Quantum (NISQ) devices [48]. This broadly defined class is intended to describe those quantum devices embedding a few tens to hundreds of qubits. Their general behaviour is already impossible to classically simulate and thus predict, within reasonable timings, even adopting the best supercomputers available [13, 48]. Therefore, even if no perfect control over each qubit is granted (i.e. they are *noisy*), so that they do not satisfy the threshold theorem (i.e. they are *pre-threshold*), such platforms are ideal candidates to falsify the extended Church–Turing thesis. The route towards this scientific milestone is to address specific computational tasks, by running instances of problems for which efficient quantum protocols are known, but simultaneously believed to be hard for classical computers [52].

Definition 1.2. A (NISQ) device would demonstrate *quantum supremacy*²⁶, if capable of running successfully a problem instance large enough to be intractable for the best classical device within

²⁴In some cases, it can be expected that $\epsilon_g \gg \epsilon_{th}$.

²⁵An outline has been provided in Sect. 1.1.3.

²⁶We observe that the wording *quantum advantage* is sometimes preferred for various reasons. We keep here the original word from [53], to avoid confusion with quantum advantage in the meaning exploited by [54], and used in the following also for quantum sensors.

a reasonable allotted run–time, even when allowing for near future technological or algorithmic enhancements [53, 54].

Notably, quantum supremacy does not imply any *practical utility* of the task being performed [52], nor it implies the achievement of a universal computational model²⁷, encompassing PTM’s. A few candidate experiments have been suggested along the years, the most promising being: boson–sampling²⁸, Arthur–Merlin games on constant depth quantum circuits and random quantum circuits [52]. However, since Def. 1.2 embeds several qualitative statements, it has been so far elusive of being bound to the success of a specific experiment. For example, dramatic improvements in the classical algorithmic toolsets for some of these problems have often redefined the size of the task instance that would demonstrate quantum supremacy, see e.g. [55] for quantum annealers, or [54] for boson sampling.

One might argue how working with instances whereby likely non–optimal quantum protocols prove significative advantages over possibly non–optimal classical protocols, signifies that NISQ devices are operating close to the so–called *entanglement–frontier* [53]. As the tasks afforded become increasingly hard, it is perhaps not surprising how very recent announcements of proven quantum supremacy have sparked not only interest, but confidence [19].

Finally, we observe how quantum sensors can also be considered members of the NISQ class. All quantum sensing demonstrations so far, indeed, have been performed using pre–threshold devices, and many sensing protocols do not explicitly require QEC, even if scenarios have been considered, where QEC could provide enhancements [23]. Also in sensing, ensembles of sensors – where each can be thought of as a qubit – might exploit purely quantum advantages in the (scaling of) the device sensitivity. Therefore, the realisation of intermediate–scale devices might enable performances out of reach for classical sensors. Nevertheless, as aforementioned and demonstrated in detail in Sect. 3.3, the *scale* of quantum sensors is not as crucial as in the computational case. Indeed, single qubit sensors can already enable the demonstration of quantum protocols like phase–estimation [56, 57], beating the SMS limit. The reason why a single qubit here suffices to observe advantages, beside pragmatic improvements due to the sensor’s nanoscopic size, is that parts of a multiple qubit quantum circuit can be replaced by measurement and classical feedback (see Chap. 2 for detailed examples) [23, 58]. Clearly, however, no demonstrations of said sensing advantages can invalidate any version of the Church–Turing thesis, as measurements of physical quantities do not represent per–se a computational model, nor a protocol for the simulation of a system²⁹.

²⁷ Therefore, quantum supremacy invalidates the *extended* Church–Turing thesis, but not its *strong* form. The latter indeed states that any universal model of computation can be simulated on a PTM with at most a polynomial overhead. I.e. PTMs would be the most general efficient universal simulators that can be designed.

²⁸I.e. the simulation of linear-optical networks.

²⁹ Which is instead the case behind e.g. recent claims in [19].

1.2 Machine learning and quantum data processing

The interplay between machine learning (ML) and quantum technologies can occur at several levels. Each instance advocates for advantages estimated according to general hypotheses about the intricacies in implementing specific computing components (or, more broadly, architectures). Drawing upon similar nomenclatures, proposed in the literature provided below, we propose here the following partition in order to provide a framework for our findings in this and later chapters:

1. Quantum (Enhanced) Machine Learning (QML): the application of quantum algorithms and hardware to machine learning tasks;
2. Quantum-inspired machine learning (QIML): the development of algorithms designed for running on classical hardware, yet inspired by the realm of quantum mechanics;
3. Machine Learning Enhanced Quantum Algorithms (MLEQA): the application of classical machine-learning methods, and computational resources, to enhance the performance of algorithms run on quantum hardware.

The key distinctive feature of QML is the usage of a quantum device for processing the data. Data, in turn, need not be natively quantum data, as they might be mapped on the quantum architecture starting from a classical dataset [59, 60]. The main interest in QML derives from its applications in basic linear algebra subroutines (BLAS), classifying data either via Support Vector Machines (SVM, [61]) or Principal Component Analysis (PCA, [62]), and least-square fitting [63]. In particular, the possibility for QML to provide exponential speedups over best-known BLAS classical routines has been shown for solving linear systems of equations [64]. QML is also expectedly advantageous for supervised labelling tasks [65, 66], recommendation systems [67] and even Natural Language Processing (NLP) [68]. Finally, quantum extensions of classical Neural Networks (NN) models have been proposed using a variety of mappings [69–71]. This opens a plethora of particular applications to traditional classical machine learning tasks [59].

On the contrast, QIML does not rely at all on quantum hardware [72], but rather on the application of concepts and methods originally quantum-mechanical. This approach originated from proposed improvements of classical genetic algorithms [73], and has since evolved to provide a toolset adopted for a variety of tasks. Exemplary applications are more efficient classical algorithms for the (approximate) estimation of matrix permanents [54, 74] originating from intuition in quantum optics experiments or in sampling schemes, or for the training of NN's [75] inspired by its QML counterpart. Interestingly, quantum inspired algorithms have even led recently to unexpected outcomes in the realm of QML's quantum advantage itself. By exploring in particular the state-preparation assumptions, it has been observed how exponential speed-ups for some QML protocols³⁰ can indeed be reduced to polynomial, when they are compared against

³⁰This is the case for the recommendation algorithm in [67], as well as low-rank instances of QML algorithms for linear algebra applications [62, 64].

randomised QIML counterparts [76, 77].

Finally, MLEQA completely reverses the research question from the possible advantages of using quantum (inspired) processing techniques to advance machine learning protocols, to the performance enhancement in quantum devices when coupled with classical ML computational resources. MLEQA approaches can target both the design as well as the operation of a quantum device. In the first instance, ML is applied to identify robust gate decompositions [78, 79] or actively decouple the system from the environment in an optimal control scenario [80]. Alternatively, ML algorithms can provide feedback during the run-time of a quantum device, to minimise the impact of noise [81, 82], as it will be shown further in this chapter. Crucially, ML can be applied to the open–problem of the characterisation of increasingly complex quantum devices, as detailed in Chap. 3, or to the reliable preparation of specific quantum states of interest, detailed in Chap. 4.

To a limited extent, proof–of–principle experiments in the QML realm have been recently performed (e.g. see [83] for SVMs and [84] for supervised learning), but the implementation using NISQ quantum devices imposes severe limitations on the addressable size of such experiments. This observation, along with the ongoing pioneering efforts in providing a quantum RAM (QRAM), an essential component for several QML protocols [72], points towards the amenability (if not necessity) of MLEQA experimental implementations over QML protocols in the short to mid–term. In this Thesis, we will consider exemplary cases exposing a basic principle: in some instances, quantum resources can be advantageously traded for classical post-processing of quantum data. Most crucially, expensive device upgrades (e.g. providing some degree of quantum error-correction) might be replaced by employing inexpensive classical computational power.

1.3 Platforms for quantum technologies

After more than 25 years since the first proof–of–concept experiments in this realm [5, 27], much has been understood about the potential and limitations of each platform, yet no single technology so far has emerged as capable of addressing simultaneously all challenges. The very problem of a meaningful, simple benchmark among architectures that are radically different has been the subject of recent proposals, see e.g. the idea of *quantum volume* [85]. Indeed, the need for hybrid devices combining the advantages of different quantum technologies has been increasingly advocated³¹. In this section we do not aim to perform a thorough review of all experimental platforms attempted to date, but to provide the reader with some essential context, before discussing in detail the two systems of our interest in Sect. 1.4 & 1.5.

Nuclear Magnetic Resonance (NMR) has been used for some of the very first experimental demonstrations of quantum algorithms, shedding light on the problems of realistic noise sources, whilst partially circumventing the fundamental issues of dealing with mixed states [5]. Neverthe-

³¹ A recent review of efforts in this direction appeared as [86].

less, the distillation techniques used to prepare pure fiducial states have afflicted its scalability to bigger instances, so that NMR is often neglected as a candidate for next generations of devices targeting computers. As NMR relies on ensembles of molecules, also its properties as a quantum sensor left space for systems with an even smaller footprint³²

Ion traps have been proposed as a suitable platform since the dawn of quantum information processing [5, 87], and remained a strong candidate. Ion traps satisfy the DiVincenzo criteria for optimal qubits [38], all the fundamental components to build a UQC have been demonstrated, and they exhibit the highest single and two-qubit gate fidelities [87]. Additionally, the physics of the device allows for arbitrary, eventually reconfigurable connectivity³³ among the physical qubits. Nevertheless, a path towards a scalable architecture needs to clear issues related to both stably trapping and addressing more than few tens of ions simultaneously [87], along with a slower gate implementation compared with other platforms. A similar technology, optical lattices of ultracold atoms, has instead provided some of the most striking simulations of quantum many-body systems [89], due to the degree of controllability of vast ensembles of quantum systems. In terms of sensing, these technologies have a major drawback from requiring cryogenic temperatures along with a vacuum environment in order to operate [5, 87, 89].

Systems relying upon technologies compatible with CMOS fabrication have the intrinsic advantage of leveraging upon decades of development and investment in micro and nano fabrication³⁴. The most immediate examples belonging to this realm are *Quantum Dots*³⁵ (QDs) and *Superconducting Qubits* (SQs). In both technologies, the operational qubits are defined within a condensed matter environment. If this avoids limitations connected with an isolated system like ion traps, it also renders the qubit states fragile, because of long-range undesired charge motion, interactions with phonons in the lattice, nuclear spin interactions [5]. In order to reduce the mechanisms leading to these noise sources, both technologies operate in cryogenic conditions, posing a major technological challenge.

Even if all the basic components for a UQC have been demonstrated in QDs [90], the noise contributions quoted above sum up leading to decoherence times hardly exceeding the μs range [90]. This leads to relatively poor two-qubit fidelities (below 90%), even when adopting optimised control techniques, and challenges in scaling the number of controllable qubits. Finally, SQs have so far provided the devices with the highest number of individually addressable physical qubits, 72 at the time of writing [91]. The main reason for the advantageous scalability of SQs compared to QDs is that qubits are usually not represented by single charge carriers in this platform. The magnetic flux in superconducting loops is used instead [5], greatly adding to robustness

³² And in particular NV-centres, explained in detail in Sect. 1.5.

³³ Considered a system of n qubits, its connectivity η_C is typically defined as proportional to the number n_E of *direct* qubit-qubit interactions addressable [88]. One might normalise this against the size of the system, visualising direct interactions as edges in a graph whose nodes are the qubits, and defining $\eta_C = n_E / \binom{n}{2}$ as the ratio of existing edges, versus the number of edges required by a complete graph, or by the connectivity of the graph itself.

³⁴Briefly mentioned at the start of this Chapter

³⁵ We will hereby consider Si electron spin qubits as part of this category, a choice based upon necessary brevity as well as similarities between the two platforms.

against a noisy background. The small parasitic capacities of the superconducting devices also contribute to very high gate speeds [5, 88]. The appeal of CMOS–compatibility³⁶ and immediate scalability in the near term has persuaded a few companies³⁷ to pursue a superconducting route to UQC. Nevertheless, significant challenges remain ahead. Let alone the required operation at cryogenic temperatures³⁸, the average two-qubit gate fidelities have been so far limited to about 95% [45], in absence of any error correcting codes³⁹. Additionally, SQs are also afflicted by limited, non-reconfigurable connectivity [88]. This is an issue common to all technologies based upon fabrication on a planar substrate, which makes long range interactions difficult to achieve. Circumventing this limitation, in the scenario of dialling arbitrary quantum algorithms on the SQ device, requires mappings that play a crucial role in the final fidelity achievable, and add classical overhead to the overall computational time [85, 88].

1.4 Integrated quantum photonics

Among the possible platforms suggested along the years to build and operate quantum devices, this Thesis will make extensive usage of integrated quantum photonics. This implementation builds on a long tradition of experiments in bulk optics, where photons had already been employed as the carriers of quantum information. Therefore, quantum photonics inherits both advantages and disadvantages of photons as information carriers: their limited interaction with the surrounding environment (inclusive of other photons) makes them stable against decoherence phenomena, but at the same time renders entangling gates harder to implement.

The main peculiarity of integrated quantum photonics is that photons do not travel in free space, but are confined in *waveguides* (WG). The working principle of a WG is total internal reflection: a channel of a higher refractive index (the *core*, n_{in}) is embedded in a medium characterised by a lower refractive index (the *cladding*, n_{out}). The channel is often rectangular in section, like in the WG displayed in Fig. 1.2. The description of a WG can be provided by solving Maxwell's vectorial equations for a medium of 0-conductivity, which leads to the simple [93]:

$$\begin{aligned}\mathbf{E}(x, y, z, \tau) &= \mathbf{E}(x, y) e^{i\omega\tau - n_{\text{eff}}z/c} \\ \mathbf{H}(x, y, z, \tau) &= \mathbf{H}(x, y) e^{i\omega\tau - n_{\text{eff}}z/c}\end{aligned}\tag{1.18}$$

³⁶ CMOS–compatible devices are those that can be readily embedded within a typical CMOS fabrication process, without being affected by subsequent manufacturing steps, nor affecting components realised in preliminary steps. E.g. devices requiring annealing at high–temperatures might provoke unwanted oxidation phenomena, or diffusion of atomic species in the fabricated wafer. Also specific materials might introduce tensile stress in the layers of the wafer, and are thus to be considered incompatible. A detailed discussion can be found in [6].

³⁷ Among the most renown, we can quote Google®, IBM® and Rigetti®.

³⁸ Which includes the related difficulties of embedding auxiliary standard CMOS components in a cryogenic environment [92].

³⁹ Modest versions of such code might boost the fidelity to 99% [45], but the same might well be advocated for other platforms.

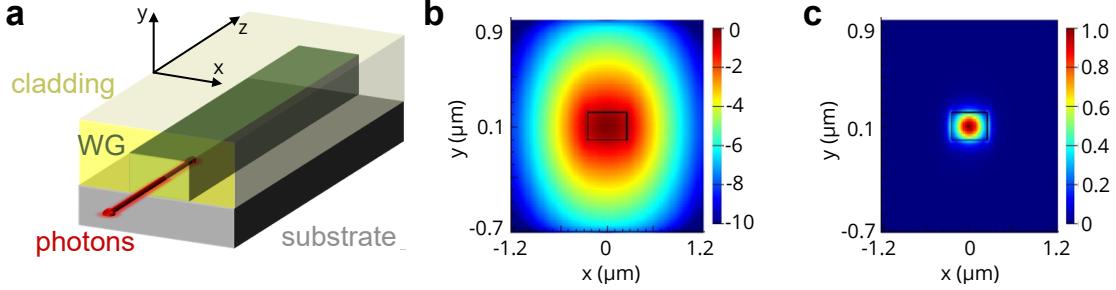


FIGURE 1.2. **a** Pictorial representation of a strip waveguide structure. **b** Lumerical MODE simulations of the Si in SiO_2 waveguide structure, seen transversely for the quasi-TE mode, in log-scale. **c** Normalised energy intensity for the same structure as in (b).

i.e. a travelling plane wave set of equations along the z direction (i.e. the axis, where c is the velocity of light, and n_{eff} the effective refractive index⁴⁰. The quantity $\omega n_{\text{eff}}/c = \beta$ is called the *propagation constant*. A fundamental design parameter for a WG is given by the contrast in refractive index between the core and cladding material, $\Delta_n = (n_{\text{in}}^2 - n_{\text{out}}^2)/(2n_{\text{in}}^2)$, which provides an indication of the strength of the confinement of the electromagnetic field within the WG. According to several factors, such as the WG's dimension along with Δ_n , the waveguide can support several propagating modes satisfying Eq. 1.18, and it is normally designed to support only the first two, namely the transverse electric (TE) and the transverse magnetic (TM) modes [94]. The naming is due to having no electric (TE) or magnetic (TM) field respectively in the propagation direction. Designing WGs and verifying supported modes is usually performed numerically, and a number of software packages are available, such as FIMMwave® or Lumerical®. The choice of the materials for the WG plays therefore an important role, influencing directly Δ_n , and in turn the compactness of the overall circuit design [92]. All chips used in this Thesis used a Silicon-on-Insulator (SOI) architecture, whereby Si waveguides are nanofabricated on top of a buried oxide (BOX) layer. An additional SiO_2 layer is then added, obtaining a *strip* waveguide⁴¹ as in Fig. 1.2. In this case, $n_{\text{in}} = 3.48$ and $n_{\text{out}} = 1.45$. Confinement of the modes is of course not perfect, as emphasised by the simulation reported in Fig. 1.2, and the field profile decays with *evanescent tails* that fall beyond the boundaries of the WG structure. The decay of the field is approximated by:

$$E \propto \exp\left(-\frac{2\pi d}{\lambda} \sqrt{n_{\text{eff}}^2 - n_{\text{out}}^2}\right) \quad (1.19)$$

where d is the distance from the core-cladding interface, and λ the wavelength of the propagating wave. The material partially dictates the latter as well, because to avoid losses, the core material needs to be transparent at λ (and within an interval around λ , for reasons that will appear clear in the following). For example, in the SOI architecture adopted for experiments in this Thesis, a

⁴⁰ Defined as the refractive index for a particular mode m , polarisation p and direction z in a medium, $n_{\text{eff}} = c/v_{mpz}$.

⁴¹ They are also known as *buried* WGs. Other possible configurations of WGs are possible, like *rib* and *ridge* WGs. The interested reader can find additional details e.g. in [94].

typically used wavelength is $\lambda = 1550$ nm, the so-called C-band for telecommunications. Beside Si transparency at this wavelength, another valid reason for choosing this value is the ready availability of commercial components when building experimental setups, as described later on.

Several degrees of freedom of a photon can be used to encode quantum information. In quantum photonics, the most common encodings use either the *path* of the photon, its *polarisation*, or *time-bins* [95, 96], even if hybrid approaches have also appeared in recent years (i.e. using *hyper-entangled* photon states [97]). The path-encoding for a single qubit relies on two spatial optical modes (often corresponding to two WGs in integrated quantum photonics). The configuration is exemplified in Fig. 1.3, and can be written in terms of photon occupation number for each WG:

$$\begin{aligned} |1\rangle &= |10\rangle \\ |0\rangle &= |01\rangle. \end{aligned} \quad (1.20)$$

The states on the right are in a *Fock state* representation, with the k -th bit referring to the k -th WG (usually counted from the top). Eq. 1.20 is named a *dual-rail* encoding [5]. A generic superposition state is thus $|\psi\rangle = c_0|01\rangle + c_1|10\rangle$. Following from the Fock state formalism, this encoding can be readily expanded to qudits, by adding additional optical modes (and their corresponding occupation numbers). The dual-rail representation has some relevant advantages[96]. An n -photons electromagnetic mode can be written as [98]:

$$|n\rangle = \frac{\hat{a}^{\dagger n}}{\sqrt{n!}} |0\rangle, \quad (1.21)$$

with \hat{a}^\dagger the *creation operator* (and equivalently a the *destruction operator*). Now, the free space evolution for such a mode can be obtained from the Hamiltonian $\hat{H}_{\text{free}} = \omega aa^\dagger$, with $\omega = 2\pi c/\lambda$. We can thus write the evolution for the generic $|\psi\rangle$ as:

$$|\psi\rangle \xrightarrow{\exp(-i\hat{H}_{\text{free}}\tau)} e^{-i\omega\tau} (c_0|01\rangle + c_1|10\rangle) \quad (1.22)$$

so that free evolution only adds an undetectable global phase [5]. Second, it implements an intrinsic error-detection code against amplitude damping channels (see Eq. 1.15). Indeed, it can be verified how $c_0|01\rangle + c_1|10\rangle \xrightarrow{\mathcal{A}(\gamma)} |00\rangle$ with probability γ . Now, the encoding does not admit a $|00\rangle$ state, so that the effect of \mathcal{A} can be immediately detected. Finally, a path encoding allows easy reconfigurability, which is less straightforward with e.g. polarisation encoded qubits [99]. Therefore, throughout experiments in this Thesis, we will keep adopting dual-rail encoded qubits (or multiple-rail encoded qudits where necessary).

1.4.1 Quantum photonic gates

Initialising a dual-rail qubit is immediate, as it is possible to control in which WG the photon is initially injected. Manipulation of the fiducial state, e.g. $|10\rangle$, requires instead additional components. Let us start with single-qubit gates (see Eq. 1.7).

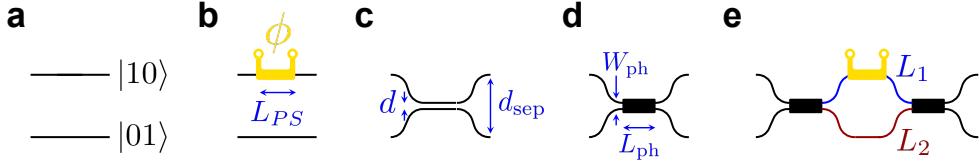


FIGURE 1.3. Schematics of some physical components of a quantum photonic chip. **a** Non-interacting modes confined in waveguides (at distance d_{sep}), used for the dual-rail qubit representation, exemplified by the Fock states reported. **b** A phase shifting element on the top waveguide. **c** A directional coupler, locally reducing the inter-waveguide distance to d . **d** An 2×2 multimode interferometer. **e** A (reconfigurable) integrated Mach-Zender interferometer assembly. The picture represents a balanced MZI ($L_1 = L_2$), as long as no phase shift is applied. However in principle, unbalanced designs are possible with $L_1 \neq L_2$ lengths of the arms.

$R_z(\phi)$ can be obtained⁴² by implementing a phase shift on the photons travelling on only one of the two rails (see Fig. 1.3). There are several routes to achieve this, and in particular here we quote:

- i thermal phase shifters, i.e. localised heaters,
- ii MEMS-based devices,
- iii electro-optic (EO) modulators.

As all components (i–iii) require electrical connection and power, they are considered *active* elements of the photonic chip. Each of these *phase shifting* technologies has peculiar (dis)advantages. In particular, (i–ii) are highly versatile and compact, but are limited to operating up to the MHz regime [92]. (i) moreover are simple to fabricate and robust, whereas the fabrication of (ii) is more involved, and they are more prone to failure. Typical EO modulators based upon $\chi^{(2)}$ nonlinearities can operate up to tens of GB/s, however, when choosing silicon as a WG material, $\chi^{(2)} = 0$ and components need to rely on the weaker $\chi^{(3)}$ nonlinearity⁴³. Integrated EO modulators investigated for Si platforms include: carrier modulation in doped Si, strain-induced phase shifting, reverse-biased PIN junctions in high-field, or the integration of high- $\chi^{(2)}$ materials in the SOI architecture⁴⁴.

In the experiments performed in Chaps. 2–4 we will make exclusive use of heaters to achieve phase shifting. Their working principle is simply based upon the Joule effect: each heater is a resistive metallic element, usually manufactured by Thin Film Deposition (TFD) on top of the waveguide, and buried⁴⁵ within the cap oxide cladding layer. The heat dissipated by the heater

⁴² Up to an unimportant global phase.

⁴³ Refer to Eq. 1.35 and the related paragraph.

⁴⁴ This discussion will be extended in Appendix D.3.

⁴⁵ Electrical connectivity to the heating element is thus provided by *vias* from the surface of the chip.

changes locally n_{in} , accruing a phase shift of approximately:

$$\Delta\phi = \frac{2\pi L_{\text{PS}}}{\lambda} \frac{dn_{\text{in}}}{dT} \Delta T, \quad (1.23)$$

with dn_{in}/dT the *thermo-optic* coefficient for the core material⁴⁶, which in general depends from T , ΔT the temperature raise, L_{PS} the length of the locally heated waveguide. The resulting component performs $R_z(\Delta\phi)$, up to a negligible global phase factor.

$R_y(\theta)$ requires the introduction of additional components. Intuitively, because R_y alters the balance of the two eigenstates of the dual-rail configuration, a mechanism is required, to distribute photons (i.e. ultimately optical power) across the waveguides composing a qubit or qudit. Photonic technologies offer two different *passive* technologies to achieve this⁴⁷:

- i directional couplers (DC) and
- ii multi-mode interferometers (MMI)

DCs make use of the evanescent field, as defined in Eq. 1.19. Waveguides for non-interacting modes are usually separated by a distance d_{sep} (see also Fig. 1.3), at which the evanescent field is negligible, i.e. $E \sim 0$. However, when the distance is reduced⁴⁸ to be $d \ll d_{\text{sep}}$, then a significant portion of the energy in one waveguide mode overlaps with the core of the other, and will increasingly excite the corresponding mode as long as the distance is kept well below the d_{sep} threshold. The modelling of the DC is usually performed via a system of coupled differential equations, that when no phase mismatch occurs, can be written [93]:

$$\frac{dE_1}{dz} = -i\kappa E_2(z), \quad \frac{dE_2}{dz} = -i\kappa E_1(z) \quad (1.24)$$

where the parameter κ captures the amount of overlap occurring between the two WGs (supposed constant for simplicity). The solution to Eq. 1.24 can be put in form of a unitary transfer matrix

$$\begin{aligned} & \begin{pmatrix} E_1(0) \\ E_2(0) \end{pmatrix} \xrightarrow{\hat{U}_{DC}} \begin{pmatrix} E_1(z) \\ E_2(z) \end{pmatrix}, \text{ with:} \\ & \hat{U}_{DC}(\kappa, z) = \begin{pmatrix} \cos(\kappa z) & -i \sin(\kappa z) \\ -i \sin(\kappa z) & \cos(\kappa z) \end{pmatrix} = \hat{U}_{DC}(r) = \begin{pmatrix} \sqrt{r} & -i\sqrt{1-r} \\ -i\sqrt{1-r} & \sqrt{r} \end{pmatrix} \end{aligned} \quad (1.25)$$

where we introduced the reflectivity r of the DC. A DC is said to be balanced when $r = 1/2$, because the energy in a waveguide becomes equally splitted after $\hat{U}_{DC}(r = 1/2)$ is applied. The photonic gate as in Eq. 1.25 is fundamental, because we can decompose the Hadamard gate in Eq. 1.5 as $H = \hat{U}_{DC}(1/2)R_z(\pi)$.

MMIs achieve a similar outcome, but leveraging upon a very different phenomenon. If in DCs the WGs to couple are taken closer, but preserve their original shape (and hence supported modes),

⁴⁶ E.g. $dn(Si)/dT = 1.8 \times 10^{-4} \text{ K}^{-1}$ at 300 K, [100].

⁴⁷ Which are the integrated equivalent of beam splitters in bulk optics [5].

⁴⁸ In the SOI platform, usually $d_{\text{sep}} \sim 10 \mu\text{m}$ is enough to achieve absence of coupling between WGs. Typical DCs have $d < 2 \mu\text{m}$.

in MMIs N^{in} *input* waveguides⁴⁹ are merged into a single *central* multimode WG. The latter then splits again into M^{out} *output* (single mode) waveguides (see Fig. 1.3, also for nomenclature). The resulting device is usually labelled as an $N \times M$ MMI, and its modes as *supermodes*. The physics of an MMI can be understood better with a simpler 2D model in (x, z) . If we name $\psi_m(x)$ the generic m -th mode, with propagation constant β_m , the propagation of the field in the central waveguide can be written similarly to Eq. 1.18, as [101]:

$$E(x, z) = \sum_m C_m e^{i\beta_m z} \psi_m(x), \quad (1.26)$$

i.e. a sum of the contributions from each mode. The expansion coefficients C_m depend on the shape of the input field $E(x, 0)$. As power must be preserved, if losses are negligible, C_m can be engineered so that the total power injected by one or more input WGs is distributed to modes of choice. The usual strategy to achieve this is the *mode matching*: the power profile of the input mode(s)⁵⁰ at $z = 0$ is decomposed into the supermodes [94]. The method is accurate when a large number of supermodes is considered, and is often implemented in commercial software⁵¹. Since, once a parametrised model for the modes in terms of the waveguide structure is known, the targeted C_m are a suitable objective function for optimisation method, automated optimal design of MMI components has also been actively investigated [102].

Within the center waveguide, the propagation constant for each mode is approximated as:

$$\beta_m \sim \beta - \frac{m(m+2)\pi}{3L_b} \quad (1.27)$$

where β holds for the fundamental mode, and $L_b \sim 4n_{in}W_e^2/3\lambda$ depends upon the *effective width* W_e , which is in turn a function of the physical WG width W_{ph} and refractive properties:

$$W_e \sim W_{ph} + \frac{\lambda}{\pi} \left(\frac{n_{out}}{n_{in}} \right)^{2\sigma} \sqrt{n_{in}^2 - n_{out}^2} \quad (1.28)$$

where $\sigma = 0$ (1) for TE (TM) modes. If all other parameters are held constant, the behaviour of the MMI is thus dictated by the ratio L_{ph}/L_b between the actual length of the central WG and the *beat length*. Let us consider for simplicity the 2×2 MMI in Fig. 1.3. Here, at $z = 6L_b$, the input field will be reproduced and light will be coupled in the *bar* output port of the device, because then $\beta - \beta_m \propto 2\pi$. On the contrary, $z = 3L_b$ has $\beta - \beta_m \propto \pi$. the output field is then the *mirrored image* of the input, respect to the symmetry axis of the central slab. For $z < 3L_b$ one observes instead a split of the optical power in both output ports [101]. In general, therefore, the unitary $\hat{U}_{MMI}(z)$ describing the output of an MMI is a function of the length of the central WG. For symmetric $N \times N$ MMI devices, with $n_{out} = 1$ and assuming single-mode inputs, it is possible

⁴⁹ Which are usually single mode.

⁵⁰ Eventually modified after introducing scattering parameters, to take into account the interface between different sections of the device.

⁵¹ E.g. the Lumerical's MODE package.

to obtain a simple rule to extract the (u, v) -th element of $\hat{U}_{\text{MMI}}(L_b)$, which reads [103]:

$$\frac{2i}{\sqrt{2N+2}} \sin\left(\frac{\pi uv}{N+1}\right) \exp\left(\frac{i\pi}{4} - \frac{i\pi(u^2 + v^2)}{2N+2}\right). \quad (1.29)$$

Using Eq. 1.29, we obtain for example that for a 2×2 MMI, $\hat{U}_{\text{MMI}}(L_b) = \hat{U}_{\text{DC}}(1/2)$, up to a negligible global factor. This shows that indeed MMIs and DCs are interchangeable components in composing a complete set of photonic gates.

A brief digression is noteworthy. The behaviour of DCs and MMIs in quantum photonics⁵² when irradiated by single photons is rather non trivial [98]. Labelling the two input channels of the SC as 0, 1 (top, bottom respectively), and the two outputs as 0', 1' we can express the DC matrix in Eq. 1.25 for $r = 1/2$ in terms of creation and destruction operators, and by moving to the Schröedinger picture obtain the following evolution⁵³:

$$\begin{aligned} |1\rangle_0 |1\rangle_1 &= a_0^\dagger a_1^\dagger |00\rangle \rightarrow \\ &\frac{1}{2} (a_{0'}^\dagger - ia_{1'}^\dagger)(a_{1'}^\dagger - ia_{0'}^\dagger) |00\rangle = \frac{-i}{2} (|2\rangle_{0'} |0\rangle_{1'} + |0\rangle_{0'} |2\rangle_{1'}) \end{aligned} \quad (1.30)$$

for an input state having one photon impinging in each of the DC (MMI) input modes $|1\rangle_0 |1\rangle_1$. The final outcome in Eq. 1.30 describes the *bunching* of both photons in either output modes, cancelling the probability of detecting one photon at each output port (named a *coincidence* event) from such experiment. This lack of coincidences from the interference of indistinguishable photons is named the Hong–Ou–Mandel (HOM) effect demonstrated experimentally for the first time⁵⁴ in [107].

In order to achieve full control on a single photonic qubit, we need the ability to perform $R_y(\phi)$ operations. This can be achieved adopting a (reconfigurable) *Mach–Zender interferometer* (MZI), which combines structures already introduced: two balanced DCs (or MMIs) and a phase shifter, positioned as in Fig. 1.3. A simple analytical model for the MZI can be provided starting from the modelling of a 50/50 DC [94], and ignoring the heater to model first a passive MZI. When light is injected in just one input port of the DC with intensity I_i , the intensities at the two output branches are expected to be $I_1 = I_2 = I_i/2$. Therefore, the electric field propagation through the MZI waveguides (as a planar wave) reads:

$$\begin{aligned} E_1 &= \frac{E_i}{\sqrt{2}} e^{-(\alpha_1/2+i\beta_1)L_1} \\ E_2 &= \frac{E_i}{\sqrt{2}} e^{-(\alpha_2/2+i\beta_2)L_2} \end{aligned} \quad (1.31)$$

where E_i is the input electric field, $\alpha_{1,2}$ phenomenologically represents losses in the two output waveguides (i.e. the *arms* of the interferometer), while $\beta_{1,2}$ is the usual propagation constant.

⁵² As well as their counterparts in bulk optics, beam splitters.

⁵³ We skip here details of the derivation for brevity, but these can be found e.g. in [104].

⁵⁴ Later on, a generalisation of the effect was observed using a MZI [105], paving the way to HOM fringe experiments [106], reported here as a preliminary characterisation task e.g. in Sect. 1.4.4.2.

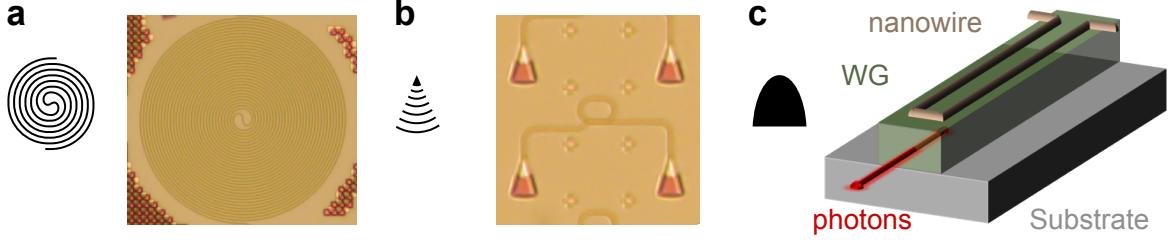


FIGURE 1.4. Auxiliary components of a quantum photonic chip. For each, on the left the symbol adopted in schematics, on the right the same component seen in an optical microscope (a–b), or a 3D rendering of the structure (c). **a** A non–resonant photon pair source, here designed as a spiralling WG. **b** A grating couplers (along with test waveguides in the optical microscope picture). **c** Single photons detector. The rendering refers to a simple SNSPD design, with the nanowire manufactured on top of the waveguide (contact pads omitted from the figure).

The two arms are recombined at the end of the MZI-interferometer. Here the two fields are mixed again and therefore we can derive for the output intensity:

$$I_o = I_i/4 \left| \sum_{i=1,2} e^{-(\alpha_i/2 + i\beta_i)L_i} \right|^2 \quad (1.32)$$

If we now assume no losses for simplicity ($\alpha_1 = \alpha_2 = 0$) and additionally suppose the material to be the same for the two branches of the interferometer ($\beta_1 = \beta_2 = \beta$) we can simplify the expression as the transfer function (T_{MZI}) for the MZI: $I_o = I_i/2(1 + \cos(\beta\Delta L)) = I_i T_{MZI}$, with $\Delta L = L_1 - L_2$ the length *unbalance* between the two arms of the interferometer. The behaviour of such a passive MZI is captured by its *Free Spectral Range* (FSR):

$$FSR = \lambda^2/(n_g \Delta L), \quad (1.33)$$

where n_g is the group refractive index in the WG, for the central wavelength λ_0 . An exemplary modelling of a passive MZI is reported in Appendix D.1, using the Lumerical® software.

The unitary transformation implemented by the generic reconfigurable MZI in Fig. 1.3 can instead be easily obtained by combining the gates already known:

$$\hat{U}_{MZI}(\phi) = \hat{U}_{DC}(r = 1/2) R_z(\phi) \hat{U}_{DC}(r = 1/2) = \begin{pmatrix} \frac{e^{i\phi}-1}{2} & \frac{i}{2}(1+e^{i\phi}) \\ \frac{i}{2}(1+e^{i\phi}) & \frac{1-e^{i\phi}}{2} \end{pmatrix} \quad (1.34)$$

The equation shows how a reconfigurable MZI satisfies the requirement, because $\hat{U}_{MZI}(\pi - \phi) = R_y(\phi)$, completing the set of gates required for arbitrary single–qubit operations [104].

1.4.2 Other photonic components

In this Section we have so far discussed some key quantum photonic components, focusing upon those that are strictly required to build a (single qubit) *linear optical quantum computer* (LOQC)

upon a photonic chip [95]. However, there are several additional components that are required in order to properly manage the quantum information carried by the photons in a photonic architecture, and in particular:

- i in/out chip light coupling structures,
- ii single-photon sources,
- iii single-photon detectors,
- iv optical filters,
- v electrical connections.

Components (ii-iv) do not strictly need to be integrated monolithically in the photonic chip. However, as discussed further below, this might be advantageous (if challenging). Interestingly, a fully integrated architecture might in principle⁵⁵ remove the need for (i), as the detectors usually transduce the optical signal from photons in electrical pulses. The scope of this section is not to provide a thorough analysis of each component (we refer the interested reader to the literature cited within), but rather to introduce key concepts that will be recalled in other sections of this Thesis.

Photons propagate easily in free space, a property that has made them an appealing choice to transfer information over long distances, and led to a plethora of key experiments in the realm of quantum optics, from the observation of the Hong–Ou–Mandel effect [98] and the first quantum teleportation experiment [108], up to very recent demonstration of the capabilities of existing infrastructure to develop a quantum communication network [109, 110]. Even if the rate of quantum information transmission across geographic distances available via satellite communication might be otherwise unachievable [110], the losses and noise introduced by scattering in an inhomogeneous medium like the atmosphere render free-space photon transmission unfeasible for more delicate quantum experiments.

Optical fibres, based upon the total internal reflection principle, have sustained the growth of the current classical communication infrastructure. They are therefore a natural option when hypothesising quantum communication networks. It is thus unsurprising that the very recent loophole-free Bell test was achieved in a fibre network [111], and experiments in bulk optics have often relied to fibres. E.g. fibres are the technology of choice for *delay lines* with reasonable losses [112], as attenuation of less than 1 dB/km is nowadays commercially available [113].

Therefore, even if at the start of this Section we advocated for integrated devices to *process* quantum information, it is still highly plausible that they must be interfaced with fibre-based components for remote *transmission* of the outcome. This is highly non trivial. E.g., in typical Si on-chip WGs, the mode is confined in a section of $\sim 500 \times 200$ nm (Sect. 1.4). The core of a single-mode fibre is usually ~ 10 μm in diameter, therefore the huge mode profile mismatch imposes the adoption of a dedicated light coupling structure, in order to keep *insertion losses*

⁵⁵ Some caveats will be made clear in the following.

low [94]. The strategies to achieve this are essentially two: *side coupling* and *vertical coupling*. In the first case, tapered fibres are used to “compress” the guided mode and then inject the light in the integrated WG, that reach out to the edge of the chip [114]. In this Thesis, this solution was adopted for the device in Fig. 1.5. In vertical coupling, the light is injected from the top of the chip. In order to guide the light into horizontally propagating in the WG mode(s), a diffraction-based component is used: *grating couplers* (GC), added to the surface of the chip. These are adopted for the device in Fig. 3.20, in this Thesis. The engineering of a GC aims to reduce scattering losses at the air–interface, as well as the light transmitted vertically across the GC and into the BOX layer [94]. Several designs are possible, according to the number of fabrication steps available⁵⁶ as well as their resolution⁵⁷, with insertion losses in the range of 4 to less than 1 dB (per GC, [94, 115]).

The generation of single photons with high purity and on–demand is often deemed as a crucial requirement for quantum information tasks [98]. In quantum optics, a considerable fraction of experiments uses non–linear photon–pair generation processes to tackle this task [104]. In brief, a laser pulse impinges on a non–linear medium, such as non–linear crystals or waveguides. This approach has a few drawbacks, the most important being that it is a probabilistic process. Therefore, non–linear photon generation might lead to failure, a photon–pair or multiple pairs generation [92]. Named $\mathbb{E}(t)$ the electric field from the input laser at time t (the *pump*), we can write the polarisation field in the medium as [101]:

$$\mathbb{P}(t) = \epsilon_0 \chi^{(1)} \mathbb{E}(t) + \epsilon_0 \chi^{(2)} \mathbb{E}^2(t) + \epsilon_0 \chi^{(3)} \mathbb{E}^3(t) + \dots \quad (1.35)$$

where $\chi^{(n)}$ is the n –th order susceptibility of the medium. The $\chi^{(2)}$ and $\chi^{(3)}$ terms are respectively responsible for the *Spontaneous Parametric DownConversion* (SPDC) and *Spontaneous Four Wave Mixing* (SFWM) non–linear processes. As SOI is the platform of choice for all photonic experiments in this Thesis, and $\chi^{(2)} = 0$ in Si due to the centrosymmetric structure of the crystal [104], on–chip sources will be all based on SFWM. In SFWM, two pump photons of frequency ω_p are annihilated, to generate a *signal* and an *idler* photons (with frequencies ω_s and ω_i), with wavelengths respectively smaller and bigger than the pump’s. This is because the phase matching conditions for SFWM read [116]:

$$\Delta\omega = 2\omega_p - \omega_i - \omega_s = 0 \quad (1.36)$$

$$\Delta k = 2k_p - k_s - k_i = 0 \quad (1.37)$$

using for the wavenumbers k the same notation as for ω . Analysing the detail of the process in second quantisation formalism [104], it is possible to estimate the probability of a photon–pair generation:

$$\xi^2 \sim \gamma^2 L_{\text{eff}}^2 W^2 \text{sinc}(\Delta k L), \quad (1.38)$$

⁵⁶ In particular, the number of *etching* steps, as well as the availability of an additional fabrication step dedicated to the insertion of a metallic, highly reflective bottom layer.

⁵⁷ E.g. the enhanced resolution of e–beam lithography over optical lithography might render photonic crystal structures available, improving the flexibility in the design of the diffraction and reflective elements.

where γ is the nonlinear parameter for the medium⁵⁸, L (L_{eff}) describes the (effective⁵⁹) length of interaction with the non-linear medium, and W is the pump optical power [92]. Therefore, the highest photon-pair generation occurs when Eq. 1.37 holds⁶⁰. It must be noted that Eq. 1.38 does not take into account multiple pair generation, and therefore for the approximation to be valid, W must be relatively low. If efficient, SFGM is ultimately non-scalable. First, as a probabilistic process $\xi^2 < 1$, and a successful photon-pair generation for m events scales as $\xi^{2m} \xrightarrow{m \rightarrow \infty} 0$. Additionally, for Si waveguides employed at $\lambda \sim 1550$ nm, a photon of frequency $\sim 2\omega$ does not fall in the transparency band of the material. The consequent phenomenon is the simultaneous absorption of a pump and a single photon (*two-photon absorptiorion*, TPA), which saturates quickly the maximum ξ^2 achievable when increasing W [92]. If TPA is absent for different materials, or operating at different wavelengths, still multiple photon-pairs generation ultimately caps the maximum W that can be adopted.

The most immediate way to leverage upon Eq. 1.38 to increase the par generation rate, given a material of choice, is thus to increase L [99, 117]. This can be achieved in a compact fashion in an integrated device, e.g. by designing spiral WGs (see Figs. 1.5 and 3.20). These structures are named *non-resonant* SFGM sources. An alternative is provided by employing *ring resonators*, i.e. *resonant* sources: as this structure is unused in this Thesis, we refer e.g. to [94, 101, 104] for details.

A certain number of different technologies have been proposed along the years to detect single photons, exhibiting the key properties required for quantum photonic technologies at different wavelengths [118]. In particular, we remind the following key performance indicators. The *dead (recovery) time* τ_{dd} indicates the minimum time interval possible between successive, successfully detected photons. The physics of the detecting mechanism as well as the physical design of the device can influence heavily τ_{dd} , that spans 4 orders of magnitude across the devices fabricated so far. τ_{dd} is an important contribution to the maximum photon countrate at which a detector can operate [118]. The *dark count rate* ν_{dk} can be understood as the background noise of the detector, as it describes the rate of false positive events recorded. Higher ν_{dk} impoverishes the SNR of the detector, imposing many detection repetitions to distinguish the signal. The *timing jitter* τ_{jt} captures the time interval between the photon absorption by the detector, and the generation of the readout signal (usually an electrical pulse). Not only τ_{jt} contributes to the maximum countrate, but in the realm of quantum optics, the resulting uncertainty in the time of arrival of the photon might render photons indistinguishable⁶¹. Perhaps the most quoted figure of merit of

⁵⁸ Defined as $\gamma = n_{\text{in}} k_0 / A$, with k_0 the vacuum wavenumber, ad A the mode area transverse to the direction of propagation.

⁵⁹ $L_{\text{eff}} = (1 - \exp[-\alpha L])/\alpha$, so that in an ideal lossless medium, $L = L_{\text{eff}}$.

⁶⁰ Eq. 1.36 must hold for the process to occur.

⁶¹ E.g. a key capability in a quantum optical device is the detection of photon coincidences, see pg. 26. As no physical detector is capable of an infinitesimal detection time slice, coincident events are operationally defined so, if they occur within the same (synchronised) coincidence window of two detectors. τ_{jt} imposes the minimum duration of such window, and a poor τ_{jt} increases the detection of false coincidence events. Currently, detectors with $\tau_{jt} < 100$ ps are routinely available.

a detector is its *efficiency*, usually provided in terms of the system detection η_{sd} . The latter is the rate of true positive events detected, once all losses within the detecting device are taken into account [119]. There are different protocols that can be used to estimate η_{sd} indirectly [118]. A high η_{sd} improves both the countrate and the SNR of the detector. Finally, the *spectral range* of operation⁶² is crucial in identifying the possible applications of a detector.

If Avalanche Photo-Detectors (APD) provide a cheap, reliable and effective solution for working in the ~ 800 nm range, the Si good light absorption properties in such bandwidth themselves prevent using this wavelength in C-band photonics [92]. A solution to the challenging problem of detecting lower energy infrared photons has been offered by the recent development of Superconducting Nanowire Single Photon Detectors (SNSPD). The basics of their detection mechanism is simple. A nanowire of a superconducting material⁶³ with critical temperature T_c is fabricated on top of a WG. When a photon is absorbed by the superconducting strip, the thermally released energy triggers a resistive *hotspot*, because T_c is overcome locally [120]. Now we assume to operate the nanowire close to its critical current density⁶⁴. Given that the nanowire has usual width below 100 nm, the hotspot generation forces the flowing current to circumvent the resistive area: if this overcomes the critical current density, the non-superconducting section expands in a reinforcing feedback, until the current is strongly shunt. This phenomenon can be easily detected by a voltage spike. As the current now falls below its critical density again, the normal superconductive behaviour is spontaneously restored. Since their proof-of-principle demonstration in the early 2000s [120], a considerable research and design effort has taken the η_{sd} of SNSPD from $\sim 20\%$ to well above 90% [121]. This and other favourable performance indicators render SNSPDs by far the most promising technology to detect photons in the mid-infrared spectral range [92, 119, 121], and they are the technology of choice for all the photonic experiments reported in this Thesis.

Leveraging upon a superconducting technology, however, introduces limitations in the range of their operating temperatures. For the current state-of-art, SNSPDs must be hosted in a cryostat, as they require $T_c < 5K$. There have been two main strategies to cope with such a severe requirement. The most pursued so far is to keep photon detectors *off-chip*. The main reason behind this choice is to avoid excessive heat load on the limited cooling power of a cryostat, due especially to the active components of a photonic chip⁶⁵, seen in Sect. 1.4, whilst relying on well-established techniques for hosting compact clusters of SNSPDs within the same cryostat⁶⁶. Clearly, there are drawbacks to this approach. First, the bulky fibre connections from the photonic chip to the cyostat-hosted SNSPD weaken one of the most favourable arguments advocating for quantum photonics: the development of compact, monolithic quantum devices. Perhaps more

⁶²In other words, the WL range wherein the detector is responsive to irradiation.

⁶³Common choices in current designs are NbN, TiNbN and WSi.

⁶⁴In practice, this is usually controlled by applying and adjusting a bias voltage of few V's across the SNSPD.

⁶⁵A detailed exemplary analysis is provided in Fig D.3.

⁶⁶At the time of writing, systems with tens of detectors are available from PhotonSpot, idQuantique, SingleQuantum among others.

importantly, having off-chip detectors introduces additional coupling and propagation losses to the weak signal at the end of the processing within the photonic chip. The alternate strategy is then to embed the detectors *on-chip*, leveraging upon the CMOS compatibility of SNSPD devices. As aforementioned, this solution is not free of challenges either, as it requires ultra-low-power active components to prevent insurmountable scalability issues (see Appendix D.3 for further details). Beside this, the integration of SNSPDs requires the availability of on-chip effective filters.

Optical filters are a stringent requirement for quantum photonic experiments, as well as any other scenario operating in the low-photon regime with sources pumped optically⁶⁷. This is because the extreme sensitivity of single photon detectors, designed to deal with counts not exceeding the GHz regime⁶⁸, makes them unsuitable to contamination of the single photons carrying actual information, with the bright light originated from the pump. Experiments in this Thesis adopt bulk filters (see Sect. 1.4.4) with *extinction-ratios*⁶⁹ beyond 90 dB for the pump, and negligible (less than 1 dB) for the signal/idler photons. Even if high-extinction integrated filters are an active topic of research, the current state of art is about 30 dB off the level required, mostly because of light scattered across the chip substrate [122].

1.4.3 A scheme for arbitrary photonic controlled-unitaries

Optical platforms for UQC could count upon well-established components for single-qubit operations [5], capable of implementing gates with high fidelity, and photonic technologies added to the stability and practicality of the devices [92]. Nevertheless, the lack of photon–photon interactions enabling multiple qubits gates mined its credibility to accomplish a UQC. The difficulty to achieve $C\hat{U}$ could have confined its applicability to the simulation of particular classes of problems (e.g. [115, 123]). The very implementation of several logical quantum circuits investigated in this Thesis⁷⁰ requires the ability to perform arbitrary $C\hat{U}$ between at least two qubits.

Such limitations were removed by the proposal to exploit ancillary photons and feedforward schemes to achieve the required interactions, without relying on the weak nonlinear coupling among optical modes [92, 124], paving the way to LOQC. With *feedforward*, the signal from detected ancillary photons is processed and used to control switches in the optical circuit *ahead* of the final measurement operations. The proposal was soon perfected into obtaining all-optical *CNOT* gates with $\epsilon_{opt} = 8/9$ error probability based upon photon–pair sources, with the possibility to post-select successful operations by observing the final measurement outcomes [95, 125].

⁶⁷As described in the beginning of Sect. 1.4.2.

⁶⁸And often operating in the kHz regime, as for the experiments in Chaps. 2-4, with state-of-art experiments for single photon counts achieving MHz rates [115].

⁶⁹Simply defined as the ratio between input and output optical power, W_{in}/W_{out} .

⁷⁰In particular, IPEA (Sect. 2.1.5), RFPE (Sect. 2.1.6), and WAVES (Sect. 4.4). The necessity for $C\hat{U}$ operations is evident observing the various quantum circuits e.g. in Fig. 2.1 & 4.2.

Clearly, non-deterministic gates with low success probability are not sufficient to achieve a realistic UQC, as concatenation of G independent gates naturally leads to an exponentially decreasing success rate, $(1 - \epsilon_{opt})^G$ [95], so that even if all errors can be detected, the clock rate of such a LOQC would become too low to be usable. Post-selected implementations can nevertheless serve as a useful testbed ahead of achieving feedforward operations. The latter are possible in principle, but have proven technologically challenging [92].

We will skip for brevity a detailed description of the $C\hat{U}$ schemes presented in [124, 125]. This is because in this Thesis we adopted instead a (post-selected) entanglement-based scheme to implement arbitrary $C\hat{U}$ gates in quantum photonics [126, 127]. Referring to Fig. 1.5, this section describes in more details such scheme, adopted for the first time in an integrated device with the experiments in Chap. 2.

We start from an initial state described by the superposition between the pairs of photons after two coherently pumped sources, that within Fock representation reads $(|0200\rangle + |0020\rangle)/\sqrt{2}$. After the two MMIs beam-splitters and the crossing element, the photons are split probabilistically so that the full state becomes $(|2000\rangle + |0200\rangle + |0020\rangle + |0002\rangle + 2|1010\rangle + 2|0101\rangle)/\sqrt{8}$. According to the post-selection scheme [99, 127, 128], we use for computational purposes only the cases corresponding to a Bell state:

$$\frac{|1010\rangle + |0101\rangle}{\sqrt{2}} \quad (1.39)$$

i.e. where the signal photon is collected from the two upper modes and the idler from the four bottom modes. This state is pictorially represented in Fig. 1.5 by the solid and faint coloured dots. We adopt the dual-rail representation of logic states as in Eq. 1.20. Within this convention, the maximally-entangled Bell state obtained at the start of the circuit is $(|0\rangle_C|0\rangle_P + |1\rangle_C|1\rangle_P)/\sqrt{2}$, where we introduced the C (P) subscript to label the *control* and *path* subspaces.

Two additional modes then extend the bottom paths, expanding correspondingly the Hilbert space. This can be thought of as the extension to an additional (*target*, T) qubit, so that we have $(|0\rangle_C|0\rangle_T|0\rangle_P + |1\rangle_C|1\rangle_T|1\rangle_P)/\sqrt{2}$. The input state $|\psi\rangle_T$ is prepared for both target states, using one R_z and one R_y gate (see also Sect. 1.4.1). The target state in the upper path is then evolved under $\hat{1}$ (the identity) and the lower one according to \hat{U} , obtaining an entangled state of the form:

$$\frac{|0\rangle_C|\psi\rangle_T|0\rangle_P + |1\rangle_C(\hat{U}|\psi\rangle_T)|1\rangle_P}{\sqrt{2}}. \quad (1.40)$$

To obtain a superposition of these two different operations, we erase the path information between the two components of the target state. This is obtained employing a waveguide crossing⁷¹ and combining the modes in two MMI beam-splitters:

$$\frac{(|0\rangle_C|\psi\rangle_T + |1\rangle_C\hat{U}|\psi\rangle_T)|0\rangle_P + (|0\rangle_C|\psi\rangle_T - |1\rangle_C\hat{U}|\psi\rangle_T)|1\rangle_P}{2}. \quad (1.41)$$

⁷¹ The final crossing in the chip represented in Fig. 1.5 mixes the two paths encoding the target registry.

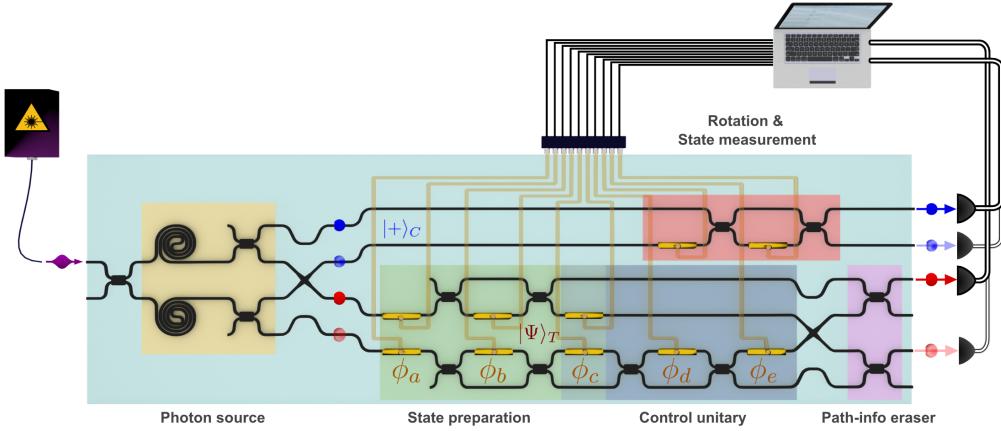


FIGURE 1.5. CW laser light is injected via lensed fibres in the photonic chip, where photon pairs (blue and red dots) are generated via spiralling integrated photon sources. Target states $|\Psi\rangle_T$ are prepared and evolved via a controlled unitary via the reconfigurable phase shifters (golden rods). The control register allows for tomographic measurements via additional phase shifters. A final crossing element erases information about the path. SNSPDs are represented as black paraboloids, with the two measurement configurations for post-selection colour-coded. An external computer elaborates the outcomes and controls the experiment. (see also Sect. 1.4.2)

Finally, projecting the third qubit into $|0\rangle_P$, the final state can be represented as

$$\frac{|0\rangle_C |\psi\rangle_T + |1\rangle_C \hat{U} |\psi\rangle_T}{\sqrt{2}} \quad (1.42)$$

which is equivalent to apply the desired arbitrary control-unitary operation. In principle, one can cascade in a non-compiled fashion other \hat{U} , for example obtaining a trotterisation for the time evolution (see Eq. 1.8). We remark that the probabilistic nature of any post-selected scheme makes it not scalable. Nevertheless, also this entangled scheme enabled a wide range of experiments previously inaccessible to integrated quantum photonics, like the ones presented in the following chapters.

1.4.4 The CU(2) quantum photonic chip

In this section we leverage upon components discussed in Sect. 1.4.1–1.4.2 to provide a thorough description of the experimental photonic device⁷² used as testbed for systematic investigation of phase estimation (Chap. 2) as well as variational algorithms (Chap. 4) performances in a noisy scenario.

The device is a silicon-on-insulator ($\text{Si}-\text{SiO}_2$) chip, fabricated via a mass manufacturable approach using 248 nm deep-UV photolithography and dry etching at Toshiba® facilities⁷³. Inside

⁷² Hereafter in brief referenced as the “CU(2) device”.

⁷³ See also the contribution box at the end of the chapter

the chip, photons propagate in integrated single-mode waveguides, having a width of 450 nm and thickness of 220 nm, covered with a 1 μm silicon dioxide upper cladding. A -40 dB crosstalk between crossing waveguides was observed, substantially negligible for all purposes within this Thesis. The device integrates key functionalities for photonic quantum information processing, such as sources to generate an entangled photon-pair and arbitrary single- and two-qubit gates.

The photons are generated via SFWM in two spiral waveguide sources (Sect. 1.4.2), with an overall 1.2 cm length to enhance the photon-pair generation. Considering a non-resonant SFWM source and a continuous-wave (CW) laser pump⁷⁴, the generation probability of multi-photon-pair events is negligible [99]. This is experimentally confirmed by the high visibility of HOM fringes [106], as shown in Fig. 1.8.

Multi-mode interferometers (MMIs) with a footprint of $2.8 \mu\text{m} \times 28 \mu\text{m}$ were used to realise integrated beam-splitters with near 50% reflectivity. The active and reconfigurable quantum operations inside the silicon device rely on thermo-optical phase shifters, which are formed by metal resistive elements ($R \sim 100 \Omega$), on top of the silicon WGs and isolated by the cladding. The heaters are depicted as golden rods in Fig. 1.5. A classical computer controls the heaters via Qontrol® commercial electronic drivers.

Light is side-coupled into and out of the chip via spot-size converters with 300 μm long inverse taper with a 200 nm wide tip beneath a $4 \times 4 \mu\text{m}^2$ polyimide WG. The loss per facet is about 8 dB, coupling single-mode lensed fibers with a 3 μm mode-diameter to the chip.

1.4.4.1 Details on the experimental setup

After describing a typical photonic chip, we enlarge our attention to include the non-integrated components used in the experiments: the setup schematic is described in Fig. 1.6. The on-chip sources were pumped using a CW bright laser at 1551.9 nm, amplified by an erbium doped fiber amplifier (EDFA), with an emerging power of ~ 10 mW. Due to vibrations, the input/output optical fibers are prone to a spontaneous drift in the coupling efficiency. This was recovered by automatically recoupling the fibres via micro-actuators, to maximize the coupling efficiency between waveguides and fibers.

To stabilise the temperature of the silicon device, prone to fluctuations due to the adoption of thermo-optical components, the chip was mounted and wire-bonded on a printed circuit board (PCB) that was glued by thermal epoxy on a Peltier device connected to a heat-sink. The temperature was controlled by a PID algorithm. Electrical cross-talk was reduced by driving the heaters in current, via a 12 bits amplified DAC driver board, controlled through RS-232 interface. A precision of approximately 0.01 rad was achieved, as reported in more details later on. Heaters have isolated voltage supply connections, but share a common ground connector, causing electrical cross-talk among them. Thermal cross-talk effects, due to heat dissipation across the SOI substrate, were partially compensated by calibrating for different chip configurations.

⁷⁴ This is further detailed in Sect. 1.4.4.1.

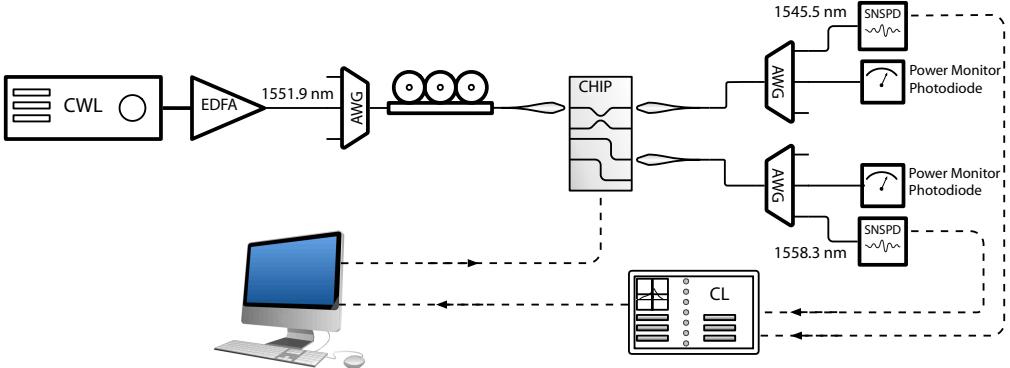


FIGURE 1.6. The CU(2) chip schematics. A CW laser light source goes through an amplifying EDFA. An AWG filter and a manual polarisation controller are used to control the laser light before using lensed-fibres to focus it inside the chip. Equivalent lensed-fibres are used to extract the light. The emerging light is again frequency-filtered to separate single photons from the pump. Standard photo-diodes monitor the filtered CW light and hence the coupling, whereas SNSPDs detect the single photons, with the pulses analysed by a coincidence counter interfaced with a classical computer. The latter controls the on-chip TO phase shifters.

The output photons, after evolving through the integrated device, were coupled into single-mode fibers, separated apart from the pump light by off-chip AWG with a 0.9 nm bandwidth and > 90 dB extinction. In particular, the wavelengths of the signal and idler photons were selected at 1545.5 nm and 1558.3 nm, respectively. Photons were finally detected using PhotonSpot™ SNSPDs, with $\eta_{SD} > 85\%$, sub-100 Hz dark counts, approximately 70 ps FWHM timing jitter and $\tau_{dd} \sim 50$ ns. The coincidence counts were recorded by the use of a time interval analyser (PicoHarp 300 by PicoQuant™) with an integration time of 10 s per data point and a coincidence window of approximately 400 ps. A maximal photon-pair rate of approximately (200 Hz) was observed.

1.4.4.2 Preliminary chip characterisation

The single-qubit gates composing the state preparation, $C\hat{U}$, and quantum state tomography are characterised both with classical light and single photons. In the first case, we recorded the injected current and dissipated electrical power, and then the mapping between electrical and optical power. In this way, a mapping between current and phase for each TO element was found. Data were fitted with a constrained Least Squares Fit (LSF). In order to compensate for thermal cross-talk, each heater was characterised across different configurations of the remaining elements. Good device performance and low residual cross-talk are confirmed by λ -classical interference and $\lambda/2$ -quantum interference, shown in Fig. 1.8.

The latter was obtained for the control qubit by setting the top reconfigurable MZI to implement a Hadamard gate (i.e relative phase shift of $\pi/2$) and collecting photon coincidences in the top two modes. By altering the phase of the other heater on the control register, we changed

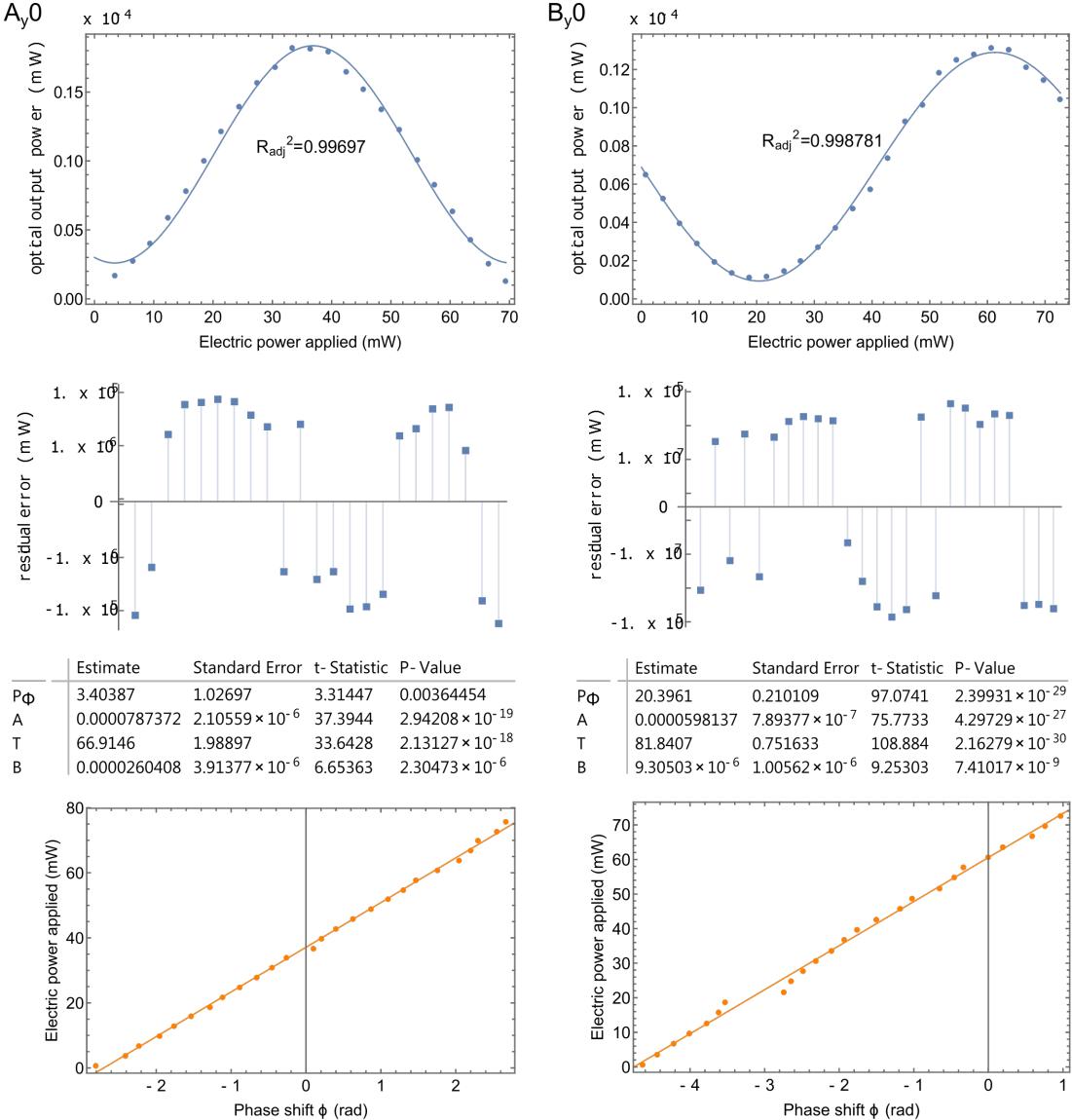


FIGURE 1.7. Characterisation for the implemented phase on three exemplary phase shifters on the device, chosen as a worst case (A_y0) and a typical case (B_y0). For each phase shifter, we report in column, from top to bottom: i) the nonlinear fit of raw experimental data, collected applying electrical power to the phase shifter, while measuring the optical power outputted at the corresponding optical port (vertical axis). The fitting is performed using the nonlinear Eq. 1.43. The value of R^2 is reported in the plot for each separate case. ii) A plot of the residuals corresponding to the nonlinear fits mentioned before. iii) A table of the parameters obtained in the same nonlinear fit is displayed. For each free parameter in the nonlinear model, we report an estimate of both the fitted value, as well as its statistical uncertainty (Standard Error), and a test of its significance (t-Statistic and P-value). iv) For clarity, we report also the linear dependence of the phase shift implemented by each heater against the corresponding electrical power applied.

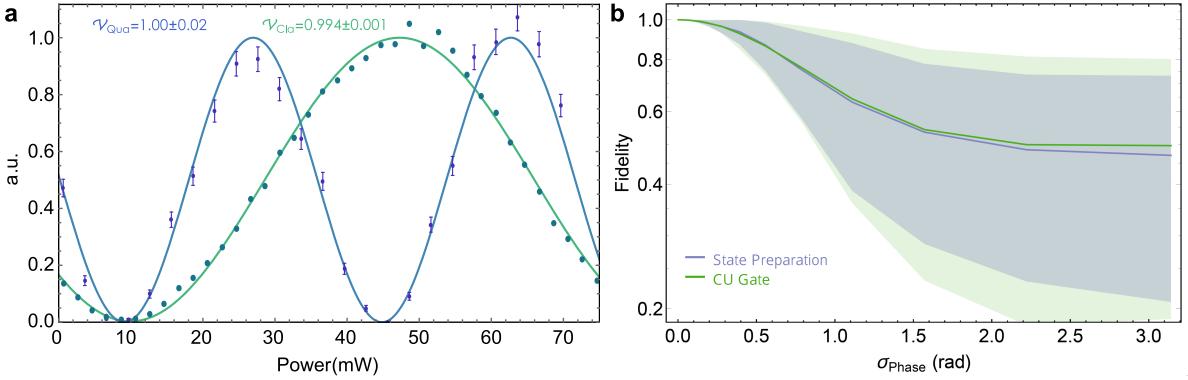


FIGURE 1.8. **a** Quantum (blue) and classical (green) interference fringes, along with their respective visibilities v_{Qua} and v_{Cla} . The fringes reported here were obtained in the device using the photon sources part and configuring the final interferometer of the control registry. The vertical axis shows the renormalised optical power, recorded via a photodetector (classical), or as photon coincidences from the SNSPD (quantum). Normalisation of the optical signal is against the peak value from the LSF: this choice allows for some experimental datapoints with optical power > 1.0 . The high visibility is essential to verify the high-performance and correct characterisation of the device. **b** Simulation of the quantum photonic device performances in terms of the average state preparation fidelity (violet) and CU-gate fidelity (green) for different levels of Gaussian noise in the phase shifters, as described by σ_{phase} spanning the full interval of values tested in the experiments reported in the main paper. The behaviour for the simulated fidelity of the two operations is very similar, with small discrepancies.

continuously a gate $\hat{R}_z(\theta)$, obtaining quantum interference with high visibility (1.00 ± 0.02 , Fig. 1.8).

1.4.4.3 Estimating and simulating phase errors in the integrated photonic device

The quantum photonic experimental setups used in this Thesis adopt thermo-optical phase shifters to implement both the quantum state preparation and unitary evolution of the qubits. When supplied with DC current, the phase shifters act as non-Ohmic resistive heaters, thus changing locally the refractive index of the waveguide [129]. Each of the heaters is driven and controlled independently using multi-channel current drivers, in order to reduce electrical cross-talk among the heaters. Cross-talk is expected because of a shared ground connection in the quantum photonic circuit. Each driver was hosted in a board featuring 12 bits DAC, controlled by the experiment software via RS-232 interface.

Thermal cross-talk among the phase shifters may also occur in a thermo-optically controlled device, because of heat flowing across the dioxide layer and the electrical connections to waveguides other than the targeted one. Thermally induced, unwanted changes in the refractive index

of other optical paths alter the implemented phase shifts in a non-controllable way, introducing systematic errors in the setup. In order to compensate for this effect, different calibrations for different chip configurations were preliminary run. For each configuration, all the heaters are constantly driven at a certain current, set to implement an appropriate state preparation and/or gate operation, except the heater to be calibrated. This heater is supplied with a range of currents (and thus electrical power, P_{el}), as oscillations in optical power P_{op} at the output of the corresponding optical path are recorded. The collected data can be fitted according to the non-linear function

$$P_{op} = B + A \cos \left[\frac{2\pi}{T} (P_{el} - P_\Phi) \right] \quad (1.43)$$

where B is a background, A is the maximum amplitude of the signal, T its period, and P_Φ is the offset power value for the heater in such a configuration, all obtained from the model fit. Detailed calibrations are reported in Fig.1.7 for three exemplary phase shifters, where we also report the relevant statistical parameters of each fit. The fits were performed using the built-in functions from *Mathematica*®, that estimate the parameters using a (non-linear) LSF approach, where the optimisation is performed using a numerical global optimisation method (“RandomSearch”). All fits show R^2 values close to one, thus suggesting that the model can adequately reproduce the data observed from the measurements. Also, we report high t-statistics and low p-values for most of the parameters, testing the significance of all the parameters adopted in the model, with the exception of the aforementioned background B parameter in those cases where the fringe visibility is particularly high (i.e. no significant background is present). In conclusion, the data analysis gives evidence of the suitability of the model in Eq. 1.43 to describe the physics of our device.

After the calibration, the targeted phase $\bar{\varphi}$ for a heater is obtained driving the heater with P_{el} such that:

$$\bar{\varphi} = \frac{2\pi}{T} (P_{el} - P_\Phi) \quad (1.44)$$

and therefore, propagation of stochastic errors in P_{el}, T, P_Φ affects $\bar{\varphi}$. From experimental non linear fits, relative statistical uncertainties are $\sigma_{P_\Phi}/P_{\max} \approx 0.2\%$ and $\sigma_T/T \approx 1.1\%$. Therefore, inaccuracies in the current supplied by the driver to each heater (± 0.005 mA), affecting the actual value of P_{el} , can be neglected as they are less than 0.04% for all the heaters, in the standard configuration used for phase estimation experiments in this paper.

Confining the propagation of errors to only T, P_Φ , and averaging over the full interval of P_{el} adopted in the experiment (appr. 5 – 80 mW, slightly different for each heater due to calibration), an average precision $\sigma_{\text{exp}} \approx 0.01$ rad. can be estimated for the phases, as experimentally implemented by the heaters. In conclusion, systematic errors in the setup can be drastically reduced via an accurate calibration procedure. Nevertheless, in our experiment the effective phase φ implemented by each of the phase shifters is affected intrinsically by a stochastic noise. This can be estimated from the fits of experimental data, into an uncertainty σ_{exp} . A natural way to investigate the role of imperfect calibration in our device is then to synthetically replace each of

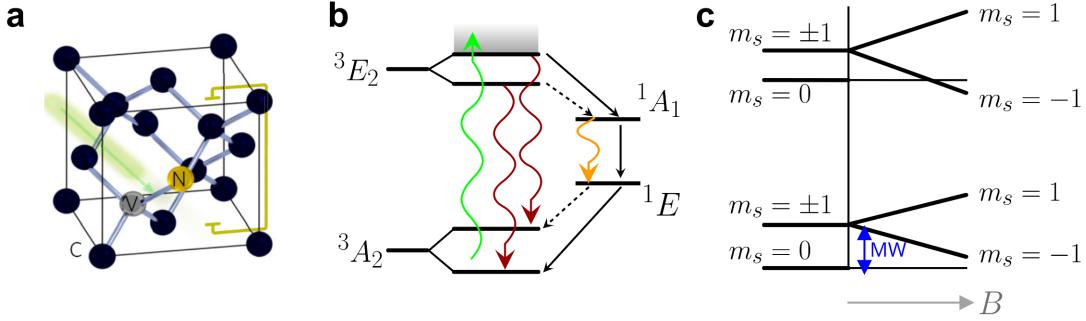


FIGURE 1.9. **a** Rendering of the Nitrogen–Vacancy (gold, grey respectively) defect in the lattice of diamond Carbon atoms (black). **b** Discrete energy levels of the NV–centre. The main levels used in the main text are as black lines, further excited states are as grey shadowing. Curve arrows indicate optical pumping (green) or radiative decay processes (red, orange). Straight arrows indicate non–radiative decay, with strong (weak) decay paths as solid (dotted) lines. **c** Zeeman splitting of the $m_s = \pm 1$ ground and excited levels in an external magnetic field of intensity B (along the NV–centre z –axis). We also highlight the transition addressed via microwave control (in blue). Further labels are explained in the main text.

the $\bar{\varphi}$, implementing the correct state and unitary preparation, with a phase value sampled from a Gaussian distribution characterised by a variance $\sigma_{\text{phase}} > \sigma_{\text{exp}}$. An increase in the stochastic phase uncertainty introduced (σ_{phase}) corresponds to a decrease in the fidelities for state preparation and gate implementation. A simulation of this dependency is reported in Fig. 1.8b.

1.5 NV–centres in diamond

In the lattice of diamond, naturally occurring point–defects are the substitution of a C with a N atom, and formation of a paired lattice vacancy (Fig. 1.9a). The 6 electrons bound to this defect form an effective spin–1 electron, that interacts with the closest $\mathcal{O}(10)$ lattice sites, before nuclear dipolar interactions become negligible [130]. In high–purity diamond samples, the species present at said sites are, beside the native nitrogen, ^{13}C isotopes, which have a natural⁷⁵ abundance of 1.1% [131]. Usually, the nitrogen–vacancy pair so formed is negatively charged (NV^-). First observed only in clusters, once detected as single defects they raised a huge interest, due to the plethora of applications that they enabled, ranging from single–photon or plasmon generation to bio–compatible sensing and quantum registers [132]. Even if NV–centres are nowadays not primarily targeted as potential qubits for quantum information processing, still NV–NV coupling and long coherence times have been demonstrated, whereas coupling to photonic waveguide is actively investigated. Additionally, NV–centres are sensitive to electric and magnetic fields, as

⁷⁵ Isotopically purified diamond samples might have about 10–fold enhancement or deprivation in the ^{13}C relative abundance.

well as strain and temperature changes in the environment, making them ideally suitable as quantum sensors [130, 133].

NV-centres occur naturally in the structure of diamond, however their random localisation would not offer a practical platform. Therefore, their (almost) deterministic fabrication has been successfully demonstrated adopting either Chemical Vapour Deposition (CVD) grown diamond thin-films, radiation damage or ion implantation in bulk diamond [132]. A crucial property of NV-centres is the sharp Zero Phonon Line (ZPL) exhibited in their absorption (emission) spectra. This property signifies that they behave as atomic defects, whose discrete levels' transitions can be addressed optically, already at room-temperature [134]. Room-temperature operation is a major advantage, when compared against other solid state platforms, e.g. quantum dots. The atomic levels of the NV-centre are illustrated in Fig. 1.9b. In particular, the ground state is the triplet 3A_2 , separated by an energy of 1.945 eV from the excited state triplet 3E_2 [132]. Usually, two of the ground spin energy levels are picked to define a qubit, e.g. in the basis $\{|0\rangle \equiv |{}^3A_2, m_s = 0\rangle, |1\rangle \equiv |{}^3A_2, m_s = -1\rangle\}$. We stress the presence of an additional doublet, 1A_1 , that plays a fundamental role in the electron spin addressing [130, 132].

A fundamental property of the NV-centre is the possibility to manipulate the electron spin state optically. This is achieved via a *confocal* microscope, delivering light within an optical volume of approximately $1 \mu\text{m}^3$, so that a single NV-centre can be addressed [133]. The initialisation in a fiducial, well-defined state occurs because of the electronic structure of the NV-centre, as detailed via a Lindbladian formalism in [130]. For brevity, we summarise here a qualitative picture. The steady-state for a given temperature is approximately given by the corresponding Boltzmann distribution, so that we can simply expect $\text{Pr}(|{}^3A_2, m_s = 0\rangle) \sim \text{Pr}(|{}^3A_2, m_s = \pm 1\rangle) \sim 1/3$, and negligible probabilities of occupation for all other states. If we now pump the NV-centre with visible, 532 nm green light, the populated ground spin states are non-selectively excited. At low-temperature, the excitation is resonant, within the ZPL, to the excited states 3E_2 , and the process is spin-preserving [135]. At room temperature instead, due to phononic spectral broadening, the excitation is non-resonant to states above 3E_2 (shown as a grey shadowing in Fig. 1.9b), from which the electron quickly decays back to 3E_2 , but the transitions do not strictly preserve the spin anymore. Further energetic decay can follow essentially two paths. One is the radiative, $|{}^3E_2, m_s = 0\rangle \rightarrow |{}^3A_2, m_s = 0\rangle$, with emission of a 637 nm photon (*photoluminescence*, *PL*), which is spin-selective. The alternative path is the decay mediated by the 1A_1 level, thus called an *inter system crossing*, which is the preferential decay path for $|{}^3E_2, m_s = \pm 1\rangle$ [130]. In this case, photons emitted are in the infrared, and the decay is not spin-preserving. Therefore, the overall effect of the optical pumping is to enrich mostly the population of the $m_s = 0$ ground and excited triplets, at the expense of the $|{}^3A_2, m_s = \pm 1\rangle$ state population, achieving a new steady-state in a few μs . Once the laser is switched off, the occupancy of $|{}^3E_2, m_s = 0\rangle$ will quickly decay radiatively, along with the residual 1A_1 population, that will decay to the ground triplet. In this way, $\text{Pr}(|{}^3A_2, m_s = 0\rangle) > 0.9$ can be obtained.

Also optical–readout is possible, by invoking a protocol based upon the same physical mechanisms. In fact, we observed how the optical pumping is spin–preserving, and excited states with $m_s = 0$ ($m_s = \pm 1$) decay preferentially via a (non–)radiative process, in the visible wavelengths range. Therefore, by optically pumping the NV–centre, the qubit state $\rho(\tau)$ at any given moment can be estimated via the extracted PL. More in detail, named μ_0 and μ_1 the rates of photon emission for the states $|0\rangle$ and $|1\rangle$ respectively, the number of emitted photons n_γ expected is governed by a Poissonian process [130]:

$$\Pr(n_\gamma|\rho(\tau)) = \frac{(P_0\mu_0 + (1-P_0)\mu_1)^{n_\gamma}}{n_\gamma!} e^{-P_0\mu_0 + (P_0-1)\mu_1} \quad (1.45)$$

where $P_0 \equiv \text{Tr}[|0\rangle\langle 0| \rho(\tau)]$ is the probability of occupation of the state $|0\rangle$. Therefore, states $|0\rangle$ and $|1\rangle$ can be distinguished via their PL *contrast* $C = (\mu_0 - \mu_1)/(\mu_0 + \mu_1)$. At cryogenic temperatures, the resonant optical excitation ensures that a *single–shot* measurement has $C \sim 1$, and a selective spin–readout is made possible [135]. At higher temperatures, the thermal broadening reduces this selectivity, and high contrast can be recovered by repeating the experiment M times⁷⁶, so that the final:

$$\frac{1}{C} = \sqrt{1 + \frac{2(\mu_0 + \mu_1)}{M(\mu_0 - \mu_1)^2}}. \quad (1.46)$$

Then, from the difference in PL one can infer the initial P_0, P_1 . In typical setups, the PL signal is enhanced and detected via an APD. Notably, if recorded the time of arrival of the photons at the APD bears additional information, as the spin difference in PL tends to smear out after ~ 700 ns [136]. Finally, we note how the recent demonstration of electrical readout might pave the way to setups more compact than those embedding confocal microscopes [137, 138].

As aforementioned, the NV–centre is not an isolated system, because it interacts with the surrounding environment of nuclear spins, each of atom species X , at lattice site χ . The corresponding Hamiltonian thus includes also terms referred to the environment, and reads⁷⁷ [131, 133]:

$$\hat{H}_{\text{nv}} = \Delta_{\text{gs}} \hat{S}_z^2 + \mu_B g \mathbf{B} \cdot \hat{S} + \hat{S} \cdot \sum_{X,\chi} (\mathbf{A}_{X,\chi} \cdot \hat{I}_{X,\chi}) + \mu_n \mathbf{g}_n \cdot \mathbf{B} \cdot \sum_{X,\chi} (\hat{I}_{X,\chi}) + P \hat{I}_z^2. \quad (1.47)$$

The *zero–field splitting* $\Delta_{\text{gs}} \approx 2.87$ GHz represents the ground–state energy splitting in absence of external fields. The second term in Eq. 1.47 is instead the *precession* of the NV electron spin in the external magnetic field \mathbf{B} , incorporating the Bohr magneton μ_B and the free electron $g = 2.0023$ factor⁷⁸. A fourth term captures the *nuclear Zeeman splitting* in the same external magnetic field⁷⁹, with μ_n the nuclear magneton and \mathbf{g}_n the nuclear gyromagnetic factor.

⁷⁶ We anticipate that repeated measurements need not be the most efficient way to retrieve such information as shown in [136] and in Sect. 3.3.

⁷⁷ Note that for ease of reference with the NV literature, we have kept the shortened notation $\hat{S}_i \equiv \hat{\sigma}_{i,e^-}$ for the Pauli matrices referred to the NV electron spin system, and $\hat{I}_i \equiv \hat{\sigma}_{i,n}$ for those referred to the generic nuclear site n .

⁷⁸ Simplifying the effective g –factor tensor [133].

⁷⁹ Considering that lattice distances are $\mathcal{O}(\text{nm})$, we assume the same magnetic field intensity across all the sites, included the electron's.

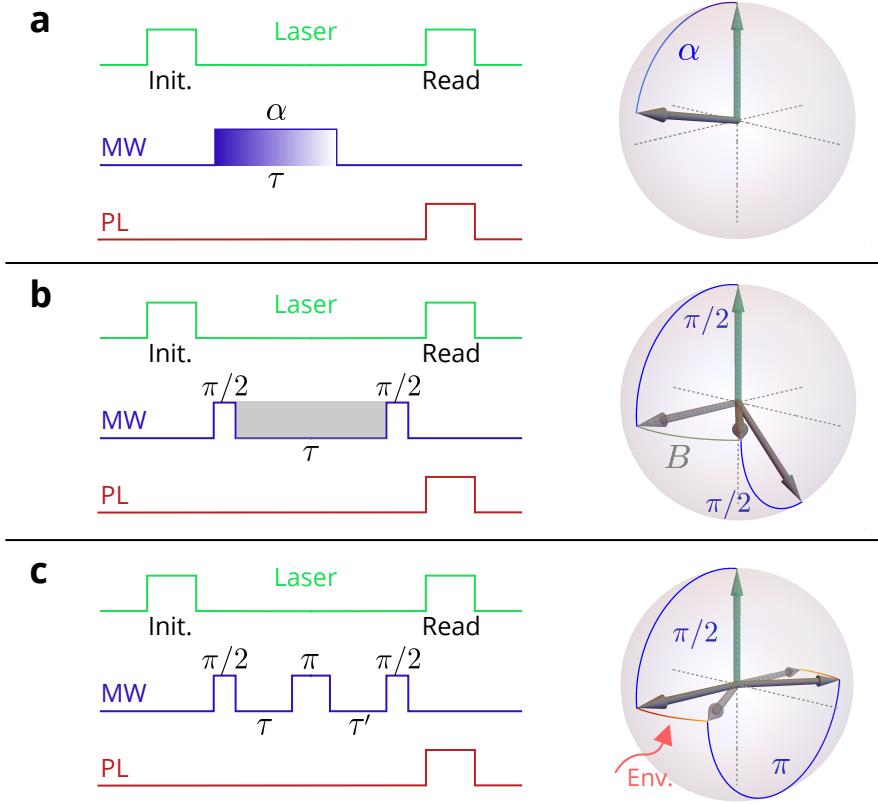


FIGURE 1.10. Exemplified coherent control of the electron state in an NV-centre. A typical: **a** Rabi sequence, **b** Ramsey sequence, **c** Hahn sequence are displayed. For each, on the left the sequence of control optical ('Laser') and microwave electrical ('MW') pulses, and the photoluminescence response from the NV-centre ('PL'). On the right, a Bloch sphere representation of the evolution of the state, following the control sequence displayed. In (c) we have omitted for clarity the last $\pi/2$ pulse.

Hyperfine terms of the form $\hat{S} \cdot \mathbf{A} \cdot \hat{\mathbf{I}}$ describe instead the interaction between the electron and the corresponding nuclear spin, combining dipolar and Fermi contact contributions [133]. The hyperfine tensor⁸⁰:

$$\mathbf{A} = \begin{pmatrix} A_{\perp} & 0 & 0 \\ 0 & A_{\perp} & 0 \\ 0 & 0 & A_{\parallel} \end{pmatrix} \quad (1.48)$$

is approximately diagonal, and the elements A_{\perp} (A_{\parallel}) are the (non)axial hyperfine parameters. Finally, the quadrupole splitting $P\hat{I}_z^2$ only applies to the ^{14}N nuclear spin (hence we drop the X, χ notation), and has $P = -4.945$ MHz.

Eq. 1.47 also hints at how coherent manipulation of the initialised $|0\rangle$ state can be achieved. Indeed, applying a continuous-wave field in the microwave (MW) range, it is possible to lift the

⁸⁰This term will be discussed further in Sect. 3.4.

degeneracy of the $|m_s = \pm 1\rangle$ ground states. The MW field is often locally applied via nanofabricated metallic striplines upon the diamond sample [134] (a pictorial representation is offered in Fig. 3.5). When the MW field is resonant with Δ_{gs} , Rabi oscillations between $|0\rangle \leftrightarrow |1\rangle$ occur [133, 139]. Therefore, by adjusting the duration of a pulsed MW field τ_{MW} , it is possible to rotate the initialised $|0\rangle$ of an angle $\theta = 2\pi/\omega_{\text{Rabi}}$ (see Fig. 1.10a), with ω_{Rabi} the corresponding Rabi frequency. The obtained state can be extracted via the projective measurement described above (i.e. an additional laser readout pulse), leading to a complete *Rabi sequence* experiment.

Coherent dynamics of the NV electron spin can also be observed in the presence of a external magnetic field B , that for simplicity we assume here aligned along the z -axis of the Bloch sphere describing the spin qubit. We can start preparing the spin state in a superposition of the $|0\rangle$, $|1\rangle$ eigenstates, by applying a MW $\pi/2$ pulse. Then, according to the second term in Eq. 1.47, the spin will precess with a *Ramsey* frequency $\omega_{\text{Ram}} = g\mu_B B$, as in Fig. 1.10b. An experiment embedding this free evolution of the spin prepared in the $|+\rangle$ state, for a time τ , is known as a Ramsey experiment.

Both Rabi as well as Ramsey oscillatory phenomena decay when the pulse or waiting times prolongate too much. In particular, when $\tau_{\text{MW}} \gg T_2^*$ (or equivalently $\tau \gg T_2^*$), the *inhomogeneous* decoherence time. We have already introduced in Sect. 1.1.2 the homogeneous decoherence time, T_2 . The difference between the two (as well as the reason of their naming) is that T_2 is referred to a single quantum system (and hence qubit), whereas T_2^* describes the dephasing process for an ensemble of quantum systems, whereby incoherent evolution among the different systems might contribute to the dephasing occurring in each system singularly, so that $T_2^* \leq T_2$. Perhaps counter-intuitively, T_2^* applies not only to ensembles of NV-centres, but also to a single electron spin. The reason lies in the form of the hyperfine term $\hat{S} \cdot \sum_{X,\chi} (\mathbf{A}_{X,\chi} \cdot \hat{\mathbf{I}}_{X,\chi})$, so that inhomogeneities here occur because of couplings with different species, each with its own hyperfine parameter, to a single quantum system.

In NV–centre experiments, therefore, T_2^* describes the dephasing process observed in Ramsey experiment. Rabi experiments are ideally subjected to T_2^* as well, however repeated sequences of strongly–driven evolution has been observed to produce a refocusing effect on the spin, leading to a quadratic instead of an exponential decay, similar to the quantum Zeno effect [140]. In order to observe the pure T_2 decoherence time, one needs to address the inhomogeneous dephasing process outlined before. The solution is to apply a *Hahn echo* sequence, completely equivalent to the scheme proposed for NMR [134], and outlined in Fig. 1.10c. In this case, after the initialisation in the $|+\rangle$ state, the free evolution for a time τ_1 (eventually in absence of any significative external magnetic field) is ended by an additional MW π -pulse. Since the hyperfine coupling is often simplified⁸¹ as $\hat{S}_z \sum_{X,\chi} (A_{\parallel}^{X,\chi} \hat{I}_z^{X,\chi})$ [140], the effect of the 180° rotation is to reverse the dephasing dynamics on the electron spin. The assumption behind a Hahn sequence is that, as long as the contributions from the bath are approximately constant within the timing of a complete sequence,

⁸¹ This approximation will be discussed in more detail in Sect. 3.4.

waiting for a time $\tau_2 \sim \tau_1$ after application of the π pulse will (almost) compensate the kick on the electron spin. In the literature, Hahn echo sequences with $\tau_1 \neq \tau_2$ are named *Hahn peak* experiments, whereas the case $\tau_1 \sim \tau_2$ goes under the name of *Hahn signal* experiments [134]. T_2 can be extracted from the envelope of the PL obtained out of Hahn signal experiments, when scaling the free evolution time $\tau_1 \sim \tau_2$. This will be discussed in detail in Sect. 3.5.2.

We have thus illustrated how it is possible to obtain a well-defined qubit from the electronic levels of an NV-centre, how to initialise it, address it and coherently manipulate it electro-optically. Therefore, they can be legitimately adopted as qubits, and in particular quantum sensors, as we will show in Chap. 3.

1.6 *Author's contributions to the Chapter*

Most of this chapter is my review of established results available in the literature, enriched by personal observations and on-purpose analyses.

The classification and nomenclature of results and research areas introduced in Sect. 1.2 is mainly my own.

The setup introduced in Sect. 1.4.4 was designed by dr. J. Silverstone and dr. D. Bonneau, and manufactured by the Toshiba research unit under supervision from prof. M. Thompson. The setup assembly (as in Sect. 1.4.4.1) was prepared by dr. R. Santagati and dr. J. Wang with help from mr. S. Paesani. I performed instead the analysis (and part of the data taking) reported in Sect. 1.4.4.2–1.4.4.3. Some of the corresponding **experimental results and analyses were published as [141] and [128]**.

2

DEPLOYING NOISE-RESILIENT BUILDING BLOCKS OF A QUANTUM COMPUTER

2.1 Phase Estimation (PE) algorithms

Since its introduction as an application of the quantum Fourier transform, phase estimation (PE) procedures have been considered a fundamental building block for quantum technologies in the broadest sense [5]. The quantum speedup expected by PE over classical counterparts [142] withstands as a key advantage in digital quantum computing protocols, such as prime–factorisation [143] and other computational algebraic number theory applications [144], partition functions approximation [145], and quantised Markov chain approaches [146]. Moreover, PE plays also a crucial role for digital quantum simulations of systems either ideal [147], or from real–world quantum chemistry [148]. Therefore, it is crucial to provide PE schemes that can cope with realistic implementations, if we are to deliver such numerous promised applications in NISQ devices.

The goal of phase estimation is, given a unitary operator¹ \hat{U} and a generic quantum state $|\psi\rangle$, to estimate an eigenvalue $e^{i\phi}$ of \hat{U} within the support of $|\psi\rangle$. An arbitrary $|\psi\rangle$ can in fact be expanded onto an eigenbasis $\{|\phi_k\rangle\}$ of \hat{U} , so that the evolution reads:

$$\hat{U}|\psi\rangle = \sum_k \alpha_k \hat{U}|\phi_k\rangle = \sum_k \alpha_k e^{i\phi_k} |\phi_k\rangle, \quad (2.1)$$

where the eigenphase $\phi_k \in [0, 2\pi)$ because of the phase periodicity. PE algorithms therefore aim at inferring ϕ_k from measurements on the evolved $\hat{U}|\psi\rangle$. In general, and in many realistic cases of interest [128, 149, 150], PE algorithms accept as input state a generic $|\psi\rangle$, to be progressively projected onto an eigenstate $|\phi_k\rangle$ along the evolution implemented by \hat{U} . However, the required

¹I.e. an operator satisfying $\hat{U}\hat{U}^\dagger = \hat{\mathbb{I}}$, and thus preserving the normalisation of the evolved state.

resources to estimate the corresponding ϕ_k can be shown to scale $\sim 1/\alpha_k$ [150], making this approach impractical to address the whole eigenspectrum of a high-dimensional Hamiltonian. Therefore, it is often assumed the input state will be prepared close enough to the eigenstate of interest $|\phi_k\rangle$, such that $|\alpha_{\tilde{k}}|^2 \ll 1 \forall \tilde{k} \neq k$. In this way, outcomes probabilistically inherent any eigenvector other than $|\phi_k\rangle$ can be neglected, avoiding any deconvolution. Unless otherwise stated (such as in Sects. 2.1.6.1&2.1.7), in the following we will therefore consider the simplified case $|\psi\rangle \simeq |\phi_k\rangle$, and therefore simplify the notation with $|\phi\rangle \equiv |\phi_k\rangle$, $\phi \equiv \phi_k$.

Standard PE algorithms work by interfering paths in which either $\hat{1}$ or \hat{U} is applied to $|\phi\rangle$, as controlled by an ancillary quantum register(s), (each) prepared as a single-qubit $|+\rangle$ state [5]. The following reasons justify the need for a controlled unitary $C\hat{U}$. First, employing a single-qubit ancillary register allows to perform easier projective measurements, abstracting from the dimension of the Hilbert space of \hat{U} . More importantly, recursive application of \hat{U} can extract additional information about ϕ as long as \hat{U} commutes with the measurement operation, and sequential PE methods exploit this property. A destructive measurement on $\hat{U}|\psi\rangle$ would invalidate this, whereas the partial information required for a single measurement can be stored in an ancillary register. The basic circuit to perform such a measurement is as in Fig. 2.1b, with control parameters M and θ , where the gates $R_z(\theta)$ and H are respectively a rotation of the ancillary qubit along the z -axis of the Bloch sphere, and the Hadamard gate (see Sect. 1.1.1).

The state evolution across the various steps can thus be written as:

$$|+\rangle \otimes |\phi\rangle \xrightarrow{C\hat{U}^M} |0\rangle \otimes |\phi\rangle + |1\rangle \otimes e^{iM\phi} |\phi\rangle \xrightarrow{R_z(-\theta)} |0\rangle \otimes |\phi\rangle + |1\rangle \otimes e^{iM\phi-\theta} |\phi\rangle \quad (2.2)$$

$$\xrightarrow{H} \left(\frac{1+e^{iM\phi-\theta}}{2} |0\rangle + \frac{1-e^{iM\phi-\theta}}{2} |1\rangle \right) \otimes |\phi\rangle \quad (2.3)$$

hence the probability to project the control registry on the computational basis state $|E\rangle$ with $E \in \{0, 1\}$, i.e. “observe a datum E ”, is conditionally dependent on the exponent M of the unitary operation and the angle applied on the ancillary θ_M :

$$P(E|\phi; M, \theta) = \begin{cases} \frac{1+\cos(M\phi-\theta)}{2} & \text{if } E = 0 \\ \frac{1-\cos(M\phi-\theta)}{2} & \text{if } E = 1 \end{cases} \quad (2.4)$$

at the end of this single step estimation procedure. $P(E|\phi; M, \theta)$ is known in general as the *likelihood function*, here describing a Bernoulli experiment, as $P(1|\phi; M, \theta) = 1 - P(0|\phi; M, \theta)$.

2.1.1 Performance metrics

Various protocols have been suggested to exploit the basic measurement operation in Eq. 2.3 to infer ϕ_k . In the following paragraphs, we will compare different PE algorithms. As with every other algorithm, different figures of merit can be considered when analysing performances.

- *Accuracy* of the expected outcome from the estimation procedure $\mathbb{E}(\phi)$, here quantitatively given² by the error $\delta = |\mathbb{E}(\phi) - \phi|$.
- *Error probability*, ϵ . If a required tolerance level in the estimate $\tilde{\delta}$ is known in advance³, the probability ϵ that a PE run provides $\delta < \tilde{\delta}$ might be preferred as a metric. In cases of interest, where the PE methodology is non-deterministic, we often provide statistical information on ϵ by cumulative results from independent runs.
- *Width* of the quantum circuit required (w), i.e. the number of quantum registries involved. At the expense of parallelism, a lower width intuitively favours experimental implementations⁴.
- *Depth* of the quantum circuit required (d), intended as the number of discretised timesteps undergone by the coherently evolved⁵ input quantum state. A timestep occurs when at least one of the quantum registries of the circuit undergoes one quantum gate. Gates that can be performed in parallel on different qubits thus count as a single timestep.
- *Efficiency* of the method (η), here loosely defined as the inverse of the number of timesteps executed, divided by the δ achieved. As many algorithms require to invoke the quantum circuit more than once (with a measurement operation concluding each of these macrostep), this figure of merit cannot be replaced by d .

Crucially, we will also compare some algorithms on the basis of their *noise resilience*. Even if we do not provide a general formal definition here, we will intuitively use the fact that in presence of an amount v of external noise altering the quantum circuit, some of the figures of merit introduced will depend on the noise as $\delta(v)$, $\epsilon(v)$, $\eta(v)$. The exact form of this dependency in general depends on the particular source of noise considered, and sometimes also on the specific experimental implementation. However, restricting to particular cases, we will be able to observe and compare how severely the noise introduced in controlled conditions can impact some key algorithm's performance metrics.

2.1.2 Information theoretic PE

Perhaps the most direct way to infer ϕ from the likelihood in Eq. 2.4 is a brute-force, randomised search in $[0, 2\pi]$, performed iterating the basic measurement step of Fig. 2.1b. In absence of a

²we express here δ in terms of an expectation value, to render our wording compatible also with probabilistic PE methods detailed in the following

³This is for example the case for quantum chemistry applications, where specific accuracies are targeted. See Sect. 2.3 and Chap.4 for additional information.

⁴Different quantum computing architectures might be sensitive to this limitation in different ways. Photonic implementations outlined in this Thesis are particularly susceptible to favouring low width circuits.

⁵We stress that the coherence in the quantum state across such timesteps is crucial for the definition of d . As example, in iterative instances of some algorithms coherence in the quantum state across iterations is not required, and therefore d applies to the quantum circuit invoked for a single iteration.

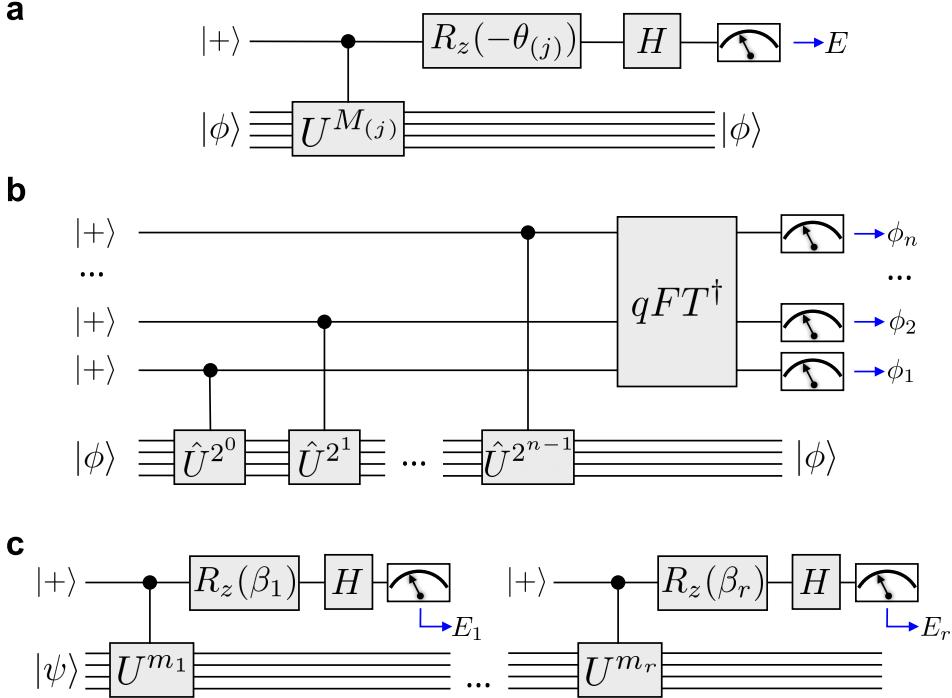


FIGURE 2.1. Quantum circuits for various PE algorithms illustrated in the text. H stands for the Hadamard gate, R_z are single-qubit rotations and the quantum Fourier Transform (qFT) is defined in App. Fig. A.1. **a** The basic circuit for ITPE, KPE, IPEA and RFPE algorithms. In particular, at step j : i) ITPE has random choices of the parameters (see 2.1.2), ii) KPE adopts $M_j = 2^j$ and $\theta_j = 0$, iii) IPEA has $M_j = 2^j$ and $\theta_j = 0.0\phi_{n-j+2}\dots\phi_n$, iv) RFPE chooses the parameters adaptively according to Eqs. 2.20 and 2.19. **b** SPEA, fully sequential circuit. **c** Generic circuit for multi-round algorithms.

priori information, the search interval can be discretised uniformly in $n \in \mathbb{N}_0$ subintervals, chosen such that their width satisfies the targeted accuracy $2\pi/n < \delta$. A single mode reconstruction from a signal discrete sampling follows, to that the algorithm is known as *Information Theoretic Phase Estimation* [151], ITPE. At each iteration j of ITPE, the control parameters are picked randomly with $M_j \in \{1, 2, \dots, n-1\}$ and $\theta_j \in [0, 2\pi]$. After J independent experiments, each obtaining outcome E_j , reconstructing the $\phi_l = 2\pi l/n \simeq \phi$ is possible solving the classical maximisation problem of:

$$P(E_1, \dots, E_J | l) = \prod_{j=1}^J P(E_j | 2\pi l/n, M_j, \theta_j) \quad (2.5)$$

for l , with each term in the product referring to the likelihood function Eq. 2.4. The classical post-processing required is $\mathcal{O}(Jn)$ to process all the probabilities from Eq. 2.5 and retrieve the peak one. It can be shown via a Chernoff bound argument how $s = \mathcal{O}(\log n)$ iterations suffice to achieve error probability $\epsilon < 1/2$ [151]. The algorithm is optimal in circuit width, as it requires only an extra ancillary quantum register to perform the whole estimation. However, assuming a non-compiled

implementation of \hat{U}^M requires $\mathcal{O}(n)$ operations at worst, ITPE accumulates $\mathcal{O}(T_n) = \mathcal{O}(\sum_{i=1}^n i)$ gates at the end of n iterations, each with a maximum depth $\mathcal{O}(n)$. This implies the necessity of a long coherent evolution across the M cascaded \hat{U} gates, which makes the algorithm prone to the impact of noisy implementations, a problem affecting all the PE algorithms reported here. At the same time, this approach allows to learn ϕ quadratically faster ($\delta \sim 2\pi/n$) and with exponentially fewer measurements, $\mathcal{O}(\log n)$, than by statistical sampling, where instead $\delta \simeq 1/\sqrt{n}$ is expected after n measurements.

2.1.3 Sequential PE Algorithm

A fully sequential PE algorithm (SPEA) can be obtained adopting the *quantum Fourier transform*, qFT [5]. Observing the circuit in Fig. 2.1a, a first crucial difference compared to the method in Section 2.1.2 is evident: SPEA expands the ancillary qubits in a register embedding as many qubits as the (binary) accuracy required n , thus increasing the circuit width to $\mathcal{O}(n + \dim[\hat{U}])$. Additionally, it is designed as a single-pass algorithm (i.e. $J = 1$), hence no classical post-processing of the measurement is required, nor any θ control angle. However, each step involves n implementations of powers of \hat{U} , so that the depth of the circuit dramatically increases to $\mathcal{O}(\sum_i^{n-1} (2^i) + n^2) \sim \mathcal{O}(T_{2^n})$, inclusive of non-compiled⁶ \hat{U}^i operations, and n^2 gates from the qFT implementation (see Appendix A.1). These features already put it at a disadvantage for experimental implementations, as it trades relatively inexpensive classical post-processing with an increase in costly quantum resources.

In detail, SPEA allocates an ancillary register for each of the n bits in the binary expansion of the eigenphase ϕ :

$$\phi = 2\pi \times 0.\phi_1 \dots \phi_n \quad (2.6)$$

and follows a predetermined schedule to target the accuracy of each single step. Unitary operations $U^M = U^{2^m}$ are performed on the register storing the eigenstate $|\phi\rangle$, choosing sequentially $m = 0, \dots, n - 1$, as controlled by the $(m + 1)$ -th ancillary register. The information about the most significant bits is extracted first. The evolution along the series of $C\hat{U}^M$ operations therefore reads:

$$|+\rangle_1 \otimes \dots \otimes |+\rangle_n \otimes |\phi\rangle \xrightarrow{CU^1} \frac{1}{2^{1/2}} \left(|0\rangle + e^{i1\phi} |1\rangle \right)_1 \otimes \dots \otimes |+\rangle_n \otimes |\phi\rangle \xrightarrow{CU^2} \dots \quad (2.7)$$

$$\xrightarrow{CU^{2^{n-1}}} \frac{1}{2^{n/2}} \sum_{M=0}^{2^{n-1}} \left(e^{iM\phi} |M\rangle \right) \otimes |\phi\rangle \quad (2.8)$$

using the shortened notation $|M\rangle = |b_1 b_2 \dots b_m\rangle$ with $\{b_1 b_2 \dots b_m\}$ the binary expansion of M . If we assume that the expansion in Eq. 2.6 is exact, then we can write the evolved state in Eq. 2.8

⁶I.e. the unitary exponentiation is intended to occur in the quantum circuit, instead of pre-computing it classically.

as:

$$\frac{1}{2^{n/2}} \left(|0\rangle + e^{2\pi i 0.\phi_1\dots\phi_n} |1\rangle \right)_1 \otimes \dots \otimes \left(|0\rangle + e^{2\pi i 0.\phi_n} |1\rangle \right)_n \otimes |\phi\rangle \xrightarrow{qFT^\dagger} |\phi_1\dots\phi_n\rangle \quad (2.9)$$

so that a measurement in the computational basis for each of the qubits in the ancillary register retrieves deterministically the n bits of the expansion. That is, $\mathcal{O}(n)$ measurements provide the phase estimate, and these can here be entirely parallelised.

We stress that requiring $\mathcal{O}(2^n)$ depth and $\mathcal{O}(n)$ measurements is not detrimental in terms of the SPEA algorithm's efficiency η , when compared to $\mathcal{O}(T_n)$ gates required by the ITPE, as they are compensated by the respective output accuracies $\delta \simeq 2^{-n}$, as opposed to $\delta \simeq 1/n$. However, the depth of the circuit is often well beyond what can be achieved with current technologies.

2.1.4 Kitaev's PE algorithm

As a corollary to his approach for solving the *Abelian stabilizer problem* within a quantum information context [152], Kitaev developed a PE algorithm (KPEA) that bears his name [5]. The idea is to use a single ancillary qubit (reducing dramatically w), and perform a measurement straight after a single controlled gate with the chosen M_j . The corresponding circuit would be as in Fig. 2.1b, with $\theta_j = 0 \forall j$. A single exponentiation of the unitary and the absence of a qFT^\dagger step contribute to reduce d as in ITPE. In this way, KPEA addresses a major drawback of SPEA that, whilst deterministic, cannot be split into a series of shorter experiments. The daunting overall size of the quantum circuit required for fully implementing SPEA is then compromised in favour of a modified version, where additional classical post-processing and measurements reduce d and w of the circuit. The exponent M is chosen so that, if we want to obtain ϕ with n bits of precision, we perform n iterations where at the j -th iteration we have $M_j = 2^{n-j}$, and the measurement outcome is used to infer a single bit in the binary expansion of ϕ (Eq. 2.6) [152, 153]. However, observing the evolution after a single $C\hat{U}^{M_j}$ operation in Eq. 2.2, and the likelihood in Eq. 2.4, we obtain:

$$\left(\frac{M_j \phi}{2\pi} \right)_{\text{mod } 1} = \left(\frac{2^{n-j} \phi}{2\pi} \right)_{\text{mod } 1} \sim 0.\phi_{n-j+1}\dots\phi_{n-1}\phi_n, \quad (2.10)$$

so that inferring each bit from a single-shot E according to the rule $E = 0 \Rightarrow \phi_{n-j+1} = 0$ (and viceversa) is probabilistic. For each bit $\phi_j \in \{0, 1\}$, we have a potentially high error rate $\epsilon = (1 + (-1)^{\phi_j} \cos(M_j \phi)) / 2$. Therefore, in order to make the algorithm useful, it is necessary to repeat each bit estimation for a number of repetitions s [5]. The most intuitive way to use the statistical information so accrued is to employ a *majority voting* scheme⁷, inferring the bit ϕ_j

⁷ We shall emphasise immediately how majority offers a natural strategy, but it is not optimal: for pairwise experiments as those outlined in this Chapter, a single bit of information is extracted from s experiment instances. Actual occurrences of $\{0, 1\}$ in the series of outcomes are disregarded. Bayesian procedures, as it will be discussed later, substantially avert this strategy by updating estimates after each single (pairwise) measurement, i.e. an *online*

according to the outcome observed in a number of instances $s_E > s/2$ out of s repetitions. In an ideal case with only binomial noise in the measurement, $\epsilon \rightarrow 0$ as soon as $\sqrt{s_E} \ll s_E - s/2$.

A slightly different approach is to choose instead at random $\theta_j \in \{0, \pi/2\}$ across the s measurement repetitions per-bit [151]. We call this for simplicity a *trigonometric scheme*. Using Eq. 2.4, it is immediate to see that $2P(0|\phi, M, \theta_j) = \cos(M\phi)\cos\theta_j - \sin(M\phi)\sin\theta_j$, and therefore the allotted choices for θ_j lead to either:

$$\begin{aligned} P_{\sin} &= \frac{(s_0|\theta_j = 0) - (s_1|\theta_j = 0)}{(s|\theta_j = 0)}, \\ P_{\cos} &= \frac{(s_1|\theta_j = \pi/2) - (s_0|\theta_j = \pi/2)}{(s|\theta_j = \pi/2)} \end{aligned} \quad (2.11)$$

where we labelled e.g. $(s_0|\theta_j = 0)$ the number of measurements with outcome $E = 0$, when choosing the control parameter $\theta_j = 0$. We can thus use Eqs. 2.11 to estimate each $\beta_j := 2^{j-1}\phi \simeq \arctan(P_{\sin}/P_{\cos})$. After processing all j 's, the bit ϕ_j of the binary expansion of ϕ can be retrieved by post-processing the data acquired, using the β_j 's and according to the rule⁸:

$$|0.0\phi_{n-j+2}\dots\phi_{n-1}\phi_n - \beta_j|_{\text{mod } 1} < |0.1\phi_{n-j+2}\dots\phi_{n-1}\phi_n - \beta_j|_{\text{mod } 1} \iff \phi_j = 0 \quad (2.12)$$

and viceversa for the case $\phi_j = 1$. As this final processing of the outcomes occurs potentially after all measurements have been performed, it follows that the part running on the quantum circuit can be effectively parallelised across copies of the same circuit as in Fig. 2.1b.

A crucial point for KPE is estimating the repetitions s sufficient to render the inference in Eq. 2.12 reliable. The same Chernoff bound argument used for ITPE suggests $s \sim \log n$, which can be further verified in simulations [151], to render negligible the probability of mistakenly infer any bit ϕ_j . Therefore, a total number of measurements $\mathcal{O}(n \log n)$ is expected to achieve 2^{-n} accuracy in the estimate of ϕ , rendering the overall procedure efficient⁹.

2.1.4.1 Enhanced KPEA

KPEA is designed to target a single bit at a time, minimising the classical post-processing required at the cost of a log-factor in the total number of measurements $\mathcal{O}(n \log n)$. In [151], it

update. Superior performance against majority voting was observed for such ideal cases in [141]. However, in some experimental setups, such as the photonic ones discussed in this Thesis, experimental outcomes occur naturally in batches, and the overheads introduced by implementing online updates may outweigh the benefits. In Sect. 3.3.2 we show how a Bayesian strategy using *probabilistic extraction* from small batches of experiments can still be beneficial versus majority voting over larger batches (operational definitions can be found in Appendix B.1). Finally, Bayesian updates over offline batches of experiments were also suggested as a further improvement in setups where updates are time consuming [136]. The latter approach was published after the experiments in this Thesis were performed, hence it is not further investigated here.

⁸Several slightly different versions of this rule appear in the literature, see [151]. The definition in Eq. 2.12 was here solely chosen for a more immediate comparison with IPEA in Sect. 2.1.5.

⁹From this, it follows also that in KPE, adopting majority voting is a choice equivalent to using the method in Eqs. 2.11 and 2.12, as long as realistic accuracies are targeted. Indeed, the estimate $s \sim 60$ for $n = 10^4$ obtained by the simulations reported in [151] for the trigonometric scheme, satisfies simultaneously the worst-case requirement $\sqrt{s} < s/2$ for majority voting.

was proposed a protocol to reduce this cost to $\mathcal{O}(n \log^R n)$, with $R < 5$ for practical reasons, by performing additional classical processing.

The idea is to split the estimation given for KPEA in R rounds. A first round proceeds exactly like KPEA, throughout n iterations, but performing only $s = 1$ repetition of the measurement. Each following round $r \leq R$ will instead perform ns_r repetitions, and at each repetition picking a set $\{i\}$ of different exponents $0 \leq i \leq n - 1$, to implement the circuit in Fig. 2.1 with:

$$M = \sum_{\{i\}} 2^i. \quad (2.13)$$

In this way, the inference involves several bits in the binary expansion at once. The extracted data $\{E_{s_r}\}$ are used to compute via multi-bit inference the β_j parameters, that enter the same rule as Eq. 2.12 for computing each ϕ_j . Also this protocol is probabilistic, but it is proven in [151] that for each given bit, $\epsilon < \prod_r e^{-cs_r S}$, with c a constant and $S := |\{i\}|$, providing an enhancement in the total repetitions required.

We omit in this Thesis the details about the classical multi-bit inference required, to be found in [151]. The reason being that this multi-round PE approach is surpassed in the enhancement of the success rate as well as the multi-bit approach, by the algorithms given respectively in Sect. 2.1.5 and 2.1.6 .

2.1.5 Iterative Phase Estimation Algorithm

The Iterative Phase Estimation Algorithm (IPEA) also uses the circuit in Figure 2.1b, inheriting the advantages of KPEA in terms of w and d of the quantum circuit required. By combining the outcomes of successive experiments via a classical inference algorithm [154], it also trades expensive ancillary quantum channels with additional iterated measurement operations, and inexpensive classical computational resources. The crucial difference, when compared against KPEA, is that each bit in the (approximate if needed) finite binary expansion of the eigenphase ϕ (Eq. 2.6) is actually estimated in sequential order. In KPEA, the ordering followed when estimating the ϕ_j 's can be altered arbitrarily. This is not the case in IPEA [152, 154], as the outcome E_j of each measurement provides a feedback for the control parameter θ_{j+1} of the following experiment. IPEA therefore loses the parallelisable approach of KPEA, but using the information progressively accrued to manipulate the ancillary state *within* the quantum circuit, the inference process following Eq. 2.10 becomes deterministic.

As in other PE implementations, IPEA controls the two parameters M and θ according to a predetermined policy. In order to benefit the inference of the most significant bits by prior information about ϕ , the procedure starts from the least significant bit ϕ_n [154]. The first iteration step is performed using $\theta_1 = 0$ and $M_1 = 2^{n-1}$, which from Eqs. 2.4 results in $P(0) = \cos^2(\pi 0.\phi_n)$.

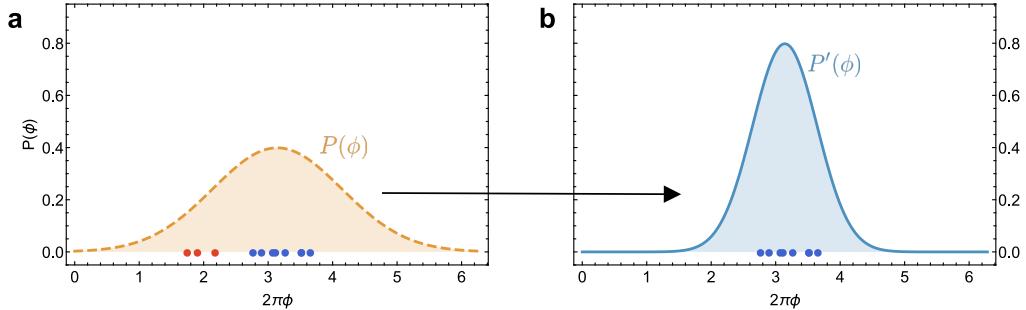


FIGURE 2.2. Intuitive representation of the rejection filtering process adopted for the RFPE algorithm in Sect. 2.1.6. **a** For each epoch, a starting Gaussian prior $P(\phi)$ (dashed orange curve) is represented by the support of the chosen *particles* (blue & red dots), each corresponding to a tentative $\bar{\phi}$ value. According to the likelihood in Eq. 2.17 and the latest obtained value E from the measurement, some particles are discarded (red dots). **b** A posterior (blue solid curve) is then refit against the remaining particles (blue dots).

Therefore, the last bit is deterministically¹⁰ implied from the measurement outcomes:

$$\begin{aligned} E_n = 0 &\iff \phi_n = 0 \\ E_n = 1 &\iff \phi_n = 1. \end{aligned} \quad (2.14)$$

The information obtained by the measurement is transferred to the quantum circuit and determines the $R_z(\theta_2)$ rotation, applied to the ancillary qubit, when estimating the more significant bit ϕ_{n-1} . Therefore, using θ_j increases the probability of obtaining the correct phase bit value at the corresponding iteration. The procedure continues extracting iteratively new bits, that are progressively more significant. To extract the n bits of the whole binary expansion, the algorithm requires n iterations of the experiment, where at the j -th iteration we choose:

$$M_j = 2^{j-1} \quad (2.15)$$

$$\theta_j = 0.0\phi_{n-j+2}\dots\phi_{n-1}\phi_n. \quad (2.16)$$

Applying the same reasoning as before, the bit ϕ_{n-j+1} is exactly determined by the datum E_j obtained via a projective measurement of the ancilla, because the feedback θ_j removes the residual in Eq. 2.10. In conclusion, IPEA is advantageous in terms of the number of total projective measurements to achieve $\delta \sim 2^{-n}$, which is reduced to $\mathcal{O}(n)$, at the cost of adding a single-qubit gate to the corresponding quantum circuit (see Fig. 2.1b). However, this additional cost can be deemed negligible whenever $M \gg 1$ and the projective measurement is slower than a single-qubit operation. This makes IPEA highly advantageous over KPEA in a broad variety of realistic cases.

¹⁰ As long as the binary expansion of ϕ is exact, see [154].

2.1.6 Rejection Filtering Phase Estimation (RFPE)

RFPE offers the same simple, intuitive approach of ITPE, combined with the advantages of sequential feedback introduced in IPEA [82], with which it shares the same quantum circuit setup (see Fig. 2.1). The fundamental differences in this approach, when compared to more traditional PE approaches, arise in how measurement data are processed, and used to make decisions about experimental control parameters. RFPE leverages on Bayesian inference, and stores the information about ϕ in a probability distribution $P(\phi)$. This *prior* distribution represents both the current hypothesis about the correct eigenphase for the injected $|\phi\rangle$ as $\mathbb{E}[P(\phi)]$, as well as the confidence in the estimate. If the distribution is characterised by its first two moments alone, such information is provided, respectively, by the mean μ and standard deviation σ of the distribution. Bayes' theorem provides the correct way to update such prior beliefs when provided with evidence. Upon a measurement outcome E , the theorem infers an updated distribution:

$$P(\phi|E; M, \theta) = \frac{P(E|\phi; M, \theta)P(\phi)}{\int P(E|\phi; M, \theta)P(\phi)d\phi}, \quad (2.17)$$

where the updated prior distribution $P(\phi|E; M, \theta)$ is known as the *posterior*. In both cases we have now made explicit the dependence upon the experimental parameters M, θ .

In general, the exact update of Eq. 2.17 is unfeasible, because ϕ is a continuous parameter, and the most effective distribution $P(\phi)$ in reproducing experimental data, inclusive of their uncertainties, might not be known in advance in a closed analytical form. In a first step, the search interval is therefore discretised with a set of *particles*, i.e. hypothetical (and here normalised) phase values $x_i \in [0, 1]$. These discrete values can be used to represent the prior as $P(\{x_i\})$, and we will hereafter refer to each probability as the *weight* associated to the particle $w_i := P(x_i)$. After each experimental outcome $E \in \{0, 1\}$, a prior update can be obtained through the corresponding likelihood function:

$$P(E|x_i; M_j, \theta_j)/\kappa = \begin{cases} \cos^2(M_j\pi x_i - \theta_j/2)/\kappa & E = 0 \\ \sin^2(M_j\pi x_i - \theta_j/2)/\kappa & E = 1, \end{cases} \quad (2.18)$$

where κ is a rescaling constant [82], to ensure $\sum_i w_i = 1$. The likelihood intuitively expresses the probability of datum E to emerge from chosen setup M, θ , if hypothetical eigenphase x_i was correct. As the values of $\{x_i\}$ are not bound to any specific bit of precision, the inference procedure in RFPE intrinsically acquires information on more than a single bit of ϕ simultaneously. This embeds the advantages of the several rounds per-measurement setup invoked by the enhanced KPEA (2.1.4.1), without the need for a strict policy about the experiment choice.

The discretisation leading to the update formulas of Eq. 2.18 still leaves open the challenge of an efficient usage of the classical resources required, as the accuracy in the PE outcome will depend crucially on the density of the particles: $\delta \simeq \mathbb{E}[|x_{i+1} - x_i|]$. An approximate estimate here indicates that a classical memory register of size $S \sim 20\delta^{-1}$ Bytes is required to store the particles' locations $\{x_i\}$ and weights $\{w_i\}$ for a naïve uniform distribution. This amounts to several GBs

for standard precisions required in quantum simulation (see also Sect. 2.3), which would be unfeasible e.g. in FPGA-run experimental setups [155]. The solution proposed in [82] was to adopt a *rejection filtering* protocol, a form of density filtering.

With this approach, at each *epoch*¹¹ j a host of particles are drawn according to the prior distribution. After the update as in Eq. 2.18 occurs, particles are discarded with probability $P(E|x_i; M_j, \theta_j)/\kappa$, i.e. according to how well each x_i was able to predict outcome E , optimising the parametrisation x among those allowed by the initial prior. The remaining samples model the posterior: for convenience, this is usually done by fitting¹² an updated Gaussian probability distribution $\mathcal{N}(\mu, \sigma)$. This represents the new prior, from which a host of fresh particles is drawn before a new experiment is performed. Intuitively, in this way a limited number of particles is constantly refocused near values where the highest likelihood of retrieving the correct eigenphase occurs, allowing the inference to be performed quickly¹³, in a memory limited environment.

Similarly to IPEA, a policy to choose the values of M_j and θ_j for each new experiment is required, in order to maximise the information gain. In RFPE, this need be extracted at each epoch from the prior distribution. It has been proven numerically [156, 157] that a near-optimal¹⁴ choice is represented by the particle guess heuristics (PGH):

$$\theta_j = M_j \mathbb{E}[P(\phi)], \quad (2.19)$$

$$M_j = \lceil 1.25/\sigma \rceil, \quad (2.20)$$

which does not require any pre-computation.

The intimate dependency on the current posterior elicits that RFPE is an adaptive protocol. By keeping the cardinality of the particles' set constant via a posterior resampling, we ensure an exponential increase in the local density of the discretisation. Indeed, as the learning progresses, the posterior will retain support on an increasingly smaller subinterval of $[0, 1]$, correspondent to the expected decrease in the width of $P(\phi) \propto \sigma$. The full RFPE algorithm is described in App. Algorithm 3, whereas we refer to [82] for numerical analyses on its stability against the implemented discretisation, as well as variance reduction strategies.

2.1.6.1 RFPE and simultaneous estimation of different eigenphases

The peculiar approach of iterative algorithms in general, and RFPE in particular, is to patch together experiments performed on a set of identical quantum states, to learn a single eigenvalue. This can also be of use to provide a strategy that tackles the more general case provided in Eq. 2.1,

¹¹As an algorithm step processing the available dataset (here, E) is customarily called within the ML community. A pseudocode is also available as Algorithm 2 in Appendix A.

¹²In practice, the fitting procedure is replaced by a more efficient computation of the first two moments of the distribution.

¹³As the update procedure requires no costly functions, requiring simple operations, see also Appendix 2.

¹⁴The asymptotic optimality result of [156], based on the Cramer-Rao bound, does not apply to RFPE where choices depend probabilistically upon prior information. However, adopting PGH and a Gaussian distribution, the PGH approximately saturates the Bayesian Cramer-Rao bound [82].

where an approximate eigenstate is not known in advance. As stated in Sect. 2.1, if the overlaps $|\alpha_k|^2$ with n eigenstates other than the one of interest are not negligible, then a proper inference step must address such eigenstates simultaneously. This requires a likelihood function of the form:

$$P(E|\phi_1, \dots, \phi_n; M, \theta) = \sum_{k=1}^n |\alpha_k|^2 \frac{1 + (-1)^E \cos(M\phi_k - \theta)}{2}, \quad (2.21)$$

to store and update information about all the n eigenvalues simultaneously, but this is of course intractable for large n .

RFPE offers a natural approach to combat this by truncating the search space. Eq. 2.21 can be immediately rewritten, by splitting the single contribution of interest $P(E|\phi; M, \theta)$ isolated from spurious contributions condensed as $P(E|\{\phi_k\} \setminus \phi; M, \theta) \leq 1 - |\alpha_k|^2$. If the prior is chosen such that $|\phi - \phi_k| < \sigma \forall \phi_k \neq \phi$, on the ground of preliminary energetic information, the approximate approach of RFPE will “classically collapse” the prepared state onto $|\phi\rangle$, unlike an exact Bayesian inference procedure [141]. In fact, if an observed datum E corresponds to ϕ , the inference will proceed according to Eq. 2.18 as usual. However, if instead the observation refers to a $\phi_k \neq \phi$, then the likelihood function corresponding to this hypothesis will be approximately¹⁵:

$$P(0|\phi, \phi_k; M, \theta) \propto \cos^2 \frac{(\phi_k - \phi)M}{2}. \quad (2.22)$$

Because of our biased prior, $|\phi_k - \phi|M \gg 1$, and therefore the evidence for an estimate close to ϕ will fluctuate randomly and be negligible in average. The evidence from Eq. 2.22 would have a net contribution for particles such that $|2\pi x_j - \phi_k|M < 2\pi$, but the chosen prior having $P(x_j) \sim 0$ for such x_j , these contributions are negligible. By penalising hypotheses belonging to other eigenvalues, the algorithm remains stable even if initial state has support in several eigenvectors of U .

Therefore, at the cost of not using effectively information from all measurements, RFPE also provides a viable alternative to knowing an approximate preparation of $|\phi\rangle$, replacing it with approximate a priori knowledge of ϕ , which is easier to retrieve, especially using approximate quantum chemistry methods [158, 159]. Also, it circumvents the costly requirement of storing multiple hypotheses at once to perform the inference in such cases. This particular application can be seen as an instance of the intrinsic RFPE noise robustness, that will be discussed in Sect. 2.2.

2.1.7 Time-series PE

Time-series PE (TSPE) was designed specifically to perform noise-robust eigenphase inference in a multi-eigenvalue scenario [150], such as presented in Eq. 2.1. Therefore, we will expose it here in this most general context, relaxing the assumption of an (approximate) initial eigenstate preparation. This latter case can be immediately recovered from what follows. The starting point

¹⁵ Because we are assuming $M \sim 1/\sigma$ and $\theta \sim \phi$

for TSPE is observing how for a generic input state $|\psi\rangle$, chosen M , phase estimation experiments will provide information on $g(M) = \sum_k^K \alpha_k e^{iM\phi_k}$. In order to resolve each single eigenphase, $g(M)$ must be decomposed in $M_{\max} \geq K$ terms. The most natural choice is to define the vector:

$$\vec{g}(M) = \sum_k \alpha_k e^{iM\phi_k} \vec{b}_k = (g(M), g(M+1), \dots, g(M+K)). \quad (2.23)$$

according to the harmonics $\vec{b}_k = (1, e^{i\phi_k}, \dots, e^{iK\phi_k})$. The coefficients for all the possible M values can then be stored in a matrix with elements $B_{Mk} = e^{iM\phi_k}$. Crucially, within the assumption $M_{\max} \geq K$, it can be observed that a matrix $T = BDB^{-1}$ exists [150], with $D_{ij} = \delta_{ij} e^{i\phi_j}$ and:

$$T\vec{b}_k = e^{i\phi_k} \vec{b}_k \quad (2.24)$$

$$T\vec{g}(M) = \vec{g}(M+1). \quad (2.25)$$

It is immediate from Eq. 2.24 that the ϕ_k components can be recovered from a diagonalisation of T . The *time shifting* property in Eq. 2.25 can instead be used to reconstruct T via an optimisation problem. To this purpose, in [150] it is suggested to build the Henkel matrices $G_{ij}^a = g(i+j+a-M_{\max})$ with $0 \leq i \leq K-1$, $0 \leq j \leq 2M_{\max}-K$ and $a \in \{0, 1\}$. Since by construction $TG^0 = G^1$, then T can be found minimising $\|TG^0 - G^1\|$.

Similarly to what discussed in Sect. 2.1.4, in order to estimate $\vec{g}(M)$ from the measurements at the end of the circuit, in [150] a fixed policy¹⁶ for both M, θ is suggested, where at each j -th step, $M = j$ and $\theta \in \{0, \pi/2\}$. Additional noise robustness can be obtained by repeating each measurement s times for each given M . Indeed, given the multi-eigenphase likelihood as in Eq. 2.21, it is possible to express each element of \vec{g} as:

$$\begin{aligned} g(M) = & P(0|\vec{\phi}, \vec{\alpha}; M, \theta=0) - P(1|\vec{\phi}, \vec{\alpha}; M, \theta=0) \\ & - iP(0|\vec{\phi}, \vec{\alpha}; M, \theta=\pi/2) + iP(1|\vec{\phi}, \vec{\alpha}; M, \theta=\pi/2) \end{aligned} \quad (2.26)$$

In order to fully deconvolute $\vec{g}(M)$, it is left to estimate the coefficients $\vec{\alpha} = (\alpha_1, \dots, \alpha_K)$. However since the set $\{\phi_k\}$ is known, $\vec{\alpha}$ can be immediately recovered minimising $\|B\vec{\alpha} - \vec{g}(0)\|$.

Also TSPE uses the quantum circuit in Fig. 2.1, so w and d are equivalent to all other iterative protocols presented in this Chapter. From Eq. 2.26, it can be estimated that the uncertainty in the estimate of any given ϕ_k scales as $\sigma_k \propto 1/(Ms^{1/2})$ [150], and therefore TSPE is expected to exhibit learning rate performances similar to RFPE in the single eigenphase case. Finally, in terms of classical computational overhead, the overall TSPE procedure requires:

1. updating the estimation of \vec{g} after each experiment, which requires a total of $\mathcal{O}(sM_{\max})$ additions,
2. performing a final minimisation to retrieve T , which has a cost $\mathcal{O}(K^2 M_{\max})$, and

¹⁶ The authors in [150] do not discuss possible improvements where adaptivity in the choice of M, θ is introduced. Considering that beside a natural approach towards multi-eigenphase estimation, the main advantage of TSPE over adaptive methods is classical computational overheads, it is possible that they prefer not to compromise this point.

| | δ | ϵ | w | d | η |
|------|--------------------|------------|----------------------------------|------------------------------|--|
| ITPE | $\mathcal{O}(1/n)$ | <1/2 | $\mathcal{O}(\dim[\hat{U}])$ | < $\mathcal{O}(n)$ | $\mathcal{O}(n \log^{-1}(n)/T_n)$ |
| SPEA | 2^{-n} | 0 | $\mathcal{O}(n + \dim[\hat{U}])$ | $\mathcal{O}(T_{2^n})$ | $\mathcal{O}(2^n/T_{2^n})$ |
| KPEA | 2^{-n} | < 1/2 | $\mathcal{O}(\dim[\hat{U}])$ | < $\mathcal{O}(2^n)$ | $\mathcal{O}(2^n \log^{-1}(n)/T_{2^n})$ |
| IPEA | 2^{-n} | 0 | $\mathcal{O}(\dim[\hat{U}])$ | < $\mathcal{O}(2^n)$ | $\mathcal{O}(2^n/T_{2^n})$ |
| RFPE | σ | < 1/2 | $\mathcal{O}(\dim[\hat{U}])$ | < $\mathcal{O}(\sigma^{-1})$ | $\mathcal{O}(\sigma^{-1}/T_{\lceil \sigma^{-1} \rceil})$ |

TABLE 2.1. Synoptic table of the characteristics of the main algorithms for noiseless PE, namely the accuracy δ , error probability ϵ , width and depth of the quantum circuit (w and d), and efficiency η (see also Sect. 2.1.1). n is here the number of algorithm steps concluded by a measurement. As δ can be made arbitrarily small in all cases, we fix each case to an arbitrary value, to show the corresponding dependence of other metrics. We remind how for all probabilistic algorithms with $\epsilon \neq 0$, we can obtain $\epsilon \sim 0$ after an appropriate number of repetitions s .

3. the diagonalisation of the obtained T , which is $\mathcal{O}(K^3)$.

Considering that only the minor of such overheads depends (linearly) on the number of repeated experiments s , and the remaining classical costs are $\mathcal{O}(1)$ for the single-eigenvalue case, TSPE can be regarded as advantageous over RFPE whenever a considerable number of repeated experiments is required to provide the final estimate¹⁷. In [150], numerical estimates show that this advantage becomes evident as soon as $s \sim 1000$.

2.1.8 Multi-round PE strategies

Observing again Fig. 2.1a-b, it appears evident how in all algorithms discussed so far, the output state from the unitary is discarded, after the projective measurement(s) on the ancillary qubit(s) has occurred. Iterative procedures like IPEA, RFPE or TSPE often target the eigenphase of a single eigenstate $|\phi\rangle$. Assuming that the cost of (eigen)state preparation is negligible compared to its evolution under U^M , there is no evident reason to recycle such output state, in a fashion outlined by the circuit of Fig. 2.1c. Indeed, decoherence phenomena would lead to an approximate eigenstate in output $|\phi'\rangle \sim |\phi\rangle$, that could invalidate any *round* r for which $r > T_2/T_U(r)$, where the cumulative evolution time for the round $T_U(r) := \sum_{i=1}^r T_{U^{m_i}}$ and T_2 is the decoherence time for the system (see Sect. 1.1.2). Even in an idealised case with no decoherence, increasing the number of rounds per experiment can be shown to worsen performances for algorithms where an efficient phase update can occur after each experiment is completed [141, 150].

This conclusion can be subverted, instead, whenever the input state $|\psi\rangle$ is only approximately prepared as an eigenstate $|\phi\rangle$, or has intentionally non-negligible overlaps with a set of K

¹⁷ Here we are assuming that a proper Bayesian update is performed after each repeated measurement, instead of using majority voting to combine subsets of repeated measurements, a hybrid approach discussed hereafter.

different eigenstates, in a multi-eigenphase estimation scenario. In such cases, additional rounds are effective in progressively projecting $|\psi\rangle$ onto one state of the eigenbasis, and the PE can leverage on this to make subsequent rounds more effective. As in this Thesis we do not address such cases experimentally, we do not discuss this point any further, but the interested reader can find a more detailed analysis e.g. in [150].

2.2 The impact of noise upon PE algorithms: a preliminary overview

So far, we have considered PE strategies within a noiseless scenario. However, unavoidable imperfections in the actual device are expected to lead to errors in the gates' implementation and outcomes' readout. A noise resilient approach is expected to handle such non-ideal circumstances in the most effective way possible. All probabilistic algorithms based upon a bitwise expansion like Eq. 2.6 embed an intrinsic, if limited, noise resilience as they acquire information from repeated measurements. As each bit is inferred from a dedicated projective measurement, a noisy implementation (whatever the cause) requires to substantially alter the expected likelihood function in Eq. 2.4, before errors become detectable as a flip in the corresponding bit. However, these algorithms still suffer from a rigid update policy for M, θ as the estimation progresses, which takes in no account the observed circuit performances, and might set unrealistic target accuracy. On the contrary, deterministic algorithms like SPEA and IPEA are not designed to deal with noisy devices, and the outcome of a single run will immediately become probabilistic, subject to the level of noise in the experimental implementation. Their usability can still be recovered adopting the same probabilistic approach of Sect. 2.1.4. In this scenario, the effect of noise is mitigated by repeating each run a multiple s times to implement a majority voting scheme [154], adopted also later in Sect. 2.3. Finally, Bayesian algorithms like RFPE are expected to be natively noise-robust: their update policy being adaptive, they learn from previous outcomes the most informative experimental choices, and are expected to recover from errors propagating up to the measurement step. Moreover, they output a well-motivated error estimate [82, 141], and not only an estimate for the eigenphases.

We will start considering a simple error model, in full generality. Called P the ideal probability of measuring the most likely datum E with a single experiment, if an error occurs with probability $P_e < 1$, the effective success probability is then $P' = P(1 - P_e) + P_e(1 - P) \geq P(1 - P_e) + P_e/2$. If we attempt to mitigate the impact of errors by repeating the measurement s times and adopting a *majority voting* to infer the correct E , we can invoke the Chernoff bound¹⁸ so that:

$$\epsilon = 1 - P' \leq e^{\frac{-s}{2P'}} (P' - \frac{1}{2})^2. \quad (2.27)$$

¹⁸Indeed, the necessary assumptions i) $P > 1/2$, ii) $P_e < 1$ and iii) P' a monotonic function of P are all met.

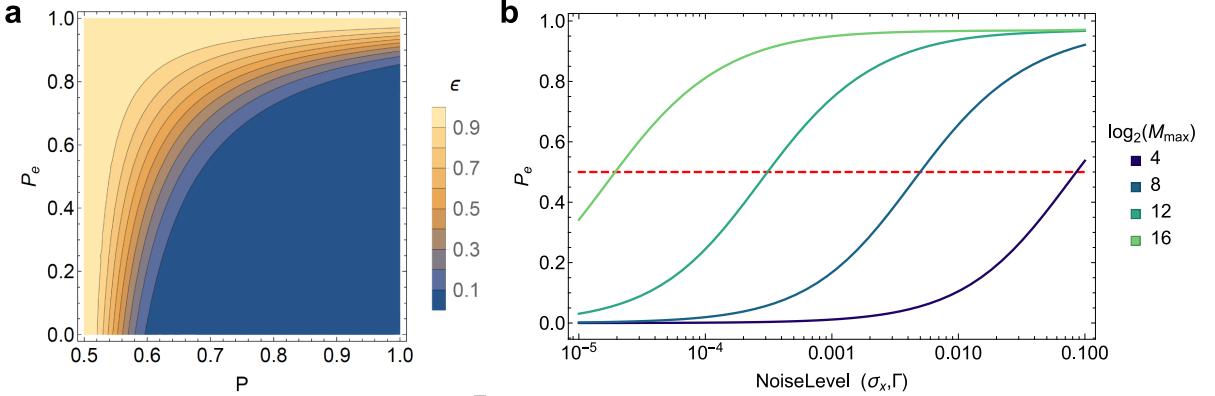


FIGURE 2.3. General modelling of noise effects in PE schemes. **a** Upper bound on error rate ϵ (colour coded), when inferring a single bit in a PE experiment, where each run has an ideal probability of success $P > 0.5$, and error rate P_e , adopting a majority voting scheme with $s = 500$ repetitions. **b** Analysis of the error probability P_e for any bit in a binary expansion up to $\log_2(M_{\max})$ bits, the generic circuit-based model of noise, Eq. 2.28, inclusive of both noisy Hadamard gates and decoherence effects. These are quantified on the x-axis by the σ_x and Γ parameters, respectively (see the main text). Here we assume only random fluctuations to affect the Hadamard gate implementation, i.e. $\theta_x \sim \mathcal{N}(0, \sigma_x)$. The curves represent averages for $\phi \in [0, 2\pi]$. A dashed red line emphasises the $P_e = 0.5$ threshold.

It is then immediate how Eq. 2.27 suggests an exponential suppression of errors with repetitions s , only as long as the success probability is bounded far from random, i.e. $P' \gg 1/2$. In a noisy case, as long as $P_e \ll 1$, the exponent is approximately $\frac{1}{2} \frac{(P-1/2)^2}{P} - O(P_e)$, and thus we must impose $P - 1/2 \gg 0$, as for small P_e . We summarise this behaviour in Fig. 2.3a, as a function of both P_e and P . We highlight the abrupt “breakdown” transition between a situation where majority voting substantially suppresses errors, and a regime where the estimate for the correct E from a set of measurements becomes rapidly unreliable.

Discussing specifically methods relying on binary expansions, we observe how in this last regime, not only the correctness of the phase estimate is affected, but most importantly also the use of statistical evidence to infer an error. As ϵ increases, according to the value of P for the various bits, some erroneous *words* of bits might become more likely than others, potentially suggesting unmotivated small error bars. Enhanced KPEA and RFPE are expected to exhibit additional robustness against these effects, as they use a specific experimental setup to infer several bits at once. If instead all the bits estimation were affected independently by the same P_e , then the probability for an overall successful n bits PE could be approximated by $\sim n \left(1 - e^{-s(P_0 - 1/2)^2/2P_0}\right)$. However, one must take into account how the longer evolutions corresponding to \hat{U}^M with $M \sim n$, and thus to the most precise bits in the expansion, are more subject to noisy implementations. Hereafter we draft a generic, semi-quantitative understanding of the

role played by this effect¹⁹.

Following [154], we model a noisy PE implementation by altering both the final Hadamard preparation of the projective measurement, and the (exponentiated) unitary gate (see Fig. 2.1). First, we replace in Eq. 1.5: $\hat{H} \rightarrow \hat{H} + \hat{R}_x(\theta_x)$, with θ_x an “error angle”, normally distributed with variance σ_x . For simplicity, we absorb in such error angle any additional imperfections occurring in the implementation of the feedback angle as $\theta \pm \sigma_z$. Also, we introduce a dephasing noise in the computational basis with rate Γ per qubit operation. This captures the unavoidably finite coherence times of many quantum computing implementations (see [82, 154]), inherently penalising circuits of increasing d . Within this error model, a single application of the circuit in Fig. 2.1b has an error probability:

$$\epsilon = \frac{1}{2} \left[1 - e^{-\sigma_x^2 - |\phi|M\Gamma} \cos(2\pi M' \theta_x) \right] \quad (2.28)$$

where M' depends on the particular PE algorithm, e.g. $M' = 2^{j-n}$ for the j -th bit in an n-bits IPEA, whereas $M' = M$ in RFPE. The exponential dependence of ϵ upon M suggests how the error rate in the bit estimation will worsen quickly as $M \sim 1/\Gamma$ (see Fig. 2.3b). This is exacerbated by the fact that all iterative PE procedures considered so far achieve an efficient estimation by scaling M exponentially in the number of epochs: in a realistic device, the exponential scaling in accuracy is therefore fundamentally limited by Γ , because as $\epsilon \rightarrow 1/2$, measurements in this regime convey no information, and $\mathbb{E}(\phi)$ can only be improved via additional statistical sampling from measurements adopting $M < 1/\Gamma$.

Combining observations from the approaches in Fig. 2.3a&b, intuition emerges about potential impracticalities of non-adaptive PE implementations. Simplistic error mitigation schemes, such as a majority voting approach, are sufficient to deliver sensible outcomes as long as a well-characterised device is operated well within its capabilities. For example, within the circuit-model discussed, one might target an accuracy in ϕ dictated by preliminary knowledge of noise parameters (σ_x, Γ) , such that $P_e \ll 1$ (see Fig. 2.3a) and errors can be easily mitigated. However, this is far from ideal: in order to achieve usable $\delta(\phi)$, it must be expected that quantum devices will operate as close as possible to the breakdown in Fig. 2.3a. In such conditions, occasional sources of additional, uncharacterised noise might quickly transition a PE algorithm implementing a fixed strategy into an unstable regime, with catastrophic degradation in the performance. In absentia of a confidence level at each epoch of the procedure, the erroneous inference of any single bit cannot recover, invalidating the rest of the measurements as well as the final output. On the contrary, adaptive algorithms such as RFPE do not make hard decisions upon the outcome of single experiments, and therefore they can recover from errors in further runs. Moreover, the Bayesian inference in Eq. 2.17 is expected to detect when the likelihood function fails to describe newly observed data, i.e. when M is big enough for flaws in the device to become apparent. At

¹⁹ We refer instead to Sect. 2.3.3, and following ones in this chapter, for a more implementation-specific studies about realistic sources of noise in the device.

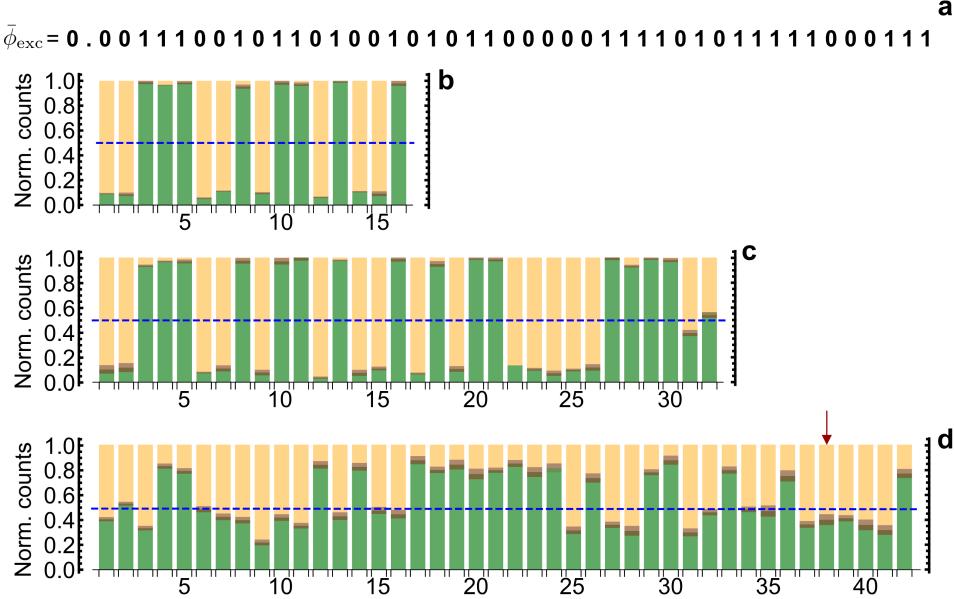


FIGURE 2.4. Experimental IPEA single runs performed on the quantum photonic chip, with majority voting on an average of ~ 500 counts per bit. **a** Binary expansion of the eigenphase for the test \hat{U} , up and including the 44th bit. **b** IPEA run up to 16 bits, with each represented as a barchart of normalised counts, with the height of the green (yellow) portion of the bar proportional to the number of times the j -th measurement is $E_j = 1$ ($E_j = 0$). The uncertainty due to Poissonian noise affecting each measurement, estimated by the number of counts obtained, is reported as a greyed area overlapping the bars. The dashed blue line emphasises the 0.5 intermediate value. **c** IPEA run as in b, up to 32 bits. **d** IPEA run as in b, up to 42 bits. The first bit erroneously estimated (overall 38th) is highlighted by the arrow.

this point, an adaptive update policy for M will autonomously avoid entering such critical regime. Instead, in eventual additional epochs it will accrue the eigenphase accuracy polynomially fast by sampling a recurrent choice of $M \sim 1/\Gamma$, along with a firm estimate in its uncertainty [82].

In conclusion, this preliminary analysis sets some expectations for experiments comparing algorithms employing a rigid policy and ad-hoc error mitigation schemes, versus adaptive approaches. Early experiments involving rigid PE strategies have evidenced their limitations, to the point of considering PE impractical in a few quantum computing platforms [160–162]. Expectations that with adaptive approaches this need not be the case were born out in simulation [82], but they had not yet been verified in practice, against a realistic device. We provide such verification below using an integrated quantum photonic device.

2.3 Testing PE strategies on a quantum photonic reconfigurable device

The device described in Sect. 1.4.4 is particularly suited for the task of implementing different PE strategies. Crucially, it satisfies the fundamental requirement of performing arbitrary two-qubit controlled unitary operations, and gives access to tomographic measurements upon the control photon. In particular, the projectors $\hat{P}(0,1)$ on the computational basis of the photonic control qubit are measured from the statistics of the coincidence counts, as extracted from the device output (see Fig. 1.5). In photonic experiments, measurements provide probability distributions rather than single-shot data [136]. A strategy to exploit this can be favourable in cases where experimental count rates²⁰ are much faster than the reconfiguration rate, even when theoretically sub-optimal [141]. Here we adopt the customary majority voting method (see Sect. 2.1.4), to obtain a binary datum value E from the projective measurement statistics.

The reconfigurability and high precision of the silicon photonic device are also key in protocols that are based upon continuous feedback from the experimental outcomes. In our experimental setup, classical post-processing of the data obtained, and control of experimental parameter can be achieved via the interface to a classical computer (Fig. 1.6). In the following, we will focus experimental tests upon IPEA and RFPE protocols. These are both compatible with the size of the corresponding available quantum circuit (unlike e.g. SPEA), and offer simultaneously an exponentially scaling η combined with $\epsilon \sim 0$ failure probabilities (unlike e.g. ITPE) in ideal conditions. Moreover, as discussed in Sect. 2.2, they are both intrinsically designed to provide a degree of robustness against noisy implementations.

As evident from Sect. 2.2, a crucial parameter in the expected noise resilience of the protocols is the targeted accuracy δ . In order to observe the experimental implementations work in realistic conditions, in the following we will refer to two different targets, derived from the applicative realm of quantum chemistry calculations²¹:

- *chemical* accuracy, equivalent to 1 kCal/mol $\sim 10^{-2}$ eV and
- *spectroscopical* accuracy, loosely defined as 0.01 kCal/mol $\sim 10^{-4}$ eV.

2.3.1 Implementation of the IPEA protocol

Firstly, we intend to demonstrate the actual capability of the photonic device illustrated before to perform IPEA, representing a more standard approach already tested in a variety of experiments [162, 165]. In order to provide a realistic testbed, we choose an excitonic molecular

²⁰In all cases presented in this Thesis, hundreds of hertz.

²¹In brief, both accuracy targets are somewhat arbitrary. The significance of chemical accuracy is related to a reasonable prediction of molecular properties, given that bond energies are known within that accuracy. Spectroscopic(al) accuracy is historically related to typical accuracies achievable for vibrational frequencies in spectroscopy, however standard definitions do not match this understanding. Details about such definitions can be found e.g. in [163, 164].

Hamiltonian \hat{H}_{exc} as an example, simplified into a double-well potential problem for an implementation amenable to the CU(2) chip. The eigenstates of \hat{H}_{exc} are mapped into the qubit basis using the Jordan-Wigner transformation, and an appropriate choice of the evolution time leads to the unitary:

$$\hat{U}_{\text{exc}} = \begin{pmatrix} 0.5125 + 0.2417i, 0.3515 - 0.7452i \\ 0.3515 - 0.7452i, 0.5125 + 0.2417i \end{pmatrix}. \quad (2.29)$$

The \hat{H}_{exc} eigenstate of choice to run PE is its ground state ($|-\rangle$ in the computational basis), acknowledging how this class of states can in principle be prepared more reliably on a quantum device [159]. Additional notes on the significance and construction of \hat{U}_{exc} are reported in Chap. 4 and Appendix C.2. In our photonic chip, therefore, the reconfigurable gates are used to implement $|-\rangle$ as the input state and \hat{U}_{exc}^M is compiled in the controlled unitary portion of the chip.

The binary expansion of the corresponding eigenphase $\bar{\phi}(|-\rangle)$ is reported in Fig. 2.4a with 44 bits of precision²². The aforementioned chemical and spectroscopical accuracy, given the choice of units in Eq. 2.29, are here respectively equivalent to ~ 9 and ~ 16 bits of accuracy in the eigenphase estimate. Each bit can be extracted implementing the protocol described in Sect. 2.1.5, with E obtained by projecting onto the computational basis $\{|0\rangle, |1\rangle\}$ (see Eqs. 2.4 and 2.14) and majority voting for the most frequent outcome E , when integrating the coincidence counts obtained in each case for 10 s. In Fig. 2.4b&c we demonstrate how bits in $\bar{\phi}$ can be successfully extracted in this way, when analysing normalised coincidence counts obtained for both $E = 0$ and $E = 1$. In particular, it is possible to observe how in-absentia of decoherence phenomena, a well-characterised photonic chip can provide most bits reliably even when Poissonian noise is taken into account. In fact, coincidence counts are expected to be distributed according to a Poissonian distribution $\mathcal{P}(n_{c,E})$, where $n_{c,E} \propto P(E)$ is the expected number of coincidence counts for each E . The variance of \mathcal{P} is reported as a shaded area on the histogram bars of Fig. 2.4b-d.

As long as such variance is small compared to the total counts, $P(0) \approx P(1)$ is prevented by the measurements, within error, and a majority voted implementation is expected to be accurate. This is the case for estimates up to 32 bits of precision for the cases shown in the same figure, as we emphasise plotting also as a dashed line the threshold value $P_t(E) = 0.5$. Experimental estimates for almost any bit in Fig. 2.4b-c are all clearly distinguishable from $P_t(E)$, with the only exception of the 32nd bit in subfigure c. However, this latest case can be understood by comparing with additional bits in the binary expansion. As the truncation to the 32nd bit is not exact, the residual unaccounted phase influences the outcome of the first measurement (as the correction angle $\theta = 0$ for the first step of the algorithm), which we remind corresponds to the last significant bit in the expansion. However, this imperfection is not sufficient to operate a bit-flip, and subsequent iterations quickly recover a clear distinction between 0's and 1's.

²²This truncated expansion is not exact, but sufficient for the scope of this test as it will appear clear in the following.

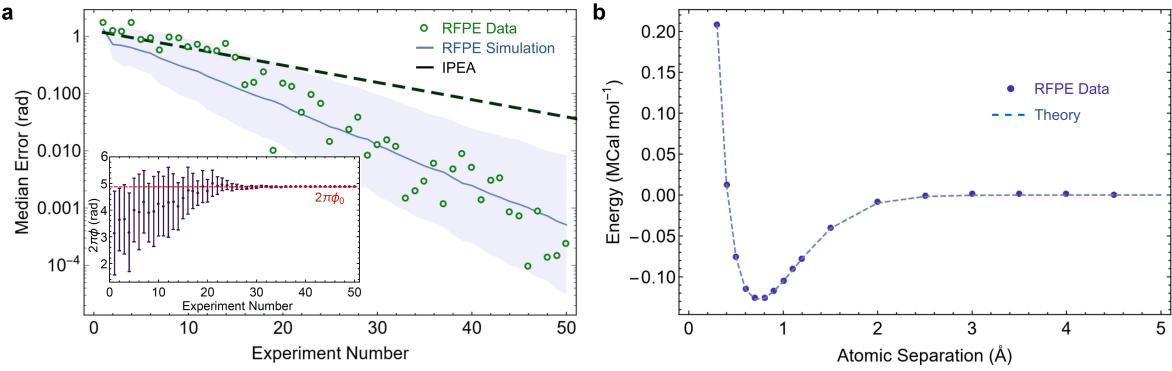


FIGURE 2.5. **a** RFPE converging to $2\pi\phi_0 = 4.8741$ rad, i.e. the energy of the dissociated H_2 molecule, starting from a prior $\mathcal{N}(\pi, \pi^2)$. Simulations are reported as a blue line, that is the average performance over 1000 independent runs of RFPE (shaded area: 67.5% credible interval). Standard IPEA behaviour is given by the dashed black line. Inset: convergence of the phase estimation to ϕ_0 (red line), where errors are evaluated using the s.d. of the prior distribution. Error bars are obtained from the standard deviation of the median. **b** Energies for the H_2 molecular bond at different inter-atomic distances, as computed theoretically (dashed line) and obtained experimentally using RFPE, after 50 steps (dots). Errors are smaller than the markers and have been omitted.

2.3.2 Implementation of the RFPE protocol

We also demonstrate preliminarily that RFPE can be successfully implemented on the CU(2) device. To do so, we use again as an example the ground state $|-\rangle$ of the excitonic unitary already used for IPEA in Eq. 2.29. We set the initial prior distribution as $\mathcal{N}(\pi, \pi^2)$, i.e. a Gaussian distribution with mean π and variance π^2 , which approximates a uniform prior. The prior was then discretised sampling 1000 particles, which is a reliable representation from previous findings [82]. Finally, the Bayesian update and changes to the controls of the quantum device required by RFPE can be calculated and fed to the quantum system using an interfaced classical computer (Fig. 1.5).

In Fig. 2.5a we report the results from a single RFPE run, having an exponentially fast convergence to the correct eigenphase $\phi = 4.8741/2\pi$ of the eigenstate $|-\rangle$. After 50 experiments, the estimated error is $2.4 \cdot 10^{-4}$ rad, in good agreement with the standard deviation of the posterior Gaussian distribution ($\approx 4.2 \cdot 10^{-4}$ rad at the same epoch – in the inset of Fig. 2.5a). Importantly, a rigorous error estimate embedded in the algorithm itself is a feature absent in other PE techniques, that fail to detect online the errors occurring in non-fault-tolerant devices [82]. An uncertainty about the estimates, using IPEA, is to rely upon repeated experiments, and then evaluate the mean and standard deviation of the different eigenphases evaluated. However, in Fig. 2.5a we can observe how RFPE outperforms IPEA, as soon as the number of repeated experiments is $\mathcal{O}(10)$ (we report the case with 10 repetitions in the plot).

A more systematic analysis is however required, to show that the results do not depend from a specific unitary configuration implemented in the chip. This is accomplished in a quantum simulation applicative scenario, i.e. estimating molecular energies as a control parameter for the molecular configuration is scanned. Here we pick as an example the well-studied H₂ molecule Hamiltonian \hat{H}_{hyd} , and the corresponding evolution unitary $\hat{U}(t) = e^{-i\hat{H}_{\text{hyd}}t}$ for a range of atomic separations between the hydrogen nuclei. Classical theoretical simulations show how this range captures both the bonded as well as dissociated molecule configurations (Fig. 2.5b, dashed line), so that the binding energy can be extracted from the scan [148, 162, 166]. In the experiment, we scanned a set of 18 different exemplary atomic separations, increasing the density of the sampling close to the bond energy. For each value, we performed a single run of RFPE along 50 iterations to evaluate the eigenphase (thus equivalently to what shown in Fig. 2.5). Each datapoint is obtained after about 20 mins. As shown in Fig. 2.5b, the estimated energies match the theoretical values with high precision. The average error for the dataset is 0.72 kCal/mol, providing a precision higher than chemical accuracy (see pg. 65).

2.3.3 Testing protocols' resilience against external noise

The main advantage of the Bayesian approach over traditional methods comes from its expected reliability on non-fault-tolerant devices. Here we investigate experimentally the robustness of the protocol against three main controllable sources of imperfections: *gate errors*, *decoherence* and *losses*. A noisy control over the quantum operations, as well as imperfections in the device fabrication or calibration lead to the well known problem of the infidelity in unitary operations implemented in quantum hardware. In our specific case, integrated quantum photonic devices tend to suffer from limitations in the accuracy of the (feedback) control, as implemented by the TO phase shifters, as well as residual thermal crosstalk (Sect. 1.4.4.2). The precise control achieved by the electronic phase-shifters driver, along with careful preliminary calibration (see Sect. 1.4.4.2), allows us to tune the noise affecting the phase gates. To do so, we replace the phases $\bar{\varphi}$ that implement the correct \hat{U} , with synthetic values φ sampled from a Gaussian distribution $\varphi \sim \mathcal{N}(\bar{\varphi}, \sigma_{\text{phase}})$. The parameter σ_{phase} therefore quantifies a synthetic Gaussian noise in the phases, which can be interpreted as infidelities in the implemented state preparation and unitary evolution. In Fig. 1.7b, we simulate in advance the amount of infidelity expected in the device, according to the value of σ_{phase} and assuming an otherwise ideal device.

Fig. 2.6a shows the outcome of both RFPE and IPEA protocols, for $0 < \sigma_{\text{phase}} \leq 0.55$ rad, corresponding to worst-case average state fidelity 94% and gate fidelity 91% expected in the chip. For each value, a single RFPE run is performed, using 100 epochs (and an equal number of experiments), with the error of the final estimate extracted from the posterior inferred. On the contrary, 16-bit IPEA is averaged over 10 independent runs to estimate the error bars, i.e. 160 total experiments per data point. Because of the different number of experiments in the two cases, the comparison emerging from Fig. 2.6a is intended to show the different robustness against

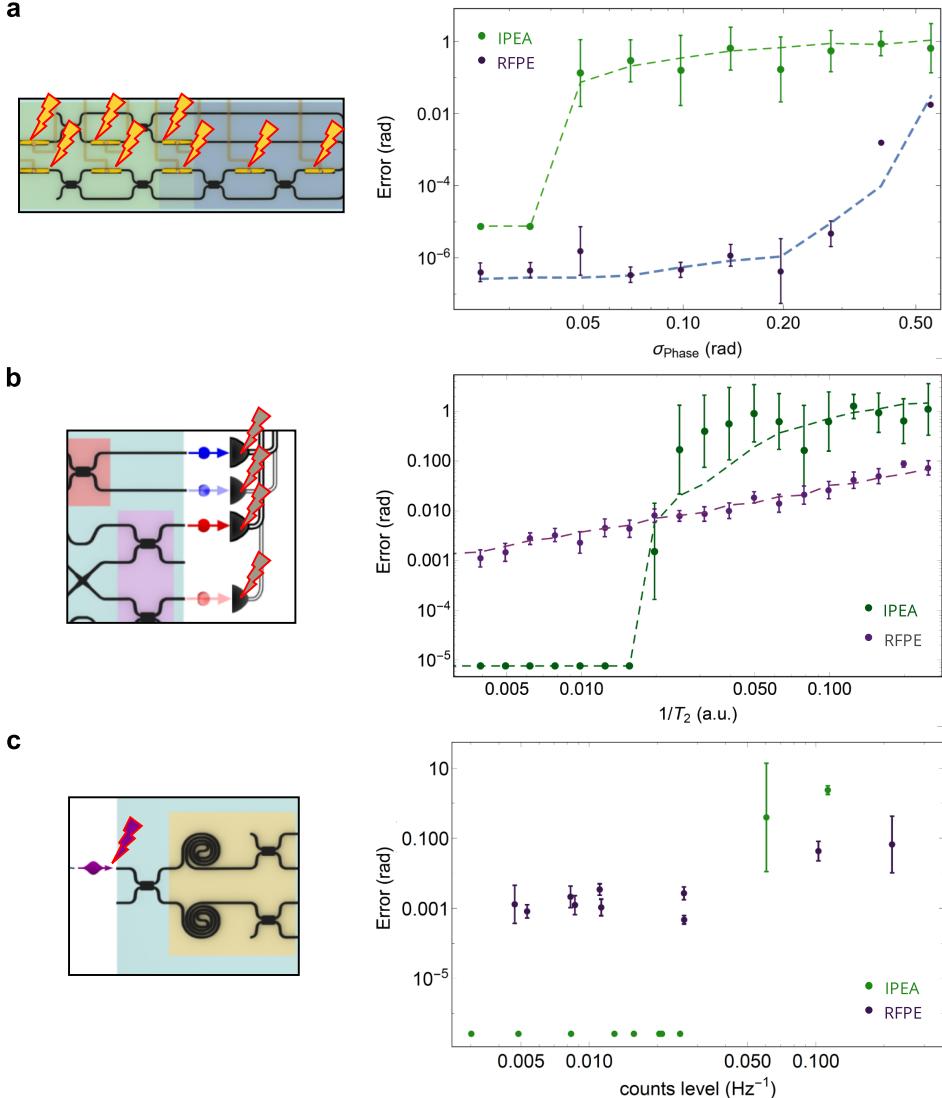


FIGURE 2.6. Effects of different experimental noises (pictorially represented on the left) on PE strategies. For IPEA data, experiments were repeated 10 times with 16-bit accuracy to evaluate median error and error bars, while the RFPE data were collected from a single run, after 100 measurements. Error bars for the estimated phase represent in both plots either a 67.5% credible region for RFPE, either a 67.5% confidence interval for IPEA. In the cases where error bars are smaller than the markers they have been omitted for clarity. Points are experimental data and dashed lines are simulations averaged over 1000 runs. **a** Infidelity of quantum operation. Each of the correct phases $\bar{\varphi}_i$ for the phase gates is synthetically replaced with a Gaussian distributed $\varphi_i \sim \mathcal{N}(\bar{\varphi}_i, \sigma_{\text{Phase}})$, where σ_{phase} models a noisy phase implementation. **b** Decoherence, simulated altering the readout according to the chosen T_2 value, as explained in the main text. **c** Losses, progressively increased by drifting via piezoelectric actuators the fibre, coupling the pump to the chip (see Fig. 1.6), starting from the optimal coupling condition (at the left of the plot). Losses are captured by the inverse count frequency (averaged for each run), reported on the horizontal axis.

noise of the two protocols, instead of comparing their η 's. We observe how when $\sigma_{\text{phase}} \geq 0.05$ rad, the accuracy for IPEA dramatically decreases. More importantly, error estimates from average runs become quickly unreliable, as it had been predicted in Sect. 2.2. RFPE, instead, is capable of providing still a high level of precision in this regime ($\delta \sim 10^{-6}$ rad), where IPEA is broken, displaying a much higher resilience to noise. In order to see the performance of RFPE substantially affected, we had to run the protocol with $\sigma_{\text{phase}} > 0.3$ rad, corresponding to operation infidelities much higher than those achieved through the characterisation outlined in Sect. 1.4.4.2. These are not only unlikely to occur in a photonic implementation, but likely to be satisfied also in many other competing quantum computing platforms [167, 168].

Decoherence is a stringent limitation in many quantum computing experiments, and therefore a crucial effect to explore in PE strategies, even if of negligible importance in quantum photonics. Therefore, we need to simulate it artificially in our setup, by flattening out the coincidence counts from the SNSPDs for the $\hat{\Pi}(0)$ and $\hat{\Pi}(1)$ projectors, via classical post-processing of the data. A Poissonian uncertainty is then included in the modified measurements. In agreement with the model for a depolarising channel (Sect. 1.1.2), we mimicked a decoherence time T_2 such that:

$$P'(D|\phi) = e^{-M/T_2} P(D|\phi) + \frac{1 - e^{-M/T_2}}{2},$$

where $P(D|\phi)$ is the data obtained from the photon coincidence counts for the outcome $D \in \{0, 1\}$. This alteration is introduced in the output data along with the iterative process, so to affect the inference for the next experiments. Post-processing data with Eq. 2.3.3, we can simulate the behaviour of RFPE and IPEA in systems that are prone to this model of decoherence.

The action of this noise model up to $T_2 = 4$ is reported in Fig. 2.6b. IPEA eigenphase estimates undergo a substantial and sharp deterioration at $T_2 \sim 32$, whereas the median error of RFPE (after 100 steps) decreases only polynomially with $1/T_2$. By maintaining an error $\mathcal{O}(10^{-2})$ even in the regime where conventional IPEA is unable to perform successfully, RFPE is confirmed as a robust choice. In the presence of characterised depolarising noise an optimised value for M is given by $\min([1.25/\sigma], T_2)$ [82]. The latter statement implies that significantly reduced T_2 will inevitably impact on the performance of RFPE significantly. Indeed, RFPE ceases to learn exponentially quickly when $1/\sigma \sim T_2$, yet continues learning at a polynomial rate, unlike IPEA, as expected from the theoretical analysis in [82].

Finally, losses in the circuit were amplified by progressively drifting the laterally coupled tapered fiber away from the optimal position, using the piezo-controllers, whilst recording the average count level at the detectors. The outcome from both algorithms in this scenario is illustrated in Fig. 2.6c.

For all cases, the mismatch in the experimental data is due to non-reproducible fluctuations in the coupling and detection efficiency, and additional losses in the setup. The simulations take into account the characterised residual phase noise in the device, used to extract the error bars reported in Fig. 2.6a–c.

2.3. TESTING PE STRATEGIES ON A QUANTUM PHOTONIC RECONFIGURABLE DEVICE

Results in this Chapter show the application of precise, controllable integrated photonics quantum devices to test and compare different quantum algorithms. We experimentally verified a Bayesian phase estimation algorithm on a fully programmable silicon quantum photonic device and demonstrated its superior performance in presence of noise. Although making use of a small-scale unitary (and hence photonic device), more complex future implementations can be efficiently performed on any scalable quantum architecture.

2.4 *Author's contributions to the Chapter*

Sect. 2.1 is my review of established results available in the literature, enriched by personal observations and on-purpose analyses.

Sect. 2.2 adapts known results to the specific case of noise impacting on some PE algorithms, as such it builds upon well-known results, to include minor ad-hoc contributions from me. Sect. 2.3 mostly reports **experimental results and analyses published as [141]**, by myself and collaborators at the University of Bristol. I have actively participated to the writing of the paper, as well as its review process. The seminal idea to test IPEA and RFPE upon the same device is due to Dr. Santagati & Dr. Wang. Contributions to design, fabrication and preliminary characterisation of the CU(2) device have already been stated at pg. 46. I have prepared the simulation code for the IPEA implementation, and written additional code to interface it with the experiment controls (this is in particular Sect. 2.3.1). Mr. Paesani has prepared the simulation code for the RFPE implementation, that was interfaced with the experiment using the same library prepared for IPEA (this is in particular Sect. 2.3.2). I have been responsible for the characterisation and simulation of noises in the CU(2) device, proposing in particular the experiment in Fig. 2.6a. Finally, I have participated to the data collection and analysis of experiments displayed in Figs. 2.4, 2.4b & 2.6.

3

CHARACTERISATION, VERIFICATION AND VALIDATION OF QUANTUM DEVICES

Recently, the evolution of data-rich methodologies, driven by increasingly powerful experimental setups, such as the large hadron collider, has met and exceeded expectations [169, 170]. Machine learning methodologies have played an essential role in delivering some of the most recent scientific discoveries, e.g. the detection of the Higgs' boson [171] and of gravitational waves [172]. Extracting usable knowledge from such huge amounts of data is thus a major scientific challenge. The automation of scientific discovery has become not just an interesting theoretical and philosophical problem, but a necessity, given that many experiments are capable of generating data much faster than they can analyse [173], e.g. the LHC facilities generating ~ 150 PB/year data rates [174]. Machine learning protocols, applied to the analysis of large experimental data sets, have shown promising results, ranging from classification and (compressed) mapping to optimisation and decision making [175].

However, solving inverse problems by distilling models from data, without incurring in excessive over-fitting, withstands as a grand challenge. This challenge is motivated by far reaching implications in the full automation of the learning process [175–177], yet frustrated by formal ill-posedness of many inverse problems of interest [178]. Within the specific realm of physical systems' modelling, research efforts evolved from optimal parametrisations of a priori structured models (e.g. for the equations of motion [179]), to graph-based searches for free-form models from the outcomes of classical physics experiments, by focusing on nontrivial conservation laws [177, 180]. However, the physical variables algorithmically combined are still based upon preliminary intuition of the phenomenon or leveraging the parallelism between representation in artificial neural networks (ANN) and physical modelling [181].

Testing reliably and efficiently quantum systems is essential for the development of quantum

technologies [182, 183]. Hereafter, we will refer to a set of tasks that specify the steps of the global testing cycle. In our terminology, we will refer to *characterisation* as a broad-scope investigation of the behaviour and response of a device, which often involves adjusting control settings within reasonable ranges. Once the capabilities of a device have been characterised, it is possible to perform a *verification* of whether its components meet the specified requirements. Finally, at a more abstract level, the device *validation* involves checking whether it is capable to perform correctly the tasks it was designed for. We have thus left intact the usual understanding of the last two tasks, as introduced originally in the computer software community, and adopted instead a hardware-specific point of view for the first [184]. We synthetically name the three steps as above “CVV”.

From a modelling perspective, in the characterisation phase we expect a comprehensive model for the device not to be used (if available at all). As Physics itself is ultimately a modelling effort, we expect that the systematic characterisation of novel quantum devices will deepen our understanding of the physics governing them. At the verification level and beyond, instead, a model for the device is in place and the testing consists of checking if, and to what extent, this model describes the behaviour of the device. As near term quantum circuits might soon provide data-throughput in the range of PB per run [185, 186], the challenge of automated characterisation is likely to become crucial as quantum circuits depth increases. ANNs have been recently applied to learn efficient representations for quantum states of interest [187, 188], to perform tomographic quantum state [181] and unitary evolution reconstruction [189], and to automatically tune quantum devices [190]. Also, supervised learning techniques have been applied to the automated design of quantum gates [191] and experimental state preparation in quantum optics [192].

In this chapter, we will discuss two families of methods addressing the open research question of CVV. First, we will present in Sect. 3.2 *Quantum Hamiltonian Learning* (QHL), a scalable solution to the problem of retrieving optimal parameters for a system whose true Hamiltonian is known to match a certain parametrised model \hat{H}^0 , provided *a priori*. QHL can also be used to qualitatively estimate how accurately \hat{H}^0 reproduces experimental measurements on the system. QHL therefore satisfies well the requirements to verify and/or validate a quantum device, and we will demonstrate the efficient usage of QHL to this extent in Sect. 3.3.1 for a nanoscopic magnetometer. However, in order to provide a thorough characterisation toolset, QHL does not suffice. To this extent, we introduce in Sect. 3.4 the Quantum Model Learning Agent (QMLA), an approach to automatically construct an approximate model to characterise the behaviour of a quantum system (Sect. 3.4). We test QMLAs in numerical simulations and we apply it to the experimental study of a two-qubit model for the open-system dynamics of an electron spin in a nitrogen-vacancy (NV) centre coupled to its spin bath (Sect. 3.5.1 and following ones), adopting both classical and quantum simulators.

3.1 Process tomography

A complete characterisation of a (sufficiently) isolated quantum system requires to address both its current state, as well as the dynamics induced by the system itself. The capability to acquire such information is crucial to certify the developments of novel quantum devices [193, 194].

Quantum state tomography (QST) has provided a golden standard to the task of identify and validate an unknown quantum state ρ^* , potentially starting from no a priori information. The key idea is to measure the expectation values for a complete set of observables, assuming that we have access to enough *exact* copies of ρ^* [5, 195]. Once this set of measurements is accomplished, classical techniques are adopted to reconstruct the density matrix, ranging from linear inversion and maximum likelihood, to Bayesian inference and penalised risk minimisation [196].

Unfortunately, this formulation encounters a major scaling problem, which is ultimately due to the intrinsic limitations of reconstructing quantum states with methods that necessarily adopt classical devices. As the Hilbert space for an n -qubits system scales as $d = 2^n$, the corresponding density matrix $\rho^* \in \mathbb{C}^{d \times d}$, so that the number of entries to estimate will intuitively be $\mathcal{O}(d^2)$.

This limitation can be mitigated whenever it is possible to make assumptions about ρ^* , introducing non-trivial a priori information. A few successful approaches focus on states whose density matrices can be described by a reduced number of entries [194], leveraging upon the observation that this class of states can cover many cases of physical interest, as discussed at length in Chap. 4. *Compressed quantum tomography* (CQT) provides a quadratic advantage over QST when ρ^* is a matrix of low-rank r [197, 198]. For example, this is the case when ρ^* is well approximated by a pure state, which has $r = 1$. In these cases, the number of measurements required by CQT is $\mathcal{O}(rd \log d)$, which is almost-optimal¹. Additionally, CQT is experimentally amenable, as the measurements required are two-outcome Pauli observables, easy to implement in several quantum computing architectures [198, 199]. Further improvements are possible when the states are well approximated by matrix product states of dimension $d_P < d$, which is generally the case for states prepared in systems whereby nearest-neighbour interactions are dominant [200]. Some architectures are particularly well-suited for this approach, as e.g. ion chains have $d_P \propto n$, so that the scaling in measurements is linear [201]. Polynomial scaling in the number of measurements as well as in the computational post-processing requirements can also be achieved when dealing with *permutationally invariant* states², or when approximating ρ^* as such [194, 202]. These methods have been successfully demonstrated experimentally, e.g. [203].

Also ML techniques have been invoked to reduce the complexity of tomographic tasks, following on the idea that a tomographic reconstruction task for ρ^* can be posed in terms of quantum state *learning*, i.e. predicting the outcome of any available measurement, given a certain experi-

¹ This is because an arbitrary rank r matrix has $\mathcal{O}(rd)$ independent entries, and therefore in principle $\mathcal{O}(rd)$ observables should suffice for a complete estimation.

² The permutationally invariant part of a n -qubit state is defined as $\rho_{PI} \equiv \sum_k (\Pi_k \rho^* \Pi_k)/n!$, so that a generic state ρ^* is permutationally invariant if $\rho^* = \rho_{PI}$.

mental setup [188]. Classical machine learning has thus been invoked in QST either to simply implement the state reconstruction phase [204], or to replace a full-reconstruction approach with an approximate learning via neural networks [181, 205], as demonstrated also experimentally [206, 207]. Finally, also the link between proper QML and tomography has been recently investigated via Quantum Boltzmann machines [208].

Even a full-reconstruction of the output state for a quantum device evidently does not exhaust its characterisation, as this tells us close to nothing about the dynamics actually implemented by the device [194]. We assumed to prepare only copies of the same output ρ^* , which is true only for a trivial dynamics. *Quantum process tomography* (QPT) attempts to address this additional task. More formally, we now aim to reconstruct the completely-positive linear map $\mathcal{M}(\rho_{\text{in}})$, describing a generic³ system's dynamics [28]:

$$\mathcal{M} = \sum_i \hat{A}_i \rho_{\text{in}} \hat{A}_i^\dagger \quad (3.1)$$

with $\{\hat{A}_i\}$ a set of operators satisfying $\sum_i \hat{A}_i^\dagger \hat{A}_i \leq \hat{1}$. A well-established classification distinguishes between *direct* and *indirect* QPT, whereby in the first case measurements are used to provide direct information about the underlying process, whereas indirect methods accomplish QPT via QST performed upon a set of probe states evolved according to \mathcal{M} [193].

Indirect methods, being based upon state tomography, suffer from the same scalability issues. For example, *standard* QPT requires to prepare d^2 linearly independent probe states, and reconstruct the corresponding outputs with QST. Therefore, the total number of measurements scales as $\mathcal{O}(d^4)$ [5, 209]. Ancilla-assisted process tomography requires instead to prepare entangled states of an expanded system–ancilla ensemble with $d' \geq d^2$ [193]. This overhead is justified by reducing the characterisation of the quantum process to an approximate quantum state tomography (hence requiring the capability to prepare a single ρ_{in}), in such a larger Hilbert space, which amounts to perform not less⁴ than $\mathcal{O}(4^n)$ measurements [193]. On the contrary, methods like the *direct characterisation of quantum dynamics* bypass completely any QST, inferring single elements of the map \mathcal{M} from the outcomes probability distributions of specific ρ_{in} states [210]. The overall cost in terms of experiments of these direct methods is also $\mathcal{O}(4^n)$, but it can be favourable whenever, in the experimental setup, measurements are more demanding than the preparation of fiducial probe states.

In order to address the scalability issues, tomography-free methods have been suggested, such as the *randomised benchmarking*, which focus on gate-implementations of the system dynamics, and replace accurate state preparation and measurement with fidelity measures and randomly chosen gates. Finally, also in the realm of QPT, machine learning methods have played a role in the process reconstruction algorithmic, both from complete [209] as well as incomplete [211] sets of tomographic data.

³See also at pg. 8.

⁴ Further expansion of the Hilbert space can actually lead to corresponding reductions in the number of experiments, see [193].

In summary, process tomography strategies outlined in this Section address the problem of the complete characterisation of a quantum process, i.e. assuming no prior knowledge about the system itself, in terms of the state mapping \mathcal{M} , described in Eq. 3.1. However, the information tomographically acquired about \mathcal{M} describes the dynamics as non-accessible, if not for the output state, evolved at a single snapshot in time. This approach does not match the most common implementations of quantum devices, which adopt *controllable* interactions, their duration being particularly prone to accurate experimental control⁵. In this way, the more compelling question of characterising the *generator* of the dynamics for a quantum system is left open. That is, if a different evolution time is chosen (e.g. a change in the duration of a pulse occurs), then performing a new process tomography is required, as the information from previous QPT provides no bootstrap for the novel characterisation task. Additionally, QPT is limited by serious scalability issues [212], as the number of input probe states and/or measurements scales exponentially with the system's size in qubits, n [193].

3.2 Quantum Hamiltonian Learning via Likelihood Estimation

In this section we discuss a fundamentally different approach called Quantum Hamiltonian Learning [157, 213, 214], which relaxes some fundamental assumptions of QPT: i) it provides a hamiltonian description for the full dynamic evolution that the quantum system undergoes; ii) it assumes some kind of a priori information about the system. The first point elucidates how QHL learns the generator of the quantum system's dynamics, assuming a closed system completely characterised via its Hamiltonian [28]. As it will appear more evident in Sect. 3.4, this limitation becomes important when the closed-system approximation is not valid. We can now introduce a more formal:

Definition 3.1. Known a parametrised Hamiltonian model for a system, $\hat{H}(\vec{x})$, where $\vec{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ the array of system parameters to characterise, a *Quantum Hamiltonian Learning* method aims to provide the optimal set of parameters \vec{x}_0 to describe the system's dynamics $\forall \xi$ experimental control setting and $\forall \tau \in [0, \tau_{\max}]$ evolution times⁶.

Therefore, when applicable, QHL tools infer the parameters of a Hamiltonian, as an inverse problem given the outcomes of experiments with control setting ξ and evolution time τ . In the following of this Chapter, we will often assume for simplicity that the only controllable parameter available in experiments is the time τ . This assumption can be immediately generalised [212], and for example in Sect. 3.4 we include explicitly as control setting the probe state, so that $\xi : |\psi\rangle, \tau$. Crucially, the Hamiltonian model assumed as a priori information renders superfluous to perform experiments $\forall \tau \in \mathbb{R}^+$, as it is expected that the information extracted from experiments at a sampled set $\{\tau_i\}$ will provide sufficient knowledge to *predict* the system's behaviour at $\tau \notin \{\tau_i\}$, as

⁵E.g. see Fig. 1.10 for exemplary sequences with NV-centres.

⁶An overview of the dynamics for a quantum system has been provided in Sect. 1.1.1.

long as $\tau_{\max} \sim \max_i \{\tau_i\}$ [157]. The cutoff time τ_{\max} therefore introduces in the QHL protocol what can be practically achieved experimentally. In a machine learning wording, therefore, the inverse problem is solved optimising \vec{x} over the *training set* of times $\{\tau_i\}$, and using \vec{x}_0 to meaningfully predict over a potentially much larger *test set* of evolution times.

The additional flexibility in terms of evolution time that QHL allows over QPT methods comes at the expense of imposing the Hamiltonian parametrisation $\hat{H}(\vec{x})$ as known in advance. However, this can be a reasonable assumption for many experimental systems, where theory or preliminary characterisation have provided an initial model, including all and only those terms that are expected from the physics of the system [212]. Also, we will normally assume that the parametrisation chosen is an efficient representation of \hat{H}^* , or more formally, we define:

Definition 3.2. A parametrisation $\hat{H}(\vec{x}) \equiv \sum_i x_i \hat{h}^i$ in the single operators \hat{h}^i as *non-degenerate*, iff any two parameters x_i, x_j are uncorrelated;

and assume $\hat{H}(\vec{x})$ to be non-degenerate unless stated otherwise. Degeneracies can be immediately identified via linear algebra in treatable cases⁷. Avoiding degeneracies in QHL can significantly simplify the task of learning an appropriate \vec{x}_0 , and will be discussed later on.

A first approach to QHL is known as Classical Likelihood Estimation (CLE). First introduced in [213], CLE combines an adaptive Bayesian inference procedure with a machine learning method and efficient heuristics, in order to appropriately design experiments acting on the quantum system to characterise. CLE stores all information about the parameters in a *prior* probability distribution $P(\vec{x})$. The latter captures how confident we are in a certain parametrisation \vec{x} representing the optimal \vec{x}_0 . Formally, the information updated progressively by CLE, through a series of experimental outcomes $\mathbb{D} := \{D_i\}$, can then be represented as a *posterior* distribution $P(\vec{x}|\mathbb{D})$, having a variance $\sigma(P(\vec{x}))$. Each epoch of the CLE protocol is composed of four main steps, which are iterated until $\sigma(P(\vec{x}))$ either converges, or shrinks below a target threshold a priori defined. A flowchart for CLE⁸ is provided in Fig. 3.4.

1. The prior distribution $P(\vec{x})$ leads to *adaptively* choosing the experimental setting (here τ) adopted for the next iteration.
2. The (closed) quantum system is prepared and evolved according to \hat{H} :
 - a) An appropriate initial *probe state* $|\psi\rangle$ is prepared as the input state, chosen such that its evolution under \hat{H} is informative.
 - b) The system undergoes the unitary evolution $\exp(-i\hat{H}(\vec{x}^*)\tau)$, dictated by the Hamiltonian \hat{H} for the chosen time τ and true parametrisation \vec{x}^* .

⁷See Appendix C.4 for some examples.

⁸The flowchart therein describes CLE applied to magnetometry purposes, as discussed later in Sect. 3.3.1.

3. A projective measurement upon the basis set $\{|D\rangle\}$ is performed on the evolved system, to obtain an outcome D with probability given by the *likelihood* function:

$$\mathcal{L}(D|\vec{x}^*; \tau) = |\langle D | \exp(-i\hat{H}(\vec{x}^*)\tau) |\psi \rangle|^2 \quad (3.2)$$

4. The probability distribution $P(\vec{x})$ of the Hamiltonian parameters is updated adopting a *simulator*:

- a) The simulator performs experiments with evolution $\exp(-i\hat{H}(\vec{x})\tau)$ as in steps 1–3, for the generic parametrised Hamiltonian $\hat{H}(\vec{x})$. In this way the corresponding likelihood $\mathcal{L}(D|\vec{x}, \tau)$ can be estimated. $\mathcal{L}(D|\vec{x}, \tau)$ represents the probability of obtaining outcome D , when \vec{x} is chosen as parameter
- b) In order to obtain the posterior distribution, we invoke Bayes' rule:

$$P'(\vec{x}|D) = \frac{\mathcal{L}(D|\vec{x}; \tau)P(\vec{x})}{P(D)}, \quad (3.3)$$

where $P(\vec{x})$ is given by the prior at the corresponding epoch, whilst $P(D)$ is a normalisation factor.

The adaptivity invoked for step 1 is not strictly necessary, as in principle it is possible to extract information also *offline*, i.e. from a set of data \mathbb{D} already obtained. However, this affects the efficiency of the protocol [213], as in this case it is not granted that experiments are performed at the most informative times⁹. Therefore, as already seen in Sect. 2.1.6 for Bayesian methods adopted for QPE purposes, also QHL natively adapts the experiment design at each epoch.

We remark how the probe state $|\psi\rangle$ (step 2) is not adaptive in CLE. By *informative*, we mean that $|\psi\rangle$ must undergo a meaningful evolution under \hat{H} , e.g. it could be a state orthogonal to the Hilbert subspace spanned by the Hamiltonian eigenstates. As the evolution in step 2.b occurs in the quantum system, it is of no concern regarding scalability.

The projective measurements in step 3 usually involve the computational basis for the system, e.g. $\{|0\rangle, |1\rangle\}$ for a single qubit. Therefore, in the following we will often contract $\mathcal{L}_0(\vec{x}, \tau) := \mathcal{L}(0|\vec{x}, \tau)$. An efficient comparison of the likelihoods in Eq. 3.2 with those obtained for trial parametrisations in step 4.a is crucial for a successful learning of \vec{x}_0 . However, as detailed in [214], this evaluation might be numerically unstable for large systems and addressing this issue requires substantial modifications of the protocol.

Step 4 is at the core of CLE, and elicits how this protocol formulates the characterisation problem in terms of a quantum simulation task: the retrieval of the optimal parametrisation $\vec{x}_0 \sim \vec{x}^*$ is obtained by recursive simulation of trial values. In CLE, a classical simulator implements the evolution $\exp(-i\hat{H}(\vec{x})\tau)$, but quantum devices might represent the only scalable choice in general. Indeed, for a general \hat{H} with no evident tensor structure, calculating this evolution is

⁹See also the discussion in Sect. 3.2.2.

expected not to be efficient on a classical computer [157, 160]. A few examples will be given in the following Chap. 4. Therefore, with the growing size of the system to characterise, unless an analytical model renders the evolution classically tractable [213], CLE does not provide an efficient solution¹⁰. In such cases, in order to make the overall inference process scalable, *Quantum Likelihood Estimation* (QLE) has been proposed [157, 214], whereby step 4.a can be run on a quantum simulator, addressing the scalability due to quantum simulation purposes. QLE has been first demonstrated experimentally by our group in [215] for the characterisation of Rabi oscillations in a diamond NV-centre, and used in this chapter in Sect. 3.5.3.

Storing the updated knowledge about the parametrisation \vec{x} as the posterior distribution $P'(\vec{x}|D)$ leads to a convenient feature of Bayesian methods: embedding a well-motivated uncertainty about the estimate \vec{x}_0 . Once a certain credibility level κ has been established as satisfactory, one can immediately introduce:

Definition 3.3. The *credible region* with level κ for the distribution $P'(\vec{x})$ is the minimal volume region \mathcal{X} in the parameter space, centred around \vec{x}_0 , such that $\int_{\mathcal{X}} P'(\vec{x}) d\vec{x} \geq \kappa$.

Credible regions are the Bayesian equivalent of frequentist *confidence regions*, signifying the multi-dimensional interval within we believe \vec{x}^* to lie, with a confidence level of κ . Therefore they will be consistently adopted in the following as the best uncertainty metric that can be provided for the outcome of the likelihood estimation procedure.

Finally, if we assume the outcomes $\{D_i\}$ obtained at the end of various experiments to be independent¹¹ and each obtained with time τ_i , one can formally write the cumulative update of the posterior as:

$$P'(\vec{x}|\mathbb{D}) = \frac{\prod_i [\mathcal{L}(D_i|\vec{x}; \tau_i)] P(\vec{x})}{\int [\prod_i (\mathcal{L}(D_i|\vec{x}'; \tau_i))] P(\vec{x}') d\vec{x}'}, \quad (3.4)$$

which summarises a complete inference process in CLE/QLE. All implementations so far invoke a classical co-processor to perform the calculation in step 4.b, so that QHL can be also considered a hybrid quantum protocol. The circuit for both QHL protocols is outlined in Fig. 3.1.

3.2.1 The prior representation problem and the SMC approximation

The Bayesian inference invoked to update the belief about a certain Hamiltonian parametrisation \vec{x} immediately poses two major computational problems: how to efficiently and reliably store, and update, the prior distribution $P(\vec{x})$. In general, the parameters are continuous, as recalled in Def. 3.1: a direct, exact experimental estimation of the normalisation factor in Eq. 3.4 is thus out of reach, as this would imply infinite experiments per chosen time τ , to compute the likelihood

¹⁰ Even if unsuitable for large scale applications, we will apply effectively CLE for the results presented in Sect. 3.3.1 and Sect. 3.4, that refer to systems composed by only a few qubits.

¹¹ We hereby intend the usual definition of statistical independence. In particular, to obtain Eq. 3.4 we need to invoke $\mathcal{L}(\{D_i\}|\vec{x}; \{\tau_i\}) = \prod_i \mathcal{L}(D_i|\vec{x}; \tau_i)$.

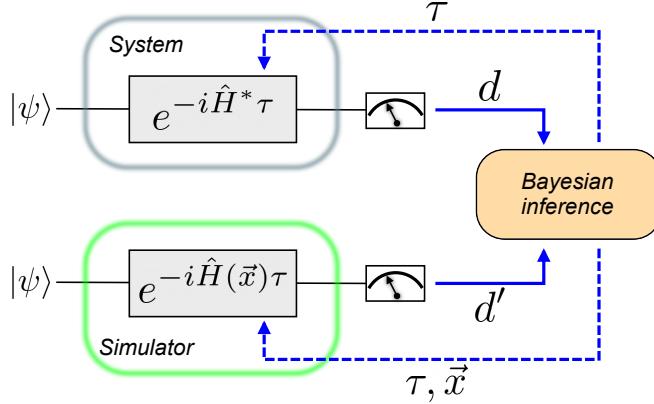


FIGURE 3.1. Logical circuits for CLE and QLE, where the *simulator* is intended to be respectively a classical and a quantum simulator. \hat{H}^* is the Hamiltonian of the system to characterise, and $\hat{H}(\vec{x})$ its parametrisation. Other symbols are explained in Sect. 3.2.

$\mathcal{L}_D(\vec{x}';\tau)$ throughout the integration range. The problem is equivalent to what already seen in Sect. 2.1.6.

In simple cases where the likelihood function is analytical, one might attempt an exact update of \mathcal{L} [57, 212]. However a closed form for the likelihood might often be unavailable, so that extracting information about the best estimate \vec{x}_0 , after \mathbb{D} data have been collected, would translate into the non-convex¹² optimisation problem:

$$\vec{x}_0 = \underset{\vec{x}}{\operatorname{argmax}} [P'(\vec{x}|\mathbb{D})] \quad (3.5)$$

and this can be computationally intensive to compute at each single step in an online experiment.

An efficient solution to this problem has been proposed since the introduction of CLE, and consists in the adoption of approximate *Sequential Monte Carlo* (SMC) methods [213]. Within this approximation, the parameter space is discretised as a set of M *particles* (see Fig. 3.2a), i.e. discrete parameter values $\{\vec{x}_m\}$ sampled from the prior distribution $P(\vec{x})$, that can thus be expressed as a convex sum of delta functions:

$$P(\vec{x}) \simeq \sum_m^M w_m \delta(\vec{x}_m - \vec{x}), \quad (3.6)$$

each with a weight proportional to the estimated value of $P(\vec{x}_m)$. This approach leads immediately

¹² In general, no proof of convexity for the problem in Eq. 3.5 can be given. Several counterexamples are known instead: already the single-parameter estimation problem illustrated in Sect. 3.3.1 for magnetometry is highly non-convex due to phase periodicities [135], and the situation exacerbates when multiple parameters are involved [216]. However, this does not preclude that some special cases might well be convex.

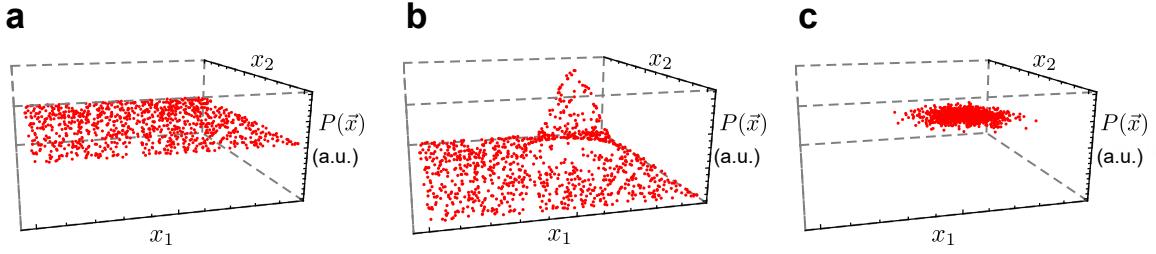


FIGURE 3.2. The SMC approximation and LWR resampling, across different steps of the inference process for a bivariate prior $P(\vec{x}) \equiv P(x_1, x_2)$. Horizontal axes report an arbitrary range of values for each parameter x_i , whereas the prior is reported on the vertical axis. **a** The initial prior as a discretised uniform multidimensional distribution (SMC). **b** The posterior after a certain number of update steps, with support on the same particles used in **a**. **c** The updated discretised representation of the posterior in **b** just after a (LWR) resampling event.

to discretised update and normalisation schedules¹³:

$$P'(\vec{x}|D) \approx \sum_m^M w'_m \delta(\vec{x}_m - \vec{x}) = \sum_m^M \left[\frac{\mathcal{L}(D|\vec{x}_m; \tau) w_m}{P(D)} \right] \delta(\vec{x}_m - \vec{x}) \quad (3.7)$$

$$P(D) = \sum_m^M \mathcal{L}(D|\vec{x}_m; \tau) w_m, \quad (3.8)$$

which can be readily computed from the set of outcomes obtained from the corresponding M experiments. In terms of cost of the outlined inference procedure, this can be split in the cost of each likelihood estimation, times the number of particles involved, times the number of epochs required for the algorithm to converge to the final estimate \vec{x}_0 . In order to estimate the likelihoods, and therefore the correct weights w_m^* , within accuracy $\epsilon \leq \sum_m |w_m - w_m^*|/M$ by statistical sampling, $\mathcal{O}(1/\epsilon^2)$ estimations are required. SMC methods require $M \propto \text{poly}[\dim(\vec{x})]$ [217]. The scaling of $\dim(\vec{x})$ with the system's size n is unknown in general [157], however it is expected to be subexponential in many cases of interest to quantum simulation (see also Chap. 4). We are thus left with the number of necessary epochs: it has been shown how efficient choices for the experimental parameters¹⁴ can provide an exponentially fast learning rate of the parameter values \vec{x}_0 [157, 212, 213]. Therefore we can conclude that the overall procedure is efficient, as long as exponential precision in the estimate of $\mathcal{L}_D(\vec{x}; \tau)$ is not required, i.e. as long as

$$P(D) \in \mathcal{O}(1/\text{poly}(n)) \quad (3.9)$$

for a random set of particles.

¹³Here reported for the single step case, equivalently to Eq. 3.3, but it can be immediately generalised to a cumulative case.

¹⁴ See Sect. 3.2.2.

Now that QHL scalability has been partially addressed, an additional remark about its practicality is due. So far, we assumed the particle positions $\{\vec{x}_m\}$ to be held constant throughout the inference process. If we consider a simple single parameter x case estimation, starting from a uniform prior with support in the interval of width $\Delta = x_{\max} - x_{\min}$, then we need approximately $M \propto \Delta/\tilde{\sigma}(x)$ particles, if we are to ensure a final uncertainty $\sigma(x)$. In this way, for highly accurate estimations, the memory usage by the classical co-processor might quickly become beyond the range provided by fast experimental control equipment¹⁵, e.g. FPGAs. Importantly, this simplistic approach is not necessary, as with the progression of the inference across epochs, the updated prior approximation will have support on only a few of particles. This is exemplified in Fig. 3.2b, where we assumed that a Gaussian approximation for the posterior holds well. Indeed, a successful learning process has $w_m \propto P(\vec{x}_m) \rightarrow 0$ for most of the initial particles, as the latter ones have been ruled out by measurements so far, and therefore storing their locations and weights provides little information.

This inefficiency can be addressed with resampling methods. These sample from the updated posterior new particles, as soon as their weights signify that the *effective size* ($\sum_m w_m^2$) has fallen below a user-defined threshold (t_{resample}). In particular, in this Thesis we rely on the *Liu-West resampler* (LWR), following implementations in [155]. A resampler essentially allows to switch between the density-based and weights-based representation of the information from the updated posterior distribution. It thus extends the approach already seen with particle filtering methods in Sect. 2.1.6. Typical SMC update events in the learning process act on the weights according to Eq. 3.7, leaving intact the locations of the particles. However, when a resampling occurs, it resets the information stored in the $\{w'_m\}$ making them uniform, at the same transferring the information from the previous set $\{w_m\}$ in the density of the new particle locations $\{\vec{x}'_m\}$ (see Fig. 3.2c). Additionally, LWR allows the user to tune how much the particle positions can be modified at each resampling event (via the a parameter), essentially compromising between the learning rate and the robustness to noisy inferences. In the Appendix we provide flowcharts for both a typical SMC update¹⁶ and LWR, respectively in Figs. 3.4 & A.2.

The definition of a credible region (Def. 3.3) can be immediately and efficiently ported into the SMC approximation. Chosen κ , it is enough to identify the subset of particles $X \equiv \{\vec{x}_m\}_{\kappa}$, whose weights satisfy $\sum_X w_m \geq \kappa$. The particles are usually selected starting from those with highest weight [155]. Then, the credible region is given by the minimum-volume enclosing ellipsoid¹⁷ with covariance matrix Σ , which satisfies:

$$(\vec{x}_m - \vec{x}_0)^T \Sigma^{-1} (\vec{x}_m - \vec{x}_0) \leq 1, \quad \forall \vec{x}_m \in X. \quad (3.10)$$

¹⁵ As we will see for a typical magnetometry case in Sect. 3.3.1, state-of-art uncertainties for the estimation of magnetic fields with quantum sensors would lead to GBs of particles' storage in memory.

¹⁶ As employed in the MFL experimental protocol

¹⁷ This can be practically found e.g. via the Khachiyan algorithm [218].

3.2.2 Efficient time heuristics

As aforementioned, an efficient inference process depends crucially from the choice of appropriate experimental controls at each epoch of the algorithm, in order to perform maximally informative experiments. We first give an intuitive explanation about this, using a simple, single-parameter Hamiltonian:

$$\hat{H}_{\text{rz}}(x) = x \frac{\hat{\sigma}_z}{2}, \quad (3.11)$$

with a probe state $|\psi\rangle := |+\rangle$ and a measurement basis $\{|D\rangle\} := \{|+\rangle, |-\rangle\}$. This case was studied in detail in [156, 213].

For this study case, we have again τ as the only experimental control parameter, and therefore choosing an experimental setup ξ here amounts to choosing an evolution time τ_i at epoch i . In Fig. 3.3 we exemplify the intimate relationship between τ and the uncertainty in the parameter estimate, as represented for example by the standard deviation $\sigma_i := \sigma(\{x_m\}_i)$ of the discretised set of particles. In the figure we report the oscillatory behaviour of $\mathcal{L}_+(x; \tau)$ with τ for the true value here chosen arbitrarily $x^* = 2$, along with a set of different representative values $\{x_m\} \in [2, 2.5] = [x^*, x^* + \sigma(\{x_m\})]$.

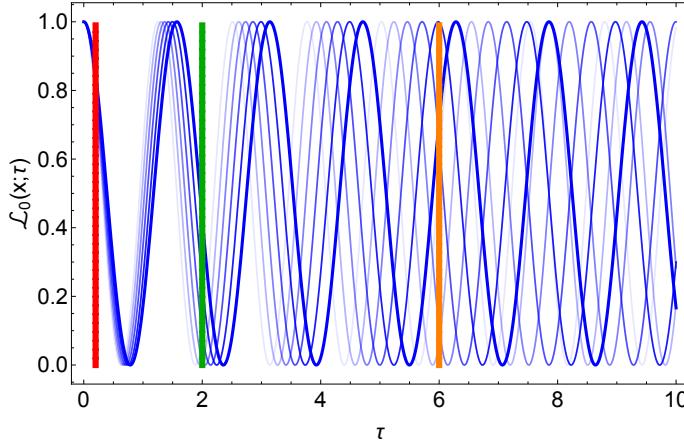


Figure 3.3: The intuition behind the Particle Guess Heuristic (PGH), for the paradigm case \hat{H}_{rz} , where the optimality of PGH can be proven analytically. The vertical axis represents the likelihood $\mathcal{L}_0(x; \tau)$ of obtaining $D = 0$ at time τ . Evolution calculated from trial values x at regular distances $|x - x^*| \leq \sigma/2 = 0.25$ from the true value x^* are represented with fading colours for higher distances. The green vertical line is in correspondence of $\tau = 1/\sigma$.

Now, the inference process outlined in Sect. 3.2 will essentially distinguish among tentative values x_m , according to how well the likelihoods $\mathcal{L}_+(x_m; \tau)$ reproduce the correct one, for the chosen experimental setup τ . In Fig. 3.3 we highlight with vertical lines three different values $(\tau_1, \tau_2, \tau_3) = (0.1/\sigma_i, 1/\sigma_i, 3/\sigma_i)$, respectively in red, green and orange. In correspondence of τ_1 , we can see how $\mathcal{L}_+(x^*; \tau) \sim \mathcal{L}_+(x_m; \tau) \forall m$, so that highly accurate likelihood evaluations are required to distinguish the performance of the particles chosen. In other words, the choice of τ_1

is too conservative, as the confidence interval about x^* stored in the prior is already too tight to be effectively distinguished for such early times. On the contrary, choosing τ_3 would lead to ambiguous outcomes, as due to the periodicities in the dynamics, $\mathcal{L}_+(\vec{x}^*; \tau)$ is accidentally better reproduced by particles which do not minimise $|x_m - x^*|$. In this case, we do not have enough prior information to perform meaningful experiments for such long times. Finally, evolution times $\tau \sim 1/\sigma_i$ appear instead an effective compromise choice, whereby the likelihoods of the various particles are well distinguished but do not lead to ambiguity in the results.

The intuitive reasoning behind Fig. 3.3 is that as the inference learns the parameters with shrinking uncertainty, longer evolution times are required to distinguish values closer in the parameter space. This simple observation can be formalised making use of the *Bayes risk*:

$$r(P(\vec{x}), \vec{x}^*, \tau) := \int R(\vec{x}, \vec{x}^*, \tau) P(\vec{x}) d\vec{x} \quad (3.12)$$

with:

$$R(\vec{x}, \vec{x}^*, \tau) := \sum_D \mathcal{L}(D | \vec{x}; \tau) Q_L(\vec{x}^*, \vec{x}_0(D, \tau)) \quad (3.13)$$

$$Q_L(\vec{x}^*, \vec{x}_0(D, \tau)) := \|\vec{x}^* - \vec{x}_0(D, \tau)\|^2 \quad (3.14)$$

respectively the *risk* of the *estimator*¹⁸ $P(\vec{x}) \xrightarrow{D, \tau} \vec{x}_0$, and the *quadratic loss* against the correct value \vec{x}^* , capturing the *estimate error*. All expressions refer to a single update step. The Bayes risk therefore is an averaged risk of how erroneous the estimate \vec{x}_0 will be, after an update is performed with control τ , obtaining datum D . Loss functions other than Eq. 3.14 are possible, the main advantage of Q_L being that adopting it, the trace of the posterior's covariance matrix equates the Bayes risk [156].

The Bayesian strategy is to minimise the objective function given by Eq. 3.12, considering that under reasonable regularity conditions for $P(\vec{x})$, at any given epoch the Bayesian estimator gives $\vec{x}_0 \equiv \mathbb{E}[P'(\vec{x}|D; \tau)]$, i.e. the expectation value of the posterior [157]. A first lower bound, in full generality for the single-parameter model in Eq. 3.11, is provided by the application of the Cramer–Rao bound¹⁹, obtaining for the risk of performing an update with time τ [213]:

$$r(P(\vec{x}), \vec{x}^*, \tau) \geq 1/\tau^2. \quad (3.15)$$

However, this lower bound has been highlighted to be loose for few exemplary characterisation cases and most importantly, does not provide a clear indication for the choice of an optimal τ . A more meaningful bound can be introduced by invoking more stringent assumptions and

¹⁸ Here the estimator is the function that infers $\vec{x}_0(D, \tau)$, i.e. the estimate of the parameter from the dataset D, τ (hence this dependence of the estimate \vec{x}_0 has been made explicit in Eqs. 3.12–3.14.), in full generality.

¹⁹ The Cramer–Rao bound states how the estimator's variance must obey $\sigma^2(\vec{x}_0) \geq 1/\mathcal{I}(\vec{x})$, with \mathcal{I} the Fisher information. Since we adopted the loss function in Eq. 3.13, we remind that we can immediately replace $\sigma^2(\vec{x}_0) = r(P(\vec{x}), \vec{x}^*, \tau)$. In general, $\mathcal{I}(\vec{x}) \equiv -\sum_D \mathcal{L}(D | \vec{x}, \tau) (\nabla_{\vec{x}} \log(\mathcal{L}(D | \vec{x}, \tau)) \cdot \nabla_{\vec{x}}^T \log(\mathcal{L}(D | \vec{x}, \tau)))$, which for this simple 1D case gives: $\mathcal{I}(x) = -\frac{\partial^2}{\partial x^2} (\sin^2(x\tau/2)) - \frac{\partial^2}{\partial x^2} (\cos^2(x\tau/2)) = \tau^2$, QED.

particularly, that the posterior is well approximated by a normal distribution²⁰. In this case, the integral from Eq. 3.12 can be expressed analytically in closed form as an oscillatory function (in τ), which is lower bounded by the envelope [156]:

$$(1 - \tau^2 \sigma^2 e^{-\tau^2 \sigma^2}) \sigma^2 \leq r(P'(x), x^*, \tau) \quad (3.16)$$

after an update starting from prior $P(x)$, characterised by the variance $\sigma^2 \equiv \sigma^2(P'(x))$ (for a compact notation). The envelope in Eq. 3.16 has a minimum of:

$$r(P'(\vec{x}), \vec{x}^*, 1/\sigma) = \left(1 - \frac{1}{e}\right) \sigma^2, \quad (3.17)$$

which corresponds to the choice $\tau = 1/\sigma$, that is thus the optimal experiment design. This proves formally the intuitive observations at the start of the section. Importantly, comparing Eqs. 3.15 and 3.17 ensures that under an online, locally optimal strategy for the experiment design, the risk is expected to shrink exponentially with a constant factor of ~ 0.6 . This is the fundamental reason behind the exponentially fast learning rates of the parameters, displayed by simulated and experimental cases shown below.

In multi-dimensional parametrisations, the strategy leading to the optimal experiment design for τ in [156] becomes analytically too convoluted. However, numerical studies highlight how exponentially reducing risks are observed for several study-cases, when choosing $\tau \sim 1/\|\Sigma(P(\vec{x}))\|$, i.e. the inverse of the norm²¹ of the covariance matrix for the estimate of \vec{x}_0 [157, 213]. This observation is rendered in a computationally efficient *particle guess heuristic* (PGH), given by:

$$\tau \sim \frac{1}{\|\vec{x}_1 - \vec{x}_2\|} \quad (3.18)$$

with $\vec{x}_1 \neq \vec{x}_2$ two particles sampled from $P(\vec{x})$.

3.3 Efficient magnetometry with NV-centres

In Sect. 1.5 we have briefly commented about the potentiality of single electron spins in NV-centres as nano-resolution electromagnetic sensors. However, delivering high sensitivity NV-magnetometers required so far high-contrast, single-shot measurements available only at cryogenic temperatures. This constitutes a major obstacle to real-world applications [135, 139], or prior knowledge of correlations existing in the estimated signal [219]. In this chapter we discuss how the QHL protocol introduced in Sect. 3.2 (and following) can be applied to single-spin magnetometers down to the one-photon readout level already at room temperature, achieving an optimal sensitivity scaling.

²⁰ This is guaranteed asymptotically, but might not be the case in early stages of the learning, nor for multi-modal distributions.

²¹ In the following, when not replaced by the PGH in Eq. 3.18, we will always adopt the *Frobenius norm* $\|\Sigma\|_F \equiv \sqrt{\text{Tr}\{\Sigma \Sigma^\dagger\}}$.

In this section we will deal mainly with DC-magnetometry in NV-centres, which relies especially upon Ramsey interferometry (see Sect. 1.5 and Fig. 3.5e–h). For an external magnetic field \vec{B} perpendicular to the electron spin magnetic moment, we have that the frequency of the Larmor precession $f_B = \gamma B / 2\pi$ depends upon the magnetic field intensity, and the electron gyromagnetic ratio $\gamma \approx 2\pi \times 28 \text{ MHz/mT}$ [220]. Therefore, for a chosen precession time τ , the initially prepared $|+\rangle$ state evolves accruing a relative phase $\phi = 2\pi f_B \tau$ between the eigenstates. This phase can be detected by a projective measurement in the $\hat{\sigma}_z$ eigenbasis of the electron spin, after the final $\pi/2$ -pulse of the Ramsey sequence. B can be thus inferred from the frequency of the fringes observed by varying τ , and at each precession time estimating ϕ .

The absolute *precision* $\bar{\eta}^2$ for a magnetometer²² takes into account the attained error in the estimate, $\sigma(B_{\text{est}})$, once a sensing time \bar{T} is elapsed:

$$\bar{\eta}^2 = \sigma^2(B_{\text{est}})\bar{T}. \quad (3.19)$$

Therefore, this figure of merit captures the rate at which B_{est} is improved. The intrinsic cost of the procedure amounts to the *total phase accumulation time* $T := \sum_i^N \tau_i$, which is the sum of every precession time τ_i used [221–223]. The scaling of the precision is fundamentally limited by the decoherence timescale T_2^* , which for an NV-centre is due to homogeneous broadening [224].

An ideal quantum sensor, controlled by an optimal sensing protocol, is expected to achieve the Heisenberg limit²³ (HL), eventually saturating at a lower precision controlled by the sensor's T_2^* . However, experimental implementations introduce unavoidable overheads, required to initialise and readout the sensor, and computationally process the outcomes. All these overheads contribute to increase the sensing time required to achieve a certain $\sigma(B_{\text{est}})$, so that the precision saturates well above the HL.

In particular, each estimate of ϕ is affected by quantum projection noise, when performing the projective measurement on the σ_z -basis, impacting on the readout process [40] independently from experimental conditions. In addition, at cryogenic temperatures, the optical transitions addressed in the Ramsey experiment (see also Fig. 3.5e&h) are spin-selective, so that in the optical readout operation, negligible photoluminescence is emitted when the electron is in the $m_s = -1$ state, and readout is possible with high-fidelity from a single Ramsey sequence [135], up to photon collection and detection efficiencies. At room temperature, instead, optical transitions are thermally broadened, and the contrast in observed PL from (non)radiative decay, after exciting different states of the ground triplet, is reduced to $\approx 35\%$ (Fig. 3.5h, refer for details to Sect. 1.5). Overcoming these sources of uncertainty is usually obtained by repeating the Ramsey sequence a large number M of times, so that sufficient statistics is accrued from the collected photons. This approach, however, introduces large overheads that impact on the achievable sensitivity. In particular, in those instances where cryogenic temperatures are not practical, such as in

²² In the following, we will also adopt the nomenclature *sensitivity*, for $\bar{\eta}$.

²³ A thorough discussion of the HL limit can be found in [56], in this Thesis we provide a qualitative understanding of it in Appendix B.2.

biological applications [23, 225], final sensitivities achievable are several orders of magnitude worse.

The choice of the sensing protocol is also crucial in contributing to the sensor performances. Naïve methods using a fixed sensing τ_f throughout the procedure do not provide optimal sensitivity, nor *dynamic range*²⁴. For example, due to the periodicity of the fringes, any strategy adopting a fixed τ_f will be intrinsically bound to sensing within an interval of width $2\pi/\gamma B$. Methods adopting an equally spaced grid of τ 's (such as Fourier–transform based methods), instead, are often too conservative in choosing a new time to perform experiments, as suggested by Eq. 3.17. A suboptimal update of evolution times bound these methods to the classical *standard measurement sensitivity* (SMS), as it will be shown in detail later on. These limitations have been overcome adapting phase estimation protocols [21, 22, 56, 57, 226, 227], with successful experimental demonstrations surpassing the SMS [135, 150, 222, 228, 229]. Some of these methods require adaptivity in the measurement basis (see Chap. 2), in order to reduce the M at which they are stable [135, 136, 230].

However, these approaches require computationally intensive updates of the corresponding probability distributions [56, 57], so that computational overheads become dominant in limiting their sensitivity. Approximate machine learning techniques, such as those described in Sect. 2.1.6 and 3.2.1, suggest an appealing alternative, that could be embedded in a computationally efficient sensing protocol [82, 141]. Nevertheless, all sensing algorithms listed so far do not offer any ad-hoc solution to mitigate the effect of noisy measurements, with Bayesian–inspired methods relying simply on their intrinsic noise robustness properties [57, 82]. In the following section, we intend to demonstrate how the CLE procedure – described in detail in Sect. 3.2 and following – can provide an effective subroutine for multi-parameter quantum sensing with Heisenberg–limited scaling of sensitivity, flexible noise–modelling, and reduced computational overheads. Therefore, the resulting sensing protocol will be named *Magnetic Field Learning* (MFL). The protocol flowchart can be found in Fig. 3.4, explained in detail in the following Section. We will show how all features as above can be accomplished in NV–centre magnetometry, without the need for cryogenic setups, nor adaptation of measurement bases²⁵.

3.3.1 Demonstrating efficient magnetometry via CLE: Magnetic Field Learning

General QHL algorithms have been introduced in Sect. 3.2. These can be readily applied to a simple instance of magnetometry with Ramsey sequences, by introducing the relevant Hamiltonian²⁶:

²⁴ The dynamic range for a sensor is defined as the largest interval $[0, B_{\max}]$, wherein the value of B can be sensed unambiguously.

²⁵ This, as well as ad-hoc noise modelling, are the ultimate reasons why we favour CLE over RFPE in designing the MFL.

²⁶ In Eq. 3.20, B refers to the magnetic field component parallel to the NV–centres' symmetry axes. However, the MFL protocol could in principle be expanded to detect arbitrary orientated magnetic fields as a multi-parameter estimation.

$$\hat{H}(B) = \omega_B \hat{\sigma}_z/2 = \gamma B \hat{\sigma}_z/2. \quad (3.20)$$

Therefore, estimating $\omega_B \equiv 2\pi f_B$ alone suffices to learn the value of B . The recursive steps of MFL, composing each epoch and represented in Fig. 3.5(a-d), are equivalent to those outlined at pg. 78. More specifically:

1. The experimental setting is here the phase accumulation time τ_i for the next Ramsey sequence, estimated from PGH as $\tau_i \approx 1/\sigma_{i-1}$, where σ_{i-1}^2 is the uncertainty embedded in the prior $P(\vec{x}_{i-1})$. By choosing each τ_i inversely proportional to the uncertainty in the last updated estimate of B , τ 's are expected to increase exponentially if MFL achieves HL-scaling in sensitivity [231].
2. The NV-centre is always prepared in a probe state $|\psi\rangle = |+\rangle$, and evolved according to a Ramsey sequence - Fig. 3.5(e-h).
3. A projective measurement is performed on the $\{|0\rangle, |1\rangle\}$ computational basis. The corresponding photo-luminescence count is then used to extract the outcome D , as in Fig. 3.5(i). In cases with limited readout fidelity, D might be extracted adopting different strategies. E.g. the projective measurement could be repeated M times and D inferred from *majority voting*, or by a weighted *binomial sampling*. Evidences detailed later suggest how choosing $M \gg 1$ is suboptimal in many cases²⁷.
4. The prior is updated using Bayes' rule, and invoking SMC methods as a practical and computationally efficient strategy (see Sect. 3.2.1).

Importantly, a quantum sensor is invoked in steps 2 & 3 alone, whereas a simulator is required by all other steps.

We tested MFL on a large data set of $\sim 60,000$ Ramsey interferometry sequences²⁸ from three different NV-centres, labelled α , β and ϵ (see Table 3.1). The final readout from sequences, each evolving the spin for a different accumulation time, can be combined into a Ramsey *fringe* (Fig. 3.5i). As explained in Sect. 1.5, fringes might report²⁹ outcomes that are averaged across several independent sequences with the same τ , to increase the contrast and offer less noisy readout. The results offer a benchmark for the performance of MFL mainly when compared with standard FFT methods, but also against former QPE algorithms. All results from the application of MFL are representative behaviour, averaged over $R = 1000$ independent protocol runs (unless otherwise stated).

²⁷ A demo notebook exemplifying this particular point was made available at: <https://figshare.com/s/740209b49b2522ad211a>, adopting the Python library written on purpose and available therein. Also, refer to the footnote at pg. 52 for a discussion about the optimality of majority voting.

²⁸ A sequence is intended as a single sequential implementation of the gates manipulating the NV electron spin, see also Fig. 1.10b.

²⁹ This will be explicitly mentioned wherever it is the case.

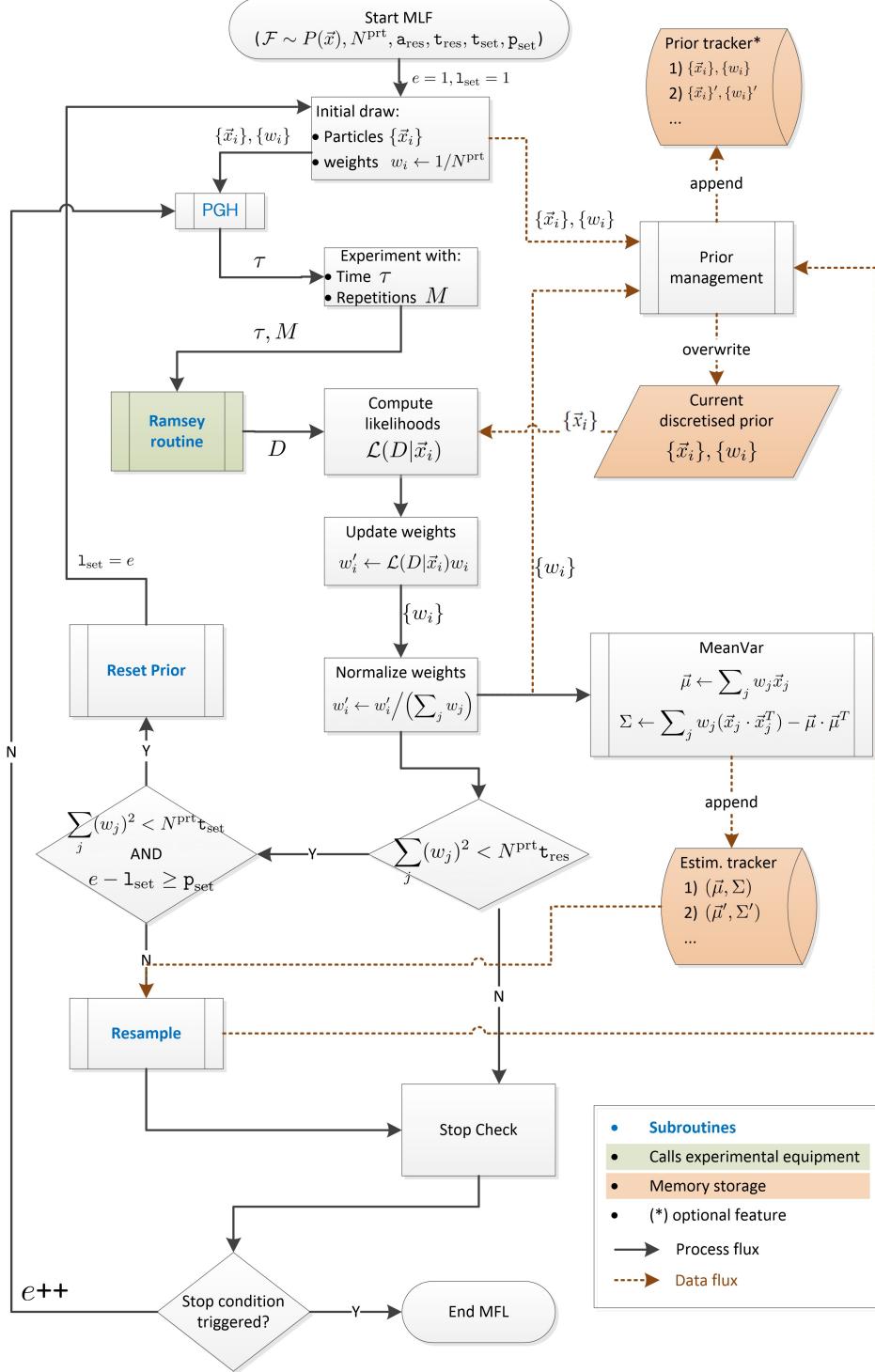


FIGURE 3.4. Flowchart for the MFL protocol. The legend in the lower box explains the colour-coding for the operations. Symbols and subroutines are defined in the main text (for details about Resample also see Appendix Fig. A.2 and about ChangeDetect, Algorithm 4).

Experiments were performed using a confocal set-up, at room temperature, with an external magnetic field of ≈ 450 G, parallel to the NV–centre axis³⁰, giving a Zeeman shift of ω_B [220, 222]. The Ramsey interferometry experiments reported here use the $m_s = 0$ and $m_s = -1$ electronic sublevels (Fig. 3.5f&h) of a few different ^{12}C isotopically purified diamond samples. For each Ramsey sequence (Fig. 3.5e), laser pulses for initialisation and readout have a wavelength of 532 nm, and the PL signal is detected with an avalanche photodiode (APD) for 350 ns.

The data processing was performed with a customised Python library, built upon open-source packages [232], and in particular QInfer [155]. As a model for the likelihood describing the Ramsey fringes collected from an NV sensor with dephasing time T_2^* , we use the function³¹:

$$\mathcal{L}(0|B, T_2^*; \tau) = \exp(-\tau/T_2^*) \cos^2(\gamma B \tau/2) + [1 - \exp(-\tau/T_2^*)]/2. \quad (3.21)$$

In Eq. 3.21, T_2^* can be considered either a known parameter, or $T_2^* = \infty$ whenever $T_2^* \gg \tau_{max}$. The various parameters of the MFL protocol, and in particular the number of particles N^{prt} , a and $t_{resample}$ of the SMC, can be optimised for a certain setup prior to the sensing operation, without adding to the sensing time overheads. For the results shown in this section, we adopted an optimised resampling threshold $t_{resample} = 0.5$ and smoothing parameter $a = 0.9$. The role of these parameters is made evident in Fig. 3.4 & Appendix Fig. A.2. By SMC, $P(B)$ can be accurately represented at any epoch, whilst using $N^{prt} \leq 1500$ particles.

| label | NV center | sets (n) | sequences (M) | $\Delta\tau$ (ns) | \bar{B}_{est} (μT) | T_2^* (from fit, μs) | η ($\bar{\eta}$) |
|--------------|------------|----------|---------------|-------------------|-----------------------------------|---------------------------------------|--|
| α_1 | α | 120 | 18500 | 20 | 52 | - | $T^{-0.99 \pm 0.02}$ (-) |
| α_2 | | 115 | 18500 | 20 | 710 | - | $T^{-0.97 \pm 0.03}$ (-) |
| β_1 | β | 67 | 30000 | 100 | 8.3 | 16 | - (-) |
| ϵ_1 | ϵ | 1 | 20275 | 20 | 58 | - | $- (T^{-0.73 \pm 0.03} \rightarrow T^{-1.0 \pm 0.02})$ |
| ϵ_2 | | 1 | 8876 | 200 | 6.0 | 64 | $T^{-0.91 \pm 0.03}$ (-) |
| ϵ_3 | | 1 | 44000 | 20 | 10 → 550 | - | - (-) |

TABLE 3.1. Synoptic view of the experimental data sets and NV–centres used for Sect. 3.3.1. Representative MFL (absolute) precision scalings η ($\bar{\eta}$) are also reported.

As a first test, we analyse the learning of B_{est} by estimating its associated uncertainty, $\sigma(B_{est})$, i.e. the variance of $P(\vec{x})$, against MFL epochs. For this purpose, we use the dataset α_1 , with 120 fringes all obtained with $M = 18500$ sequences. As this test was run offline, at every MFL epoch we replace the adaptively chosen $\tau_i \simeq 1/\sigma_{i-1}$, with the experimental τ minimising $(|\tau - \tau_i|)$. Figure 3.6a shows $\sigma(B_{est})$ decreasing exponentially, until about 50 epochs are reached. After this

³⁰ This intensity leads to hyperpolarisation of the NV–centre, preventing deconvolution of different frequency responses from the magnetometry signal [132].

³¹ Modelling of the decay of Ramsey fringes, as well other experiment where no refocussing techniques are employed, employs in general a *stretched exponential* factor $\exp(-\tau/T_2^*)^p$. According to the lineshape of the NV defect resonance when probed with optical light, one assumes $p = 1, 2$ for idealised cases [233]. Here we found phenomenologically an excellent agreement of the decays with $p = 1$ and adopted it, also to match the theoretical investigation for QHL purposes performed in [213].

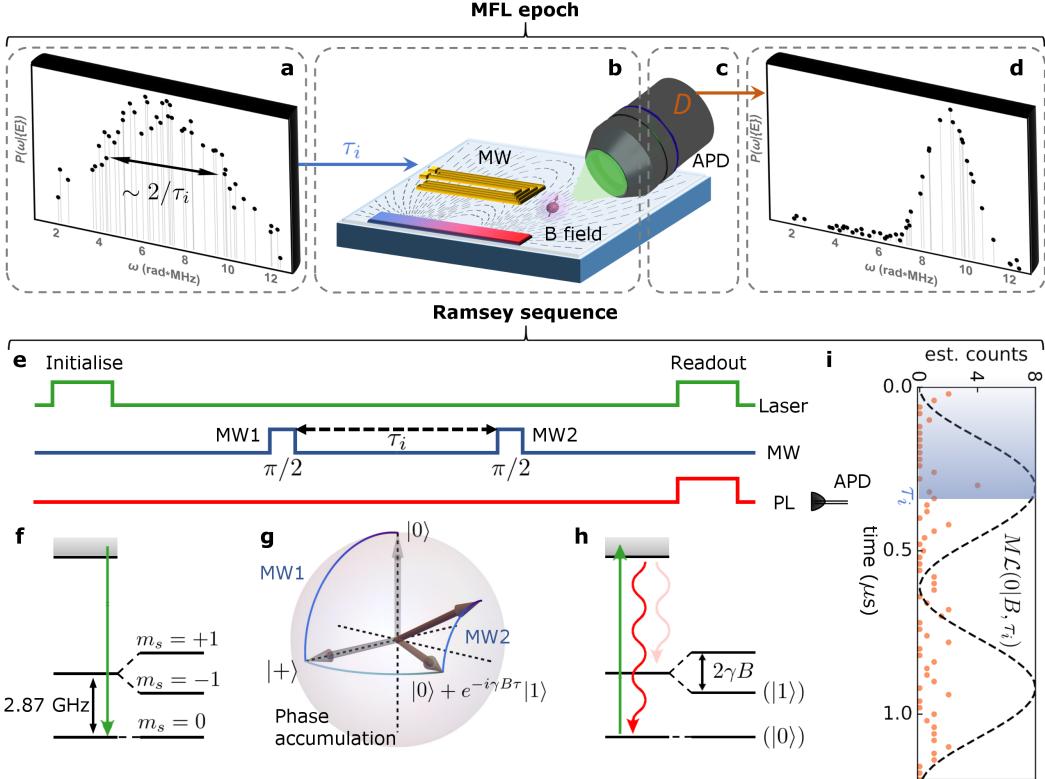


FIGURE 3.5. An MFL algorithm epoch, along with a Ramsey sequence and readout. **a** An experimental phase accumulation time τ_i is chosen, according to the uncertainty encoded in the prior P_{i-1} . **b** M Ramsey sequences of duration τ_i are implemented. Permanent magnets generate a B -field driving the precession. A confocal setup focuses coherent light on the NV–centre site. MW pulses are generated via planar Cu wires fabricated on the surface of the bulk diamond sample. **c** The outcome D from the Ramsey sequences is measured. **d** An update of the prior $P_{i-1} \rightarrow P_i$ occurs, leading to a novel τ_{i+1} . **e** The NV spin vector is initialised with laser light, rotated with MW pulses, and, using a second laser pulse, readout from photoluminescence (PL) with an avalanche photodiode (APD). **f** MW manipulation occurs between the $m_s = 0$ and $m_s = -1$ states, encoding $|0\rangle$ and $|1\rangle$, respectively, see also Sect. 1.5). **g** Visualisation of the effect of MW rotations and Larmor precession on the electronic spin state. **h** Optical detection via the pump-and-probe PL mechanism described in Sect. 1.5. **i** A representative PL fringe (theory plotted as dashed line) with orange data-points representing the number of detected photons for $M = 8$.

point, the precession times τ selected by MFL saturate at the highest available experimental time $\tau_{\max} = 10 \mu\text{s}$ for the sequences labelled as α_1 . As a consequence, further reduction of $\sigma(B_{\text{est}})$ occurs polynomially fast, by accumulating statistics for times $\tau \sim \tau_{\max}$.

The sensitivity η of a magnetometer can be obtained simplifying Eq. 3.19:

$$\eta^2 = \delta B^2 = \sigma^2(B_{\text{est}})T, \quad (3.22)$$

so that only the total phase accumulation time T is taken into account (when neglecting

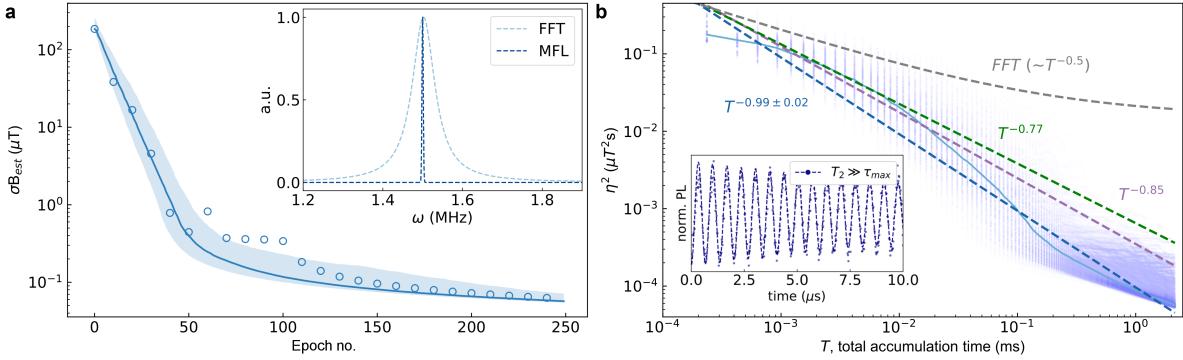


FIGURE 3.6. Experimental results showing the scaling in uncertainty and precision. Lines represent median values, and performance within the 68.27% percentile range is shown as shaded areas. **a** Estimated uncertainty $\sigma(B_{\text{est}})$ against epoch number; data from one sample run is shown as blue circles. In the inset, a Lorentzian fit (FFT) and a Gaussian fit (MFL) providing $\sigma(\omega_{\text{est}})$ in the Ramsey frequency at the end of a typical protocol run. **b** The scaling of precision with total phase accumulation time T , excluding all overheads. Density plots are single runs, and as a blue dashed line the LSF to the data, representing a typical MFL performance. Results from FFT are as a grey dashed curve. In green and violet, the scaling for phase estimation algorithms from Refs. [222, 228], respectively. The HL-scaling (T^{-1}) is omitted as non distinguishable from typical MFL performance. The inset shows normalised PL in an experimental Ramsey fringe, with a 20 ns sampling rate, up to $\tau_{\text{max}} \sim 0.14 T_2^*$. Super-imposed, a LSF with a decaying sinusoid (blue dashed line).

overheads). In Fig. 3.6b we plot how η^2 scales with T , for each epoch, again for the α_1 set. $\eta^2 \propto T^{-0.99 \pm 0.02}$ for MFL (the error is estimated from a bootstrapping procedure³²) thus overlapping with HL scaling (that is $\propto T^{-1}$). Processing the dataset with FFT, instead, $\eta^2 \propto T^0$ is quickly approached, i.e. the SMS. Here, where overheads are neglected, the scaling reported for previous QPE methods behaves in a manner that is qualitatively comparable to MFL, however the additional post-processing required becomes evident when analysing $\bar{\eta}$ (see Appendix Fig.B.1).

As aforementioned, the learning rate slows when $\tau_i \geq \tau_{\text{max}}$ are chosen by the PGH. This slowdown is an interesting feature, and can be observed by plotting either the $\sigma(B_{\text{est}})$ or the η scalings (shown in Fig. 3.6a&b respectively). We stress how the latter case has $\tau_{\text{max}} \ll T_2^*$, i.e. the maximum accumulation-time budget is met well before any decoherence can be observed. In this situation, when times generated by the PGH exceed the τ_{max} limit, an accumulation of data-points with $\tau \approx \tau_{\text{max}}$ occurs, and we observe $\eta^2 \propto 1/\sqrt{\tau_{\text{max}} T}$. To verify this, in Fig. 3.7b the scaling of $\sigma(B_{\text{est}})$ is plotted alongside the ratio τ/τ_{max} , using instead sequences drawn from the

³² In particular, the bootstrapping here involves a sampling, with replacement, from the available sets in each dataset used – Table 3.1. Each resample generates a sample of cardinality here chosen to be 1000, and MFL is run on each sample independently. The median scaling performances for each sample are collected in an approximate population (which is thus of size 1000), and the standard deviation from this approximate population of scaling performances is provided as the precision scaling error.

set ϵ_1 . Evidently, as soon as a typical run meets the $\tau \geq \tau_{max}$ condition, $\sigma(B_{est})$ plateaus, which in turn saturates the increase in τ .

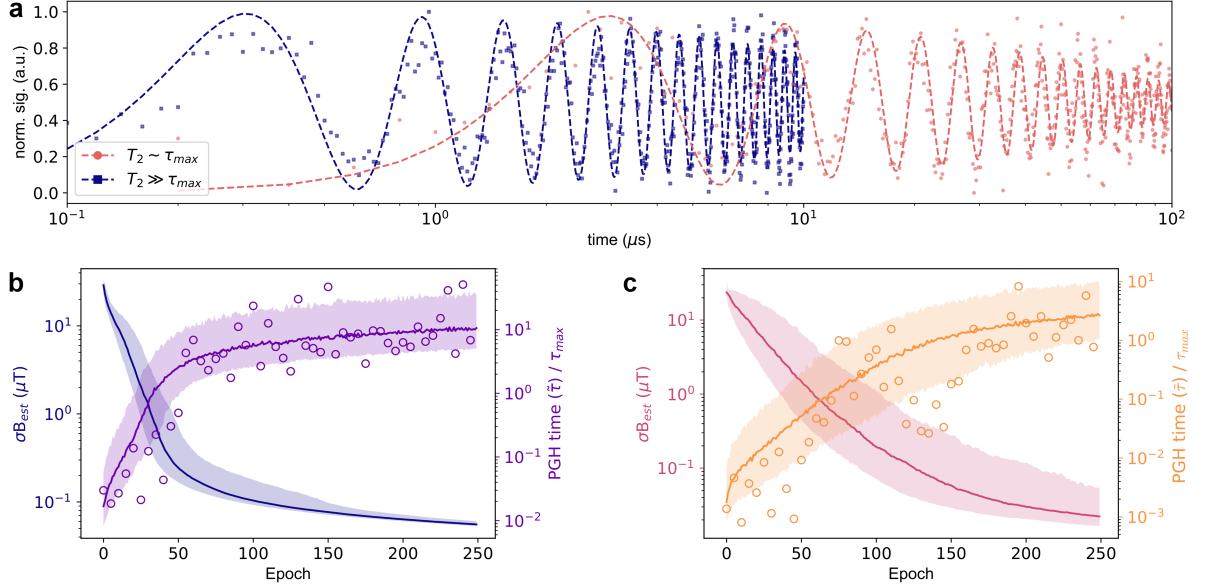


FIGURE 3.7. Analysis of the behaviour of the PGH for datasets $\epsilon_{1,2}$ as in Table D.1, chosen such that $T_2^* \gg \tau_{max}$ or $T_2^* \approx \tau_{max}$ (reported in the plots in darker and brighter colours, respectively). **a**, Renormalised photon counts along two different Ramsey experiments with the same NV-centre (scatter plots). The different frequencies of the fringes originate from different magnetic field intensities ($\approx 58 \mu\text{T}$ and $\approx 6 \mu\text{T}$). Superimposed a least-square fit (dashed lines), adopting the oscillatory function with depolarising noise as in Eq. 3.21. **b**, Estimated uncertainty $\sigma(B_{est})$ and ratio between PGH-generated time and τ_{max} as available from the first dataset, plotted against each epoch of the MFL algorithm. A majority voting method is adopted, under the hypothesis that $T_2^* \gg \tau_{max}$. Solid lines are median values calculated over 1000 independent runs, whereas shaded areas are 68.27% percentile ranges centred around the median. Superimposed as a scatter plot, a sample of times generated by the PGH during an representative run. **c**, Same as in **b**, for the case where $T_2^* \approx \tau_{max}$. No majority voting is in place, and data from the experiment are extracted probabilistically from the experimentally estimated likelihoods.

Note, this artefact deriving from the artificial choice of a maximum time budget τ_{max} is equivalent to the phenomenon exhibited in correspondence of dephasing noise [213], reducing the contrast in the fringes from a Ramsey sequence (see red-coloured data in Fig. 3.7a). We can prove numerically this statement with Fig. 3.7c. Here, experimental datapoints are from the set ϵ_2 , where $\tau_{max} = 100 \mu\text{s} \approx 1.6 T_2^*$ (estimated from a least-squares fit). Also, to approximately preserve the number of periods in the Ramsey fringes of this and α_1 datasets, we reduced $B = 6 \mu\text{T}$. Now, these data display an evident decay. Majority voting schemes tend to hide this, especially for high numbers repetitions, hence they are here removed and replaced with probabilistic extraction of

the datum D at each epoch, from the experimentally estimated likelihood (see Appendix B.1).

By not leveraging upon the majority voting scheme, we are using less effectively³³ the information available in $\mathcal{L}(B; \tau)$: this produces a slower scaling of $\sigma(B_{\text{est}})$. Indeed, with this readout processing, the extraction of each D is equivalent to the one from single-shot measurements from a setup with comparable readout fidelity. Hence, the same binomial uncertainty affects each D obtained. This is why the information accumulated about $\mathcal{L}(B; \tau)$, by combining $M \gg 1$ sequences, is not made completely available by the bipartite sampling of D to the inference process. Now that decoherence effects are important, $\sigma(B_{\text{est}})$ plateaus already when $\tau \approx T_2^*$ in average. Nevertheless, important statistical fluctuations in τ are still possible within a single run (see Fig. 3.7c). A corresponding plateau can be observed also for the precision scaling: $\eta^2 \propto 1/T_2^*$ (omitted for brevity).

Finally, we discuss the frequency resolution of MFL and FFT methods in the inset of Fig. 3.6(a). In both cases, the estimation is halted when the maximum time budget τ_{max} is reached. Because of the different nature of the two estimation procedures, the fits are performed against two different distributions: a Gaussian and a Lorentzian. When $\tau \approx \tau_{\text{max}}$, we find that $\sigma(B_{\text{est}})$ is ~ 40 times smaller for MFL. In order to have a meaningful comparison, we allowed for approximately similar T to be cumulated for both MFL and FFT protocols. Therefore, in this case FFT adopts 500 equally spaced datapoints in the interval $[0, \tau_{\text{max}}]$, hence $T_{\text{FFT}} = 2.5$ ms, whereas the representative MFL run has $T \approx 2.2$ ms after analysing 250 datapoints. In the inset of Fig. 3.6, we pick a random subset from α_1 as example (Table 3.1). Then, we perform FFT using all the available 500 τ_i once, obtaining the final estimate shown in the Figure.

3.3.2 Resource efficient magnetometry with MFL and noise modelling

For a true measure of absolute sensitivity, experimental and computational overheads must be accounted for into the total running time \bar{T} , as in Eq. 3.19. In this experiment, for an MFL epoch using a single Ramsey sequence these overheads are:

- the time τ^{comp} spent to obtain the next experiment with the chosen algorithm (for MFL ~ 0.4 μ s per step, per particle on a single-core machine),
- the initialisation and readout times (i.e. two laser pulses for $\tau^{\text{las}} \sim 3$ μ s in total),
- the τ^{wait} relaxation time (1 μ s),
- the TTL pulse τ^{TTL} duration, to enable the photodetector (20 ns)
- the MW control pulses, which last $\tau^{\text{MW}} \sim 50$ ns.

Communication times τ_{comm} are instead not considered here. As the overheads listed are for each sequence, we need to analyse how statistics is accrued from a number of sequences M . The

³³ It must be emphasised how this is true here only because $M \gg 1$ sequences were already combined in a single fringe, from where single binary data D are extracted. However, explained in the following Section, offline accumulating many sequences before an update is performed can be a very inefficient choice in itself, see also [234].

dataset ϵ_1 comprises Ramsey sequences for precession times increasing from τ_1 to τ_{500} in steps of 20 ns. For each τ_i , 20275 sequences were performed, and data were stored such that the results from each individual sequence could be retrieved³⁴. It was thus possible to run MFL against data belonging to different fringes, each obtained combining a fixed numbers of sequences $M \geq 8$. It is noteworthy that when combining an increasing number M of sequences, the number of luminescent photons used to infer the readout at each epoch increases linearly ($n_{\text{phot}} \propto M$). Due to collection efficiencies, on average we detect one photon per $M \approx 8$ sequences (raw counts are displayed for this case in Fig. 3.5i).

The total running time of the protocol, allowing for both variable and constant overheads, is thus:

$$\bar{T} = \sum_i^{N_e} (M\tau_i + \tau_i^{\text{comp}}) + N_e M (\tau^{\text{las}} + \tau^{\text{wait}} + \tau^{\text{TTL}} + \tau^{\text{MW}}) \quad (3.23)$$

after N^{exp} epochs, each using data extracted from M combined sequences.

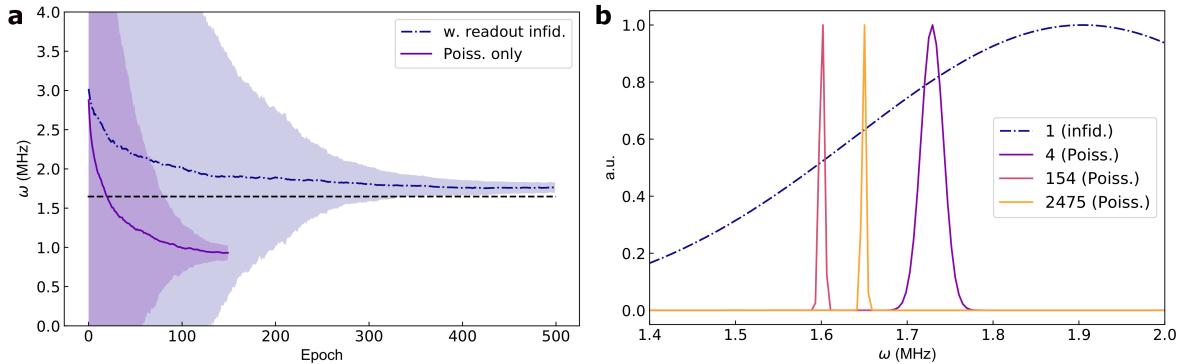


FIGURE 3.8. Compensation of noise in MFL. **a**, The precession frequency ω is estimated from the set ϵ_1 , using $M = 8 \Rightarrow n_{\text{phot}} = 1$ average photons collected per step. In violet the result using the likelihood model³⁵ in Eq. 3.21. In blue, results from the modified likelihood Eq. 3.24. Shaded areas here represent the median $\sim 68.27\%$ credible interval provided intrinsically by MFL at each epoch, averaged over 1000 runs. **b** Gaussian fit over the posterior learnt after 150 epochs, for a representative MFL run and the two cases in **a**. Other cases with different n_{phot} (in the legend) are also displayed (compare with Fig. 3.9). The Gaussian fit for $n_{\text{phot}} = 1$ was truncated to visualise better the remaining cases.

Quantum projection noise and readout infidelities pose a severe challenge against a successful usage of room temperature set-ups. As the readout fidelity decreases along with the fringes' contrast, it is particularly important to address these shortcomings when M is reduced. Here, at first we rely upon the intrinsic noise-robustness of Bayesian methods, adapting the SMC parameters as needed, and taking into account the Poissonian noise affecting the photon counts

³⁴ Therefore, $\binom{20,275}{M}^{500}$ subsets of data from ϵ_1 could be selected and combined to construct fringes comprised of M sequences.

for each datum (Appendix B.1). However, when $n_{\text{phot}} < 4$, readout infidelities and losses become the dominant noise mechanisms, such that our model likelihood \mathcal{L} cannot describe effectively outcomes extracted from smaller batches of experiments. This failure is exemplified in Fig. 3.8a, where adopting \mathcal{L} as in Eq. 3.21 and incorporating only Poissonian noise sources³⁶, the true value of ω (as a dashed line) isn't retrieved within $\sigma(B_{\text{est}})$. More in detail, our model \mathcal{L} assumes implicitly binomially distributed $D \in \{0, 1\}$. However, the readout fidelities³⁷ ξ_0 and ξ_1 , respectively for the $|0\rangle$ and $|1\rangle$ states, are asymmetric due to photon losses. In particular, using the raw data in Fig. 3.5i for $n_{\text{phot}} = 1$, we obtain that $\xi_1 \approx 0.54$ when assuming $\xi_0 \sim 1$. The corresponding unbalance in output probabilities conflicts with our strict Poissonian modelling assumption and therefore, CLE noise robustness alone does not suffice to protect against such “poisoned” modelling [155].

We now intend to demonstrate that the obstacle preventing MFL to succeed lies exclusively in the incorrect model, and not in any fundamental limitation of the approach. To this extent, we adopt a modified likelihood:

$$\mathcal{L}'(1|\vec{x}, \tau) = \xi \mathcal{L}(1|\vec{x}, \tau) \quad (3.24)$$

where $\xi \in [0, 1]$, $\mathcal{L}'(0|\vec{x}, \tau) = 1 - \mathcal{L}'(1|\vec{x}, \tau)$. When $\xi = 1$, the usual $\mathcal{L}'(0|\vec{x}, \tau) = \mathcal{L}(0|\vec{x}; \tau)$ can be recovered. We also improve our initial estimate for ξ , by a preliminary CLE run³⁸ with $M = 20275$, ω is assumed known, and $\vec{x} = \{\xi\}$ is the only free parameter. Such numerically optimised outcome ($\xi = 0.72$) is then adopted as a fixed parameter in MFL runs using $\mathcal{L}'(B)$. The likelihood in Eq. 3.24 expands ad-hoc the original model, to incorporate possible photon losses with the additional parameter ξ . This might either be user-tuned, preliminary learnt in a characterisation experiment, or learnt simultaneously with \vec{x} . We chose the second strategy here, having observed losses to be constant in the setup, within the timing of the magnetometry experiment.

The effectiveness of this noise modelling procedure is reported in Fig. 3.8, emphasising how the final estimate for ω is greatly improved, and is now well compatible with the target value ω^* . Intuitively, with this approach measurement outcomes $D = 0$ are not trusted by the inference process, as they might be spurious reads due to losses. By potentially discarding informative experiments, the learning rate slows down, but a correct behaviour and estimates are restored.

Finally, we combined all results so far using Fig. 3.9 to display the scaling of $\bar{\eta}$ with \bar{T} when using different n_{phot} per epoch (and up to 250 epochs). Plots using a lower n_{phot} are characterised by an initially increasing precision (along with the phase accumulation time, because $\tau_i \propto 1/\sigma_{i-1}$), followed by a HL-scaling, before the saturation discussed before occurs. Increasing n_{phot} reduces the range of epochs where a slower learning occurs. However, this advantage during the initial stages of the learning³⁹ is compensated by worse absolute sensitivities upon convergence. Indeed,

³⁶ By invoking the binomial sampling of each extracted datum, as outlined in Appendix B.1.

³⁷ Taking all sources of noise and loss into account.

³⁸ In principle, it is also possible to estimate ξ using a multi-parameter inference model as in Sect. 3.3.3.

³⁹ Along with the consequent higher precision scaling, that increases from $\bar{T}^{-0.73}$ for $n_{\text{phot}} = 4$ to Heisenberg limited for $n_{\text{phot}} \geq 20$, see the inset of Fig. 3.9b.

experimental overheads increase approximately linearly with n_{phot} , as emerging from Eq. 3.23. Notably, once the particles positions start collapsing towards a valid estimate of B , an exponential decrease in uncertainties occurs, as shown by Fig. 3.9b.

In the $n_{\text{phot}} \approx 1$ case, the final $\sigma(B_{\text{est}}) \approx 0.45 \mu\text{T}$ after 500 epochs, and $\bar{T} \approx 18 \text{ ms}$, that is $\bar{\eta} \approx 60 \text{ nTs}^{1/2}$. In the $n_{\text{phot}} = 20$ case, after ~ 150 epochs we have $\sigma(B_{\text{est}}) \approx 0.3 \mu\text{T}$, and $\bar{T} \approx 78 \text{ ms}$. The precision scaling in the latter case is essentially Heisenberg limited, with a final sensitivity $\bar{\eta} \approx 84 \text{ nTs}^{1/2}$ and 12.8 Hz repetition rates. This analysis has a favourable comparison against former results leveraging upon cryogenic setups [135]. Also, we find a consistent overlap with HL scaling for the scaling in η (from linear least squares fitting) for update rates up to 13 Hz. We stress that approximate Bayesian inference methods, as described in Sect. 3.2, keep computational overheads low. In turn, this favourable characteristic plays a crucial role in the precision performances.

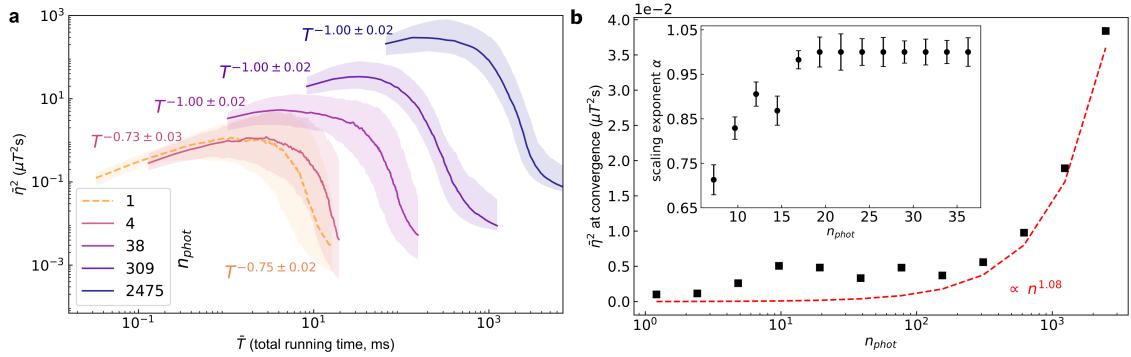


FIGURE 3.9. Experimental results for scaling of absolute precision. Lines represent median values, and performance within the 68.27% percentile range is shown as shaded areas. **a** The representative scaling of precision $\bar{\eta}^2$, inclusive of overheads, plotted against total running time \bar{T} . Cases with different n_{phot} (hence M , see main text) are colour-coded. Each protocol run for $n_{\text{phot}} > 1$ comprises $N = 150$ epochs, and only Poissonian noise is modelled in the likelihood function. For $n_{\text{phot}} = 1$, each run comprises $N = 500$ epochs and the model is in Eq. 3.24. **b** Final $\bar{\eta}^2$ and scaling exponent α for the absolute precision (inset), $\bar{\eta}^2 \propto T^{-\alpha}$, displayed for a set of different n_{phot} analysed. The red dashed line attempts a fit of the obtained datapoints with a power-law having n_{phot} as independent variable.

3.3.3 Multi-parameter estimation in MFL

Former sensing algorithms encountered a consistent degradation of the sensitivity, as soon as the procedure invoked times $\tau_i \gg T_2^*$, where no information can be retrieved and sensing time is essentially wasted, as observed e.g. in [222]. By learning T_2^* , we can ensure adaptively that all $\tau_i \leq T_2^*$, guaranteeing a worst-case SMS scaling for η , when the sensing becomes limited by decoherence.

Estimating T_2^* simultaneously with B might be beneficial for unstable systems, and can be achieved in a multi-parameter $\vec{x} = \{B, T_2^*\}$ learning strategy [97, 235]. MFL inherits the multi-

parameter estimation capabilities outlined for QHL methods. Here, we capture the uncertainty in the joint probability distribution $P(\omega, T_2^*)$ by a normalised covariance matrix Σ , and as discussed in Sect. 3.2.2, we pick $\tau \sim 1/\|\Sigma(P(\omega/w, T_2^*/t_2))\|_F$. The parameters $w = \gamma \max_{B:P(B) \neq 0} B$ and $t_2 = \min_{T_2^*:P(T_2^*) \neq 0} T_2^*$ are introduced to render $\|\Sigma\|_F$ dimensionless, with P the prior at epoch $N = 100$, when both parameters start to be learnt simultaneously. In particular, we use here $w = \gamma 11 \mu\text{T}$ and $t_2 = 20.2 \mu\text{s}$. A simultaneous learning of both parameters can incur an initial slow learning period, due to shorter τ_i being initially most useful in estimating ω (and thus B), whilst longer precession times are better for an estimation of T_2^* . In order to address this limitation, we start MFL by estimating B alone, and at epoch⁴⁰ $N = 100$ a simultaneous learning of T_2^* is introduced.

Figure 3.10 shows the outcome of a multi-parameter learning with MFL, using the β_1 data set⁴¹. The generalised uncertainty scales exponentially with the number of epochs, even if it can be observed a faster learning rate for B than for T_2^* . Interestingly, the estimate of T_2^* from MFL (3.10a) does not match the one from a (non-weighted least-square) fit to the decaying sinusoid shown in 3.10d. Such discrepancy is due to a preferential choice of $\tau_i < T_2^*$ via the PGH, so that the corresponding T_2^* estimate weights as more informative the data-points at shorter time scales. On the other hand, the fit estimates a shorter T_2^* because it mediates equally over all data-points, inclusive of those where almost no residual fringe contrast is visible.

3.3.4 Applications in magnetic field tracking

In a typical sensing experiments we expect a varying strength of the magnetic field with time (see Sect. 1.5). The Bayesian inference process is conceived to detect conflicting experimental likelihoods $P(E|\vec{x})$ and prior information, adjusting the latter on-line. Therefore, it is possible to leverage upon our protocol processing speed and adaptivity to achieve a successful tracking of time-varying magnetic fields. Introducing minor controls in the Bayesian inference procedure⁴², MFL not only accounts for a fluctuating B , but it can detect also stepwise changes that invalidate the credibility region, as estimated a-posteriori at any given epoch. Here, we test an algorithm that tracks a B_{set} -field using the ϵ_3 dataset, where B_{set} was experimentally modulated by changing the position of the permanent magnet (see Fig 3.5b). No data recording is performed as the magnet position is adjusted, so that transitions in B are stepwise (with changes up to ≈ 30 -fold), each intensity staying stable for a period of 100 – 1000 ms. The corresponding experimental dataset might well represent actual applicative scenarios. An example is the raster scan of magnetic nanoparticles functionalising a surface [224], as sketched in Fig. 3.11a.

⁴⁰ Chosen empirically.

⁴¹ Which has $\tau_{\max} > T_2^*$.

⁴² This scenario is sufficiently demanding, that it requires the standard CLE inference to be modified ad-hoc with additional heuristics. The latter ones are needed to detect when the sensed quantity has changed enough to (almost) completely invalidate the posterior at a certain epoch. CLE would naturally react to these changes, but without triggering immediately a resample event, it would first attempt Bayesian inference steps. These are not useful, because after a stepwise change, the new B has little support on the posterior, and a more aggressive update needs to be performed. Details of the modified inference procedure are provided as pseudo-code in Appendix 4.

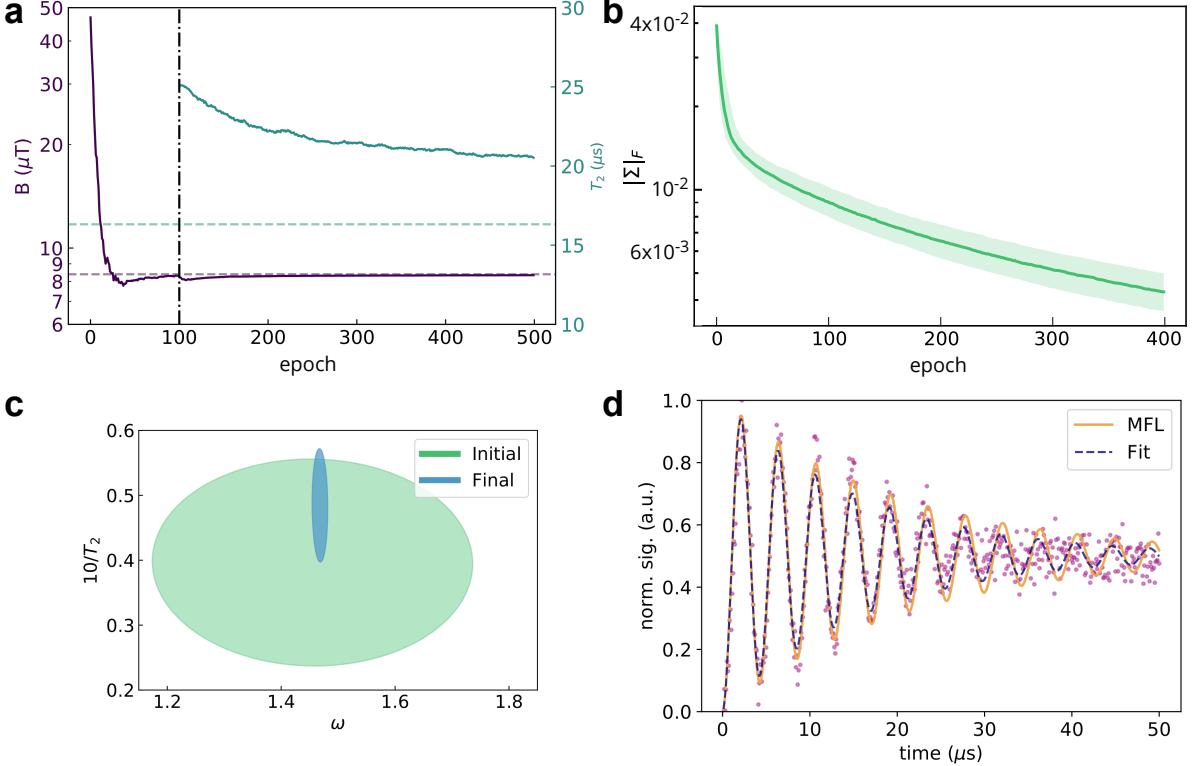


FIGURE 3.10. **a** Simultaneous estimate of magnetic field B (purple solid line) and decoherence time T_2^* (green solid line) for a single run. Simultaneous estimation occurs only for epochs higher than 100. Dashed lines are a LSF to experimental datapoints in (d). **b** Uncertainty in simultaneous estimates of B and T_2^* , represented as $\|\Sigma\|_F$, against epoch number. The median performance is shown as a solid line, with a shaded area representing the 68.27% percentile range. **c** 68.27% credible region at epoch 100 (green) and 500 (blue) for ω and (scaled, inverse) T_2^* . The smaller area of the region at epoch 500 indicates an improved estimate for both parameters. **d** Renormalised experimental PL data for a Ramsey fringe with $\tau_{max} > T_2^*$, along with a LSF and an MFL-learned decay function.

Results are shown in Fig. 3.11b–c. MFL detects when the latest data become non-representative of the posterior distribution, and via a complete resampling increases the uncertainty, $\sigma(B_{est})$. After approximately 10 epochs (observe the inset in Fig. 3.11c), the estimate converges to the new value set for B . In the case of FFT, new data-points are simply, cumulatively added to the dataset. FFT has performances similar to those in Fig. 3.6 during intervals where B is held constant. However, FFT estimates B from the prominent peak in the spectrum, i.e. the value observed for the longest, not the most recent times, failing to track changes as they occur. The analysis in Fig. 3.11c also emphasises the high dynamic range $B_{max}/\sigma(B_{est})$ achievable with the protocol, that for this dataset is as high as 12300. Figure 3.11(b) splits \bar{T} per epoch (≈ 10 ms) into the different computational and experimental contributions. The computational overhead

is $\tau^{comp} \approx 0.21$ ms, whereas experimental routines add $\tau^{exp} = 16.28$ ms and the average phase accumulation time is $\tau = 0.41$ ms. These estimates point to the computational efficiency of MFL as one of its strengths.

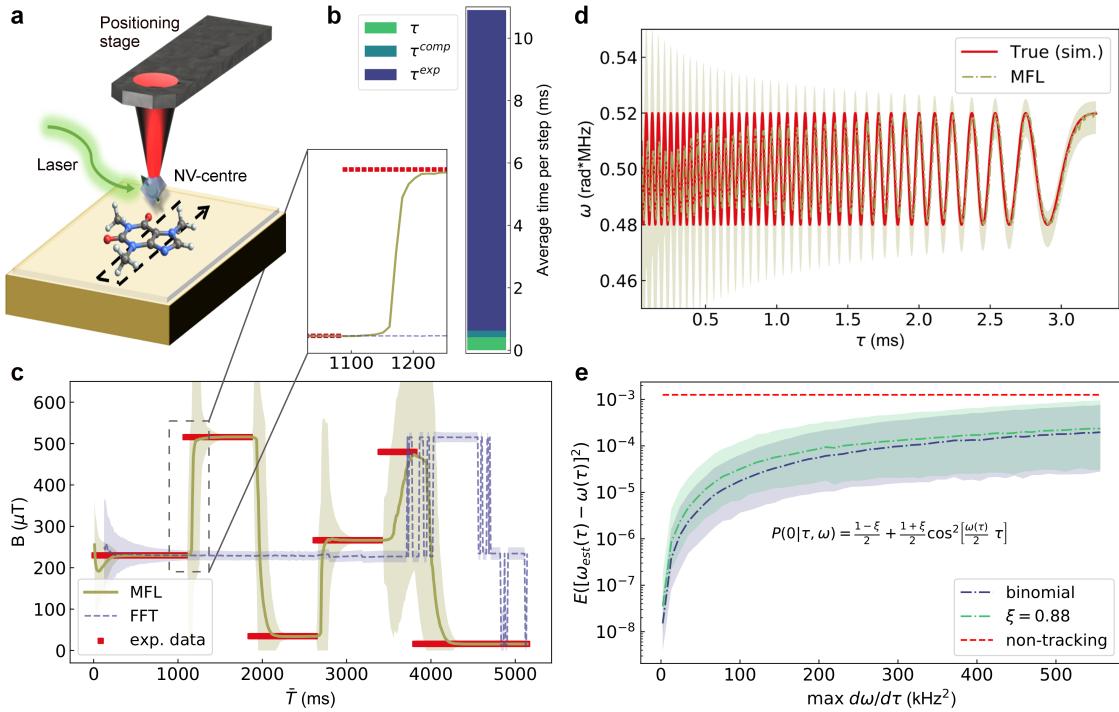


FIGURE 3.11. Magnetic field tracking. **a**, Pictorial representation of an NV-centre positioned at the end of a scanning microscope tip, scanning the magnetic field B in the proximity of a molecule absorbed on a substrate. **b**, Itemisation of the contributions to the average running time \bar{T} taken into account in (c): the precession time τ , computational (τ^{comp}) and experimental (τ^{exp}) overheads. **c**, Experimental demonstration of (offline) magnetic field tracking (here $M = 4000$). The intensity of B is reported in red bars. The solid golden line represents typical performances of MFL, whereas shaded areas represent the usual $\sim 68.27\%$ percentile range. For comparison, a dashed purple line indicates an FFT protocol applied cumulatively to all data available up to time \bar{T} , with the corresponding uncertainty from a Lorentzian fit as a shaded area of the same colour. Results using less than 10 data-points are omitted for FFT. **d**, Simulation of tracking a chirped sinusoidal magnetic field, with no experimental overhead and Poissonian noise alone (i.e. high-fidelity readout). The frequency ω is linearly increased after each update step. **e**, Numerical study of MFL performance in tracking sinusoidally time-dependent magnetic fields $B(\tau) = \omega(\tau)/\gamma$, under ideal conditions ($T_2^* = \infty$, $\tau^{exp} = 0$). The y-axis gives the nms_ω (see main text) in the Ramsey frequency estimate, against the peak speed at which B changes along each simulated Ramsey sequence ($\max d\omega(\tau)/d\tau$). The blue dashed line refers to the case including only binomial noise in $\mathcal{L}(B; \tau)$, while the green line is the case with limited readout fidelity ($\xi = 0.88$), as defined in [234]. A non-tracking strategy is reported for comparison as a red dashed line.

Figure 3.11d–e shows numerical results for the tracking of an A.C. oscillating field $B(\tau) = \omega(\tau)/\gamma$, with $\omega(\tau) = \omega_0 + w \cos(v\tau)$, constant v and $w \ll \omega_0$, so that the dynamic component has increasing frequency. As a figure of merit for the tracking, we adopt a time-dependent normalised squared error $\text{nms}_\omega := E[\omega_{\text{est}}(\tau) - \omega(\tau)]^2/\omega_0^2 = \sum_i^N (\omega_{\text{est}}(\tau_i) - \omega(\tau_i))^2/(N\omega_0^2)$. nms_ω is thus an average of the estimate error across all epochs. We obtain that average estimates differ for less than 3% from the actual value, for dynamic components up to 18 $\mu\text{T}/\text{ms}$.

3.4 Characterising an (almost) blackbox quantum device: the model learning agent

So far, by adopting QHL we always assumed a model Hamiltonian $\hat{H}(\vec{x})$, pre-selected in advance⁴³, to be trusted as the most appropriate in describing the system. Whenever knowledge of such a model is limited or uncertain, QHL can only flag an unsuccessful learning attempt by a non-shrinking uncertainty of the posterior distribution⁴⁴. At this point, improving the model is left to ad-hoc modifications from the user [130, 215], or to the application of ANNs for inferring it, from a pre-established set of local terms [236].

We can make this point more clear with an example. Assuming that two quantum systems (labelled 1, 2) can be described via a non-transverse Ising chain interaction such as $\hat{H}_{\text{Is}} := j_{12}\hat{\sigma}_1^z\hat{\sigma}_2^z$, QHL can give us information on the parameter j_{12} governing the interaction. However, it is not capable to identify a better modelling, e.g. inferring if an equivalent Heisenberg model $\hat{H}_{\text{Hs}} := \sum_{i \in \{x,y,z\}} j_{12,i}\hat{\sigma}_1^i\hat{\sigma}_2^i$ corresponds to a best approximation to the real system.

In this section, we describe how a *Quantum Model Learning Agent* (QMLA) can provide approximate, automated solutions to the inverse problem of characterising a quantum system, by inferring a Hamiltonian which is not provided by the user, nor necessarily known in advance. Therefore, QMLA is capable of relaxing the requirement of QHL for additional a priori information, when compared to QPT. At the same time, as it will be made more clear in the following, QMLA still attempts to learn the generator of the dynamics, an approach that we discussed already in Sect. 3.2 as advantageous for the efficient and complete characterisation of controllable quantum devices, which are key to digital and analogue quantum technologies [237, 238]. We define our method as a learning agent [12], because it uses its cumulative knowledge not only to decide which experiments to perform on the quantum system under study, but it is also capable of generating new models to be tested from a library of *primitives*, \hat{h}_p . In brief, QMLA combines a model search through a tree-like graph, and a systematic evaluation of models' performances. The performances are evaluated and compared in a Bayesian inference framework, using the outcomes from experiments designed by the agent as a feedback. The scheme for a *macrostep* in QMLA is shown in Fig. 3.12a–d. In the following, we present in detail each phase.

QMLA grows dynamically a *structural* Directed Acyclic Graph (sDAG)⁴⁵, whose nodes are each representative of a candidate Hamiltonian \hat{H}_i (Fig. 3.12a). The learning process thus involves a search through the sDAG, for the Hamiltonian that optimises the simulation of the target system \hat{H}' . In the realm of automated experiment interpretation, this is reminiscent of the graphic

⁴³See Def. 3.1.

⁴⁴This behaviour follows from Sect. 3.2.2, and will be shown in Figs. 3.13 & 3.14.

⁴⁵A DAG is a graph with directed edges, embedding no cycle - i.e. any path whereby following the directionality of the edges, $A \rightarrow B \rightarrow \dots \rightarrow A$ occurs, with A, B generic nodes in the graph. As it will be more evident from the following, in particular the sDAG is essentially a tree, with one minor exception: considering the way the QMLA generates new models, in most cases the sDAG has more than one initial root node (even if likely only one will be left *active* by the time the second layer of the sDAG is generated). Hence we prefer using the more general definition of DAG.

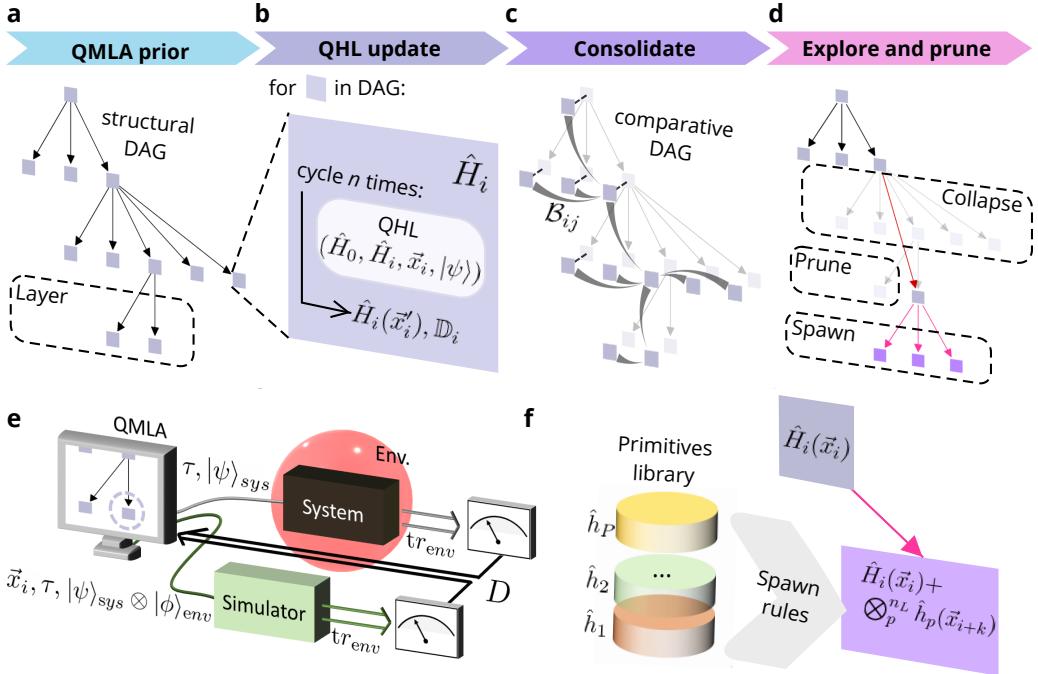


FIGURE 3.12. Schematic representation of a learning macrostep for QMLA. **a** Starting from a prior tree-like structure inclusive of all the test models already explored, **b** each model $\hat{H}_i(\vec{x}_i)$ proposed within the latest spawned layer undergoes a cycle of n QHL epochs, using probe states $|\psi\rangle$ from a set of random probes. Upon convergence, each model instance stores an updated parametrisation $\hat{H}_i(\vec{x}'_i)$, along with a set of results $\mathbb{D}_i = \{D\}_i$ from the experiments performed on the system. **c** The different models are compared in pairs $\{i,j\}$ using Bayes factors \mathcal{B}_{ij} , and the results stored as a comparative DAG. **d** The Bayes factors values are used to select the best models and remove those with less predictive power according to collapse and prune moves. A spawn move generates a new tree layer if appropriate, expanding the QMLA prior ahead of a new macrostep. **e** Schematics of a generic QHL instance. QMLA chooses random probe states $|\psi\rangle_{sys}$ and adaptive evolution times t for both the system and the simulator. In order to study the effect of the environment, the probes and dynamics of the simulator are extended accordingly (see main text). A projective measurement traces out the environment subspace and returns a binary outcome D for the Bayesian update. **f** Spawn rules in this work generate new nodes in layer L , by the *inheritance* from layer $L-1$ of the champion model $\hat{H}_i(\vec{x}_i)$. A single term parametrised by x_{i+k} is added, composing operators from the set of primitives $\{\hat{h}_p\}$ up to dimension $n_L = n_{L-1}$. If the models already explored saturated all those compatible with the constraints of dimension and spawn rules, then $n_L = n_{L-1} + 1$, and the inherited model: $\hat{H}_i(\vec{x}_i) \otimes \hat{I}$.

searches in [180, 216], and more in general, of Bayesian classification trees [239, 240]. However, unlike these former approaches, here all nodes themselves are representative *classification labels*⁴⁶, entertained by the learning process. Introducing one only category of nodes simplifies

⁴⁶ Bayesian classification trees expand upon *Classification and Regression Trees* (CARTs). In the latter ones, the

the visualisation of the sDAG. Leveraging the graphic structure, the search occurs in *layers*, whereby each layer μ includes models characterised by the same Hilbert space dimension n_μ , and the same cardinality of parameters $\dim(\vec{x})$.

3.4.1 The QMLA protocol

The pseudocode for the QMLA protocol is given in Algorithm 1 (we refer to *Lines* therein in the following text), whereas its subroutines can be found in Appendix A.4, Algorithms 5&6. The backbone of QMLA is to perform QHL, for each explored model $\hat{H}_i(\vec{x}_i)$, for a variable number of epochs N_i^{exp} until convergence (Lines 23-26), obtaining cumulatively a set of observed data \mathbb{D}_j from the quantum system, and updating the parameters \vec{x}_i accordingly (Fig. 3.12b,e and Lines 30-31). When launching QHL instances⁴⁷, we assume that the QMLA is able to access key experimental controls in the system to characterise: the evolution time τ and the initially prepared probe state $|\psi\rangle_{\text{sys}}$. As in QHL, we also require access to a trusted (quantum) simulator to (efficiently) test trial Hamiltonians forms \hat{H}_i and parametrisations \vec{x}_i [157, 213, 215]. However, so far we focused on characterising a closed system, entirely described by its full Hamiltonian \hat{H}_{glo}^* , on a Hilbert space \mathcal{H}_{glo} , by preparing input (and measuring output) states: $|\psi\rangle_{\text{glo, in (out)}} \in \mathcal{H}_{\text{glo}}$. This need not be the case for characterising an unknown quantum system, that might be open and therefore affected by the contribution of an environment in \mathcal{H}_{env} , so that the effective Hamiltonian can be written as [28]:

$$\hat{H}_{\text{glo}}^* = \hat{H}_{\text{sys}}^* + \hat{H}_{\text{env}}^* + \hat{H}_{\text{int}}^*, \quad (3.25)$$

with \hat{H}_{int}^* describing the interaction between system and environment.

In order to address this more general task, we introduce the following approach. First, in the rest of this Chapter we assume that we are able to prepare states $|\psi\rangle_{\text{sys}}$ in the quantum system independently from its environment, so that input states are separable, of the form:

$$|\psi\rangle_{\text{glo, in}} = |\psi\rangle_{\text{sys}} \otimes |\phi\rangle_{\text{env}}. \quad (3.26)$$

In general, the probe state will be a function of both the model instance tested and of the current epoch: $|\psi(\hat{H}_j, e)\rangle_{\text{glo, in}}$ (Line 24). In most cases, the environment will not be directly accessible⁴⁸, so that likelihood functions as in Eq. 3.2 are in general not achievable, and a preliminary partial

tree is explored according to observations represented at each leaf, whose outcome guides the branching. In discrete classification trees, the closest to QMLA, the final outcome classifies the observed “system” into one of a set of *labels*, dependent upon the tree path traversed [239]. The final label retrieved is reminiscent of QMLA’s suggesting the most likely Hamiltonian model. Bayesian CARTs add tree-building capabilities, by adjusting the partition probabilities for the predictor variables [240], a feature which has similarities with the tree-growing described later on for the QMLA. However, there are also substantial differences with Bayesian CARTs: the graphical model of QMLA represents candidate labels as nodes, and embeds the probabilistic inference for classification in an additional tree structure, so that edges in CARTs and QMLA have completely different meanings.

⁴⁷ Refer to Sect. 3.2.

⁴⁸ And in particular, this will be the case for the experimental implementation of QMLA discussed starting from Sect. 3.5.1.

trace over the environmental degrees of freedom needs to be taken into account. The likelihood for the Bayesian inference is thus here modified as:

$$\mathcal{L}(D|\hat{H}_{\text{glo}}^*, \tau) = \langle D | \text{tr}_{\text{env}} [\rho_{\text{glo}}(\tau)] | D \rangle_{\text{sys}}, \quad (3.27)$$

with $|D\rangle_{\text{sys}} \in \{|D\rangle_{\text{sys}}\}$ the chosen measurement basis for \mathcal{H}_{sys} , and

$$\rho_{\text{glo}}(\tau) = \exp(-i\hat{H}_{\text{glo}}^*\tau) |\psi_{\text{glo,in}}\rangle\langle\psi_{\text{glo,in}}| \exp(i\hat{H}_{\text{glo}}^*\tau) \quad (3.28)$$

the global density matrix evolved under each \hat{H}_{glo}^* .

As the algorithm progresses, the credible region for the parameters, starting from a suitable prior distribution, will be reduced in volume V [155] (Lines 21 & 27). An efficient way to halt the QHL subroutine for \hat{H}_j is to assign N_j^{exp} dynamically, by checking periodically for saturation in the reduction of V_j , as a signature of a converged Hamiltonian learning instance (Line 23 referring to the Algorithm in Appendix A.4). Once all models for the latest entertained layer have converged, a *consolidation* phase starts (Fig. 3.12c & Lines 32-35). In order to preserve the Bayesian framework in our protocol, we adopt here for a meaningful pairwise (\hat{H}_i, \hat{H}_j) comparison between models the *Bayes factor* (BF):

$$\mathcal{B}_{ij} = \exp[\ell(\mathbb{D}_{ij}|\hat{H}_i) - \ell(\mathbb{D}_{ij}|\hat{H}_j)], \quad (3.29)$$

where $\mathbb{D}_{ij} = \mathbb{D}_i \cup \mathbb{D}_j$ is the joint pairwise dataset, and

$$\ell(\mathbb{D}_{ij}|\hat{H}_i) = \sum_{D \in \mathbb{D}_{ij}} \log P(D|\hat{H}_i(\vec{x}_i)) \quad (3.30)$$

is the cumulative log-likelihood⁴⁹. BFs are known to be a statistically robust measure for comparing the predictive power of different models, whilst favourably weighting less structure to limit overfitting [241]. Eq. 3.30 captures such predictive power of the model, cumulatively for each datum $D \in \mathbb{D}_{ij}$, the tested dataset. Intuitively, overfitting is mitigated because the training dataset \mathbb{D}_{ij} is equal for both models in Eq. 3.29: a more complex model must train a higher number of parameters with the same abundance of data. BFs have been successfully used in the context of resolving multi-modal distributions in [216]. The resulting \mathcal{B}_{ij} are stored as a *comparative* Directed Acyclic Graph (cDAG) representation across the same nodes, where the edges' directionality maps the sign of $\mathcal{B}_{ij} - 1$, pointing towards the model favoured by statistical evidence (Fig. 3.12c & Line 35). If all BFs composing the graph were computed using the same dataset, i.e. $\mathbb{D} = \mathbb{D}_{ij} \forall (i,j)$, generating a DAG would be immediately granted by Eq. 3.29. This is not necessarily the case when $\mathbb{D}_{ij} \neq \mathbb{D}_{jk}$, which is most likely to happen if we run an independent instance of QHL per model. However, as we expect the graph to converge to a DAG as soon as enough statistical evidence is collected (i.e. once $\dim(\mathbb{D}_{ij}) \gg 1 \forall (i,j)$, as this should make

⁴⁹ The reason to adopt a logarithmic likelihood is here that likelihoods are prone to numerical instabilities as the number of sample measurement grows [155], so that adopting a log-difference instead of a ratio can help reducing artefacts.

differences across datasets negligible), we explicitly exclude cycles of ambiguous interpretation in the graph, by removing the edge which minimises $|\mathcal{B} - 1|$ for each accidental cycle. Performing this step and comparing systematically all model pairs leads naturally to the selection of a *layer champion*, corresponding to the node with highest indegree (Line 36).

An *exploration* stage follows, whereby a new layer ($\mu + 1$) is generated (Lines 14-18). Generating layers that progressively increase the complexity of the models entertained is a key feature of QMLA, as it affects directly the *interpretability* of the models. Indeed, we have already introduced recent works that aimed at the characterisation of e.g. quantum states via NN [80, 181, 187]. This strategy, when targeting an ideally black–box quantum device, has the limitation of employing implicitly *dimensionality reduction* schemes, to map the interactions characterising the system onto the possibly smaller set of weights between the neurons within the NN⁵⁰.

Now these kinds of mapping, if effective, are hard to be checked against intuition (or more quantitative predictors) that a human researcher might have or acquire about a system. As worries arise for the impact of artificial intelligence in the crisis of reproducibility affecting science [243], our intention is here to design an automated protocol bootstrapping the human characterisation process of a system, rather than replacing it. In a sense, a model learnt by a device to reproduce another device’s behaviour, that cannot be used to distillate general properties nor understanding of the first, advances us from an “untrusted black box” to a “trusted one”, without ultimately solving the characterisation problem. Therefore, we decided in favour of a protocol exploring models constructed from a reduced set of primitives, that are easily interpretable, and namely single-qubit Pauli operators $\hat{\sigma}_\alpha \in \mathbb{S}_P$, i.e. $\alpha \in \{1, x, y, z\}$ with $\hat{\sigma}_1 \equiv \hat{1}$. In fact, it is known that any n -dimensional Hamiltonian can be written as the sum of tensor products of σ_α [244]:

$$\hat{H}^* = \sum_{i,\alpha} h_{\alpha}^i \hat{\sigma}_{\alpha}^i + \sum_{i,j,\alpha,\beta} h_{\alpha\beta}^{ij} \left(\hat{\sigma}_{\alpha}^i \otimes \hat{\sigma}_{\beta}^j \right) + \dots + \sum_{i,j,\dots,n,\alpha,\beta,\dots,v} h_{\alpha\beta\dots v}^{ij\dots n} \left(\hat{\sigma}_{\alpha}^i \otimes \hat{\sigma}_{\beta}^j \otimes \dots \otimes \hat{\sigma}_v^n \right), \quad (3.31)$$

with Latin indices labelling the i –th subsystem on which operator σ^i acts. A generic layer μ can then be interpreted as encompassing potentially all truncations of Eq. 3.31, that include exactly μ terms (if the sDAG started with a root encompassing one term models). Most importantly, k –sparse, row–computable Hamiltonians⁵¹ $\mathbb{H}_{k\text{comp}}$ have been shown to admit a decomposition of the form in Eq. 3.31, with at most $\mathcal{O}(k^2)$ terms [245, 246]. The class $\mathbb{H}_{k\text{comp}}$ is surprisingly wide, as it covers the cases of electronic Hamiltonians in quantum chemistry⁵², as well as lattice model

⁵⁰ The closest equivalent of this approach, within our Bayesian framework, would be to assume the dynamics of the quantum system to be entirely described via the special unitary group $SU(n)$, with n known, or lower–bounded via the adoption of a *dimensional witness* [242]. Then, QMLA would be posed the task to learn which elements of the generator basis (e.g. the set Γ_n of generalised Gell-Mann complex $n \times n$ matrices γ_n) to include in the model, and with which parameters. That is, QMLA shall explore among subsets $\Gamma'_n \in \Gamma_n$ models of the form $\sum_{\hat{\gamma}_i \in \Gamma'_n} g_i \hat{\gamma}_i$ for the best Hamiltonian to reproduce measurements on the system. In this way, the new layer $\mu + 1$ would simply e.g. include an additional γ_n from the set.

⁵¹ That is, Hamiltonians whose matrix has at most k non–zero elements per row or column, the locations of said elements being retrievable with an efficient classical algorithm.

⁵² See Chap. 4.

Hamiltonians (such as Ising and Heisenberg lattices [160]). In all these cases, we can thus expect the depth of the *ideal* sDAG to grow polynomially in k , *independently* from the size of the system n , an observation which is key in estimating the worst case scaling of QMLA.

In general, we assume n not known in advance, even if dimensional witnesses might bootstrap the guess for the root model(s) of the sDAG. Therefore, QMLA is designed to preferentially introduce models \hat{H}_j in the *spawned* layer ($\mu + 1$) that embed additional terms as linear combinations of appropriate tensor products of primitives $\hat{h}_i \in \mathbb{S}_P$ (see Fig. 3.12f), exploring the same Hilbert space dimension n_μ as the consolidated layer μ . However, if μ has saturated the maximum number of independent operators allowed by n_μ (Line 18, can be replaced by user-defined rules), QMLA will generate a layer where $n_{\mu+1} = n_\mu + 1$. Now, even if we admit the system Hamiltonian $\hat{H}^* \in \mathbb{H}_{k\text{comp}}$, the exploration phase is still affected by a major scalability issue: if the nodes of the ideal sDAG might scale polynomially, this is not true for a sDAG that explores brute-force also other potentially valid models. This is most easily seen by studying a specific case, e.g. truncating the expansion in Eq. 3.31 to only include pairwise interactions, i.e. the first non-trivial terms, and also the most crucial for a wide variety of systems (see Chap. 1 & 4). The number of all possible terms Θ that can be generated in this instance scales as $\mathcal{O}\binom{n}{2}$, i.e. the sDAG depth grows combinatorially with the system's size. Additionally, also the maximum width of the DAG grows combinatorially, as the generic layer μ includes models with any of $\mathcal{O}\binom{\Theta}{\mu}$ terms' combinations. This intractability is well known in the field of structure learning via graphical models, and the most direct way to deal with it is the introduction of a *greedy* exploration phase [12, 247, 248]. That is, local optimisation is favoured against global optimisation, to keep the size of the global search DAG manageable.

In QMLA, the greediness in the search is imposed via the *inheritance* rule here adopted as the spawn rule \mathcal{R} (Line 15&18): all models in layer ($\mu + 1$) expand upon the Hamiltonian form of their common parent node⁵³, that is the champion node in μ (Fig. 3.12a&f). This greediness dramatically reduces the global number of models considered in the exploration. Adopting again the pairwise model, the overall number of nodes explored in the worst case⁵⁴ is expected to be downsized:

$$\mathcal{O}\left[\prod_{\mu=1}^{\Theta}\binom{\Theta}{\mu}\right] \rightarrow \mathcal{O}\left[\sum_{\mu=1}^{\Theta}\binom{\Theta}{\mu}\right] \quad (3.32)$$

when the inheritance rule is adopted. The huge improvement exemplified in Eq. 3.32 is clearly not enough to make QMLA tractable already for relatively small instances, and highlights the discrepancy between the expected polynomial scaling of terms and a search through a

⁵³ Eventually, after a dimensional matching if $n_{\mu+1} \neq n_\mu$. E.g. terms with the operator \hat{o}_z^1 will be interpreted as $\hat{o}_z^1 \otimes \hat{I}^2$.

⁵⁴ These estimates are intended as worst cases, because they describe the situation where: i) the correct model is the most complex one (so that the sTree construction is not halted, until all available models have been explored, see below), and ii) a priori information is completely absent, so that the tree root attempts the characterisation starting from the hypothesis that all the qubits in the system are isolated, which is not the case e.g. when bootstrapping the exploration with a dimensional witness.

combinatorially growing models' space, which evidently goes beyond the hypothesis of $\hat{H}^* \in \mathbb{H}_{k\text{comp}}$. However, the inheritance rule becomes a powerful heuristic, when combined with a priori knowledge about the system to characterise: we exemplify this with a particular case. A reasonable hypothesis is to take into account only (as first) nearest neighbours interactions, and assume that the index labelling is representative of neighbouring conditions. In this case, the first layer of the DAG chooses one out of n choices, and the overall models explored in the worst-case is $\mathcal{O}\left[\sum_{\mu=1}^n(n-\mu+1)\right] = \mathcal{O}(n^2)$. One can use the last layer of the learnt DAG to then progressively explore alternative hypotheses like index inversions, or longer-range interactions, without incurring straightforward in the intractable scaling as in Eq. 3.32.

Now that issues in the scaling of the sDAG nodes have been discussed, we are left with the scaling of the number of edges in the cDAG. If we were to compare each model \hat{H}_i against all others explored, the cDAG would be a *complete* graph, whereby the number of edges, each corresponding to a costly BF calculation, grows combinatorially with the sDAG depth [249]. In order to avoid generating overall a complete graph, the nodes within a layer form complete subgraphs of the cDAG (Lines 32-35), but the comparison among different layers occur in the QMLA only via their respective champions (Fig. 3.12 & Lines 40-43). In this way, all nodes other than the champion node for a given layer are essentially discarded from the learning process, as they stop being compared with any other competitive model (Lines 37-39). Additionally, if the comparison between the parent champion and a child champion \hat{H}_c leads to $\mathcal{B}_{pc} \gg 1$, a *collapse* rule reallocates the subtree rooted at \hat{H}_c (if any) to \hat{H}_p as a parent node, discarding the whole layer of \hat{H}_c (Lines 494-45). This collapse rule also aims at mitigating overfitting that might arise from the greediness introduced with the inheritance rule, as terms provisionally inherited might prove superfluous once the model is expanded, and would be discarded invoking a new instance of the inheritance rule after the collapse, against the new parent node \hat{H}_p . Inheritance rules can be seen as an extreme *pruning* rule for all the other nodes in μ (Fig. 3.12d), similarly to what adopted for a graphical model exploration in [216]. Therefore, another way to mitigate the greediness of the approach is to discard only those particularly unsuccessful models \hat{H}_r , that have: $\mathcal{B}_{rj} < b$, $\forall j$ in the same layer and b a user-defined threshold (Line 39). Intuitively, this kind of stratagems increases the explorative nature of the QMLA, along with the likelihood of approximating the global optimum, at the expense of the overall cost of the procedure.

A collapse event also offers naturally a *termination* condition for the growth of the sTree, because it signifies that more complex models are not justified by a better predictive power, given the observed dataset \mathbb{D} . One might terminate the protocol after a single collapse, again in favour of cost-saving local optima, or after a certain number of collapse events has been recorded, to prolongate the exploration⁵⁵.

In conclusion, the cumulative knowledge of the QMLA is represented as a multi-layer DAG (see Fig.3.12c), combining a sDAG tracking the generative process of new models and a cDAG

⁵⁵This latter option should be favoured whenever N^{exp} nor N^{prt} are chosen adaptively, because intuitively models with a lower $\text{dim}(\vec{x})$ will learn quicker a suitable parametrisation \vec{x}_0 , as we show for simulations in Fig. 3.14a

that embeds the information about how effective a given model is in replicating the experimental dataset, compared to neighbouring nodes in the cDAG. The efficiency in the QMLA's exploration of the models' space is intrinsically dependent from the extent to which an approximate solution to the characterisation problem can be deemed acceptable, and therefore how greedy towards local optimisation are the choices performed at each generation stage in the process. This trade-off resembles the limitations in the accuracy of the predictions that can be attained via graphical classification modelling [12, 247], versus their human interpretability [239].

Algorithm 1: Quantum Model Learning Agent (QMLA) protocol. Subroutines Converged, BfBatchUpdate, DAGupdate, FindChampion, CollapseLayer in Appendix A.4.

```

1: Input:  $\{\hat{h}_k\}$ , a set of primitives, i.e. operators used to construct each instance  $\hat{H}(\vec{x})$ .
2: Input: a set of spawn rules  $\mathcal{R}$  dictating the construction of each successive layer
3: Input: user defined thresholds  $b_c, b_f, b_e > 1$  to trigger pruning operations
4: Input: user defined  $\epsilon, N_{min}, N_{max}$  to check convergence in parametrisations learning
5: Input: user defined maximum allowed Hilbert space dimension  $n_{max}$  and sTree depth  $d_{max}$ 
6: Input: a heuristic  $f : \text{dim}[\vec{x}] \rightarrow N_{prt}$ , the number of particles to be used in the QHL instance [213, 215].
7: Input: a prior distribution  $\mathcal{F} \sim P(\vec{x}) \forall \hat{H}(\vec{x})$  initialised model.
8: Input: The resampling parameters  $a_{res}$  and  $t_{res}$ .

9:  $n \leftarrow 1, d \leftarrow 0$        $\triangleright$  initialise the spawned Hilbert space dimension and explored sTree depth
10:  $\{\hat{H}_i\}_T \leftarrow \emptyset$      $\triangleright$  initialise the explored models as the empty set

11: function QMLA ( $\{\hat{h}_k\}, \mathcal{R}, \mathcal{F}, f, b_f, a_{res}, t_{res}, n_{max}, d_{max}$ ) :
12:   while  $n < n_{max}$  and  $d < d_{max}$ :
13:      $d \leftarrow d + 1$ 
14:     if  $\{\hat{H}_i\}_n \setminus \{\hat{H}_i\}_T \neq \emptyset$ :     $\triangleright$  check not all available models for  $n$  have been explored
15:        $\{\hat{H}_i\}_\mu \leftarrow \mathcal{R}(d, n, \{\hat{h}_k\})$      $\triangleright$  populate new layer  $\mu$  with all applicable models
16:     else:
17:        $n \leftarrow n + 1$ 
18:        $\{\hat{H}_i\}_\mu \leftarrow \mathcal{R}(d, n, \{\hat{h}_k\})$      $\triangleright$  otherwise explore higher dimensional models

19:     foreach  $\hat{H}_i(\vec{x}_i) \in \mu := \{\hat{H}_i\}_\mu$ :
20:        $N^{prt} \leftarrow f(\text{dim}[\vec{x}_i])$      $\triangleright$  assign an appropriate  $N^{prt}$  for the QHL subroutine
21:        $\vec{s} \leftarrow (V(\mathcal{F}_i(\vec{x}_i)))$      $\triangleright$  initialise uncertainties' array according to prior volume

22:      $e \leftarrow 1$ 
23:     while Converged ( $\vec{s}, \epsilon, N_{max}, N_{min}$ ) is False:
24:       do prepare  $|D\rangle_{sys}$   $\triangleright$  prepare the probe state
25:          $\triangleright$  call QHL to obtain posterior  $P'(\vec{x})$ 
26:       do QHL (  $|D\rangle_{sys}, \hat{H}_i(\vec{x}_i), N^{prt}, N^{exp} = 1, \mathcal{F}_i, a_{res}, t_{res}$  )  $\rightarrow D_e(\tau_e), P'_i$ 
27:       do append  $V(P'_i(\vec{x}_i))$  to  $\vec{s}$ 
28:        $e++$ 
29:     end while
```

```

30:    $\mathbb{D}_i = \{D_e(\tau_e)\}_i$   $\triangleright$  record the dataset  $\mathbb{D}$  for  $i$ -th model
31:    $\vec{x}_i^* \leftarrow \mathbb{E}[P'_i(\vec{x})]$   $\triangleright$  extract an estimate  $\vec{x}^*$  from the posterior  $P'(\vec{x})$ 
32:   foreach  $\hat{H}_i(\vec{x}_i^*), \hat{H}_j(\vec{x}_j^*) \in \mu$ , with  $j > i$ :  $\triangleright$  each couple of models in layer  $\mu$ 
33:      $\triangleright$  obtain Bayes Factor  $\mathcal{B}_{ij}$  updating the learnt parametrisations
34:      $\mathcal{B}_{ij} \leftarrow \text{BfBatchUpdate } (\hat{H}_i(\vec{x}_i), \hat{H}_j(\vec{x}_j), \mathbb{D}_i, \mathbb{D}_j, P'_i(\vec{x}_i), P'_j(\vec{x}_j), f)$ 
35:     cDAG  $\leftarrow \text{DAGupdate } (\text{cDAG}, \mathcal{B}_{ij})$   $\triangleright$  add directed edges of weight  $\mathcal{B}_{ij}$ 
36:    $\hat{H}^\mu \leftarrow \text{FindChampion } (\text{cDAG} \cap \mu, b_c)$   $\triangleright$  identify layer champion
37:   foreach  $\hat{H}_k \in \mu \setminus \hat{H}^\mu$ :  $\triangleright$  activate intra-layer pruning rule(s) for sTree
38:     remove  $(\hat{H}^{\mu-1}, \hat{H}_k)$  from sTree  $\triangleright$  champion rule, can be replaced by
39:     (remove  $(\hat{H}^{\mu-1}, \hat{H}_k)$  from sTree iff  $\mathcal{B}_{\mu k} > b_f$   $\triangleright$  flooring rule)
40:    $\triangleright$  activate inter-layer comparison against  $\mu$ 's parent layer
41:    $(\hat{H}_i, \hat{H}_j) \leftarrow (\hat{H}^{\mu-1}, \hat{H}^\mu)$ 
42:    $\mathcal{B}_{ij} \leftarrow \text{BfBatchUpdate } (\hat{H}_i(\vec{x}_i), \hat{H}_j(\vec{x}_j), \mathbb{D}_i, \mathbb{D}_j, P'_i(\vec{x}_i), P'_j(\vec{x}_j), f)$ 
43:   cDAG  $\leftarrow \text{DAGupdate } (\text{cDAG}, \mathcal{B}_{ij})$ 
44:    $\triangleright$  check if any layer in the sTree is collapsible
45:   sTree, cDAG  $\leftarrow \text{CollapseLayer } (\text{sTree}, \text{cDAG}, b_e)$ 
46: end while
47:  $\hat{H}^0 \leftarrow \text{FindChampion } (\text{sTree} \cap \{\hat{H}^\lambda\}, \text{cDAG})$   $\triangleright$  identify the tree champion
48: return  $\hat{H}^0$ 
49: end function

50: Output: the most successful model  $\hat{H}^0$  in describing the cumulative set of observations  $\{\mathbb{D}_i\}$ 
 $\forall i \in \text{sTree}$ , according to the Bayes factors  $\{\mathcal{B}_{ij}\}$  stored in the cDAG.

```

3.5 Experimental CVV of NV-centres and their environment via Hahn-echo sequences

In this section, we provide extensive testing of the QMLA procedure against an experimental case, to check the behaviour of the protocol in a realistic case. We complement the analysis with contextual simulations of equivalent models, in an ideal case, to highlight crucial differences.

A first observation of how inaccurate models affect the outcome of a QHL procedure occurred during the experimental demonstration of CLE and QLE, obtained in our group [215]. In this experiment, the scope was to analyse Rabi oscillations⁵⁶, to extract an estimated Rabi frequency ω_0 as a characterisation and verification task for a negatively-charged NV-centre electron spin. The progress of the CLE learning procedure against experimental data is shown in Fig. 3.13. Here, we choose to display as a figure of merit the s.d. of the posterior distribution $\sigma(\omega)$. For a single-parameter case, we remind how $\sigma^2(\omega)$ equates the Bayes risk at each update, and $\sigma^2(\omega) \propto V[P(\omega)]$,

⁵⁶ Introduced in Sect. 1.5.

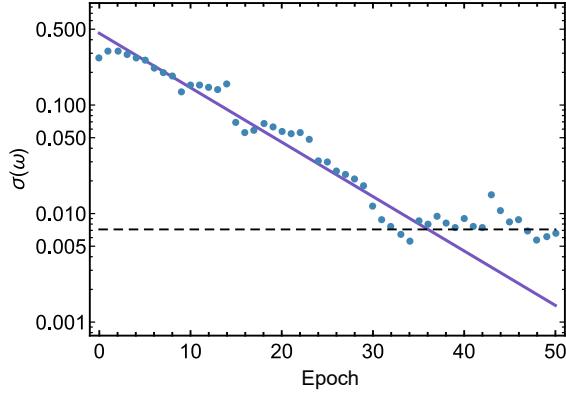


FIGURE 3.13. Observed saturation in the reduction of the uncertainty $\sigma(\omega)$, for a single-parameter experimental QHL instance. Here ω is the Rabi frequency obtained via QHL from experimental Rabi oscillations (Sect. 1.5), measured with the electron spin of an NV–centre [215]. Uncertainty data from a single experimental run as dots, linear fit of the data at epochs 1 – 35 as a solid blue line, and at epochs 36 – 50 as a dashed line.

the volume of the distribution. It appears evident how the exponential suppression of the risk expected from Eq. 3.17 plateaus after a certain number of experiments. This artefact is a signature of the inaccuracy of the model introduced to interpret the data⁵⁷, here given by the Hamiltonian $\hat{H}_{\text{Rabi}} \equiv f_{\text{Rabi}} \hat{\sigma}_x / 2$, with f_{Rabi} the Rabi frequency. Overcoming this limitation was possible in [215], by adopting a phenomenological time-dependent Hamiltonian to model the observed *chirp* in the Rabi frequency. However, this ad-hoc approach does not investigate any further the form of the effective Hamiltonian (thus ultimately the nature of the interaction with the bath), and reduces to verify the effectiveness of the chirped Hamiltonian modelling.

Motivated by these preliminary results, achieved for QHL by approximating the NV–centre as an isolated system, we decided to use QMLA to learn the fundamental terms of the Hamiltonian describing the same NV–centre system (Sect. 1.5), now interacting with the surrounding nuclei of its spin-bath. In particular, we use a bulk diamond sample, with a ^{14}N defect in the carbon lattice. Therefore, we expect the main interaction for the electron spin (here the *system*) to occur with the ^{14}N donor nucleus, and surrounding ^{13}C atoms [131, 132], that can be considered the environment, or *spin bath*, for the system. These system–bath interactions are all known to be well described by hyperfine terms, according to the Hamiltonian \hat{H}_{nv} , as introduced in Eq. 1.47. The main research question that we address in this section is thus if QMLA can identify an appropriate form of the hyperfine interaction automatically, using the greedy search introduced in Sect. 3.4, through Hamiltonians expanded according to the general form in Eq. 3.31, and access to a class of experiments addressing the NV–centre. We start by discussing the terms in \hat{H}_{nv} . Within the QMLA framework, this amounts to an analysis of what constitutes preliminary information about the system’s dynamics, and what instead we aim to learn automatically.

⁵⁷ The same effect will be later observed in Fig. 3.16b.

The isolated electron–spin dynamics is captured by the first two terms. Here, we choose to omit the zero-field splitting ($\Delta_{\text{gs}}\hat{S}_z^2$), as this provides a constant shift on the fine electron states, not contributing directly to the interaction which is our main interest to study. Moreover, the zero–field splitting is in the GHz range (Sect. 1.5), so that its dynamics can be averaged out, in the ms timescales of the experiments in this section, by adopting the rotating frame of $\Delta_{\text{gs}}\hat{S}_z^2$. For a well–characterised NV–centre setup, and a wide range of experiments⁵⁸, it is customary to assume the external magnetic field well aligned perpendicular to the z –axis of the electron spin, so that only the term $\gamma_e B_z \hat{S}_z$ is retained [250, 251]. However, this presupposes a characterisation which is the task of QMLA, and therefore we will keep also terms on the (x, y) –axes when exploring the sDAG of trial Hamiltonians.

The crucial term, as anticipated, is the hyperfine interaction between the electron spin, and the nuclear spin of each atom species X at lattice site ζ [131]. As per Eq. 1.47, this interaction is described in terms of the electron (nuclei) spin operators⁵⁹ $\hat{\mathbf{S}}_{\text{sys}}$ ($\hat{\mathbf{I}}_{\text{env}}$). In order to avoid including degenerate terms⁶⁰ in the trial Hamiltonians \hat{H} , here we replace the expansion:

$$\hat{S} \cdot \sum_{X,\chi} \mathbf{A}_{X,\chi} \cdot \hat{I}_{X,\chi} \rightarrow \hat{S} \cdot \mathbf{A} \cdot \hat{I} \quad (3.33)$$

that is, all nuclear contributions to the electron spin dynamics are mapped onto a single *environmental* qubit. Eq. 3.33 retains the original hyperfine term, whilst conveniently requiring a lower computation time in classical simulators⁶¹ and splitting the evaluation of $\dim(\xi)$ into a separate learning problem (see Sect. 3.5.1). These advantages come at the cost of introducing finite–size effects in the bath, that will be discussed in Sect. 3.5.2 and more in detail with Appendix B.4. Finally, we expand the term on the left of Eq. 3.33 as⁶²:

$$\hat{\mathbf{S}} \cdot \mathbf{A} \cdot \hat{\mathbf{I}} \simeq A_{\parallel} \hat{S}_z \hat{I}_z + A_{\perp} (\hat{S}_x \hat{I}_x + \hat{S}_y \hat{I}_y) + A_{xz} (\hat{S}_z \hat{I}_x + \hat{S}_x \hat{I}_z) \quad (3.34)$$

$$\simeq A_{\parallel} \hat{S}_z \hat{I}_z + A_{\perp} (\hat{S}_x \hat{I}_x + \hat{S}_y \hat{I}_y) \quad (3.35)$$

$$\simeq A_{\parallel} \hat{S}_z \hat{I}_z \quad (3.36)$$

with A_{\parallel} (A_{\perp}) the (non)axial parameters [252], and A_{ik} the off–diagonal parameters of the hyperfine tensor. The latter are often neglected [133, 250] as in Eq. 3.35, because \mathbf{A} is diagonal in the point–dipole approximation [253]. The off–diagonal term appearing in Eq. 3.34 describes nuclear spins coupled along the principal axis of the NV defect, which is expected to be the strongest coupling due to the system’s symmetry [251]. Finally, Eq. 3.36 is a further simplification

⁵⁸E.g. in Ramsey interferometry, studied in detail with Sects. 3.3.1 and following. Interferometric protocols are explained in Sect. 1.5.

⁵⁹In the following we will omit the “sys, env” notation wherever evident, to simplify the expressions.

⁶⁰Note how the correlations among the $\{A_{X,\chi}\}$ parameters here would not arise if we had access to the full output $|\psi\rangle_{\text{glo,out}}$, because the various $\{\hat{I}_{X,\chi}\}$ act upon different subspaces. However, the effect of each site upon the traced–out final $\rho_{\text{sys,out}}$ depends only on the hyperfine parameter characterising the corresponding interaction.

⁶¹As all simulations will involve not more than 2 qubits, instead of the ~ 15 estimated later to actually compose the nuclear bath.

⁶²Tensor product symbols omitted for clarity.

including only the axial contribution, obtained by invoking the 0th order approximation of the Floquet theory [250, 251], also known as the *secular approximation* [254].

Finally, we decide here to drop completely the bath dynamics *per se*, i.e. both the precession in the external magnetic field ($\gamma_n \mathbf{B} \cdot \hat{\mathbf{I}}$) and the contributions from the quadrupole splitting (PI_z^2). Again, the aim of this first QMLA application is to characterise the system via a suitable bath model, rather than the opposite⁶³. Therefore, it is reasonable to adopt here the a priori approximation⁶⁴ of a stationary bath, when decoupled from the system. A stationary bath is a typical working-hypothesis adopted in the realm of open quantum systems to be valid within the relaxation timescale of the system [28]. Also, for PI_z^2 similar considerations to those outlined for the zero-field splitting apply, with the exception of the quadrupole splitting being in the MHz range (Sect. 1.5).

| $\dim(\hat{H}_{\text{glo}})$ (qubits) | Operator form | Abbreviation |
|--|--|---|
| 1 | \hat{S}_α | S_α |
| 1 | $\sum_{i \in \{\alpha, \beta\}} \hat{S}_i$ | $S_{\alpha, \beta}$ |
| 1 | $\sum_{i \in \{x, y, z\}} \hat{S}_i$ | $S_{x, y, z}$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1}) + \hat{S}_\alpha \otimes \hat{I}_\alpha$ | $S_{x, y, z} Y_\alpha$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1}) + \sum_{i \in \{\alpha, \beta\}} (\hat{S}_i \otimes \hat{I}_i)$ | $S_{x, y, z} Y_{\alpha, \beta}$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1} + \hat{S}_i \otimes \hat{I}_i)$ | $S_{x, y, z} Y_{x, y, z}$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1} + \hat{S}_i \otimes \hat{I}_i) + \hat{S}_\alpha \otimes \hat{I}_\beta$ | $S_{x, y, z} Y_{x, y, z} T_{\alpha\beta}$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1} + \hat{S}_i \otimes \hat{I}_i) + \sum_{i, j \in \{\alpha\beta, \gamma\delta\}} (\hat{S}_i \otimes \hat{I}_j)$ | $S_{x, y, z} Y_{x, y, z} T_{\alpha\beta, \gamma\delta}$ |
| 2 | $\sum_{i \in \{x, y, z\}} (\hat{S}_i \otimes \hat{1} + \hat{S}_i \otimes \hat{I}_i) + \sum_{i, j \in \{xy, xz, yz\}} (\hat{S}_i \otimes \hat{I}_j)$ | $S_{x, y, z} Y_{x, y, z} T_{xy, xz, yz}$ |

Table 3.2: List of all models explored in the QMLA implementation, along with the corresponding number of qubits and the abbreviation used in the main text. Greek indices here do not include the case of the identity, $\hat{1}$.

We observe how dropping some terms from the usual form of the Hamiltonian is here a fundamental step in ensuring that QMLA is implemented efficiently, compromising against its immediate interpretability. As the first, the decomposition in Eq. 3.31 ensures that some of the terms would lead to degeneracies⁶⁵ in the Hamiltonian, as e.g. $\hat{S}_z^2 \in \mathbb{S}_P$. These limitations can be addressed by clustering methods (see the footnote at pg. 126), at the cost of introducing additional computational overheads to the procedure. Moreover, including degenerate terms increases the cardinality of the parametrisation $\dim(\vec{x})$, thus demanding more particles to make the learning stable⁶⁶ without introducing any advantage in the modelling. Such degeneracies are easy to verify in the models of Sects. 3.5.1 & 3.5.3 which have $n = 2$, therefore we deemed appropriate to remove degenerate terms a priori. Importantly, in systems with n big enough to

⁶³This will be instead the scope of Sect. 3.5.2.

⁶⁴This approximation will be confirmed also a posteriori, from the excellent agreement with experimental datasets for the NV-centre studied, as well as from observations about the behaviour of $V[P(\vec{x})]$ during the learning, Fig. 3.16b.

⁶⁵Now intended strictly in the meaning of Def. 3.2.

⁶⁶See Sect. 3.2. The number of particles is crucial in the photonic implementation outlined hereafter in Sect. 3.5.3.

render intractable the linear algebra of the corresponding matrices, an a priori guess for \hat{H}^0 is not expected to be available, and this preliminary check would play no role.

In conclusion, our brief analysis of the system's expected Hamiltonian suggests how different experiments might highlight terms, that can be safely neglected otherwise, by invoking some of the various approximations introduced. Therefore, the corresponding characterisation of the NV–centre and surrounding nuclear spins, using QMLA, is a study case of non–trivial interpretation. In the following sections, we will test the capabilities of the protocol to automatically build a model \hat{H}^0 for the experimental dynamics of the system, using a single environmental ancillary qubit. In Table 3.2 we list all the model instances that can be generated according to methodology described in Sect. 3.4, and the a priori information discussed in this Section, along with a brief naming convention adopted for each instance, to ease the reporting of the results hereafter. Also, we will discuss how \hat{H}^0 provides insight about which, among the contributions to Eq. 3.34, suffice to provide a statistically motivated model.

3.5.1 QMLA with a classical trusted simulator

In a first test of QMLA, we decided to adopt *Hahn peak* experiments⁶⁷, and analyse them invoking a classical device as the trusted simulator for each CLE instance. The choice of the experiment typology is due to our focus on modelling hyperfine interactions. In Rabi and Ramsey interferometry experiments, the action of the nuclear environment amounts essentially to decoherence phenomena and chirping in the Rabi (Ramsey) frequency. However, both cases are known to admit excellent phenomenological descriptions⁶⁸, making a first–principle analysis less compelling. On the contrary, data from Hahn peak experiments stress the role of the spin bath in the fast decoherence occurring when $\tau_1 \neq \tau_2$, as well as the particular contribution of weaker terms in Eq. 3.34. The latter are emphasised when $\tau_1 \sim \tau_2$, in correspondence of the peaks in the data, because then the Hahn sequence protects against drifts along the Bloch sphere equator of the evolved electron state $|\psi\rangle'_{\text{sys}}$ (see Fig. 1.10). Indeed, such drifts are primarily due to the $\hat{S}_z \hat{I}_z$, if assuming negligible off–diagonal terms in the \mathbf{A} tensor. A typical experimental outcome of a Hahn peak sequence is displayed in Fig. 3.16d, and used for the results in this section. The plot shows the expectation value $P_0 \equiv P(0 | H_{\text{glo}}^*; \tau)$ for the electron spin to be in $|0\rangle_{\text{sys}}$ at the end of the Hahn sequence⁶⁹, against evolution times τ . As described in Sect. 1.5, P_0 can be efficiently extracted from the normalised photoluminescence (PL) signal recorded from the NV–center.

⁶⁷ The operations on the NV electron spin for each of the experiments quoted in this section, along with the required control sequences, are illustrated in Sect. 1.5. We refer there for all definitions and notation.

⁶⁸ See in particular Sect. 3.3.1 for the dephasing time, and [215] for chirping.

⁶⁹ Refer to Eq. 3.27. For all the experiments in this chapter, we assume no experimental control over any of the nuclear spins. Therefore, the final measurement involving exclusively the electron spin corresponds effectively to tracing out the evolved $|\phi\rangle'_{\text{env}}$. The unknown true Hamiltonian for the combined system–environment, H_{glo}^* generically described in Eq. 3.25, is here believed to be approximated by \hat{H}_{nv} .

As a simulator, here we invoke a classical device, and specifically the *BlueCrystal* cluster⁷⁰ (BCc), running an implementation of QMLA in Python language, using an Anaconda installation, customised with several free-software packages [232] to perform efficient linear algebra operations (including evolution of quantum states), statistical inference, task parallelisation, and plotting.

Before moving to the analysis of experimental data obtained from the NV–setup, described in Sect. 1.5 , we first perform a sanity check, testing that the QMLA methodology can retrieve a Hamiltonian of the form $S_{x,y,z}Y_{x,y,z}$, along with its multi–parameters $\vec{\omega}$ and \vec{A} :

$$\hat{H}_{\text{test}}^*(\vec{\omega}, \vec{A}) \equiv \sum_{\alpha \in \{x, y, z\}} (\omega_\alpha \hat{S}_\alpha) + \sum_{\alpha \in \{x, y\}} (A_{\perp, \alpha} \hat{S}_\alpha \hat{S}_\alpha) + A_{\parallel} \hat{S}_z \hat{I}_z, \quad (3.37)$$

when searching through a tree generating (potentially) all models listed in Table 3.2. $S_{x,y,z}Y_{x,y,z}$ was chosen, because it is representative of the experimental case, by corresponding to a model inclusive of the electron spin precession in external fields, as well as the hyperfine interaction approximated according to Eq. 3.35. Note, it allows to verify both *underfit* as well as *overfit*⁷¹ in the model \hat{H}^0 learnt by QMLA, compared with \hat{H}_{test}^* , because according to the growth rules imposed for the sTree, $S_{x,y,z}Y_{x,y,z}$ is going to be generated at layer 6, independently⁷² from the performances of models belonging to layers 1–5. In order to collect sufficient statistics also for the deepest layers in the cTree, for this implementation of QMLA we do not include the termination rule outlined before, and we let the protocol exhaust all available layers in each run.

The choice of the probe state to dial before the evolution, $|\psi\rangle_{\text{glo, in}}$, plays here a crucial role (see Sect. 3.4). In simulations, we have complete access over both the system and its environment, so that we can in principle choose $|\psi\rangle_{\text{glo, in}}$ optimally. However, preparing with high–fidelity the global system of an NV–centre and its surrounding spins in an arbitrary state is challenging (see Sect. 1.5 and 4.5.4). In order to test how these experimental limitations can impact upon the success of the protocol, we run the simulations with a set of different probes, labelled according to:

$$|+r\rangle \equiv |+\rangle_{\text{sys}} \otimes |\psi\rangle_{\text{env}} \quad (3.38)$$

$$|++\rangle \equiv |+\rangle_{\text{sys}} \otimes |+\rangle_{\text{env}} \quad (3.39)$$

$$|rr\rangle \equiv |\psi\rangle_{\text{sys}} \otimes |\psi\rangle_{\text{env}}, \quad (3.40)$$

where we adopted the two–qubits approximation in Eq. 3.37, ‘sys’ refers to the NV electron spin, ‘env’ to the (single) environmental nuclear spin, $|+\rangle_{\text{env}} \equiv (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$ and $|\psi\rangle$ is intended

⁷⁰ This is provided by the University of Bristol as a remote computing service. Here in particular we refer to the *phase 3* unit. The simulations were run parallelising CLE instances on BCc’s base blades, each provided with “16 x 2.6 GHz SandyBridge cores, 4GB/core and a 1TB SATA disk” [255].

⁷¹ We here intend intuitively as underfitting models, those lacking some terms that are instead present in the true \hat{H}^* (known with certainty in a simulation). Therefore, named μ^* the layer including \hat{H}^* , overfitting models belong to layers $v < \mu^*$, closer to the root of the sTree. On the contrary, overfitting models will include terms absent in \hat{H}^* , and occur in layers $v > \mu^*$.

⁷² This is made evident by Fig. 3.18.

to be a random 1-qubit state. $|+r\rangle$ closely mimics the likely outcome of the initial preparation sequence for Hahn peak experiments in Fig. 1.10: only the electron spin is fully addressed and prepared constantly as $|+\rangle$ to keep a simple experimental setup, whereas surrounding nuclear spins are expected to be randomly distributed [254]. $|rr\rangle$ assumes that arbitrary preparation⁷³ of the electron state is available, thus removing any bias towards those terms in the Hamiltonian, to which a $|+\rangle_{\text{sys}}$ system probe is particularly sensitive. In this way, $|rr\rangle$ provides a benchmark for a QMLA protocol run under ideal conditions. Finally, $|++\rangle$ is motivated by both numerical and experimental reasons. The first is the observation that running QHL over probes that are varied continuously appear to consistently reduce the learning rate of the parameters⁷⁴. On the contrary, the choice of a static probe is a worst-case during the sTree exploration, as it biases the model learning towards some instances.

We emphasise that measurement-induced backaction to prepare nuclear spin states (Sect. 1.5) can lead inadvertently to polarisation of both ^{13}C [254, 256] as well as the host ^{14}N nuclear spins [250]. E.g. using a Ramsey-type sequence can efficiently polarise the ^{14}N spin as $|\uparrow\rangle$ conditionally on P_0 , as long as the free-evolution time is carefully chosen as $\tau_F \simeq \pi/A_{\parallel}$, which is $\mathcal{O}(\mu\text{s})$ [250]. In our experiment, the free evolution times τ_1, τ_2 are changed across a range of few μs , so that a consistent polarisation across different epochs should not be expected. However, this could occur instead for experiments addressing a single ^{13}C nuclear spin in weak external magnetic fields: the optimal τ_F to map the electron state on the nuclear state is $\mathcal{O}(\text{ms})$ for this system [254]. Therefore, along evolution times of few μs , the environmental spins might be prepared in a constant $|\phi\rangle_{\text{env}}$, $\forall e$ epochs. The $|++\rangle$ probe is therefore intended to investigate the behaviour of QMLA in such instances.

The essential figures of merit for this simulated test are intended to verify independently:

1. The learning of a correct parametrisation $\vec{x} \equiv (\vec{\omega}, \vec{A})$ for \hat{H}_{test}^* – i.e. validate the single CLE instances, and particularly the one for $S_{x,y,z}Y_{x,y,z}$;
2. The successful retrieval of the correct model $S_{x,y,z}Y_{x,y,z}$ – i.e. validate QMLA overall.

As described in Sect. 3.2, the signature of a successful QHL run is given by an exponential reduction of the estimated uncertainty for the values of each parameter, which in turn can be estimated via the SMC approximation and the volume of the particles representing the posterior, $V(P'(\vec{x}))$. In Fig. 3.14a we report $V(P'(\vec{x}))$ for a significative subset of models generated in the sTree. In particular, we report beside the correct one, a slightly underfitting model, $S_{x,y,z}Y_{x,y}$, and two models including superfluous terms, $S_{x,y,z}Y_{x,y,z}T_{xz}$ and $S_{x,y,z}Y_{x,y,z}T_{xy,xz,yz}$. All models display an exponential reduction of $V(P'(\vec{x}))$ during the epochs of the respective QHL instances, however it is interesting to highlight two behaviours. Simpler models benefit initially from a reduced number of parameters to learn, and therefore their learning rate is higher. However,

⁷³ This amounts to have access to arbitrary MW signal generators.

⁷⁴ Results omitted for brevity.

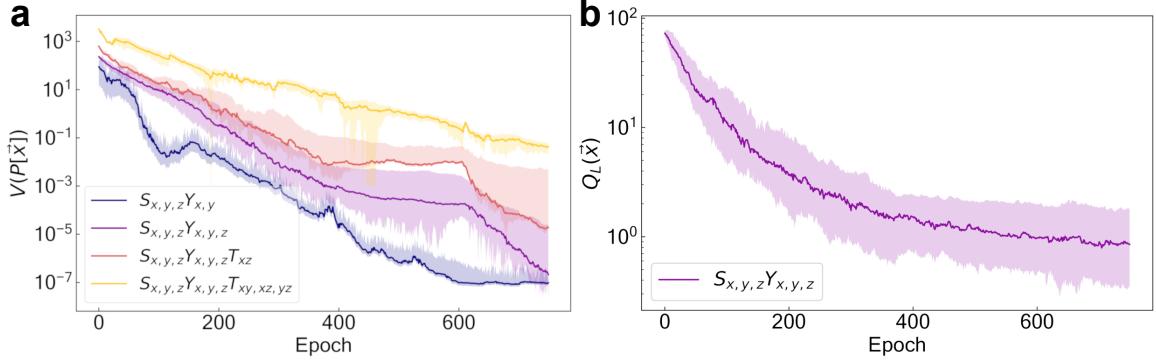


FIGURE 3.14. Simulated testing of QMLA. Shadowed areas represent the 68.27% credible region. **a** Volume of the posterior for the selected subset of models explored by QMLA, plotted at each epoch of the corresponding QHL instance in the sTree. Volumes are estimated from the SMC approximation, Eq. 3.10. **b** Quadratic loss $Q_L(\vec{x})$ for the *correct model* $S_{x,y,z}Y_{x,y,z}$, chosen as the target of the simulated QMLA run, at each of the epochs as in **a**.

underfitting models eventually saturate (in this case for epochs beyond 600), once the limits in the form of the Hamiltonian overcome any further parameter optimisation. Saturation is not observed, instead, for the correct nor overfitting models, as they are both capable of reproducing appropriately the true system dynamics. As a simulated test, here we have certainty about the correct model along with its parametrisations, therefore we can also check that \vec{x} is being learned exponentially fast. Therefore, we display $Q_L(\vec{x})$ in Fig. 3.14b for the correct model⁷⁵ $S_{x,y,z}Y_{x,y,z}$. From the same plot, we observe the appearance of a slowdown in the learning rate, due to a cap in the longest available evolution time that can be requested from the quantum system. This artefact has been discussed at length at pg. 94. Here, we imposed a cap at $3\tau_{\max} \approx 14\mu\text{s}$, in order to provide simulations with a consistent advantage compared to the experimental test⁷⁶, but at the same time not unrealistically benefiting from an unlimited decoherence time.

The overall QMLA performance in simulations is instead extracted running 100 independent instances of QMLA, each leading to the generation of a single cDAG similar to the one displayed in Fig. 3.16a. Spawning the corresponding sTree is non-deterministic: according to the times chosen via the heuristic, and the outcomes of the measurements \mathbb{D}_j , the computed \mathcal{B}_{ij} can differ in various QMLA instances, leading to slightly different sTrees as well. This aleatory behaviour is emphasised in Fig. 3.18a, where we report a *cumulative* cDAG. Here, the colour of each edge (i,j) maps the median \mathcal{B}_{ij} obtained across the runs, whereas the edge thickness is proportional

⁷⁵ One might extend the definition of $Q_L(\vec{x})$ for an overfitting model \hat{H} , by imposing that any term in \hat{H} , parametrised by x_i , that does not appear in \hat{H}^* must satisfy $x_i = 0$. On the contrary, there is no equivalent definition of a single metric such as $Q_L(\vec{x})$ for underfitting models. The value of the parameters for terms that are actually present in \hat{H}^* may be altered, as an artefact of invoking a wrong Hamiltonian in the first instance [155]. Therefore, here we decide to adopt $Q_L(\vec{x})$ only when we know that $\hat{H} = \hat{H}^*$.

⁷⁶ Compare for example with experimental data displayed in Fig. 3.16d.

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

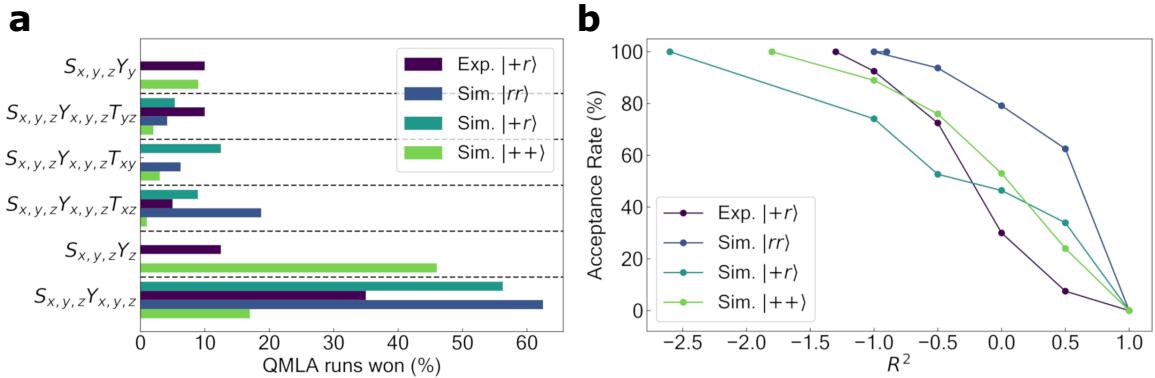


FIGURE 3.15. Experimental QMLA with NV-centres. The cases referring to the various probe states (Eqs. 3.38–3.40) are colour-coded in the legends, reporting both numerical simulations (*Sim.*) as well as using experimental Hahn peak datasets from the NV-centre (*Exp.*). **a** Instances won per model, out of 100 independent QMLA runs, displayed in percentage. Only the 6 most successful (overall) models are reported. **b** Number of QMLA runs that would be accepted as correct (in percentage), against the chosen threshold level for the coefficient of determination R^2 . The value of R^2 is obtained comparing the dynamics predicted by the model identified as correct via QMLA, against the simulated (experimental) data, similarly to Fig. 3.16d.

to the frequency with which that edge has been generated. Therefore, nodes connected via thin edges represent models that are generated less frequently, and hence their exploration is not favoured statistically. Also, the colour-scale highlights how statistically significant a comparison is, with light-blue corresponding to the least decisive comparisons, and viceversa.

The final models suggested as the best in reproducing the data via QMLA are summarised in Fig. 3.15a, for the various $|\psi\rangle_{\text{glo,in}}$ introduced in Eqs. 3.38–3.40. We can observe how in simulations, both $|+r\rangle$ as well as $|rr\rangle$ lead to success rates respectively above 55% and 60%, with a final *information gain*⁷⁷ $\mathcal{I} > 2.5$ in both cases. Overfitting is limited to models including a single term not present in the correct \hat{H}^* , such as $S_{x,y,z}Y_{x,y,z}T_{xz}$. Simpler models with a single hyperfine term are instead not represented in the histogram for simulations with such probes. On the contrary, a simplistic $|++\rangle$ probe state is not sufficient to distinguish among some of the models and leads to consistent underfitting. This limitation is shown analytically for some of the \hat{H}_j in Appendix C.4. Finally, we adopt an alternative, frequentist metric to further comment on the capabilities of QMLA. We use the R^2 , calculated for a subset of datapoints $\{P(0|H_{\text{glo}}^*; \tau_i)\}$ in correspondence of a linearly spaced set of evolution times $\{\tau_i\} \in [0, \tau_{\max}]$, compared with the predicted dataset adopting the learnt champion model, $\{P(0|H_{\text{glo}}^0; \tau_i)\}$. In Fig. 3.15b, we display

⁷⁷ We remind that in this case, $\mathcal{I} = S' - S_0$, the difference between the Shannon entropy at the end of the learning $S' = -\sum_i p_i \log p_i$ (with i the model index and p_i its likelihood), and its corresponding start value S_0 . At the beginning, we pick $p_i = 1/N^m$, the inverse of the maximum number of permissible models in the tree (N^m), as we assume no bias towards any of the models that can be generated.

how many of the champion models obtained in QMLA would be accepted, if we established the threshold R^2 value reported on the x -axis. Two observations are interesting. First, QMLA is much more efficient than a naïve strategy, attempting to simply achieve high values of $R^2 > 0.9$. The latter invalidates a vast majority of runs that identified instead the correct model. Second, the behaviour for different probes is somewhat counter-intuitive. From Fig. 3.15a we noticed how choosing a less biased probe such as $|+r\rangle$ is advantageous over $|++\rangle$ in retrieving the correct model. However, in Fig. 3.15b it also appears how a static probe leads to a higher learning rate, and therefore to achieving a better agreement of the learnt models with the dataset. We believe this enhanced performance to occur because of an artefact: the training of QHL instances is facilitated by optimising over a single probe $|\phi\rangle_{\text{glo,in}}$, rather than a set of probe states. This artefact was mitigated in simulations by adopting a finite set of random states, here with $\dim(\{|r\rangle_i\}) = 20$. Then, at each epoch the experiment is performed using one of the states $|r\rangle_i$ in a rota (refer also to Fig. 3.12e).

Now that we have gained insight in the performances of QMLA in controlled simulations, a meaningful discussion of the protocol's behaviour when operated using realistic experimental data is possible. The learning is performed offline over the Hahn peak dataset in Fig. 3.16d (red dots), with a $\tau_{\max} \simeq 4 \mu\text{s}$. Each datapoint is affected by the noise sources already discussed in Sect. 3.3.2, here mitigated by adopting $M = 1000000$. Even if a proper error bar for each datapoint cannot be precisely estimated here⁷⁸, the fluctuations in P_0 – when the signal is completely decohered – make evident how for times $\tau \sim \tau_{\max}$, the maximum error in the normalised PL signal could be as high as $\Delta P_0 \sim 0.1$.

We start commenting the differences in Fig. 3.15a with simulated cases. Interestingly, QMLA is highly successful in discarding one-qubit models, that are never retrieved as the best model in any of the 100 independent runs (see Appendix Fig. B.3 & B.4 for a complete list). $S_{x,y,z}Y_{x,y,z}$ is found as the champion model by the QMLA search in a clear majority of the cases ($> 30\%$), but this is less evident than simulated cases. Indeed, the average information gain out of the search is here $\mathcal{I} \simeq 1.15$. Competitive underfitting models appear among the sTree champions, and in particular $S_{x,y,z}Y_z$ and $S_{x,y,z}Y_y$. There are several possible explanations for this. Eq. 3.35 can be well approximated via the secular approximation as Eq. 3.36, which justifies the appearance of $S_{x,y,z}Y_z$ as a likely champion model. At the same time, residual polarisation of the nuclear spins might instead lead to selecting $S_{x,y,z}Y_y$, as for simulations with $|++\rangle$. Finally, QMLA converging to Hamiltonians with transverse hyperfine terms $S_{x,y,z}Y_{x,y,z}T_{\alpha\beta}$ is of uncertain interpretation: QMLA could either be sensitive to the small off-diagonal hyperfine contributions, or these finds might be an artefact of slight overfitting, as already observed in simulations adopting $|+r\rangle$ probe states.

Additional information can be extracted from Fig. 3.16. In subfigure a, we report an exemplary search through the cDAG, here leading to $\hat{H}^0 \equiv S_{x,y,z}Y_z$ being selected as the final champion

⁷⁸ A precise estimate was not possible because only the averaged outputs were recorded in this experiment, instead of the PL signal out of each sequence, as was done e.g. in the MFL experiments.

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

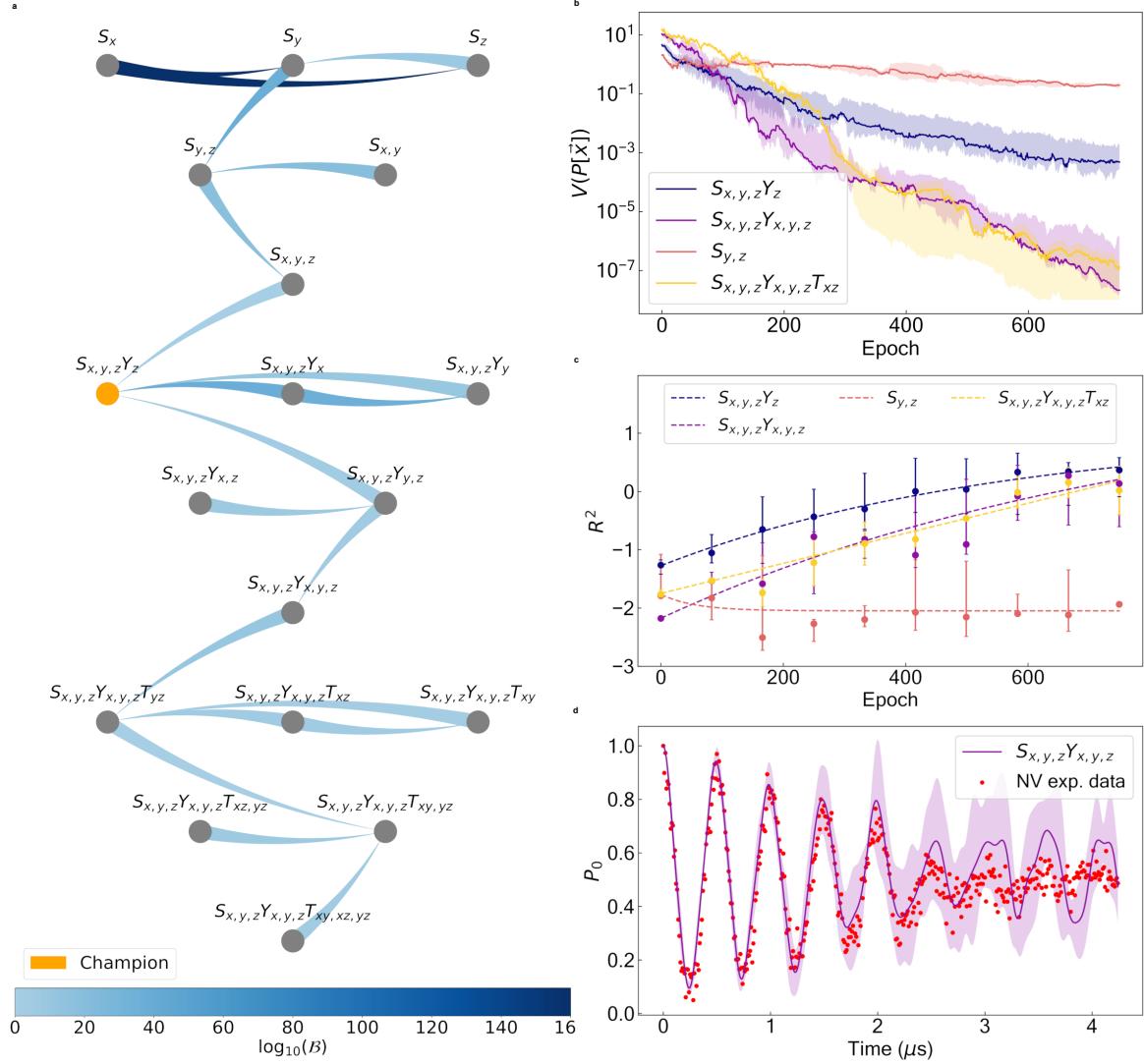


FIGURE 3.16. Experimental QMLA with NV-centres and a classical simulator. **a** The output cDAG at the end of a single QMLA run. Hamiltonian models are represented as dots (with labels referring to Table 3.2). The model selected by the run is colour-coded in yellow. Edges colours represent the corresponding \mathcal{B}_{ij} as from the legend at the bottom. **b** Evolution across epochs of the median particles' volumes $V(P[\vec{x}])$ for a selection of models. **c** Evolution across epochs of the agreement, as estimated by the R^2 , between the entire experimental dataset in (d) and predicted values for $P_0(\tau)$ by each colour-coded model, with parametrisation $\mathbb{E}[\vec{x}]$, at epoch e . **d** Experimental datapoints obtained performing a Hahn-peak sequence as described in the main text (red dots). Values predicted by the median learnt parametrisation, for the most successful model (see Fig. 3.15), are as a solid violet line. Shadowed areas in (b&d), and error bars in (c) represent the 68.27% credible region, as obtained upon completion of 100 independent QMLA runs.

model. In this case, where the learning is performed against the experimental dataset, it appears evident how \mathcal{B}_{ij} are very significative when comparing the simplest models, but get progressively less clear, as soon as the search moves through to terms including at least a hyperfine terms. Even if $\mathcal{B}_{\hat{H}^0, \hat{H}_j} \gg 1$, $\forall \hat{H}_j$ spawned from \hat{H}^0 , the evidence is not enough to completely collapse the layer. The same is true in other cases, such as $S_{x,y,z}$, where the corresponding layer is worse than both the parent and child layers. Observing the volumes for a subset of models in Fig. 3.16b, we observe how the exponentially fast learning of the parameters proceeds across all epochs for all models more complex than $S_{x,y,z}Y_z$. Too simple models, such as $S_{y,z}$, have a very poor learning rate and their $V(P(\vec{x}))$ saturates quickly. Models effective such as $S_{x,y,z}Y_z$, instead, register mostly a slowdown in the learning rate, eventually saturating at the final epochs in our range. As there is no such sign e.g. for $S_{x,y,z}Y_{x,y,z}$, more than an artefact due to the bounds in τ_{\max} we interpret this as a sign of $S_{x,y,z}Y_z$ not being completely sufficient to capture the actual \hat{H}^* .

An analysis complementary to the one regarding the volumes concerns the times suggested by the heuristics, to perform a meaningful experiment at each epoch. In Sect. 3.2.2 we have introduced the PGH as an effective solution to the problem of identifying an appropriate τ_i to evolve the system at each experimental step. However, performing QMLA, the speedup in the learning rate for simpler models must be taken into account, when performing comparisons across layers in the cDAG. We exemplify the different learning rates in Fig. 3.17a, containing the distribution of PGH generated times, for the model champions in the same exemplary run displayed in Fig. 3.16a. Evidently, when learning a single parameter the maximum number of epochs allocated here ($N^{\exp} = 750$) is sufficient to saturate the experimental τ_{\max} available, but this is not the case for the models with the highest number of parameters. Such a behaviour can exacerbate the bias towards simpler models when calculating edges in the cDAG: as each \mathcal{B}_{ij} descends from \mathbb{D}_{ij} (see Sect. 3.4), the more complex model will be exposed to a range of evolution times that it did not get the chance to be trained upon. In a machine learning wording, the two models compared will have dramatically different training and test sets, which is expected to lead to poor performances. A way to address this is to assign a variable number of epochs, as described in Sect. 3.4, however this was impractical with our implementation⁷⁹. Therefore, we prefer here instead to partially lift the PGH approach in the experiment choice. That is, after the training adopting PGH for N^{\exp} epochs, we include an additional training phase with linearly spaced times in $[\tau_{N^{\exp}}, \tau_{\max}]$, for a total of $N^{\exp'}$ epochs, before starting the consolidation phase. The new distribution of times for an exemplary run, for layer–champion models, is reported in Fig. 3.17b.

In Fig. 3.16c–d we analyse instead the performances of the learnt models, in reproducing the experimental datasets. Subfigure c shows how the learning progresses across epochs, with

⁷⁹ This is because the access to the BCp3 cluster requires to have an accurate estimate for the running time of each task queued for execution. A dynamic allocation of N^{\exp} would render it difficult to predict in advance the running time, with corresponding jobs being terminated before completion.

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

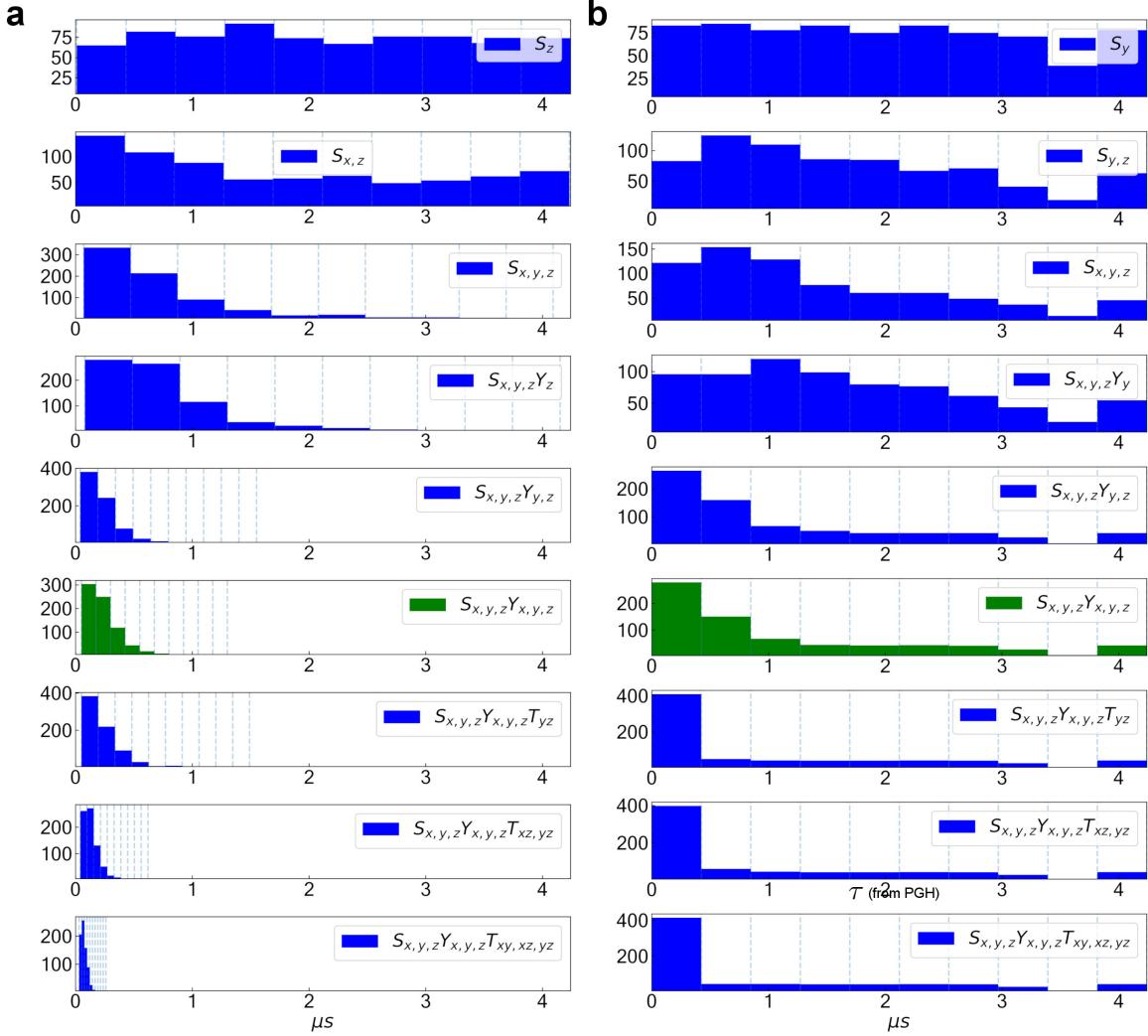


FIGURE 3.17. Experimental QMLA with NV-centres. We report the distributions of evolution times, along a total of $N^{\text{exp}} = 750$ epochs, as generated by the PGH approach alone (a), or adjusted via a linearly spaced additional training phase (b), see also main text. The model instances for each case are shown in each legend.

R^2 improving substantially for trial \hat{H}_j instances, that are highly successful among searches. The poor performances of models unsuccessful in reducing the prior volume $V(P(\vec{x}))$ is here confirmed by a poor final fit with the experimental data. Even if the values of R^2 after the learning is performed might not appear high, they are to be put in context of a highly non-linear optimisation process. Fig. 3.16d makes this evident, by plotting the evolution of the predicted $P_0(\hat{H}^0, \tau)$ for the most successful model out of QMLA searches with NV-centre data, $S_{x,y,z}Y_{x,y,z}$. A vast majority of experimental points lie within the 68.27% credible region, and the remaining ones are still reasonably well approximated. It also appears evident how the decohered portion of

the Hahn peak data is actually contributing the most in reducing the final R^2 . The challenge in reproducing correctly the noisy dynamics for times close to τ_{\max} appears also from the plot in Fig. 3.15b, as a vast majority of runs gets rapidly invalidated as soon as the threshold R^2 is increased, regardless of the champion model chosen. Such failure in using fit performances alone to infer an appropriate model shows qualitatively the superiority of our approach.

Finally, in Fig. 3.18b we report the cumulative cDAGs obtained averaging 100 independent QMLA instances, using the same experimental dataset discussed above. Comparing with the simulated run adopting $|\psi\rangle_{\text{glo,in}} \equiv |+r\rangle$ in subfig. a, we highlight how the BFs obtained for the earliest single-qubit layers generated are less pronounced. At the same time, the stronger preference towards hyperfine models that include the Y_z term can be interpreted both as Eq. 3.36 providing a reasonable approximation of the dynamics, as well as the bias introduced by possibly imperfect preparation of the probe states $|+r\rangle$. Finally, if overfitting appears reliably ruled out in the simulated case, as the chosen $S_{x,y,z} Y_{x,y,z}$ has no outgoing edges, in Fig. 3.18b we observe instead how the comparison against $S_{x,y,z} Y_{x,y,z} T_{xz}$ tends to be less stable, which might point again to the various degrees of approximation in Eq. 3.34. Details about the numerical stability in the values of the BFs retrieved by QMLA for this experimental run is reported in Appendix Fig. B.4.

For brevity, we skip a discussion about the values retrieved for the chosen parametrisation $(\vec{\omega}, \vec{A})$ in Eq. 3.37. This is because retrieving parameter values belongs to the exquisite realm of QHL, already discussed at length, and also the mapping in Eq. 3.33 renders $(\vec{\omega}, \vec{A})$ less immediate to interpret.

3.5.2 Environment characterisation via hyperparametrised QMLA

The analysis so far has focused on mapping the action of the NV-centre environment upon a 2-qubit Hamiltonian, retrieving the most effective model \hat{H}^0 , in reproducing the experimental statistical evidence. This approach has left aside a more in-depth study of the bath, which was mapped onto a single environmental qubit. Such an approximation is valid for the rapid dynamics of the fast decay in Hahn peak experiments, however the effects of a finite-size bath are expected to be evident⁸⁰ at the longer times probed by Hahn signal experiments [28, 134]. Therefore, we now intend to investigate how to implement an automated characterisation of Hahn signal experiments, based upon the QMLA principles, and in particular we aim to learn the number of nuclear spins in the environment n^s , coupled to the NV electron spin. The experimental dataset used for this analysis is reported in Fig. 3.19b.

A straightforward extension of QMLA to characterise in detail the bath, via the full set of parameters $(\{\vec{\omega}\}_{X,\chi}, \{\vec{A}\}_{X,\chi})$, thus dropping the approximation in Eq. 3.33, however, poses challenges. First, the number of interacting nuclear spins has been estimated to be potentially in the order of 1000s, via a priori knowledge of NV-centre devices [134]. QMLA requires a complete

⁸⁰ This point is discussed more in detail in Appendix B.4.

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

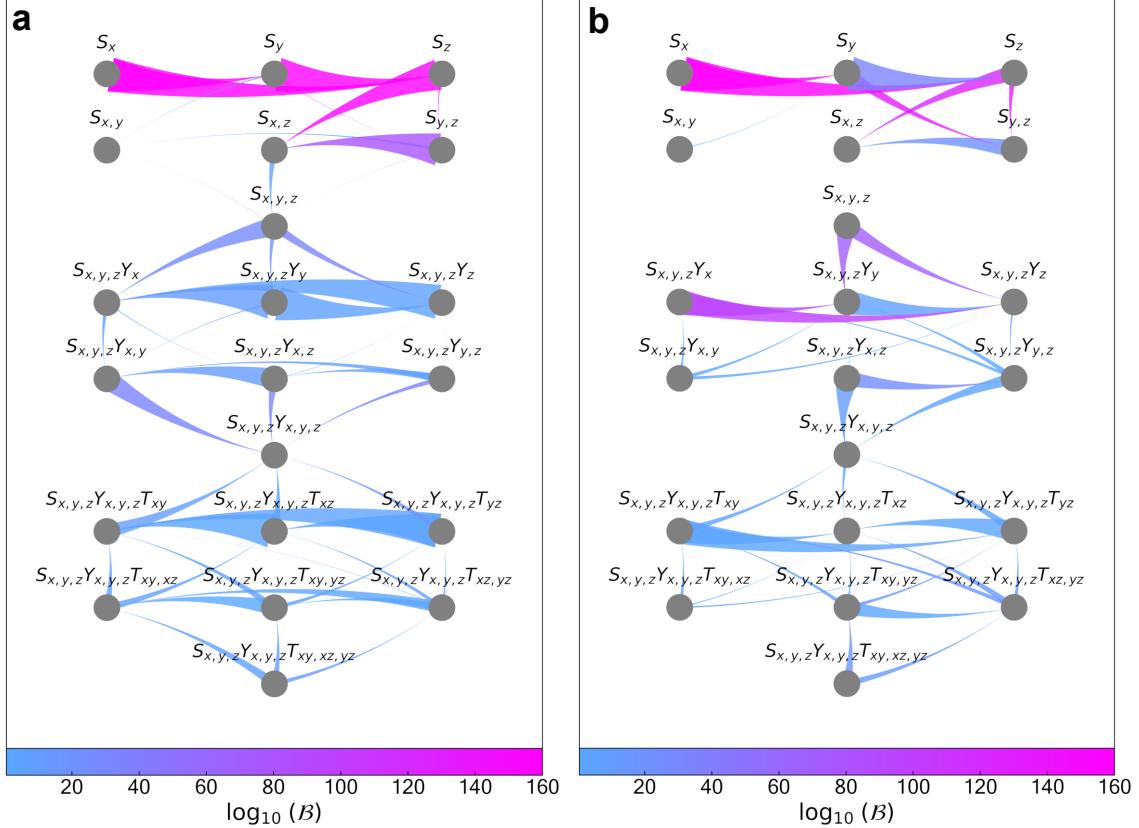


FIGURE 3.18. cDAG searches in QMLA, displayed as average cDAGs obtained from 100 independent instances of QMLA. For each edge (i, j) , the thickness is proportional to the number of times \mathcal{B}_{ij} has been estimated across the 100 runs. Colour-coded and legended in the bar at the bottom, the median value of \mathcal{B}_{ij} , obtained averaging all Bayes factors at the end of each QMLA instance, where (i, j) had been generated in the cDAG. Additional details in the main text. **a** QMLA run with simulated data of a typical Hahn peak sequence, adopting the probe state $|+r\rangle$. **b** QMLA run against experimental NV-centre data, and the same probe in the simulator.

quantum evolution⁸¹ of the (simulated) global system, to accurately learn its dynamics. When adopting the BCp3 classical simulator of Sect. 3.5.1, this limits effectively the practicality to about ten-qubit systems. Even if further experimental studies have highlighted how only a very small fraction (~ 10) of those sites can be observed interact with the NV electron spin [131], using this information would defy the purpose of an automated characterisation of the bath structure.

In order to address the classical simulability issue posed by CLE, in this Section we decide to move from generating models out of the Pauli primitive set (Eq. 3.31), in full generality, to a dedicated analytical approach. The idea is that, once an interpretable model for the system has been suggested via the QMLA approach described so far, one can bootstrap further analyses, leveraging upon the information already obtained about the system. Using the approach in Sect. 3.5.1, we have confirmed how the NV–centre dynamics is effectively described via the hyperfine interaction with a spin bath. For Hahn-signal experiments within such a system, an analytical model is known [134], describing the hyperfine coupling of the electronic spin to each χ^{th} nuclear spin, as a combined two-qubit system $\hat{H}_{e-\chi}$, physically equivalent to each contribution in Eq. 3.33. The unitary evolution originated from $\hat{H}_{e-\chi}$ can be analytically solved [257] to retrieve the outcome of Hahn-echo signal experiments, in terms of an effective magnetic field \mathbf{B}_χ at each site:

$$\Pr(0|\tau; \hat{H}_{e-\chi}) = 1 - \frac{\Pi_\chi(S_\chi) + 1}{2}, \quad (3.41)$$

via the *pseudospins*:

$$S_\chi = 1 - \frac{|\mathbf{B}_0 \times \mathbf{B}_\chi|^2}{|\mathbf{B}_0|^2 |\mathbf{B}_\chi|^2} \sin^2(\omega_0 \tau/2) \sin^2(\omega_\chi \tau/2), \quad (3.42)$$

with \mathbf{B}_0 the external magnetic field and $\omega_{0(\chi)}$ the (bare) Larmor frequency.

Qualitatively, Eq. 3.42 describes how the ground state of $\hat{H}_{e-\chi}$ precesses at a rate ω_0 , whereas the excited states incur in a splitting ω_χ . In free evolution, after the initial $\pi/2$ pulse of the Hahn sequence, the nuclear and electron spin become progressively more entangled at a rate dictated by the hyperfine interaction, getting maximally correlated at times $\tau \propto \pi/A$ [250]. When the two spins get disentangled again, revivals can be observed in the experimental Hahn-echo signal (see Fig. 3.19b). Decays and revivals in the PL signal $\Pr(0|\tau; \{\mathbf{B}_\chi\}, \{\omega_\chi\})$ can then be interpreted respectively as beatings among the modulation frequencies at different sites ω_χ , and re-synchronisation when $\tau = 2\pi/\omega_0$.

By adopting Eq. 3.42, we obtain a huge computational advantage in estimating each likelihood, as the scaling of corresponding overheads with the bath size n^s is now linear and not exponential. Nevertheless, the parameter space is still of dimension $\mathcal{O}(10n^s)$, and in addition, the sets of parameters $\{\mathbf{B}_\chi\}$ and $\{\omega_\chi\}$ are highly correlated, as it is evident from the shape of Eq. 3.41. This breaks the QHL hypothesis that at any given point the prior distribution is unimodal⁸², a problem

⁸¹ In particular, we here refer to the step exponentiating test Hamiltonians, as in Eq. 3.2.

⁸² I have made available a tutorial in Python discussing how multimodal prior distributions can undermine the performance of QHL, and how clustering methods can recover this, see <https://figshare.com/s/5721438de7c307947ea9>.

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

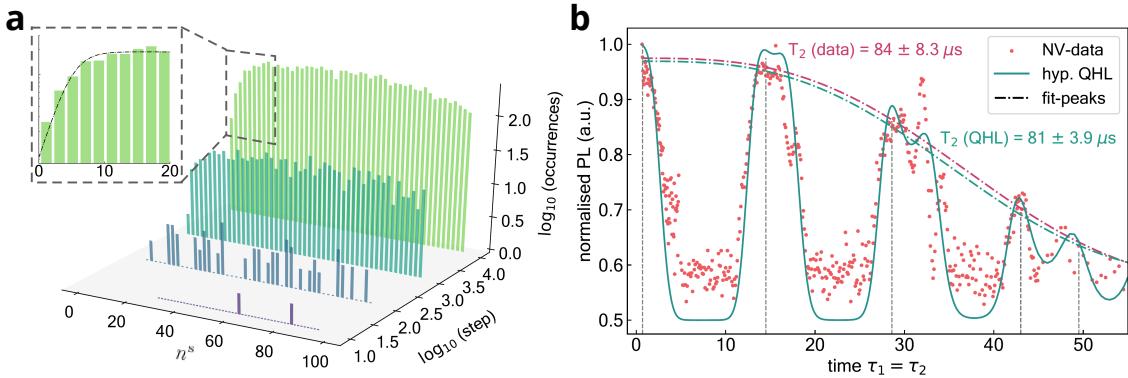


FIGURE 3.19. **a** Outcome of Metropolis-Hastings algorithm (MHA) sampling from a distribution proportional to the absolute log-likelihood $|\ell(n_c)|$, with n_c the number of environmental spins included in the model from Eq. 3.41 [253]. Exemplary steps out of a single MHA run up to 10000 steps are displayed. **Inset** Detail of the distribution behaviour for $n_c \leq 20$. A dash-dotted black line indicates the fit of the logistic mock function to the output sample from MHA. **b** Normalised PL response for a Hahn-signal experiment, against evolution time $\tau_1 = \tau_2$ (see Sect. 1.5). Red dots represent experimental data, whereas the green line is the expected PL obtained from a representative hyperparametrized model with $n_c = 20$ (Eq.3.42), whose parameters are the average outcome of 100 independent QHL runs. Estimates for the decoherence time T_2 via a peak-fitting algorithm run either on experimental data or the simulated model are also reported and colour coded correspondingly.

recently highlighted in [216]. Therefore, if each likelihood can now be estimated efficiently, the optimisation problem solved by SMC is still too challenging.

We solve these shortcomings via an ad-hoc method: a *hyperparametrisation* of the problem, whereby we replace the learning task for the sets $\{\mathbf{B}_\chi\}, \{\omega_\chi\}$ with learning the moments of the corresponding normal distributions $\mathcal{N}(\mathbf{B}^*, \sigma_B)$ and $\mathcal{N}(\omega_0 + \delta_\omega, \sigma_\omega)$, from where a tentative number $n^{s,i}$ of \mathbf{B}^* and ω_χ are drawn. In this way, for each tentative $n^{s,i}$ a CLE instance can be performed against a reduced (hyper)parameter set: $\vec{x} := \{\mathbf{B}_0, \mathbf{B}^*, \sigma_B, \omega_0, \delta_\omega, \sigma_\omega\}$. Inferring n^s is then left to a Metropolis–Hastings procedure [247], that approximates the distribution of $P(n^s)$. At each algorithm (macro)step i , a new tentative $n^{s,i}$ is sampled, and CLE is performed along N^{\exp} epochs. The probability of accepting $n^{s,i}$ as representative of the distribution is taken as the Bayes factor $\mathcal{B}_{i,i-1}(\mathbb{D})$, with \mathbb{D} the cumulative set of experimental data collected throughout all past epochs. In this way, higher values of n^s can be considered if they are statistically motivated by a better reproduction of the data. We stress how this *hyperparametrised model learning* can still be formalised within the general framework of QMLA. Indeed, the only crucial difference is that primitives are here pseudospins, and are combined via multiplication in Eq. 3.41. Therefore, we can represent the search along an sTree which embeds a single model of tentative dimension $n^{s,i} \equiv \dim[\{\chi\}]$ per layer. The sTree is thus a chain here, and only inter-layer comparisons are operated. As the hyperparametrisation introduces additional probabilistic behaviour via

the sampling of $P(n^s)$, we also increase the resilience to local minima in the sTree search by effectively allowing *jumps* across layers, instead of moving progressively to higher $n^{s,i+1}$, as well as reconsidering layers previously discarded. Indeed, the decision of which layer to entertain next depends solely upon the current $P(n^s)$. Finally, the crucial information is not stored in which node corresponds to the most successful model after a number of steps, but rather in the $P(n^s)$ distribution.

In Fig. 3.19a we show the progressive outcome of this hyperparametrised QMLA, up to 10000 algorithm (macro)steps. In the inset, we emphasise how the approximated distribution starts to plateau for $n^s \sim 15$, so that for this system, there is no compelling evidence to consider many additional spins in the bath. Interestingly, our estimate is well below the number of nuclear sites employed in initial simulations of Hahn-echo experiments with NV-centres [134], but agrees in order of magnitude with the number of ^{13}C nuclei in the so-called *first-shell*, that is known to be hyperpolarizable [258]. For completeness, we also show in Fig. 3.19b the expected normalised PL signal Pr_0 , estimated from the model in Eq. 3.41, with $n^s = 20$, and the hyperparameters as learnt after 400 CLE epochs. Simulated data accurately reproduce experimental findings, including the revival peak positions, allowing an independent estimate of the decoherence time for this system ($T_2 = 81 \pm 3.9 \text{ } \mu\text{s}$) from the envelope of the revived PL signals.

3.5.3 QMLA with a photonic chip as trusted quantum simulators

In the last sections, we have demonstrated how QMLA can retrieve crucial information about the Hamiltonian describing an (open) system dynamics, starting from realistic experimental data, and in particular an NV–centre in bulk diamond. However, so far the trusted simulator in Fig. 3.12e was a classical device, whereby each trial evolution \hat{H}_j can be implemented with arbitrary⁸³ fidelity. Also, as discussed at length in 3.2, CLE suffers from intrinsic scalability issues, as the implementation of the generic unitary $\exp(-i\hat{H}_j t)$ is not efficient in a classical device. Therefore, in the same spirit of [215] for QHL, in this section we intend to show that QMLA is indeed feasible within a realistic, non–error–corrected quantum simulator, and in particular a photonic device.

Specifically, we adopt here a quantum photonic chip to demonstrate a QMLA protocol implementation, invoking QLE instances. The platform is thus similar to Chap. 2, however it is immediate to note how the chip used in Sect. 2.3.1 being capable of dialling an arbitrary $\hat{U} \in \text{CU}(2)$, it is insufficient to implement even the reduced mapping proposed for Hahn peak experiments in Eq. 3.33, which describes a 2–qubits system (including the nuclear spin). We therefore needed an upgrade of the photonic chip, which is illustrated in the rendering of Fig. 3.20a. This device can prepare arbitrary states and dial arbitrary $\hat{U} \in \text{CU}(4)$, it will therefore be labelled “CU(4) chip” hereafter. Beside the components already outlined for the CU(2) chip, we notice

⁸³ As long as the precision of the 64-bit computational architecture employed allows. A more detailed study about the limitations imposed on the current software implementing QMLA is in preparation as [259].

3.5. EXPERIMENTAL CVV OF NV-CENTRES AND THEIR ENVIRONMENT VIA HAHN-ECHO SEQUENCES

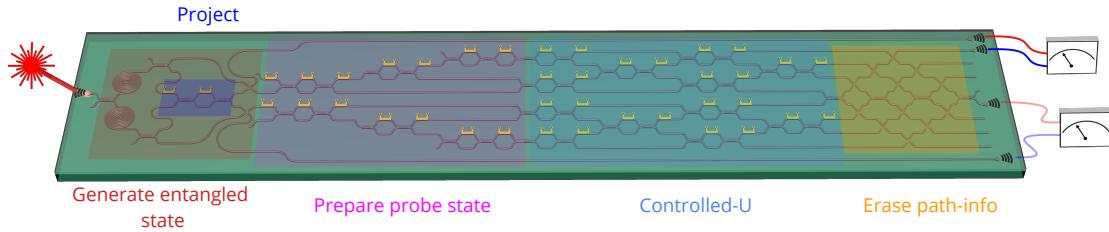


FIGURE 3.20. A rendering of the photonic chip, employed here an experimental QMLA implementation with a quantum simulator (**b**). Photons pairs generated on chip path-encode the control qubit (signal photon, blue) and the target ququart (idler, red). Coincidence events are collected by superconducting single photon detectors off chip, after frequency-filtering (not shown). Operations available with the reconfigurable device are reported, with the corresponding parts colour coded. Thermo-optic phase shifters are reported in gold. Coupling into and off chip is achieved by grating couplers.

as the most obvious difference that the target registry is now expanded as a ququart. A trivial mapping can be shown between 2-qubit and 1-ququart states and operations , so that introducing a ququart-based target registry is here an effective way to expand the operational Hilbert space, while at the same time preserving the entanglement-based scheme already adopted in Chap. 2.

The most crucial operation in the photonic chip is to retrieve the $\mathcal{L}(D|\hat{H}_j, \tau)$ in Eq. 3.27, for each probe state $|\psi\rangle_{\text{glo,in}}$ and tentative model \hat{H}_j . The preparation of random probe states $|\psi\rangle_{\text{glo,in}}$ and evolution according to a random 2-qubit unitary $\exp(-i\hat{H}_j\tau)$ can be achieved in the CU(4) chip, by means of the decomposition first formalised in [260], and used for the design of the chip in Fig. 3.20. Beside a characterisation equivalent to what already illustrated for the CU(2) chip in Chap. 1.4.4.2, in this case we also estimated the fidelity in state and unitary implementation via a full process tomography performed on the chip. Key results, detailed in Appendix Fig. D.2, show that fidelities well above 90% in most cases can be expected, so that the photonic chip can be actually assumed to be a trusted simulator, within the overall QMLA implementation (Fig. 3.21a).

Finally, $\langle D | \text{tr}_{\text{env}} [\rho_{\text{glo}}(\tau)] | D \rangle_{\text{sys}}$ is obtained via the entanglement based scheme proposed in [261], and already implemented for QHL purposes with a photonic chip in [215].

With our quantum photonic implementation, the evaluation of the likelihood for each parametrisation (i.e. particle), for each model requires about 5 s , which is about 3 orders of magnitude slower than the corresponding implementation on a classical simulator. Therefore, in order to keep the total run-time of QMLA reasonable, we had to reduce consistently $N^{\text{prt}} = \dim(\vec{x})$, N^{exp} as well as the maximum number of models entertained. We adopted a simplified tree embedding a single model per layer for the search, i.e. a tree with a branching factor of 1. The models selected are listed in the legend of Fig. 3.21a, and do not include any models embedding transverse terms for simplicity. In order to improve the resilience against the less decisive \mathcal{B}_{ij}

expected from reduced $N^{\text{prt}}, N^{\text{exp}}$, we also replaced the efficient intra-layer and inter-layer model comparisons with a complete cDAG, so that all pairwise comparisons are computed. Finally, in order to reduce the contribution from degeneracies in the parameters⁸⁴, we bootstrap the prior for the hyperfine parameters \mathbf{A} (whenever present in the model), leveraging upon preliminary knowledge about the NV–centre physics [253, 262]. In particular, we choose at epoch 0: $P(A_{\perp,x}) = P(A_{\perp,y}) = P(A_{\parallel}) = \mathcal{N}(\tilde{A}, \sigma_A)$, with $\tilde{A} = 2.5$ MHz and $\sigma_A = 0.3$ MHz. In this way, the residual multi-modality of the prior does not pose a major issue to the learning, because within the support of each prior, the distribution can be well approximated by a Gaussian.

The results for an exemplary run are outlined as logarithmic radar plots⁸⁵ in Fig. 3.21. Here we report each computed pairwise \mathcal{B}_{ij} , by colour-coding each \hat{H}_i model as an inner polygon, and reporting at the edge of the outer black polygon the \hat{H}_j model for reference. The distance from the centre of the \hat{H}_i ’s inner polygon, in direction of the j -th model, represents $\log(\mathcal{B}_{ij})$. Therefore, whenever \hat{H}_i is statistically favoured in the search, it verifies $\log(\mathcal{B}_{ij}) > 0$ and the corresponding radar vertex lies outside the circle of radius $\mathcal{B}_{ij} = 1$.

Fig. 3.21b shows the values for each \mathcal{B}_{ij} at epoch 1. All vertices being close to the circle of radius 1, no model has been learnt as a better description of the experimental Hahn peak dataset. Fig. 3.21c shows instead the progress of QMLA after 30 epochs of QHL for each model, and an equal number of Bayes factor updates, see Sect. 3.4. Now, a polygon encompassing all others clearly emerges as the one referred to $S_{x,y,z} Y x, y, z$, matching the findings obtained in Sect. 3.16 with a classical simulator.

3.6 Discussion and final remarks

The scope of this chapter was to exemplify how techniques enhanced by machine learning can provide an effective solution to the problem of CVV in NISQ devices. We first summarised the peculiarities of the QHL approach to *verify* a parametrised Hamiltonian model, and characterise the value of its parameters, briefly discussing how it can be advantageous in situations whereby standard tomographic methods are not suitable.

We then proceeded to demonstrate how an efficient implementation of QHL can greatly enhance the performances of a quantum sensor in terms of sensitivity and responsivity. In particular, we addressed magnetometry with NV–centres as the application of choice, *validating* their deployment as room–temperature magnetometers. We achieved this by overcoming the limitations of a noisy readout via the adaptive data–processing techniques offered by QHL,

⁸⁴ We remind that even if we excluded explicit degeneracies in the \hat{H}_j models, by restricting to Eq. 3.33, experimental data used as a training set are generated by the NV–centre \hat{H}^* . In Sect. 3.5.2, we estimated that \hat{H}^* describes the interaction with $\mathcal{O}(10)$ sites in the nuclear spins environment, each with a different \mathbf{A} , see also [131]. Therefore, when mapping \hat{H}^* onto an effective 2-qubit model, the beatings among these different contributions are equivalent to degeneracies in the model.

⁸⁵ The choice of a radar plot is here deemed more appropriate, compared to the tree-plots adopted in Figs. 3.18. This is because the graph search leads here to a complete cDAG storing each \mathcal{B}_{ij} computed, making a graph visualisation difficult to interpret.

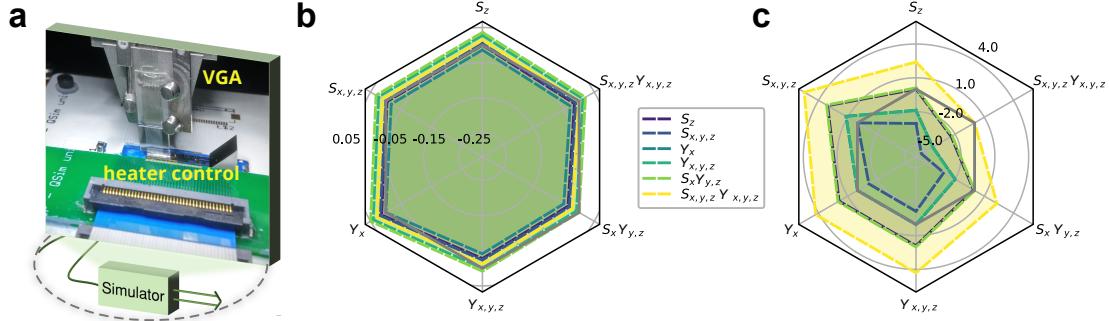


FIGURE 3.21. Experimental results of QMLA implemented on a photonic chip, here employed as a quantum simulator (a). \mathcal{B}_{ij} for all pairs of models entertained in the QMLA run are represented as a logarithmic radar-plot: at the beginning (b) and end (c) of the run. $\log \mathcal{B}_{ij}$ is plotted for model i (colour coded) versus model j (labelled on the vertex). For reference, the case $\log \mathcal{B}_{ij} = 0$ where the two models are equivalent is indicated by an inner grey polygon. In pairwise comparisons, the model at the vertex of a radius is thus favoured by statistical evidence against all those models lying within the inner polygon.

and focusing upon optimising the related computational overheads. We named the resulting protocol Magnetic Field Learning, MFL. Its performances are comparable to other protocols in low thermal–noise cryogenic environments[135].

Additionally, we showed how MFL naturally provides multi-parameter estimation in a quantum sensor, as well as allowing tracking of dynamic phenomena with high repetition rates. Multi-parameter estimation has a crucial importance in the broader context of quantum metrology, as genuine quantum effects may arise from simultaneous estimation of multiple parameters, e.g. in quantum state estimation [263]. Combined with the high-sensitivity room-temperature performances, these capabilities might pave the way for important applications ranging from the biological to the material science realms.

Finally, we extended the possibilities of automated *characterisation*, by proposing a novel protocol, the Quantum Model Learning Agent (QMLA), with the intent to a priori suggest a Hamiltonian model for a quantum system, statistically well motivated. When introducing QMLA, we aimed to learn more than a matrix representing the map for a quantum system dynamics, but rather an approximate, full decomposition in its fundamental terms. This format is vital to let a human user formulating insights on the physics governing the system dynamics [264], as every component of the Hamiltonian corresponds to a specific interaction. QMLA approaches this task as a data-centric method, developing trial models from limited prior knowledge, i.e. only a set of primitive operators, and growth rules to graphically explore the models’ space.

The method is naturally designed to grant flexibility in compromising between the amount of prior knowledge provided, and the corresponding computational cost of the protocol. Intuitively, a higher degree of prior knowledge provides a faster, more reproducible convergence to a suggested model, yet at the same time, it introduces biases which may prevent the search from considering

more appropriate models.

We have tested QMLA's robustness and capability to learn the form of the Hamiltonian describing an actual physical system, i.e. the hyperfine \hat{H}_{nv}^* of an NV-centre coupled to its environment. The experimental training data-sets, Hahn-echo sequences, embedded realistic measurement uncertainties for a bulk diamond NV-setup, such as Poissonian noise, losses, and a finite decoherence time. We further implemented the protocol also adopting a non-error-corrected quantum optical simulator, and training it to learn successfully the key terms of \hat{H}_{nv}^* . We also investigated thoroughly the impact on QMLA, of accessing only a limited subset of possible trial states that can be prepared as input for the system. For example, in experimental scenarios this might be due to limitations of the setup, or non-i.i.d. sources of noise. We have shown how biases in the probe states adopted can favour some models, altering the outcome of QMLA, but in all instances considered, such disruption is limited and the protocol is still capable of identifying fundamental components of \hat{H}_{nv}^* . Adopting solely Hahn peak datasets, QMLA was also capable of inferring *meaningful* values for the hyperfine parameters, that are usually estimated with completely different setups, or via computationally intensive ab-initio calculations of the NV-centre structure.

The model obtained gives a physical explanation of the NV-centre dynamics occurring at evolution times $\tau \ll T_2$, but the approximate 2-qubits model inferred does not explain the system's behaviour and *revivals* occurring at longer timescales, investigated via Hahn signal experiments. In order to address this case, we adapt our method to use the characterisation outcome gathered with Hahn peak datasets, as evidence to adopt an analytical growth rule, making computationally feasible a modelling of the environment including up to 100 additional spins. We run a probabilistic version of QMLA, based upon the Metropolis–Hastings approach, to train against experimental Hahn signal datasets. In this way, we gather evidence of the presence of weakly coupled impurities (in this case, ^{13}C spins) in the bath, alongside the host ^{14}N atom. The results reproduce exceptionally well the observed revivals and T_2 for the NV-system, when considering ~ 15 additional spins, and therefore an extensive, first-principle analysis of this case via QMLA would require an efficient 15-qubits simulator, which was unfortunately beyond our current capacity.

In summary, QHL (with MFL as a particular case) and QMLA methodologies have the potential to become invaluable tools for characterise, verify and validate new devices, or provide automatically more faithful and reliable models for systems currently under study, giving direct insights about their physics.

3.7 Author's contributions to the Chapter

Sects. 3.1–3.3 (re)formulate ideas and results already known from published literature, merely enriched by personal observations and on-purpose analyses.

The seminal idea of the experiments in Sect. 3.3.1 and 3.3.3 originated from dr. Santagati & dr. S. Knauer. I proposed instead the experiments leading to the results in Sect. 3.3.2 and 3.3.4 (the latter along with prof. A. Laing). A first experimental test of Sect. 3.3.1 was performed by dr. Knauer on the NV setup built by himself, and the data analysis by mr. Paesani and myself. Then, I participated to discussions concerning the details of performing the experiments with prof. Jelezko's group in Ulm. All experimental data shown in Sect. 3.3 were obtained by mr. S. Schmitt, under supervision of prof. F. Jelezko and dr. L. McGuinness. The data processing, and testing of the algorithm performances in Sect. 3.3.1 and 3.3.3 was performed by myself and dr. Santagati, with feedback from dr. Wiebe and dr. Granade. The latter also provided the library for a few routines used in the MFL.

Later, I proposed the experiment in Sect. 3.3.2 to investigate the absolute sensitivity of the MFL approach. I developed with dr. Granade the noise modelling presented in Fig. 3.8, and performed the entirety of the data analyses presented in the corresponding Section. Myself and prof. Laing suggested to test the dynamic tracking capabilities of CLE methods. A first simulation code was developed by myself and mr. Paesani. I later expanded such code to cover all experimental cases presented in Sect. 3.3.4.

The outcomes of this project were published as [265].

Concerning Sect. 3.4, the idea of the protocol was originally conceived by dr. Santagati, Wiebe, myself and mr. Paesani. I then implemented with dr. Santagati a first general-purpose version of the protocol, in simulation. I built most of the wrapper library connecting native QInfer functions with the classes required in QMD, expanding QInfer functionalities where needed. I developed instead a first version of the code dedicated to the analysis of experimental data as in Sect. 3.5.1. This code was then parallelised to run on the BCp3 facilities and further developed to include transverse terms in the tree exploration, by mr. B. Flynn. I entirely ideated, developed and tested the approach in 3.5.2, with feedback and suggestions from dr. Wiebe and Granade.

The photonic setup in Sect. 3.5.3 was prepared by me, dr. Santagati and dr. Wang. The chip was designed by dr. Santagati, mr. Paesani and myself. The entire experiment automation and control, data collection and analysis was performed by me, leveraging also upon some libraries shared by dr. J. Adcock.

Preliminary results from this Section have been accepted for oral contributions at QTML 2019 (SK) and the Symposium on the Science of Light 2019 in Erlangen (DE). Part of the outcomes has been currently arxived as [266].

4

QUANTUM CHEMISTRY WITH QUANTUM SIMULATORS

The development of the field of quantum simulation originated from a crucial observation: when adopting conventional classical methods, the resources required to simulate a quantum mechanical system scale intractably with its size. The seminal idea to adopt quantum devices in order to tackle this problem is due to Feynman [267], but it was only formalised in the following decades with the appearance of a series of algorithms dedicated to the specific field of quantum simulation [148, 268, 269]. These were followed by a plethora of experimental demonstrations [160, 162, 165, 166, 270–275], attempting to show the feasibility of these approaches to provide meaningful results in actual devices.

The variety in approaches led to the introduction of some nomenclature to categorise them. The most fundamental distinction is perhaps between “analogue” and “digital” quantum simulators [59]. The first category captures special purpose devices, whose controllable dynamics mimics the behaviour of the quantum system to be simulated, whereas the second describes general-purpose quantum computers dedicated to simulation tasks (see also Chap. 1). Another crucial classification separates methods addressing properties at equilibrium (i.e. concerned with computing stable states) from those describing and simulating the time evolution of the quantum system in conditions of interest [276].

In this Chapter, we will mostly deal with the *eigenproblem*, i.e. the analysis of eigenspectra in a Hamiltonian framework¹. This particular problem is of paramount importance, because an efficient eigensolver has a range of applications, ranging from material science [276] and biology [277, 278] to data science and information retrieval [279]. Here, we highlight in particular those belonging to the exquisite realm of quantum chemistry, like the modelling of many-body interactions in electronic systems [187, 280], or the understanding of reaction rates and light-

¹For the reason to adopt this framework, see below Sect. 4.1.2.

harvesting transfer processes. The latter, in particular, require a characterisation of the system's excited states [281, 282].

Solving the eigenproblem in the most general case is believed to be hard not only for classical, but also for quantum computers [148], as it would imply a collapse in complexity classes, QMA=BQP [283]. This applies as well to worst-case quantum chemistry models [284]. However the latter finding does not preclude successful applications of quantum simulators, as several chemical configurations of practical interest are believed to fall within a subset of lesser complexity [269, 276, 285]. This conjecture is supported by empirical observations, described in Sect. 4.3.2. Therefore, in this chapter we will focus on strategies to extract the ground state, and in some cases the full eigenspectrum, of some instances of interest where quantum computers are expected to enable scalable solutions [148, 268]. The importance of addressing excited states with quantum methods is not only driven by the realm of applications, but also by the investigation of quantum advantages. Indeed, the complexity of an eigenproblem is known for quantum systems to depend not only on the corresponding Hamiltonian, but also on the class of states addressed. For example, ground states can be prepared efficiently with classical methods in weakly interacting Hamiltonians, but this is often not true for the corresponding excited states [286, 287].

The chapter is organised as follows. At first, the application of PE-inspired methods is briefly discussed, drawing on conclusions achieved in Chap. 2. Then, we will review techniques traditionally employed to address ground-states (in particular variational methods), and see how they can be extended to the whole eigenspectrum. We will then expose a novel, intrinsically quantum strategy to prepare excited states in realistic quantum systems, that we have recently introduced [128]. In conclusion, we will provide an outlook of additional methods that stemmed from our idea, providing various improvements in achieving the same goal.

4.1 Mapping quantum chemistry Hamiltonians onto a quantum computer

We have outlined how a quantum device may offer a more natural testbed for the simulation of chemical structures and processes, when compared to its classical counterpart. However, problems in the realm of quantum chemistry often require non-trivial mapping operations, before they can be addressed with a quantum simulator. Indeed, some instances of such problems that might benefit the most from quantum simulators², namely the *electronic structure* problem, and quantum lattice models [159, 160, 276], are formulated as many-body molecular wavefunctions [148], whereas digital quantum simulators operate on qubits. A convenient way to map one framework to the other is thus required. A summary of the various steps providing such mapping is presented in Fig. 4.1. In the following sections we will introduce in some detail only

²See also Chap. 3.

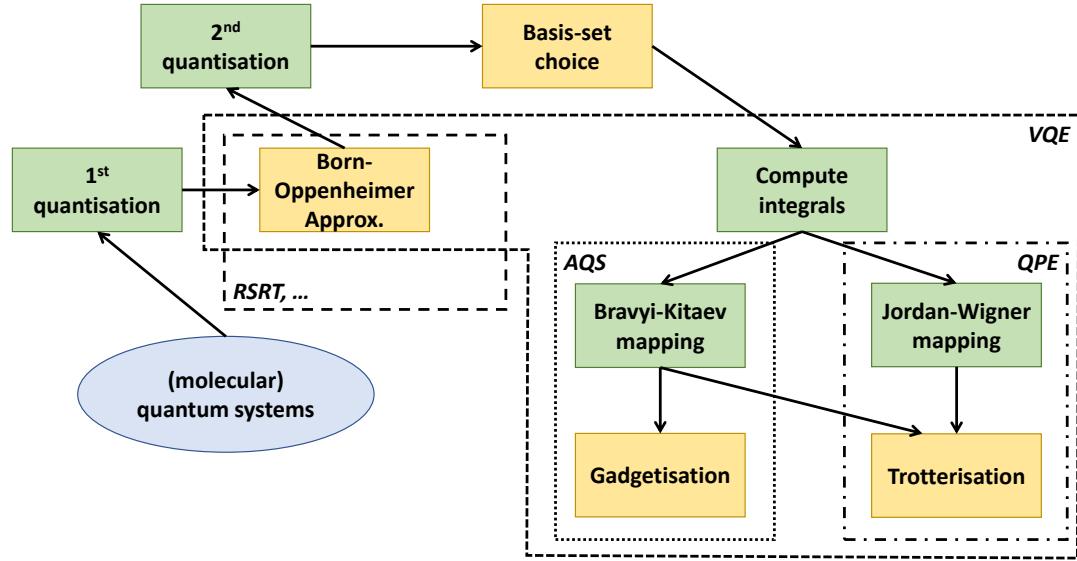


FIGURE 4.1. Summary of steps and techniques to map quantum chemistry problems onto quantum simulators. Operations to perform various steps of the mapping are in coloured rectangular squares, with those introducing approximations in orange, the other steps in green. In dashed rectangular frames the techniques proposed to solve the corresponding mapped problems. Abbreviations are introduced in the text.

techniques that rely upon a 2nd quantised Hamiltonian, hence neglecting those methods natively designed to address models that invoke solely the Born–Oppenheimer approximation (e.g. the Rear-Space-Real-Time dynamics (RSRT) in [288]).

4.1.1 Spin-orbital configurations and qubit states

In order to describe this in further detail, we will pick a specific example covering most cases in the remaining of this chapter: the electronic structure problem in molecules. This can be described in full generality by an N-electron wavefunction Φ_N . A well-known approach in quantum chemistry is to express Φ_N as a sum over *determinants* [289, 290], i.e. products of single-particle wavefunctions:

$$\Phi_N = c_0 \Psi_0 + \sum_{ra} c_a^r \Psi_a^r + \sum_{\substack{a < b \\ r < s}} c_{ab}^{rs} \Psi_{ab}^{rs} + \dots \quad (4.1)$$

Here, each $\Psi = \prod_i^{n \leq N} \chi_i$ is a determinant involving simultaneously n single-electron *spin-orbitals*: each $\chi(\vec{r}, \sigma) = \psi(\vec{r})\alpha(\sigma)$ is the product of a spatial orbital depending on the electron position³ and

³ The spatial wavefunctions are solution of the fermionic Hamiltonian of the electrons moving in the Coulomb potential, within the Born–Oppenheimer approximation, and usually further approximated by Hartree-Fock calculations [159]. In the latter case they are therefore approximated as non-interacting.

the spin state represented by $\sigma \in \{\uparrow, \downarrow\}$. The number of superscripts of Ψ customarily represents the number of electrons occupying excited (virtual) orbitals [289], starting from the corresponding occupied (hole) orbitals identified by the subscript. In this way, the first term in the expansion, Eq. 4.1, corresponds to the ground state, the second to single-excited states, etc. We omit terms higher than doubly-excited configurations.

In principle, an exact calculation should include all unique ordered sets of virtual spin orbitals, resulting in an infinite set of determinants $\{\Psi_0, \Psi_a^r, \Psi_{ab}^{rs}, \Psi_{abc}^{rst}, \dots\}$. However excellent approximations can be obtained by using instead a finite set of K spin orbitals $\{\chi_i\}$ [289], i.e. an incomplete basis of the N -electron Hilbert space, leading to the calculation of up to $\binom{2K}{N}$ determinants⁴. If all such determinants are used, the corresponding calculation is named a *full configuration interaction* (FCI) [276].

The most intuitive choice of the basis-set $\{\chi_i\}$ is the *occupation number basis*: we map the state of the N -electron system as a tensor product of $K = N$ single-electron spin orbitals [290], in a Fock-space representation $\{f_0, f_1, f_2, \dots, f_K\}$ whereby the occupation of each orbital χ_p is labelled by the corresponding bit $f_p \in \{0, 1\}$. This direct representation also offers an immediate mapping onto the quantum simulator, by interpreting the binary string above as the corresponding N -qubit state:

$$|Q\rangle = |q_0 q_1 q_2 \dots q_K\rangle, \text{ with:} \\ q_p = f_p \quad (4.2)$$

This implies that the occupancy of an orbital can be immediately inferred from single qubit measurements, being a local information in this basis [276]. Even if intuitive to adopt, and advantageous in terms of the size of controlled operations required by the corresponding quantum circuit [148], *compact representations* are often preferred. These reduce the number of qubits required by the circuit at the expense of increasing the cardinality of the basis-set $\{\chi_i\}$, such as the STO-3G (used for simulations in this chapter), 6-31G and cc-pVTZ bases [148, 164, 291, 292].

4.1.2 The mapping of Hamiltonians

Once a basis-set and eventually a certain (sub)set of determinants has been chosen to represent the states of the molecular system, we are left with finding a representation of the interactions between electrons, and therefore of the molecular system's dynamics, that can be suitably implemented on a digital quantum simulator. Even if a number of proposals have addressed the quantum simulation of Hamiltonians directly in their first-quantised form [159], in this Thesis we shall focus upon the more practical choice of second-quantisation. Hamiltonians describing the interaction among N -body systems, including at most two-particle interactions, can be constructed in the most general way by combining creation and annihilation operators

⁴The 2 factor resulting from considering the spin in counting the orbitals.

(respectively a^\dagger, a) [290, 293]:

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \quad (4.3)$$

with h_{pq} and h_{pqrs} overlap integrals, that are pre-computed classically and depend upon the basis-set chosen to represent the quantum states [159]. The second-quantised Hamiltonian reported in Eq. 4.3 is particle-conserving⁵ but at the same time it is evidently not expressed in terms of operators acting on qubits. In order to have a form implementable in a quantum simulator, we are still required to choose a suitable mapping, which will impact crucially on the final performances of the algorithm. The most intuitive mapping can be inferred from the usage of the *occupation number basis*, and rewriting the creation and annihilation operators in terms of qubit raising and lowering operators [148]:

$$\begin{aligned} a_p^\dagger &\rightarrow \hat{\sigma}_p^+ \\ a_p &\rightarrow \hat{\sigma}_p^- \end{aligned} \quad (4.4)$$

with p the generic index of a spin-orbital (qubit), and the $\hat{\sigma}^\pm$ notation introduced in Chap. 1. However, the corresponding operators would not obey the fermionic commutation relations that are valid for the original operators [276]. In order to address this limitation, various mappings have been proposed, but the ones mostly used in the community are currently the *Jordan–Wigner* and *Bravyi–Kitaev* mappings [159, 293, 294]. Each of them has peculiar advantages and disadvantages [293], however recent studies have emphasised the substantial savings in circuit depth offered by the Bravyi–Kitaev choice for molecules of size close to intractable [294].

The Jordan–Wigner transformation has been both historically proposed⁶ [295], and also experimentally adopted first [148]. The mapping transforms:

$$\begin{aligned} a_p^\dagger &\rightarrow \left(\bigotimes_{m < p} \hat{\sigma}_m^z \right) \otimes \hat{\sigma}_p^+ \otimes \left(\bigotimes_{m > p} \hat{I}_m \right) \\ a_p &\rightarrow \left(\bigotimes_{m < p} \hat{\sigma}_m^z \right) \otimes \hat{\sigma}_p^- \otimes \left(\bigotimes_{m > p} \hat{I}_m \right) \end{aligned} \quad (4.5)$$

so that fermionic anti-commutation rules are preserved, as expected for an appropriate representation of fermionic models in qubits. Observing Eq. 4.5, it appears evident how $\hat{\sigma}^\pm$ operators still act on a single qubit, thus preserving the locality of the occupation information across the evolution. At the same time, the additional $\hat{\sigma}^z$ operators act non-locally on the parity of the occupancy [276], in order to preserve anti-commutation relations. This approach has a major drawback, as k -local interactions⁷ are introduced, with k scaling linearly with the mapped system's size. This is detrimental to the intended quantum simulation in two ways. First, it

⁵ Indeed each term has as many annihilation as creation operators.

⁶ Even if outside of the proper realm of quantum simulation.

⁷ A k -local term in the Hamiltonian being defined as a Hermitian matrix, acting on not more than k qubits at once.

introduces a $\mathcal{O}(N)$ overhead in depth of the circuit, per electronic operator, for an N -electron system [294]. Additionally, further steps in the mapping⁸ request a control precision in the gates that scales exponentially in k , ultimately making the Jordan-Wigner approach non-scalable [276, 293].

The Bravyi-Kitaev mapping addresses this shortcoming by adopting a different basis for the states, and dropping the simplicity of a local orbital occupation information [285, 296, 297], as well as the local parity information invoked by other mappings [276, 294]. The simple rule in Eq. 4.2 is thus replaced by a more convoluted rule for computing each qubit in the mapped multi-qubit state $|Q\rangle$:

$$q_p = \begin{cases} f_p & \text{if } p \text{ is even} \\ \sum_{i \leq p} f_i \pmod{2} & \text{if } p \text{ is odd and } \log(p+1) \in \mathbb{N}_0 \\ \sum_{i=p-1,p} f_i \pmod{2} & \text{otherwise.} \end{cases} \quad (4.6)$$

It is now evident that excluding even-indexed qubits, the others store delocalised information about the parity and occupation of the other orbitals. Therefore, the transformation leads to some subsets of qubits with particular significance for each index p [285, 294]:

1. the *update set* $U(p)$, i.e. the set of qubits that depend on the occupation state of orbital p ,
2. the *parity set* $P(p)$, i.e. the set of qubits that are required to determine the parity of the p -th orbital (so all qubits indexed $j < p$),
3. the *flip set* $F(p)$, whose qubits determine whether $q_p = f_p$ or not,
4. the *remainder set* $R(p) = P(p) \setminus F(p)$, defined as a set complement.

The final transformation of fermionic operators depends explicitly upon these sets, as it can be written [296]:

$$\begin{aligned} a_p^\dagger &\rightarrow \frac{1}{2} \left(\bigotimes_{m \in U(p)} \hat{\sigma}_m^x \right) \otimes \hat{\sigma}_p^x \otimes \left(\bigotimes_{m \in P(p)} \hat{\sigma}_m^z \right) - i \left(\bigotimes_{m \in U(p)} \hat{\sigma}_m^x \right) \otimes \hat{\sigma}_p^y \otimes \left(\bigotimes_{m \in D(p)} \hat{\sigma}_m^z \right) \\ a_p &\rightarrow \frac{1}{2} \left(\bigotimes_{m \in U(p)} \hat{\sigma}_m^x \right) \otimes \hat{\sigma}_p^x \otimes \left(\bigotimes_{m \in P(p)} \hat{\sigma}_m^z \right) + i \left(\bigotimes_{m \in U(p)} \hat{\sigma}_m^x \right) \otimes \hat{\sigma}_p^y \otimes \left(\bigotimes_{m \in D(p)} \hat{\sigma}_m^z \right), \end{aligned} \quad (4.7)$$

with $D(p) = P(p)$ if p is even, and $D(p) = R(p)$ if p is odd (identity operators are here omitted for simplicity). The fundamental property that emerges from Eq. 4.7 is that in order to simulate each fermionic operation, the qubit operations required act on a limited subset of the global space spanned by the chosen basis. We do not write here explicitly the sets $U(p), P(p), F(p)$ and $R(p)$, however, it can be formally proven that the cardinality of these sets scales as $\mathcal{O}(\log N)$ [285, 296]. Therefore, the scaling of this transformation is quite advantageous over the Jordan-Wigner one,

⁸ See Sect. 4.1.3.

and the advantage is preserved also when additional steps to make the mapping operational⁹ are considered [294].

4.1.3 Gadgetisation

A generic physical Hamiltonian is expected to exhibit long-range couplings involving more than two subsystems at once [298]. However, this behaviour is normally restricted to exotic condensed matter systems, and for fermionic Hamiltonians describing electrons moving in a lattice of nuclei, two-body Hamiltonians such as the second quantised form in Eq. 4.3 were found to be an excellent approximation to the corresponding many-body problem [299].

A broad class covering many cases of physical interest is given by k -local Hamiltonians¹⁰ (with fixed k), which e.g. arise naturally from lattice Hamiltonians with nearest-neighbours interactions [300]. Several important cases in the quantum chemistry realm are even more restrictive, as they can be naturally cast into 2-local Hamiltonians, like the aforementioned electronic many-body problem. Unfortunately, this property is not retained once mapped onto digital quantum simulators adopting e.g. the Jordan–Wigner or Bravyi–Kitaev transformations (Sect. 4.1.2), so that the input of the quantum simulation will generally be a k -local \hat{H} .

Indeed, digital quantum simulators architectures are usually confined to 2-local programmable couplings among the simulator qubits, and the corresponding mismatch with the system (mapped) Hamiltonian must then be addressed. This is of particular importance when a Bravyi–Kitaev mapping is adopted, because the additional couplings between qubits introduced by the transformation depend upon the state of other qubits (see Eq. 4.7). The standard approach to make such interactions experimentally manageable is known as [276]:

Definition 4.1. *Gadgetisation*, the class of methods that render a k -local Hamiltonian \hat{H}_{sys} , acting upon n qubits in total, with a 2-local *gadget* Hamiltonian \hat{H}_{gad} , that acts on an expanded basis of $n + a$ ancillary qubits.

In order to avoid a full construction of \hat{H}_{gad} , perturbative methods are the usual choice to perform an approximate construction [283]. Because of the systems and mapping chosen for the simulation and experimental results of this chapter (Sect. 4.5 and following), gadgetisation will here play no role. Therefore, we will skip further details.

4.1.4 Trotterisation

In order to have an efficient, reliable quantum simulator to analyse the dynamics of a Hamiltonian system, one must consider a simulation algorithm for $e^{-i\hat{H}t}$, as introduced in Chap. 1. Here we additionally make some specific statements that apply to the realm of quantum chemistry. In

⁹ See Sect. 4.1.3.

¹⁰I.e. a Hamiltonian that can be expressed as the sum of polynomially many Hermitian terms, that describe k -local interactions.

particular, we observe that for \hat{H} of the form in Eq. 4.3, describing a system of n spin orbitals, at most $n^{8+o(1)}$ primitive sequential exponentiations are required [33].

However, empiric evidence has shown how low-order Trotter decompositions might actually have a less demanding scaling $\mathcal{O}(n^{5.5})$, that can be further improved with more involved methods [167]. In conclusion, the literature supports the efficient trotterisation of the unitary evolution required by VQE protocols, a result that will be used in Sect. 4.4.2 to estimate the overall complexity of our variational method.

4.2 Quantum eigensolvers with PE-based strategies

Given an approximate eigenstate, the quantum phase estimation (QPE) algorithm can efficiently estimate the corresponding eigenvalue, this being experimentally demonstrated with various quantum systems, such as nuclear magnetic resonance, photonic and superconducting systems [162, 166, 273]. The potentiality of this approach, and various specific algorithms implementing it, have already been discussed at length in Chap. 2. Here, therefore, we will limit the discussion to some considerations of interest within the realm of quantum simulation.

The most important point is that QPE embeds no state preparation subroutine, i.e. the eigenstate, whose eigenvalue is to be estimated, is intended as a given input. This poses a severe problem, as for a Hamiltonian with unknown spectrum, it is likely that the eigenstates are unknown as well. A first approach is to use classically scalable quantum chemistry approximations, to provide an initial guess to the QPE algorithm. Indeed, given an input state $|\Psi\rangle$ having sufficient overlap with a targeted eigenstate $|\bar{\Psi}\rangle$, QPE is known to be able to collapse the estimate to the appropriate eigenvalue with high success rate¹¹. Additionally, one might employ QPE strategies to provide simultaneous estimation of various eigenvalues simultaneously, if the input state is in a superposition of few eigenstates¹². Therefore, a simplistic solution would be to adopt Hartree-Fock approximated wavefunctions $|\Psi_{HF}\rangle$ as the input state [289]. It has been demonstrated that this approach works well when searching for the ground state in the limit of small molecules, and weak electron interaction strength [148]. However, it is also known that for bigger molecules $\langle\bar{\Psi}|\Psi_{HF}\rangle^2 \sim 0$ already for the ground state, and the approximation is worse for excited states [289]. The same applies to highly-excited states of even small molecules.

4.2.1 Adiabatic state preparation

A more systematic approach to address this limitation, as long as ground states alone are of concern, introduces the *adiabatic state preparation* (ASP) [148, 159, 276]. ASP relies on the *adiabatic theorem*:

¹¹ See Chap. 2.

¹² See Sect. 2.1.6.1.

Theorem 4.1. *If the ground state $|\Psi\rangle_{g,0}$ of a Hamiltonian \hat{H}_0 undergoes evolution whilst \hat{H}_0 is changed to another \hat{H}_1 “slowly enough”, then it will evolve into the ground state of the new \hat{H}_1 , $|\Psi\rangle_{g,1}$.*

In practice, this is usually done by adopting a parametrisation of the initial Hamiltonian, capable of expressing the target one. The most immediate example is a linear interpolation [301]:

$$\hat{H}(s) = A(s)\hat{H}_0 + B(s)\hat{H}_1, \quad (4.8)$$

where it is customary to consider a normalised parameter $s \in [0, 1]$, with $\hat{H}(0) = \hat{H}_0$ and $\hat{H}(1) = \hat{H}_1$. The rapidity in time t of the change is thus governed by how quickly $s \xrightarrow{t \rightarrow t_{\max}} 1$. The figure of merit is $t_{\max}/\Delta\mathcal{E}_{ge}$, i.e. the ratio between the longest evolution time available and the energy gap of the ground and 1st excited state [276].

The rate at which s is changed is key to an effective and efficient preparation of $|\Psi\rangle_{g,1}$. In ideally noiseless conditions, under a unitary evolution, knowing the path $\mathcal{F} : t \rightarrow s$ is sufficient to predict the final state adiabatically prepared [301]. However, as long as:

$$t_{\max} \gg \frac{1}{\Delta\mathcal{E}_{ge}}, \quad (4.9)$$

the adiabatic theorem ensures that any path will prepare $|\Psi\rangle_{g,1}$ exactly, so that path optimisation plays an important role only when e.g. experimental constraints limit the available duration of coherent evolution [159]. We exemplify the importance of the duration in time of the adiabatic schedule in Appendix C.3, adopting for simplicity a linear adiabatic schedule like in Eq. 4.8.

The potential requirement for long coherent evolution and deep circuits, in order to prepare the input states for QPE, are a major obstacle against widespread adoption of ASP in near-term implementations on quantum computers. The adoption of more complex adiabatic schedules, enriching the simple scheme of Eq. 4.8 to include multiple parameters and non-linear functions, has been proposed to mitigate such issues [159, 302]. However, for large molecules, the classical cost of optimising such schemes against the available decoherence time in the quantum simulator might prove non-trivial. Indeed, the preliminary path optimisation is usually performed on classical devices, with the fidelity against $|\Psi\rangle_{g,1}$ as the objective function. Whenever the size or features of the quantum system prevent finding an effective adiabatic path within the approximations that make it possible to classically simulate such system, ASP cannot be used to bootstrap the quantum simulation.

4.3 Variational approaches to Quantum eigensolvers

Variational quantum eigensolvers (VQE) have proposed a hybridisation of quantum and classical approaches, in order to deliver hardware-agnostic quantum chemistry protocols with a reduced depth of the quantum circuit, as well as of the daunting overheads to implement traditional error-corrections schemes [159, 160, 162, 303–305]. The shortcomings of noisy protocols affect many

QPE eigensolvers, limiting crucially their applicability in NISQ devices, as it appeared evident in recent experimental results, where gate fidelity and decoherence phenomena play a major role [162]. The VQE strategy is inspired by the same principle already discussed in Chapters 2–3, whereby experimentally demanding quantum computational resources are systematically replaced by inexpensive classical counterparts. Specifically, in Chap. 2 we have seen how this approach can improve the noise resilience of QPE, and hence of its applicative scenarios, such as quantum chemistry.

Here, we discuss how this strategy can be successfully employed leveraging upon the variational principle (see below Sect. 4.3.1). In detail, each VQE protocol has two fundamental components: the preparation of probe states that are evolved according to the Hamiltonian \hat{H} of interest, and the estimation of the energy of such probe states. Therefore, like PE-based strategies, also VQE relies on the possibility to implement efficiently unitary evolution \hat{U} operators, related to \hat{H} , in the quantum circuit¹³. Provided such components, the eigenproblem can then be addressed as a (variational) optimisation problem in terms of the energy as a figure of merit. The optimisation schedule is performed on a classical co-processor [160], according to an algorithm of choice. Exemplary optimisation methods successfully implemented can thus range from simplistic gradient descent [160, 275], to Nelder-Mead [159, 160], classical Gaussian-approximated Bayesian [306], dividing rectangles algorithms [307], and swarm optimisers [128]. As the energy landscape of high-dimensional Hamiltonians is expected to be rich in local minima, global minimisation algorithms are recommended [159].

As optimisation schedules rely on sampling for the energy estimation, they are expected to provide quadratically less precise estimates than QPE for the same number of measurements. At the same time, the sampling approach grants unique robustness to certain error classes [162, 308]. In the following we analyse in depth each of these aspects.

4.3.1 Energy estimation and the variational principle

A long known strategy to render the eigenproblem as an optimisation problem is the Rayleigh-Ritz quotient method [292]. This is based upon the *variational theorem*:

Theorem 4.2. *Called $\vec{\theta}$ the set of real parameters that describe the generic state $|\Psi(\vec{\theta})\rangle$ (e.g. via the decomposition in an orthonormal basis for the system), the variational theorem states that if λ_g is the eigenvalue for the ground state $|\Psi\rangle_g$, described by the set of parameters $\vec{\theta}_g$, then:*

$$\langle \Psi(\vec{\theta}) | \hat{H} | \Psi(\vec{\theta}) \rangle := \langle \hat{H} \rangle_{\vec{\theta}} \geq \langle \Psi(\vec{\theta}_g) | \hat{H} | \Psi(\vec{\theta}_g) \rangle := \lambda_g, \quad \forall \vec{\theta}. \quad (4.10)$$

Leveraging upon the variational theorem, the quotient method reduces the problem of finding the ground state of \hat{H} and its energy to identifying the approximate $\vec{\theta}$ which minimises the expectation value in Eq. 4.10. λ_g is then immediately given by the corresponding minimum $\langle \hat{H} \rangle_{\vec{\theta}_g}$.

¹³ Refer to Sect. 4.1.4.

Here, VQE simply replaces the classical device with a quantum circuit, when estimating each $\langle \hat{H} \rangle_{\vec{\theta}}$, until the optimisation schedule chosen converges to the minimum.

Taken a normalised quantum state $|\Psi\rangle$, in principle a quantum circuit can efficiently provide an estimate for the expectation value of an operator $\langle \hat{O} \rangle_{\Psi} := \langle \Psi | \hat{O} | \Psi \rangle$ (see also Chap. 3). VQE leverages on this, and invokes the quantum circuit to obtain $\langle \hat{H} \rangle_{\Psi}$. In practice, for an efficient implementation, it is additionally required that \hat{H} can be decomposed into a polynomial number of local operators $\hat{H} = \sum_{\alpha} p_{\alpha} \hat{h}_{\alpha}$. The locality prevents complex architectural requirements for the quantum device, whilst the polynomial scaling avoids issues concerning the scalability of VQE [159, 160, 162]. This holds for a wide variety of Hamiltonians describing physical systems¹⁴, and hereafter we will always assume it verified by the system of interest. If these conditions are satisfied, and eventually invoking an appropriate gadgetisation (Sect. 4.1.3), one can decompose the evaluation of:

$$\langle \hat{H} \rangle_{\Psi} = \sum_{i,\alpha} p_i^{\alpha} \langle \hat{\sigma}_i^{\alpha} \rangle_{\Psi} + \sum_{i,j,\alpha,\beta} p_{ij}^{\alpha\beta} \langle \hat{\sigma}_i^{\alpha} \otimes \hat{\sigma}_j^{\beta} \rangle_{\Psi} \quad (4.11)$$

with $\hat{\sigma}^{\alpha} \in \{\hat{I}, \hat{\sigma}^x, \hat{\sigma}^y, \hat{\sigma}^z\}$ and roman indices identifying the qubits in the quantum simulator. The decomposition in Eq. 4.11 can be fundamental for the practicality of VQE, as estimating $\langle \cdot \rangle_{\Psi}$ is usually achieved without invoking any entangling gate as it is customary in other QC applications [309, 310]. Therefore, expectation values are extracted by measurements of the output, and Eq. 4.11 ensures that this can be accomplished with a number of single-qubit measurements that is $\mathcal{O}(\text{poly}(n))$ [311].

4.3.2 Probe state parametrisation

We are now left with the problem of generating trial states $|\Psi(\vec{\theta})\rangle$ entering the optimisation procedure. This is far from a trivial problem, as it requires compromising the capability to “explore” a sufficient portion of the Hilbert space spanned by the eigenbasis of \hat{H} , with the efficiency of the variational search. Indeed, a search occurring across a substantial sampling of the Hilbert space is out of reach already for modest-sized systems, even adopting a quantum device, as it encounters an *exponential wall* in the size of $\vec{\theta}$ required to describe the corresponding states. For example, the wavefunction of $N = 100$ interacting electrons requires $\dim(\vec{\theta}) \sim 10^{150}$ [312], making the corresponding minimisation unfeasible. This limitation, stated in the introduction to this chapter, has been formalised by the proof that the problem of finding the ground state for an arbitrary k -local Hamiltonian, with $k \geq 2$, is in general QMA-complete [283].

Fortunately, as premised in the introduction to this Chapter, several results point out that such an exhaustive search might be unneeded in many cases of interest. From an information theoretical point of view, the set of states that can be practically prepared in an ideal quantum device does not differ dramatically from those realistically accessible to a class of well-behaved

¹⁴See the introduction of this chapter, as well as Sect. 4.1.3. For example, Hermitian operators can always be expressed as a linear combination of tensor products of Pauli matrices (see Eq. 3.31).

Hamiltonians [300]. In particular, assuming a time-dependent, k -local Hamiltonian \hat{H}_{ph} with a finite bound to the strength of any of its interaction terms and k independent from n , we may consider *physical* only those states that are the outcome of fiducial states evolved by \hat{H}_{ph} within a time, that is at most polynomial in the number of particles involved n . Crucially, the assumptions on \hat{H}_{ph} capture those Hamiltonians that can be efficiently mapped onto a quantum device provided with a universal set of one and two-qubit gates [300], along with model Hamiltonians for ensembles of spins in many-body systems [160]. Here, imposing a limited duration to the time evolution is key in avoiding, and thus not contradicting, the worst-case result of QMA-completeness cited above.

However, the most generic quantum chemistry models (e.g. those obtained via density functional theory) are known to require $k \propto \log(n)$ [293], when the fermionic problem is mapped onto a qubit system (see Sect. 4.1.2). Therefore, the result in [300] cannot be directly invoked to postulate an efficient eigenstate preparation on a quantum device. In such cases, the promise for quantum computers to provide an advantage as eigensolvers is debated in terms of algorithmic efficiency [313], and relies on empirical considerations.

First, the limited evolution time argument, cited above, captured quantitatively the intuitive observation that real quantum systems occurring in nature collapse in a state of minimal energy on fast timescales [276]. Therefore, they do not appear affected by the consequences of a search through an exponentially growing Hilbert space.

Additionally, VQE methods rely on efficient parametrisations, known as *ansätze*, to prepare test states in the quantum circuit, as elicited in the following. These *ansätze* have proven highly successful in describing the eigenstates of several systems in quantum chemistry simulations running on classical hardware [159]. At the same time, they are designed to easily extend to explore states that are hard to prepare classically [314], which makes them a promising candidate to show a clear advantage using quantum simulators. Different types of *ansätze* have been proposed for the variational search. Whenever possible, an *ansatz* leverages upon pre-existing knowledge available about the system of interest to simplify the parametrisation [292]. The ASP outlined in Sect. 4.2.1 can itself be considered a single-parameter¹⁵ *ansatz*, and it has been therefore considered for adoption also in VQE protocols [159].

Coupled-cluster (CC) approaches are a long-known way to express wavefunctions of interest in the quantum chemistry community. CC methods traditionally employ a *reference state* $|\Psi\rangle_0$, chosen among those easy to classically compute for practical reasons, such as Hartree-Fock solutions [315]. Works targeting a QC scenario, instead, tend to propose

$$|\Psi\rangle_0 = |00\dots 0\rangle \tag{4.12}$$

on the same ground of easy accessibility, here intended for a quantum device [159, 160]. Once a

¹⁵ The only parameter actively regulated in the quantum circuit being indeed the evolution time t . Additional parameters describing the path in terms of t are used only by the classical computing schedule, and hence are not considered here.

reference state is chosen, CC uses such reference to explore other states *close* to $|\Psi\rangle_0$. The shape introduced in Eq. 4.12 suggests to introduce the ideas of CC in terms of qubit *flips* [159], i.e. the proximity of a target state $|\bar{\Psi}\rangle$ to $|\Psi\rangle_0$ might be intended as the number of qubits F that must be flipped in order to express $|\bar{\Psi}\rangle$. All states with $F = 1$ can then be parametrised as:

$$|\bar{\Psi}_{\text{CC}}(\vec{\theta})\rangle_1 = \sum_{p1} \theta_{p1} \hat{\sigma}_{p1}^+ |\Psi\rangle_0, \quad (4.13)$$

with $p1$ labelling which qubit has been flipped. This parametrisation is known as the *configuration interaction* (CI)¹⁶. One might expand the reasoning to any number of spin–flips, thus parametrising any possible state in the Hilbert space of the system:

$$|\bar{\Psi}(\vec{\theta})\rangle_F = \sum_{f \leq F} \left(\sum_{\{p1, \dots, pf\} \in \mathbb{S}_{F,f}} \theta_{p1} \hat{\sigma}_{p1}^+ \dots \theta_{pf} \hat{\sigma}_{pf}^+ |\Psi\rangle_0 \right), \quad (4.14)$$

with $\mathbb{S}_{F,f}$ the set of $\binom{f}{F}$ possible combinations of f indices, each with elements $p1, \dots, pf$. The number of operations required to express a target state F flips away thus grows combinatorially with F : a solution to this issue is to truncate the summation in Eq. 4.14 to small F (and thus f). CI has another significant disadvantage: the energy of states expressed by the truncation is not a monotone function of f , so that approximate CI’s lack of *extensivity* [314]. Also, it is of reduced interest to us, because the ansatz for a truncated f is already classically efficient, so that a quantum simulator offers no advantage in this phase [159].

The traditional CC (TCC) approach solves the extensivity problem by employing exponentiated operators in the ansatz [315], that writes accordingly to first order as:

$$|\bar{\Psi}(\vec{\theta})\rangle_{\text{TCC},1} = \exp\left(\sum_{p1} \theta_{p1} \hat{\sigma}_{p1}^+\right) |\Psi\rangle_0 \equiv e^{\hat{T}} |\Psi\rangle_0, \quad (4.15)$$

to compare with the corresponding CI case in Eq. 4.13. A crucial advantage of TCC consists in the expansion being finite, as it can be seen by expanding to higher orders and using the Baker-Campbell-Hausdorff (BCH) commutator to evaluate a generic $\langle \bar{\Psi} | \hat{H} | \bar{\Psi} \rangle_{\text{TCC},1}$ for a two-body \hat{H} in closed form [315]. Unfortunately, even if applied successfully to other scenarios, TCC-generated states suffer from not being variationally bound to the ground state energy [314]. Not respecting the variational theorem thus makes TCC not applicable to VQE.

The most intuitive solution to reinstate a variational bound is to renormalise the expectation values for the energy obtained as $\langle \bar{\Psi} | \hat{H} | \bar{\Psi} \rangle_{\text{TCC},1}$, using the normalisation factor $\langle \Psi | e^{\hat{T}^\dagger} e^{\hat{T}} | \Psi \rangle_0$. In this approach, known as Variational Coupled Cluster (VCC), the normalisation factor does not truncate naturally via BCH, making it computationally expensive to estimate [316]. Also, it requires the implementation of non-unitary evolution on the final quantum simulator, which is highly non-trivial and normally requires significant overheads (see e.g. [317] for a photonic implementation).

¹⁶ Not to confuse with the FCI introduced before for Hamiltonian mappings.

A better solution is offered by modifying the ansatz in Eq. 4.15 to introduce unitary operators [159, 160, 314, 315]. This can be achieved with an anti–Hermitian exponential such as:

$$\left| \bar{\Psi}(\vec{\theta}) \right\rangle_{\text{UCC},1} = \exp \left(\sum_{p1} \theta_{p1} (\hat{\sigma}_{p1}^+ - \hat{\sigma}_{p1}^-) \right) |\Psi\rangle_0, \quad (4.16)$$

for the first order expansion, and so on. The ansatz in Eq. 4.16 is known as *unitary coupled cluster* (UCC). The benefit of dealing with unitary operators, and therefore to generate variationally bound states, has one major drawback: the BCH series has no natural truncation, and therefore infinite terms should be ideally considered [314]. However, several studies noted empirically how already a truncation to the second order offers a sufficient approximation for many cases of interest [159, 314, 315], so that the scaling in the parameters required by the state preparation is in worst–case: $\dim(\vec{\theta}) \sim \mathcal{O}(n^2) \sim \mathcal{O}(N^4)$, i.e. quadratic in the size of the mapped quantum system n , and quartic in the number of orbitals N [161]. In practice, the number of parameters required for an appropriate ansatz does not scale with N (if $\|\hat{H}\|$ is held constant), but rather with the molecular properties of the system under consideration, such that far better scaling than this might be achievable [161, 318]. In conclusion, UCC can be used effectively in the context of VQE.

To improve the flexibility of the UCC ansatz, recent ideas have emphasised how, if the reference state in Eq. 4.12 is a natural output from a quantum simulator initialisation, it is also true that the starting state of a variational procedure $|\Psi\rangle$ could have been generated via an adiabatic bootstrap, or other guesses from classical methods, and not necessarily be a separable state [292]. In this regard, one can extend the ansatz correspondingly, to add flexibility, including the whole single–qubit Pauli group [159]:

$$\left| \bar{\Psi}(\vec{\theta}) \right\rangle_1 = \exp \left(i \sum_{p1,\alpha} \theta_{p1}^\alpha \hat{\sigma}_{p1}^\alpha \right) |\Phi\rangle, \quad (4.17)$$

again restricting to the first order for simplicity. The construction in Eq. 4.17 introduces a slight overhead in the parameter scaling, $\dim(\vec{\theta}) \sim \mathcal{O}(4n^2)$, but it makes evident how CC ansätze can explore states hard to prepare classically, as now $|\Phi\rangle$ can in principle be an arbitrarily entangled state [159]. The generalisation of Eq. 4.17 can be written in a more compact fashion as:

$$\left| \bar{\Psi}(\vec{\theta}) \right\rangle_{\text{UCC}} = e^{T - T^\dagger} |\Phi\rangle, \quad (4.18)$$

where $T = T_1 + T_2 + \dots$ is the cluster operator, whose term $T_1 = i \sum_{p1,\alpha} \theta_{p1}^\alpha \hat{\sigma}_{p1}^\alpha$ consists of single excitations away from the reference state, T_2 consists of double excitations etc. with $T \approx T_1 + T_2$ a valid approximation (see above).

In conclusion, UCC methods offer an efficient parametrisation of states of interest in the field of quantum chemistry. The UCC ansätze can also be mapped efficiently onto a quantum computer, as they prepare states according to unitary operators (see Eqs. 4.16, 4.17) that can be conveniently trotterised.

4.3.2.1 Parametrised probe states in molecular simulations

Moving back now to the molecular problem described by Eq. 4.3, once chosen a reference $|\Psi\rangle_0$ state, the description in terms of qubit–flips operated before translates into transitions between orbitals, so that a typical UCC ansatz has the form:

$$\hat{A}(\vec{\theta}) = \exp \left[\sum_{ij} \theta_{ij} (a_i^\dagger a_j - a_j^\dagger a_i) + \sum_{ijkl} \theta_{ijkl} (a_i^\dagger a_j a_k^\dagger a_l - a_l^\dagger a_k a_j^\dagger a_i) \right]. \quad (4.19)$$

An essential point which captures the flexibility of $\hat{A}(\vec{\theta})$ is which orbitals, labelled with the roman subscripts, are allowed to be evaluated in Eq. 4.19.

The most straightforward approach is to include all and only those terms that are included also in the original second quantised $\hat{H}(\vec{\theta})$. This would preserve the *symmetries* of the original problem, as no interactions among arbitrary orbitals are introduced, and is known as a Parametrised Hamiltonian (PH) ansatz. However, one might desire additional terms to be included, eventually breaking the symmetry of the original problem, in order to allow a more efficient variational search, or to introduce resilience to symmetry-breaking noises¹⁷, at the cost of searching through a wider parameter space (i.e. $\dim(\vec{\theta})$ increases).

4.3.3 VQE beyond ground-state estimation

QPE methods can be natively extended to excited states (Sect. 4.2), even if this is not true for their state–preparation protocols (Sect. 4.2.1). On the contrary, extending variational methods to target excited states is less direct. The two main directions investigated so far are to introduce linear response methods, or to rely on spectral folding¹⁸.

A linear response methodology is effective in preserving the advantages of VQE in terms of circuit depth, yet it requires additional sampling measurements and cannot refine approximate excited states [149, 308].

The Folded Spectrum (FS) method is also a variational search for the state $|\bar{\Psi}\rangle$, minimising the energy *folded* around an energy shift ϵ , as in the modified eigenproblem [160, 319, 320]:

$$(\hat{H} - \epsilon)^2 |\Psi\rangle = (\mathcal{E} - \epsilon)^2 |\Psi\rangle, \quad (4.20)$$

so that the solution is expected to be the eigenstate closest in energy to the chosen ϵ [319]. The choice of the specific classical optimisation schedule is again here completely independent (see Sect. 4.3). Unfortunately, the effectiveness of FS methods is known to rely heavily on the choice of ϵ , as values that are legitimate in principle can lead to highly non-convex folded energetic landscapes [319, 320]. Therefore, global optimisation methods are necessary to escape the local minima, corresponding to superpositions of excited eigenstates. Another limitation known in the FS literature is the quadratic increase in the number of terms of the effective Hamiltonian,

¹⁷ This will be discussed in Sect. 4.6 and following.

¹⁸ See also Sect. 4.7 for a survey of additional recent developments.

due in turn to the squaring of the relevant operator $(\hat{H} - \epsilon)^2$ [319, 320]. This overhead might be worsened in pragmatic implementations of the procedure in a quantum device, as classical matrix multiplications involved in Eq. 4.20 are replaced by cascading trotterised \hat{H} , so that more generally a polynomial increase in the cost of a single probe–state evaluation is expected, when FS is compared with standard VQE.

4.4 A Witness-Assisted Variational EigenSolver

In this section, we intend to present a novel method capable of targeting natively also excited states and energies, that like variational methods embeds a parametrised state–preparation routine, and yet does not require preliminary knowledge of the energetic landscape like the FS method [128]. We named this method witness-assisted variational eigenspectra solver (WAVES), as its cornerstone is the introduction of an “*eigenstate witness*”.

As already outlined in the Introduction to this chapter, we cannot expect WAVES to provide an efficient solution to the broadest classes of eigenproblems, but to be applicable to the same cases of interest for which efficient Hilbert–space parametrisations are known from quantum chemistry (Sect. 4.3.2). Crucially, the advantages of WAVES made possible the first successful experimental search for excited states on a quantum device (Sect. 4.5.4), anticipating similar results later accomplished with other methods [149].

4.4.1 The WAVES protocol

The approach proposed here is divided into three main steps (see Fig. 4.2A):

1. an ansatz-based variational search for the ground state,
2. a witness-assisted variational search for excited states, starting with an initial guess obtained from the ground state reference as outlined below, and
3. IPEA for the accurate energy estimate of the eigenstates found.

The quantum logic circuits for WAVES are shown in Fig. 4.2b and Fig. 4.2c. The variational search proceeds by preparing trial states $|\Psi\rangle_T$ in the target register, according to the ansatz and reference state $|\Phi\rangle$ (similarly to Sect. 4.3.2), and setting the control qubit to $|+\rangle_C$ (Fig. 4.2b). Trial states can always be written in the eigenbasis of \hat{H} as

$$|\Psi\rangle_T = \sum_j \lambda_j |\lambda_j\rangle \quad (4.21)$$

such that $\hat{H}|\lambda_j\rangle = \lambda_j|\lambda_j\rangle$. The combined state $|+\rangle_C \otimes |\Psi\rangle_T$ is then evolved through a controlled unitary ($C\hat{U}$) operation which embeds the unitary $\hat{U} = e^{-i\hat{H}t}$ for the evolution of $|\Psi\rangle_T$ according to the Hamiltonian \hat{H} , for a time t . The emerging control qubit state $\rho_C = \text{Tr}_T(\rho)$ is then analysed

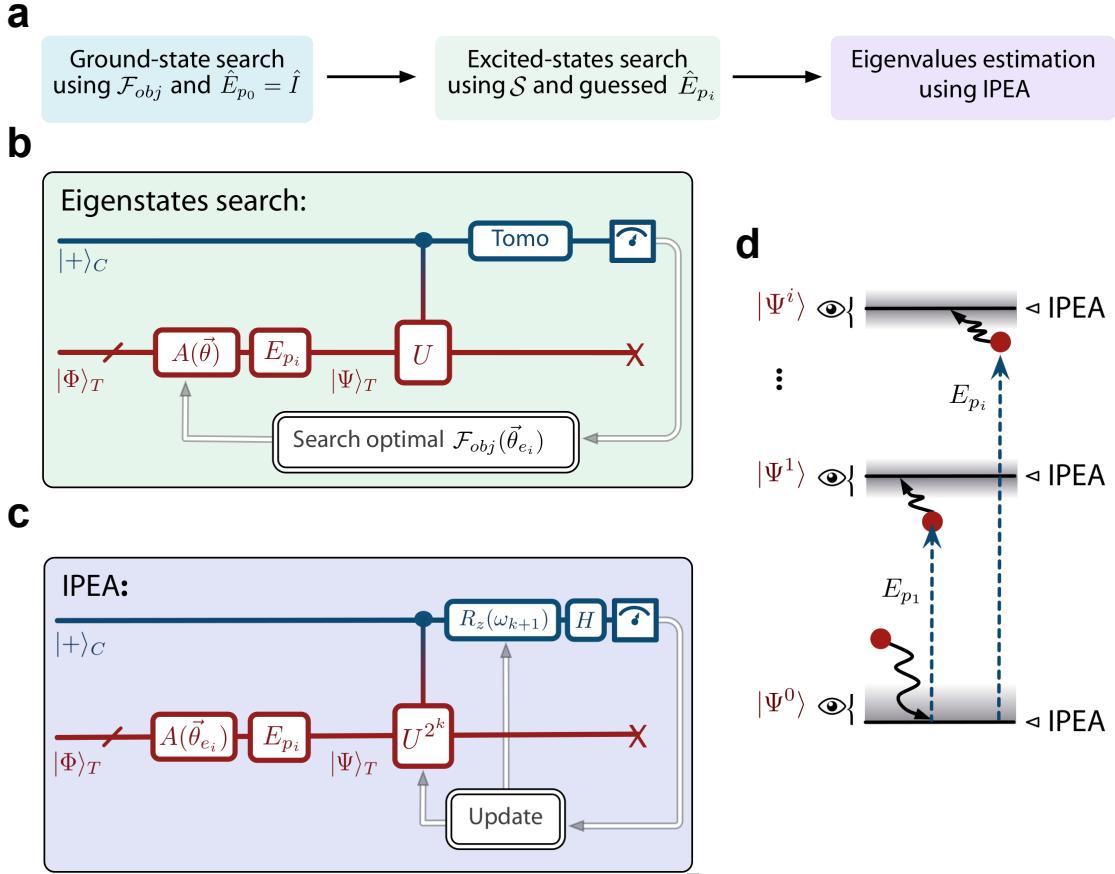


FIGURE 4.2. The WAVES protocol. **a** Flowchart describing the macrosteps of the protocol. **b** The quantum circuit invoked to estimate $\mathcal{F}_{obj}(\vec{\theta})$ for each trial state $|\Psi\rangle_T$. $\hat{A}(\vec{\theta})$ is the parametrised ansatz employed. In ground state searches, $\hat{E}_{p_0} = \hat{I}$ (i.e. no perturbation is applied) and $\mathcal{F}_{obj} = \mathcal{E} + T\mathcal{S}$. For excited state searches, $\mathcal{F}_{obj} = \mathcal{S}$, and the aprioristic perturbation \hat{E}_{p_i} is applied (see the main text for details). **c** Scheme for the quantum circuit performing the iterative phase estimation algorithm (IPEA) in WAVES (see also Sect. 2.1.5 for details). $\vec{\theta}_{e_i}$ refers to the parametrisation retrieved for the i -th eigenstate. The different colour coding in **b** and **c** is reminiscent of the different frequencies of the photons in the control (blue) and target (red) registries with our setup. **d** The intuition behind WAVES visualised in the energy diagram, with initial guesses for excited states, variational refinement using the witness, and IPEA projection on the eigenvalues.

by single-qubit state tomography, from which is possible to calculate the von Neumann entropy¹⁹ $\mathcal{S}(\rho_T) = \mathcal{S}(\rho_C)$, as well as an energy estimator $\mathcal{E} = -\text{Arg}[\langle \Psi | e^{-i\hat{H}t} | \Psi \rangle_T]/t$, evaluated using the off-diagonal elements of ρ_C . As an approximation of \mathcal{S} , it is also possible to adopt the linear entropy

$$\mathcal{S} = 1 - \mathcal{P} = 1 - \text{Tr} [\rho_C^2], \quad (4.22)$$

¹⁹ We remind that $\mathcal{S}(\rho) \equiv -\rho \log(\rho)$.

Named \mathcal{P} the purity of the reduced density matrix of the control qubit.

Measuring the entropy S enables us to variationally target excited states, and not only the ground state, because S behaves as an *eigenstate witness*. In particular, for almost any t , S is upper bounded by a monotonic function of the support of $|\Psi\rangle_T$ in the eigenbasis of \hat{H} :

$$S \leq 1 - \sum_j |\lambda_j|^4, \quad (4.23)$$

where the α 's coefficients are the same of the expansion in Eq. 4.21, while at the same time being lower bounded by:

$$S \geq 1 - \frac{[1 + \cos(\sum_{j,k} |\lambda_j|^2 |\lambda_k|^2 |\lambda_j - \lambda_k| t)]}{2} \quad (4.24)$$

for reasonably short t and maximum energy gap $\max_{j,k} |\lambda_j - \lambda_k|$. Details of the derivation for both boundaries is provided in Appendix C.1. Using these results, it can be seen that the linear entropy exhibits a local minimum ($S = 0$) if the target state is an eigenstate of the Hamiltonian, and it is bounded by well-behaved functions in an ϵ -neighbourhood about such optimal points, which justifies its adoption in the variational protocol.

The variationally optimal ground state simultaneously minimises the entropy $S(\rho_C)$ (being an eigenstate) and the energy estimate \mathcal{E} (because of the variational principle). Therefore, we have mapped the ground state retrieval into an optimisation problem, that adopts a physically motivated \mathcal{F}_{obj} combining these two figures of merit²⁰. Multiple choices could here be possible, but we focus on a two parameters, linear objective function case:

$$\mathcal{F}_{\text{obj}}(\mathcal{P}, \mathcal{E}) = A\mathcal{E} + T S \sim A\mathcal{E} - T\mathcal{P} = A\mathcal{E} - T \text{Tr}[\rho_C^2], \quad (4.25)$$

with $A, T \in \mathbb{R}^+$, and the approximation neglects constants. In the remaining of this chapter, the free parameter

- $A = 1$ for ground-state searches, to ease the interpretation of \mathcal{F}_{obj} in terms of a *free energy*, with T the only parameter trading off between energy and entropy optimisation,
- whereas $A = 0$ for excited-state searches, so that energy is effectively not used in such cases.

and T was also fixed to values optimised via noiseless numerical simulations. However, both A, T may be left as additional parameters to learn during the optimisation, running an adaptive version of the WAVES algorithm.

The key feature of our approach is therefore the optimisation of \mathcal{F}_{obj} of the form in Eq. 4.25, to variationally find a parametrisation for the input state that makes it as close as possible to an eigenstate, prior to refining this guess with QPE. Similarly to methods introduced in previous chapters, we call an *epoch* each iteration leading to an estimation of \mathcal{F}_{obj} . Defined an initial

²⁰ As outlined in Sect. 4.3, limiting the optimisation to \mathcal{E} alone is possible as long as ground states are of concern, however we stress in Sect. 4.6 how introducing S can improve the overall noise resilience of the protocol.

reference state $|\Phi\rangle$ and a complex $\vec{\theta}$ describing the ansatz-based state preparation $\hat{A}(\vec{\theta})$, i.e. $|\Psi\rangle_T = \hat{A}(\vec{\theta})|\Phi\rangle$, we can now detail the phases of our algorithm as follows (refer to Fig. 4.2):

1. Variationally search for the state parameters $\vec{\theta}_g$ that minimize the objective function \mathcal{F}_{obj} , thus obtaining the unitary for the ground state $\hat{A}_g = \hat{A}(\vec{\theta}_g)$.
2. Construct a unitary for an approximate i^{th} target excited state via $\hat{E}_{p_i}\hat{A}(\vec{\theta}_g)$, with \hat{E}_{p_i} a system dependent perturbation. Variationally search for the $\vec{\theta}_{e_i}$ that minimises $\mathcal{F}_{\text{obj}} = \mathcal{S}$ (i.e. choose $A = 0$ in Eq. 4.25) obtaining the unitary for the target excited state $\hat{A}_{e_i} = \hat{E}_{p_i}\hat{A}(\vec{\theta}_{e_i})$.
3. Using \hat{A}_g for the ground state or $\{\hat{A}_{e_i}\}$ for the excited ones in the state preparation, perform the IPEA which further projects each state onto the closest eigenstate and refines the energy estimate.

Using IPEA as the last step of the algorithm, as we will see in more detail in Sect. 4.4.2, is not crucial for the efficiency of WAVES, but it provides a quadratic speedup in achieving the same eigenvalue accuracy ϵ , which might prove practical in some circumstances. Perhaps most importantly, QPE protocols provide a projector²¹ onto the eigenvectors of the Hamiltonian, which cannot be achieved in general through a polynomial sized variational ansatz [321], a fundamental assumption for the overall scalability of WAVES. IPEA could be replaced potentially by any of the QPE protocols discussed in Chap. 2, inheriting the related properties. In particular, RFPE²² could be an interesting choice to increase the noise resilience of the overall WAVES algorithm, yet this replacement has not been investigated any further in this Thesis.

On the contrary, relying on a variational first step provides outcomes of easier interpretation²³ in situations where the initial input state $|\Phi\rangle$ has consistent overlaps with several eigenstates. In this way, WAVES addresses the state preparation problem in QPE on a non-fault tolerant quantum computer, beyond what can be achieved for ground-state preparation via adiabatic schemes (Sect. 4.2.1).

4.4.2 Complexity analysis of the WAVES protocol

The computational complexity of the WAVES protocol in finding eigenvalues of the Hamiltonian is essentially captured by the number of controlled unitary operations required to converge to the optimal parametrisation $\vec{\theta}$. This is dependent on several factors: the ansatz $\hat{A}(\vec{\theta})$ adopted, the prior information used for the initial guess $\vec{\theta}_0$, the proximity in the parameter space of other eigenstates²⁴ and finally the optimisation method used.

²¹ Note how IPEA, as well as other iterative QPE methods do not obtain a projected eigenstate as an *output state* of the quantum circuit, but still provide a *mapping* of such a projection as the list of applied gates and corresponding outcomes.

²²Sect. 2.1.6.

²³ The difficulty of this scenario for QPE algorithms was discussed in Sect. 2.1.6.1.

²⁴The role of some of these factors will appear more clear with the numerical analyses in Sect. 4.5.3.

As the first three factors are highly system-specific, we focus initially on achieving a generic understanding of the scaling of resources required by WAVES with the size of the system. Simulated and experimental results shown in this Chapter are obtained via a particle–swarm optimisation method²⁵. At each iteration \mathcal{F}_{obj} is measured for a swarm of trial states (*particles*), each corresponding to a randomly sampled $\vec{\theta}$ from a prior distribution. The outcome of the measurement is used to infer a posterior with lower mean \mathcal{F}_{obj} (see also Appendix Algorithm 7). It is thus interesting to provide the complexity of the protocol for such a swarm optimisation. This is provided by the following:

Theorem 4.3. *Let $\hat{H} \in \mathbb{C}^{2^n \times 2^n}$ be Hermitian and assume that, after $k \in \{1, \dots, N_{\text{iter}}\}$ epochs, the trial state is $|\Psi\rangle_T(k)$, expanded as in Eq. 4.21. We introduce a particle swarm $\Xi(k) := \{\vec{\theta}_j\}$ of constant size N_ξ , i.e. $\dim(\Xi(k)) = N_\xi \forall k$. Also assume that prior information and the physics of the problem bound the variance of the particles to a finite value: $\max_{\vec{\phi} \in \Xi(k)} \|\vec{\phi} - \mathbb{E}_{\vec{\theta} \in \Xi(k)}(\vec{\theta})\|_{\max} \leq x_{\max}$. Finally, we define: $\{\epsilon_\mu^2(k)\}$ ($\{\epsilon_\Sigma^4(k)\}$) as the tolerance in the (variance of the) trace of the covariance matrix of the sample mean, and $\Gamma := \max_k(x_{\max}(k)/\epsilon_\mu(k), x_{\max}^2(k)/\epsilon_\Sigma^2(k))$. Then, the number of applications of controlled $e^{-i\hat{H}t}$, for $[0, \pi/(2\|\hat{H}\|)] \ni t \in \Theta(\|\hat{H}\|^{-1})$, required by a particle swarm optimisation to learn an eigenvalue within error ϵ with probability at least 1/2 is:*

$$O\left(N_{\text{iter}}N_\xi \dim(\vec{\theta}) \left(\frac{\|\hat{H}\|^2}{\min_k \sum_i |\alpha_i(k)|^4} + T^2 \right) \left[\frac{\Gamma}{\delta} \right]^2 + \frac{1}{\epsilon} \right) \quad (4.26)$$

where δ is the maximum error in the evaluation of \mathcal{F}_{obj} allowed.

We omit the demonstration, which can be found in [128] along with equivalent theorems for the complexity of gradient-based methods. The above theorem implies that the cardinality of the parametrisation $\dim(\vec{\theta})$ represents a fundamental contribution to the scaling of the WAVES cost per iteration. Indeed, the term $\|\hat{H}\|^2$ depends on the nature of the problem and $\min_k \sum_i |\alpha_i(k)|^4$ on the quality of the ansatz: some cases of interest are known to perform well in practice [322], but it is beyond the scope of this Thesis to discuss in full generality the scaling of these terms with n . The quadratic dependence on Γ/δ is instead specific to the optimisation schedule chosen. Finally, the effective number of particles and iterations required N_ξ, N_{iter} can be intuitively expected to grow with $\dim(\vec{\theta})$ (as well as prior knowledge about the solution). Direct swarm methods have only been successfully applied to a parameter space of not more than few hundreds of independent variables, however it must be stressed how algorithmic improvements have been developed to extend such performances to thousands of variables with only polynomial²⁶ increase in N_ξ, N_{iter} [217, 323].

The main research problem that the application of WAVES relies upon is thus an appropriate ansatz, a problem discussed already in Sect. 4.1. In other words, the efficiency of WAVES is strictly dependent upon the conjecture of a polynomial scaling $\dim(\vec{\theta}) \propto n^\alpha$ in the number of

²⁵ See Algorithm 7.

²⁶ Empirically observed for a set of test functions in large multimodal optimisation [323].

spin-orbitals, for physically relevant systems [148, 159, 161, 187, 269, 303, 324]. Again, the idea is that the subspace of interest to capture the essential phenomena of the quantum system investigated involves only a handful of the available orbitals (often named the *active component* of the system [324]).

A problem closely related to the choice of the ansatz is the choice of the excitation operators \hat{E}_{p_i} used to target excited states (see Fig. 4.2a,b). An intuitive approach relies on the observation that interacting many-body Hamiltonians can often be decomposed into $\hat{H} = \hat{H}_0 + \hat{V}$, where $\hat{H}_0 = \sum_i \epsilon_i \hat{a}_i^\dagger \hat{a}_i$ is the dominant one-particle term and \hat{V} can be considered a perturbation term describing the interactions [308, 325]. A first approximation for \hat{V} is provided by a sequence of Hartree-Fock single-excitation operators $\hat{a}_i^\dagger \hat{a}_j$, each with a corresponding unitary $\hat{E}_p = \exp [\pi/2(\hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i)]$, similarly to the approach discussed in Sect. 4.3.2 for the preparation of states within a UCC ansatz. Whenever an increase in accuracy is required (e.g. for hard instances where electron correlations are strong enough), excitation operators can be readily computed for a set of chemical systems of interest [308], using advanced methods such as multi-configuration self-consistent-field (MCSCF) approximations [326].

The QPE algorithm phase adds a simpler contribution $1/\epsilon$ to Eq. 4.26, in absentia of any trotterisation [327]. Previous work by Wecker et al [161] highlighted how VQE requires a number of measurements to estimate the ground state energy that scales as $\mathcal{O}\left([\sum_j |h_j|]^2/\epsilon^2\right)$, given a Hermitian $\hat{H} = \sum_j h_j \hat{H}_j$ and $h_j \in \mathbb{R}$. This is the origin of the quadratic speedup obtained by invoking QPE protocols as the final step of the algorithm, once the guesses all have a major overlap with a single eigenstate²⁷.

In realistic implementations, a trotterisation²⁸ or linear-combination step should also be taken into account [328, 329], introducing an overhead of $\mathcal{O}(n^{5.5})$, to be multiplied by the above costs. The preliminary conclusion is thus that WAVES is efficient for all those cases where VQE is efficient, as it requires $e^{-i\hat{H}t}$ to be efficiently simulatable, the norm of the Hamiltonian to be bounded, a constant precision in the estimation outcomes, and a polynomial dependence of the ansatz and optimisation schedule parameters from the size of the system. It must be stressed how even for classes of systems of interest, this latter dependence is an open problem in VQE, and only empirical evidence of it has been provided so far for selected test cases [159–161].

4.5 Testing the WAVES protocol

In the following paragraphs, we test the behaviour of our protocol. We first introduce two classes of different systems that can be mapped onto the quantum simulator. We then perform both numerical simulations, and a full algorithm run on a photonic quantum simulator. Finally, we investigate the noise resilience exhibited by WAVES against both experimental sources of noise, as well as imperfect ansätze. A discussion of the outcomes concludes this section.

²⁷Excluding degeneracies, that are discussed within detailed numerical simulations in Sect. 4.5.3).

²⁸See Sect. 4.1.4.

| | truncated PH dim($\vec{\theta}$) | standard PH dim($\vec{\theta}$) | unrestricted UCCSD dim($\vec{\theta}$) | dim($\{\hat{E}_{p_i}\}$) | $ \Phi\rangle$ |
|-------------------|---------------------------------------|--------------------------------------|---|----------------------------|--------------------|
| H_2 (4 qubit) | 13 | 15 | 110 | 3 | $ 1100\rangle$ |
| H_3^+ (6 qubit) | 60 | 66 | 462 | 15 | $ 111000\rangle$ |
| H_3 (6 qubit) | 112 | 114 | N.A. | 15 | $ 110000\rangle$ |
| H_4 (8 qubit) | 181 | 185 | N.A. | 22 | $ 11110000\rangle$ |

TABLE 4.1. Ansätze used for numerical simulations, along with the corresponding molecular systems investigated, and the cardinality of their parametrisation - $\text{dim}(\vec{\theta})$. Further details are in Sect. 4.5.1. Finally, $\text{dim}(\{\hat{E}_{p_i}\})$ is the cardinality of the set of excitation operators, and $|\Phi\rangle$ the initial fiducial state.

4.5.1 Molecular Hydrogen test case

As a standard testbed for the performances of WAVES, we have used ground and excited electronic states of molecular Hydrogen systems H_2 , H_3^+ , and H_4 in a STO-3G basis, using the Jordan-Wigner representation (see Sect. 4.1.2). These represent 4-, 6-, and 8-qubit Hamiltonians respectively.

In our investigations we looked mainly at two different ansätze $\hat{A}(\vec{\theta})$, see Table 4.1. First, we utilised the PH ansatz, with the parameters $\vec{\theta} := \{t_{ij}\} \cup \{t_{ijkl}\}$ allowing for the variational searches. In some cases highlighting the limitations of the simple PH parametrisation, we also tested an unrestricted UCC-derived ansatz (UCCSD), allowing transitions among many more orbitals, compared to those restricted by the corresponding original $\hat{H}(H_x)$ (refer to Sect. 4.3.2.1). In all cases, the reference state $|\Phi\rangle$ on which $\hat{A}(\vec{\theta})$ acts is taken to be the Hartree-Fock state with the correct number of electrons. Finally, below we will refer synthetically to the single terms \hat{A}_i of each ansatz, such that:

$$\hat{A}(\vec{\theta}) = \exp \left[\left(\sum_i \theta_i \hat{A}_i \right) t \right], \quad (4.27)$$

making implicit the dependence of $\hat{A}(\vec{\theta})$ upon the evolution time t , as it is here a fixed parameter (see Appendix C.2.1).

Similarly, the approximate excitation operators used are defined in this basis as:

$$\hat{E}_{ij} = \exp \left[\frac{\pi}{2} (a_i^\dagger a_j - a_j^\dagger a_i) \right] \quad (4.28)$$

where we take i (j) to index the (un)occupied orbitals of the Hartree-Fock reference state. Note how the set of \hat{E} adopted does not depend upon the $\hat{A}(\vec{\theta})$ chosen, nor they are modified anyhow by the guessed $\vec{\theta}$ at any given point of the search.

4.5.2 Excitonic Hamiltonians test case

In Sect. 4.3 we have emphasised how many systems of interest in quantum chemistry, such as the electronic structure in molecules (Sect. 4.5.1) or spin lattice models, are represented

by interacting fermions. Nevertheless, interesting cases are offered by systems of interacting quantum particles and quasi-particles that are not fermions. It is thus interesting to approach more general systems, that might be crucial to perform quantum simulations in a variety of fields ranging from physics to chemistry, biology and materials science.

As one such example, here we show a bosonic Hamiltonian \hat{H} , later adopted for the experimental demonstration of WAVES (Sect. 4.5.4). Mapping the resulting evolution $e^{-i\hat{H}t}$ onto a quantum computer is a non-trivial task, because there is not a simple analogue of the transformations that map electronic occupation numbers to qubits (Sect. 4.1). In principle, it is possible to invoke broad-scope simulation methods that express \hat{H} as a sum of (at most) $O(n^6)$ one-sparse matrices, provided that \hat{H} does not contain interactions higher than two-body [159, 166]. However, the circuit depth and overheads in qubits demanded by such methods²⁹ makes them impractical for present day experiments.

Notwithstanding this open challenge, we select to experimentally demonstrate our WAVES approach via a 2×2 double-well potential Hamiltonian, offering a simple description for the coupled system of two Chlorophyll units in the 18-mer ring of the LHII complex, as found in the light harvesting complexes of purple bacteria [330]. This is a bosonic Hamiltonian, that for a single localised exciton on each Chlorophyll unit, can be written as:

$$\hat{H} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \quad (4.29)$$

where $\alpha = 1.46$ eV is the energy of the exciton on one of the Chlorophyll units and $\beta = 0.037$ eV represents the dipole-dipole interaction between the excitons. The qubit representation of this two-state Hamiltonian is obtained using compact mapping [159] and is

$$\hat{H}_{\text{qubit}} = \alpha \hat{I} + \beta \hat{\sigma}^x \quad (4.30)$$

where \hat{I} and $\hat{\sigma}^x$ are the usual Pauli matrices in the computational basis. Finally, we pick $\hat{H}' = \hat{H} - \ell \hat{I}$ redefining the set of eigenvalues by an offset $\ell \approx 1.24$ eV to ensure better performances of our algorithm³⁰.

The energy estimation in WAVES adopts the form:

$$\mathcal{E} = -\text{Arg}[\langle \Psi | e^{-i\hat{H}t} | \Psi \rangle_C]/t \quad (4.31)$$

t must then be chosen to prevent the 2π periodicity of the Arg function from folding the eigenvalues. This is a limitation already known from IPEA, where t is usually chosen small enough to avoid cases where the eigenvalues are estimated mod 2π (Sect. 2.1.5).

However, in the WAVES protocol additional boundaries for t emerge from considerations about the \mathcal{P} estimator (and ultimately \mathcal{S} as an eigenstate witness). The span in purity within the

²⁹Ultimately based upon a coherent implementation of a graph colouring method

³⁰Additional remarks can be found in Appendix C.2.

accessible Hilbert space dominated the choice³¹ of the evolution time $t = 26$, as described in Appendix C.2.1.

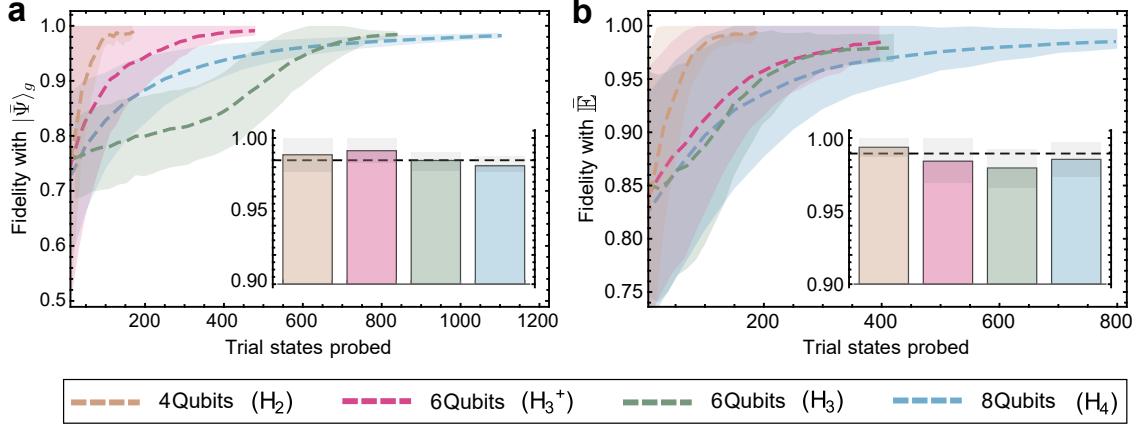


FIGURE 4.3. Numerical simulations for molecular-Hydrogen Hamiltonians. The cases studied refer to molecular Hydrogen systems (H_2 , H_3^+ , H_3 , H_4) with the full parametrised-Hamiltonian ansatz. **(a)** Ground state variational search. **(b)** Variational search for the targeted subspace of degenerate excited states. On the x-axis we refer to the cumulative number of trial states probed, i.e. particles in the swarm times variational steps. For ease of comparison, the x-axis origin has been shifted in **(A)** for the various cases to have equivalent fidelity for the average initial guess. Dashed lines report average fidelities with the shaded areas indicating a 67.5% confidence interval. The average fidelities achieved by the particle swarm optimisation for both ground and excited states, are calculated for 100 independent runs of WAVES. In all simulations a binomial noise model has been taken into account. Insets: bar-charts summarising final fidelities obtained by each search. The fidelity of 0.99 is highlighted by the dashed black line.

4.5.3 Simulated performances

We start the tests of the WAVES algorithm with a set of numerical simulations using the standard cases introduced in Sect. 4.5.1, i.e. modest-sized fermionic Hamiltonians. The scope is to investigate the capability of WAVES to detect different eigenstates for systems of dimensionality as high as 8 qubits. Interestingly, these examples exhibit several degeneracies in the spectrum, providing insight into the behaviour of our approach in such cases. We will not report for brevity the details of the analyses for the H_3 and H_4 systems (as they offer no additional insight about the performance of WAVES, compared with cases shown here).

³¹ The paragraph might suggest that preliminary knowledge of λ_g, λ_e is thus required in order to choose an appropriate t . However, a rough estimate $\Delta_{\max} \gg (\lambda_e - \lambda_g)$ would be already a satisfactory bootstrap for the procedure, which might then be refined after observing a reduced span in \mathcal{P} covered by the trial states used in the search.

The simulations of the WAVES algorithm are performed at a logical circuit level³². Uncertainty from binomial noise is nevertheless taken into account for tomographies of the control qubit (conceptually equivalent to the Poissonian noise model for a photonic setup, see Sect. 2.3.1). The level of noise chosen in these simulations is consistent with the experimental implementation³³ of WAVES (Sect. 4.5.4). The variational protocol proceeds as described in Fig. 4.2, with \mathcal{F}_{obj} optimised via the particle swarm method aforementioned. For the molecules H_2 , H_3^+ , H_3 and H_4 , we increased the number of particles to 8, 16, 30 and 50 respectively, which follows approximately linearly the number of parameters involved in the corresponding “parametrised Hamiltonian” ansatz (see Table 4.1).

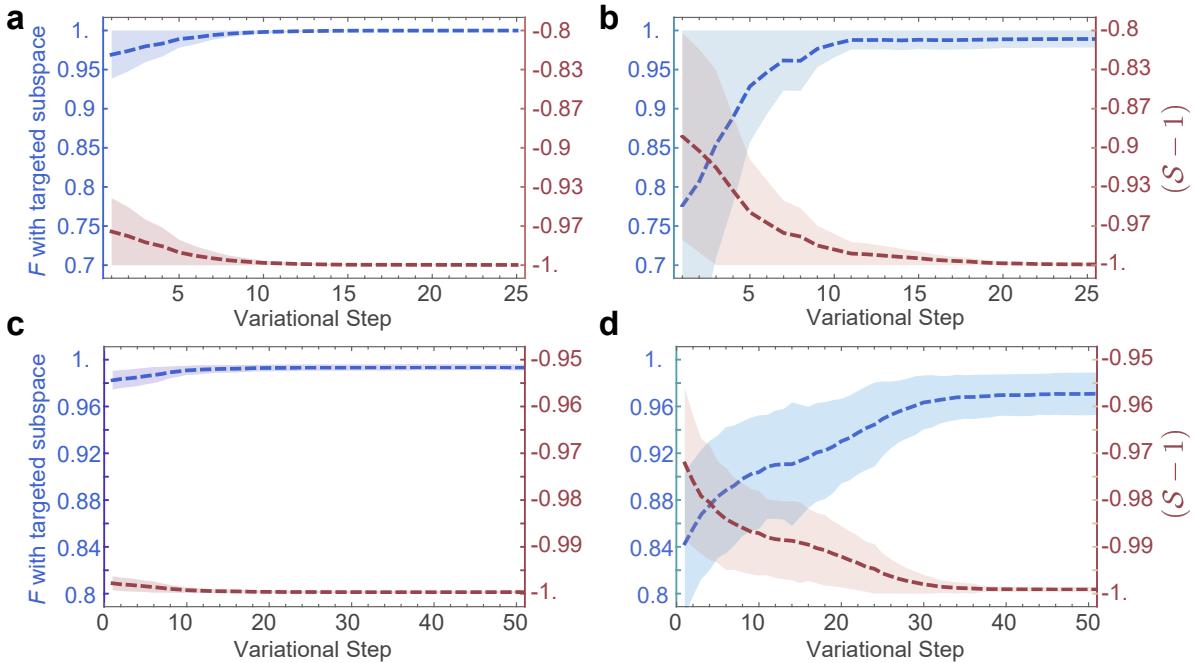


FIGURE 4.4. Convergence of the WAVES algorithm to a subspace of excited states in the case of the H_2 STO-3g (**a-b**, upper row) and H_3^+ STO-3g (**c-d**, lower row). On the left (**a&c**) we report simulations averaged over 100 runs of the algorithm, adopting the standard PH ansatz. On the right (**b&d**, emphasized by fainter colors), the algorithm is run with identical parameters, but adopting a truncated version of the same ansatz (refer to Algorithm 8). Respectively 2 and 6 operators were truncated in cases **b&d**.

For each system, we first variationally optimise the parametrization $\vec{\theta}$ provided by one of the ansätze in Table 4.1, optimising $\vec{\theta}_g$ for the ground state. In this step, no a-prioristic knowledge is assumed available, so that the initial guess is provided by the reference state: $\hat{A}(\vec{\theta}_0)|\Phi\rangle \equiv |\Phi\rangle$ (see also Table 4.1). The swarm of particles is drawn from an initial Gaussian distribution

³²Specifically, we mean that the simulations are implementation-agnostic, and therefore, the gates employed do not take into account e.g. trotterisation schemes, or a limited gate-set, nor gate infidelities.

³³In brief, it considers 1500 measurement repetitions, for each control qubit tomography

$\mathcal{N}(0, \sigma_{in})$ for each θ_i , with $\sigma_{in} \gg \max_i\{|\theta_i|\}$ but otherwise arbitrary. This justifies different average initial overlaps with the targeted state $|\Psi\rangle_g$, here known from direct diagonalisation of the Hamiltonian. Precisely, $|\langle\Phi|\Psi\rangle_g|^2$ is in the range $0.7 - 0.8$ for H_2 and H_3^+ , but as low as 0.2 for both H_3 and H_4 . WAVES demonstrates the capability to greatly improve even poor initial guesses, achieving fidelities higher than 95% with $|\Psi\rangle_g$. Whenever a better initial overlap bootstrapped the variational search, final fidelities are in excess of 99% (see Fig. 4.3a). Even if the limited amount of cases available cannot be used to draw general conclusions, it is worth noting that from the inset of the same figure, there appears to be no clear trend instead, with the growing dimension of the systems.

A more compelling case is given when WAVES attempts to learn a target excited state $|\bar{\Psi}\rangle$, by applying a set of different excitation operators of the form \hat{E}_p to the ground state parametrization $\hat{A}(\vec{\theta}_g)|\Phi\rangle$. Indeed, this variational search is the main novelty of the protocol. For this second phase, the initial set of swarm particles was initially drawn from a Gaussian distribution centred around θ_{i_g} for each parameter θ_i , with an initial standard deviation equal to $\max_i\{|\theta_i|\}$. When targeting a subspace of degenerate excited states, we replace the fidelity $F = |\langle\Psi|\bar{\Psi}\rangle|^2$ for the generic trial state $|\Psi\rangle$ with a generalised expression, given by the projection onto the Hilbert subspace $\bar{\mathbb{E}}$:

$$F = \left\langle \Psi \left| \hat{\prod}_{\bar{\mathbb{E}}} \right| \Psi \right\rangle = \left\langle \Psi \left| \sum_k (\bar{\Psi}_k \rangle \langle \bar{\Psi}_k) \right| \Psi \right\rangle \quad (4.32)$$

where $|\bar{\Psi}_k\rangle$ is a basis of $\bar{\mathbb{E}}$. In the following, therefore, we will refer to each set of degenerate excited states as the *excited subspace* \mathbb{E} that they span. We also assume for the scope of this work that QPE up to 32 bits of accuracy will refine the eigenvalue estimate provided by the first step of the WAVES protocol. Therefore, we choose to ignore energy differences smaller than $10^{-9} \text{ Ha} \approx 3 \cdot 10^{-8} \text{ eV}$, and consider eigenstates differing by smaller gaps degenerate, as these would be not distinguishable under spectroscopical accuracy³⁴. Finally, we remind the adoption of the simpler form $\mathcal{F}_{\text{obj}} \equiv -\mathcal{P} \equiv \mathcal{S} - 1$ for excited state searches.

In Fig. 4.5a we report a summary of the results for excited state searches in the H_2 system (whereas exemplary runs are detailed in Fig. 4.4a), using 3 different excitation operators \hat{E}_{p_j} , and measuring projectors onto all the 9 non-degenerate excited subspaces that characterize the corresponding Hamiltonian. The set of \hat{E}_{p_j} is expected to provide access to a portion of the spectrum with most of its support lying onto low-energy excited states, and a limited support onto higher energy excited states. We notice how the optimisation leads to a final fidelity F with one of the excited subspace(s) exceeding $99\%+$ in all cases (colour-coded in the figure). In particular, \hat{E}_{p_3} produces an initial guess that is a superposition of eigenstates belonging to the 1st and 4th excited subspaces. This is a highly non-trivial case: because of the initial standard deviation adopted in the algorithm, the initial swarm of particles spans fidelities that are as low as 0.5 with

³⁴See Sect. 2.3.

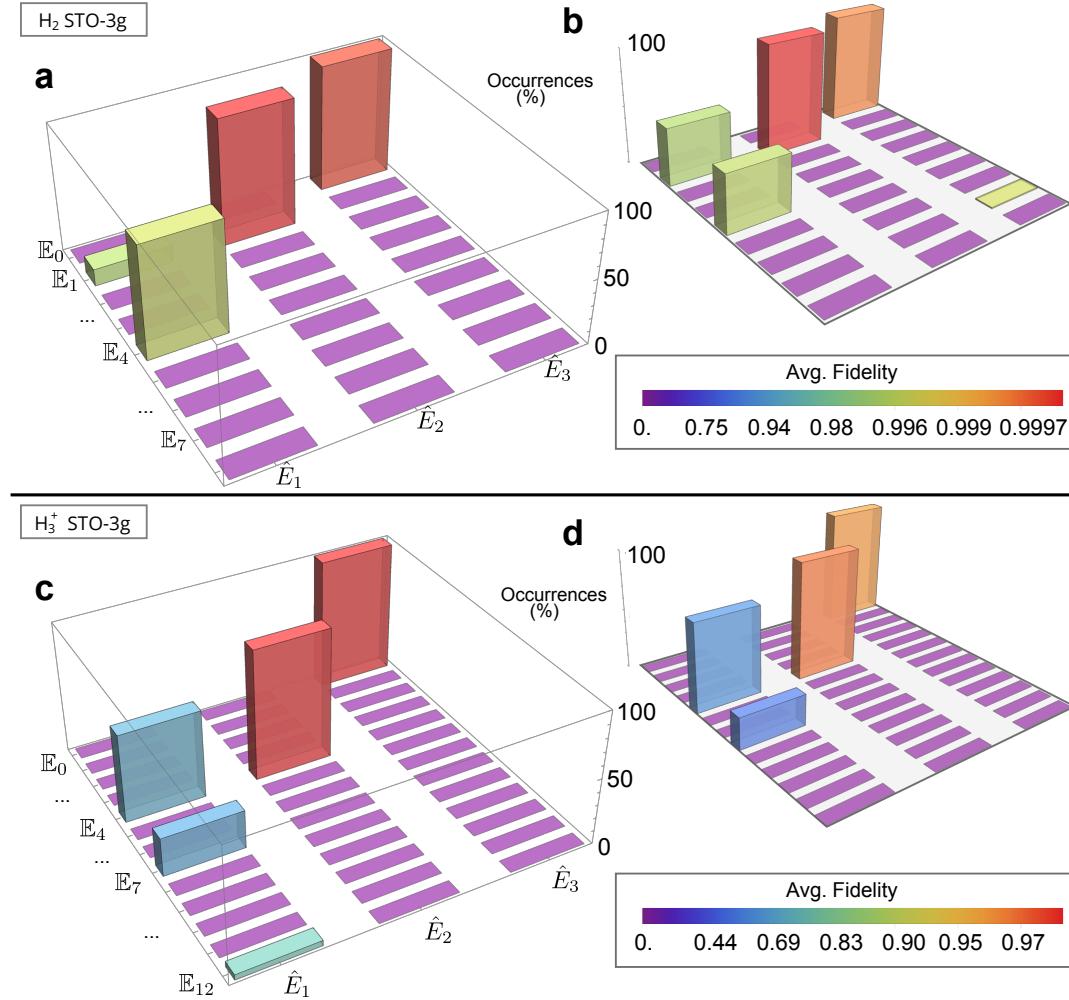


FIGURE 4.5. Synopsis of numerical simulations of excited states searches for the molecules H_2 and H_3^+ , adopting the full PH ansatz (**a&c**). Results for the synthetically reduced ansatz are in the inset outlined in light gray (**b&d**, axis labels not shown as they are the same of a&c). On one horizontal axis are reported the non-degenerate subspaces spanned by the eigenstates of the corresponding Hamiltonian (\mathbb{E}_i), which are 9 for H_2 and 23 for H_3^+ . On the other axis are the 3 different excitation operators of the form \hat{E}_p tested. Height of the histogram with coordinates $\{i,j\}$ corresponds to the relative frequency with which the swarm of particles collapses onto the i -th excited subspace, when \hat{E}_{p_j} is adopted in the second phase of the search. We accept the swarm has collapsed onto subspace \mathbb{E}_i , whenever the corresponding projector $\hat{\Pi}_i$ exhibits the highest overlap with the estimate $|\Psi\rangle$ at the end of the search (see Eq. 4.32). Fidelities reported for a certain \mathbb{E}_i and \hat{E}_{p_j} are calculated averaging the final fidelity achieved by all and only those algorithm runs, where the swarm had collapsed onto \mathbb{E}_i . The bottom-right bar legend explicates the colour coding.

either subspace. Performing IPEA directly with this initial guess would perform poorly³⁵. WAVES

³⁵ This can be qualitatively inferred observing the analysis presented in Fig. 2.3. Additionally, as emphasised in

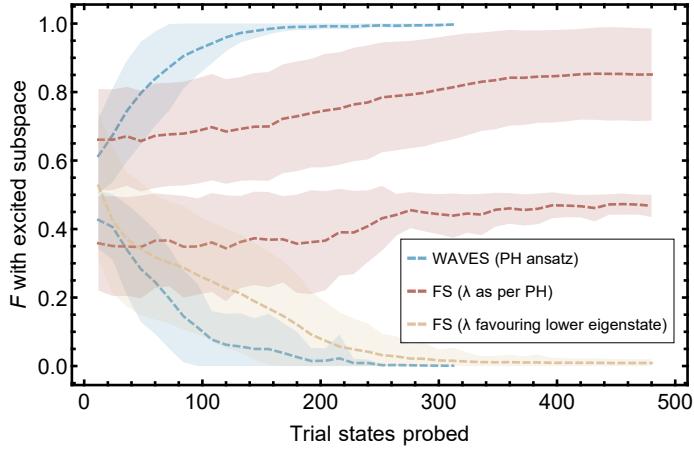


FIGURE 4.6. Behaviour comparison of the first part of WAVES, and an equivalent implementation of the Folded Spectrum (FS) method, when applied to the initial guess provided by the \hat{E}_{p_3} excitation operator for the H_2 system. FS simulations were run adopting two different values of the energy shift λ , colour-coded as in the legend. The fidelity reported is with the 4th excited subspace (F_4). We remark how all cases for both WAVES and FS implementation that converge to states with $F_4 \approx 0$ have $F_1 > 0.99$ (not displayed). Therefore, the convergence can be considered successful, even if towards a different excited subspace (the 1st). In all simulations, binomial noise was included (see Sect. 4.5.3). Ansätze in the legend refer to Table 4.1.

instead refines the initial guess detecting 87% of the times the 4th subspace. This preferential collapse approximately matches what expected from the initial guess $|\Psi_0\rangle = \hat{E}_{p_3}\hat{A}(\vec{\theta}_g)|\Phi\rangle$, that has 0.78 overlap with the 4th subspace.

The variational search in WAVES introduces advantages against the standard FS method³⁶. In particular, the original number of terms is retained and there is no user-defined parameter playing a decisive role such as the shift parameter λ [159, 160]. It is interesting to observe how these advantages can become evident already for systems as simple as the H_2 molecule, see Fig. 4.6. The comparison shows how the FS optimisation finds states with poor overlap with any true eigenstate if an accidentally problematic choice for λ is adopted, without providing any metric signalling the failed convergence. In this case, the choice for λ is reasonably dictated by the excitation operators derived by the PH ansatz (Table 4.1), but this points to an energy in between the E_3 and E_4 excited subspaces. Therefore, in order to proof-test the result of a single run, whenever the energy spectrum is not known in advance, FS methods shall require scanning different energy shifts, a strategy that can prove ill-posed whenever the energy gaps are expected to be small compared to the noise in the device. Additionally, already in Fig. 4.6 the

Appendix B.3, adopting a Folded Spectrum method would also be of little help in this case, if insufficient information about the energy spectrum is available.

³⁶Introduced in Sect. 4.3.3.

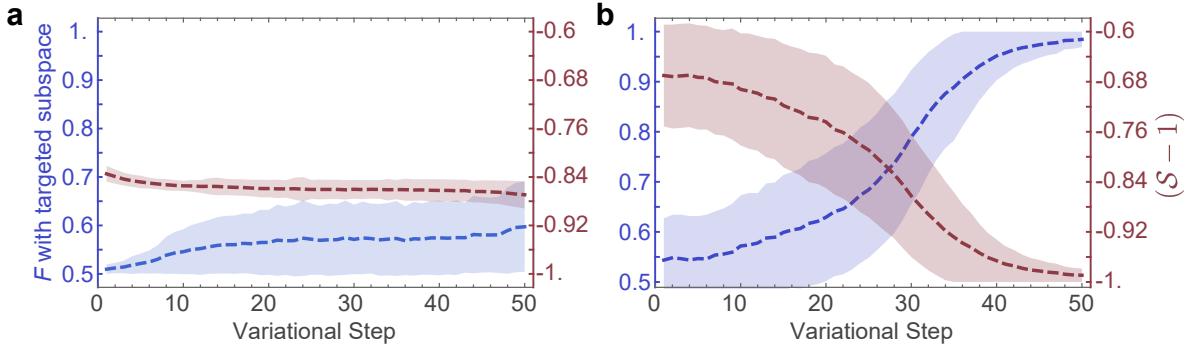


FIGURE 4.7. Comparison between different ansätze adopted in the search for excited states in the H_3^+ system. In both figures, the excitation operator is \hat{E}_{p_3} , and the fidelity plotted against $\bar{\mathbb{E}}_4$. In **a**, we report the average results from 100 runs of the first part of the WAVES protocol adopting the PH ansatz in the H_3^+ system and $t = 10$ in Eq. 4.27. The average entropy for the final estimate is $S > 0.12$. In **b** an equivalent optimisation adopts the unrestricted UCCSD ansatz from Table 4.1, but the same swarm-optimisation of the Algorithm 7. The average entropy for the final estimate is $S < 0.02$.

faster convergence of WAVES appears clear, due to the more convenient parametrisation of the problem not requiring an increase of the necessary terms.

The results for the H_3^+ molecule require additional effort, as they highlight the importance of the ansatz. This system is described by a bigger Hilbert space, with 23 non-degenerate excited subspaces. For brevity, we select but 3 different excitation operators $\hat{E}_{p_{1-3}}$, and compare the results with equivalent cases for the system H_2 . In particular, \hat{E}_{p_3} produces an initial guess that is a superposition of eigenstates belonging to 7 different excited subspaces. The initial $|\Psi_0\rangle$ has the most support (and hence overlap) on the 4th excited-subspace. In this case, the simple PH ansatz does not suffice to successfully converge to a single excited subspace - see Fig. 4.7a. However, in the same figure, a plot of the variational optimisation can provide a crucial sanity check of the eigenstate search, recalling Eq. 4.23 and the significance of the eigenstate witness S . A poor convergence rate for $S \rightarrow 1$, and its consistent degradation in single runs, when the evolution time t is increased flag a failure of the protocol³⁷, : either the ansatz has to be enriched, or the initial guess perturbed, for WAVES to achieve a successful convergence to a single $|\bar{\Psi}\rangle$ or $\bar{\mathbb{E}}$.

We remark how a successful variational search for H_3^+ can be immediately recovered via the adoption of a more accurate ansatz (the UCCSD in Table 4.1). This is shown in Fig. 4.7b, where a clear convergence of S to its optimal value occurs. Comparing Figs. 4.7a and b, a difference can be

³⁷ As it can be seen by comparison with Eq. 4.24. The role of increasing t , along with the boundaries imposed by the adopted \mathcal{T}_{obj} , is exemplified in Appendix C.2.1. However, we assumed to operate within a regime where trotterisation errors and decoherence phenomena are negligible, so that they are not expected to degrade the outcome quality, when t is increased.

noticed in the 67.5% confidence intervals, for both F and S , at “step 0”. This is due to attributing the same initial uncertainty to each variational parameter θ_i for both ansätze, but $\dim \vec{\theta}$ for the unrestricted UCCSD being about 8 times larger than the standard PH case.

Hence the prior guess spans a bigger portion of the Hilbert space. The final outcomes for all $\hat{E}_{p_{1-3}}$, adopting the extended UCCSD ansatz, are synthetically shown in Fig. 4.5c. The optimisation leads to a final fidelity F with one of the excited subspace(s) exceeding 99%+, in each case where the initial guess accurately targets a single excited subspace. For the excitation operator E_{p_3} , the preferential collapse of the particle swarm into either of the subspaces 4, 7 and 23 follows well what expected from the initial overlaps of $|\Psi_0\rangle$.

Finally, Figs. 4.3a-b summarise the final average outcomes of simulated variational searches with the WAVES approach for both the ground state, and exemplary excited states. Note that the second (IPEA) part of WAVES is not shown in the Figure, and therefore the final fidelities can be made higher³⁸, an option that is absent in previous VQE implementations. In Table 4.2, instead, we list synthetically the behaviour observed for the various scenarios investigated, according to the crucial figures of merit³⁹: i) whether the ansatz adopted allows the variational search towards the targeted $\bar{\mathbb{E}}$, and ii) how accurate it is, the initial guess provided by the excitation operator \hat{E}_p .

| | $ \Psi_0\rangle \xrightarrow{\hat{E}_p, \hat{A}(\vec{\theta})} \mathbb{E}_i, \bar{\mathbb{E}}$ | $ \Psi_0\rangle \xrightarrow{\hat{E}_p, \hat{A}(\vec{\theta})} \mathbb{E}_i, \bar{\mathbb{E}}$ |
|---|--|--|
| $\forall \mathbb{E}_i \neq \bar{\mathbb{E}} : \langle \Psi_e \hat{\Pi}_{\mathbb{E}_i} \Psi_e \rangle \ll \langle \Psi_e \hat{\Pi}_{\bar{\mathbb{E}}} \Psi_e \rangle$ | ✓ (convergence to $\bar{\mathbb{E}}$) | - |
| $\exists \mathbb{E}_i \neq \bar{\mathbb{E}} : \langle \Psi_0 \hat{\Pi}_{\mathbb{E}_i} \Psi_0 \rangle \simeq \langle \Psi_0 \hat{\Pi}_{\bar{\mathbb{E}}} \Psi_0 \rangle$ | ✓ (convergence to either $\bar{\mathbb{E}}, \mathbb{E}_i$) | ✗ (no convergence to any \mathbb{E}_i) |

TABLE 4.2. List of possible situations occurring in numerical simulations, when WAVES is performed with different ansätze and excitation operators, targeting a certain excited subspace $\bar{\mathbb{E}}$ from an initial guess $|\Psi_0\rangle$, using the ansatz $\hat{A}(\vec{\theta})$ and the excitation operator \hat{E}_p . To simplify readability, we use the notation: $|\Psi_e\rangle = \hat{E}_p \hat{A}(\vec{\theta}_g) |\Psi_0\rangle$ and $\hat{\Pi}$ as in Eq. 4.32. A generic excited subspace, other than the targeted, is indicated as \mathbb{E}_i . The case with outcome “-” cannot occur (see main text).

4.5.4 Experimental demonstration

The experimental demonstration of WAVES was performed on the same two-qubit silicon quantum photonic processor, already described in Sect. 1.4.4 and 1.4.4.2. We refer there for details concerning the fabrication and preliminary characterisation of the device. Also in this case, the

³⁸IPEA would further project the states prepared by the optimal $\vec{\theta}$ into the eigenstate with the highest overlap.

³⁹The case not reported in the Table would contradict the condition imposed about $|\Psi_e\rangle$: if $\langle \Psi_e | \hat{\Pi}_{\bar{\mathbb{E}}} | \Psi_e \rangle \sim 1$, then states prepared by $\hat{E}_p \hat{A}(\vec{\theta}_g) |\Psi_0\rangle$ must lie approximately in $\bar{\mathbb{E}}$.

WAVES algorithm, including the control of quantum gates, the data collection and real-time analysis, was automated interfacing the classical computer with the quantum photonic chip.

More in detail, $C\hat{U}$ operations are achieved also for this experiment with the entanglement based scheme introduced in Sect. 1.4.3. When implementing WAVES, the state preparation and unitary implementation are obtained respectively via the pre-compiled decompositions:

$$\begin{aligned}\hat{A} &= e^{i\phi_a} e^{i\phi_b \hat{\sigma}_z/2} e^{i\phi_c \hat{\sigma}_y/2} \\ \hat{U} &= e^{i\phi_c \hat{\sigma}_z/2} e^{i\phi_d \hat{\sigma}_y/2} e^{i\phi_e \hat{\sigma}_z/2}\end{aligned}\quad (4.33)$$

where the labels ϕ_i for the phases refer to the thermal phase shifters in Fig. 1.5. We remind that also in this case, no dedicated trotterised decomposition⁴⁰ into multiple two-qubit gates was performed. Finally, the single-qubit operations available for the control qubit allow both for the tomography required to estimate \mathcal{P} (and thus \mathcal{S} and \mathcal{F}_{obj}), as well as to perform the projective measurements needed by IPEA 2.3.1.

We used the quantum photonic chip to implement WAVES for calculating the eigenspectrum of the simplified exciton transfer Hamiltonian introduced in Sect. 4.5.2. We remark that such simplified model is not intended to provide an accurate description of the corresponding system, yet it serves as a useful demonstration and test for our algorithm. The minimisation of the objective function was performed again adopting the particle-swarm method⁴¹. Finally, for this Hamiltonian the perturbing unitary for the excited state corresponds to

$$\hat{E}_p = e^{i\pi\hat{\sigma}_z/2}. \quad (4.34)$$

We first describe the experimental results of the WAVES approach for the ground and excited state variational searches. In Figs. 4.8a-b we report colour-coded evolution of the swarm, achieving rapid convergence of the particle distribution towards the expected eigenstates of the Hamiltonian: the ground state $|-\rangle$ and the first excited state $|+\rangle$. For the ground state search, pessimistically assuming no pre-existing knowledge about the \hat{H}_{exc} , the prior is initialized to span uniformly the sub-section of the Hilbert space identified by the ansatz. For the excited state, instead, the search is initialized with the guessed state obtained by applying \hat{E}_p (Eq. 4.34) to the ground state.

As shown in Figs. 4.8c-d, \mathcal{F}_{obj} converges in 10-13 search steps to either eigenvalue. Figs. 4.8e-f report the fidelities with the ground and excited states: $99.48 \pm 0.28\%$ and $99.95 \pm 0.05\%$ respectively, as obtained from the mean of the particles distribution. All uncertainties are given by the variance of the posterior distributions: a well motivated error bar is among the amenable features deriving from the adoption of a swarm optimization method. Experimental fidelities were obtained evolving the dialled input state $|\Phi\rangle$ in a classical simulator, reproducing the photonic circuit in Fig. 1.5, according to the experimentally retrieved optimal phases $\vec{\phi}$. The simulator logical circuit is equivalent to the scheme in Fig. 4.2b.

⁴⁰ As defined in Eq. 1.8 and further commented in Sect. 4.1.4.

⁴¹ As outlined in Sect. 4.4.2, and provided as pseudocode Algorithm 7.

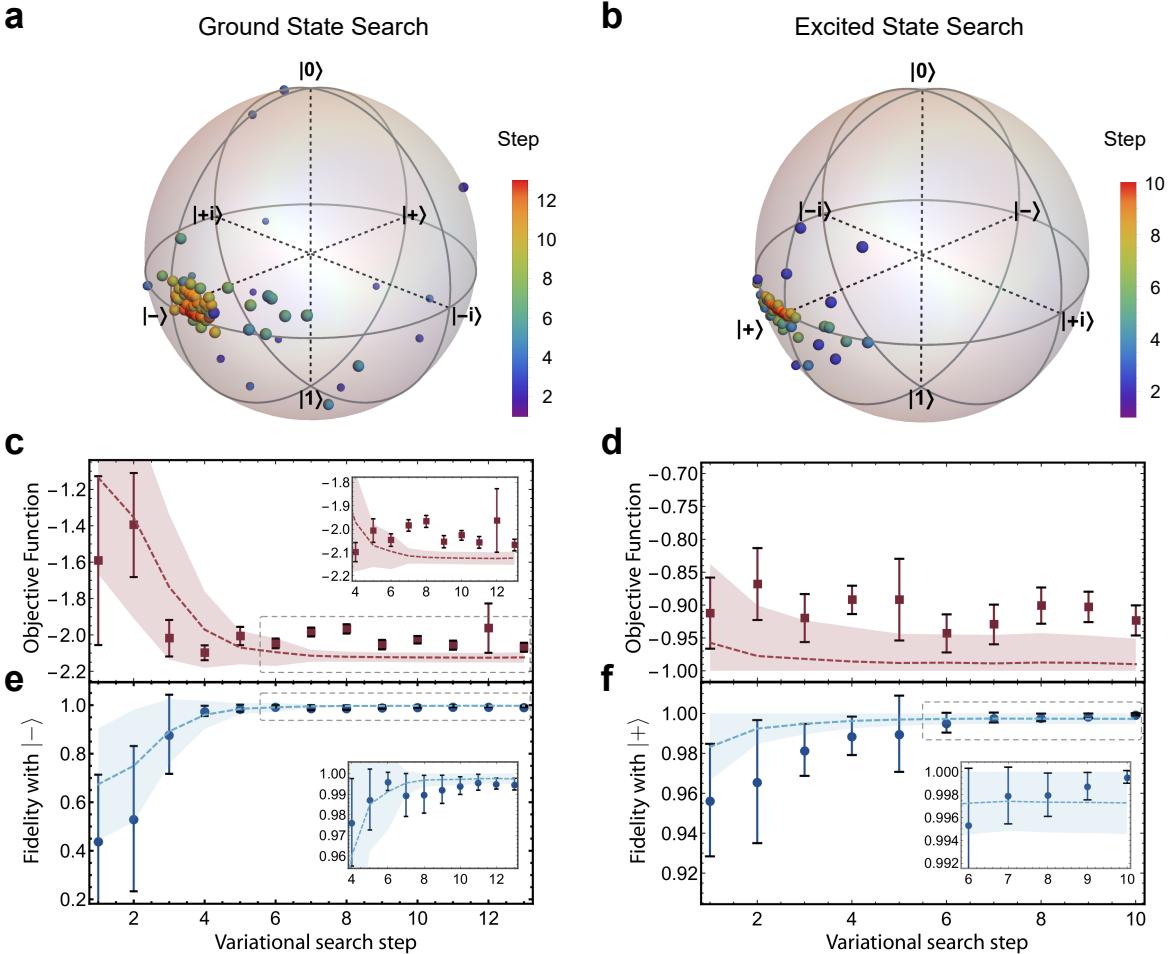


FIGURE 4.8. Experimental results for the variational search in WAVES. A Hamiltonian representing a single-exciton transfer between two chlorophyll units is implemented on the silicon quantum photonic device for an experimental test of the protocol. (a) and (b) Colour-coded evolution of the particle swarm for the WAVES search of the ground state ($|-\rangle$) and excited state ($|+\rangle$) shown on the Bloch spheres. Different colours correspond to different steps of the search protocol. For the ground and the excited state searches we report in: (c) and (d) the evolution of \mathcal{F}_{obj} ; (e) and (f) the fidelity ($F = |\langle \Psi | \bar{\Psi} \rangle|^2$) versus search steps, converging to a final value of $99.48 \pm 0.28\%$ and $99.95 \pm 0.05\%$ respectively. Error bars are given by the variance of the particles' distribution and the uncertainty originated from photon Poissonian noise. Dashed lines are numerical simulations of the performance of the algorithm, averaged over 1000 runs, with shaded areas representing a 67.5 % confidence interval. Insets: behaviour close to convergence.

The successful convergence of $|\Psi\rangle_T \rightarrow |\bar{\Psi}\rangle$ is achieved by optimizing the \mathcal{F}_{obj} function. In particular, for the ground state search we used a small value of T^{42} in \mathcal{F}_{obj} , while for the excited

⁴² $T = 1.25$, as optimised via preliminary numerical simulations.

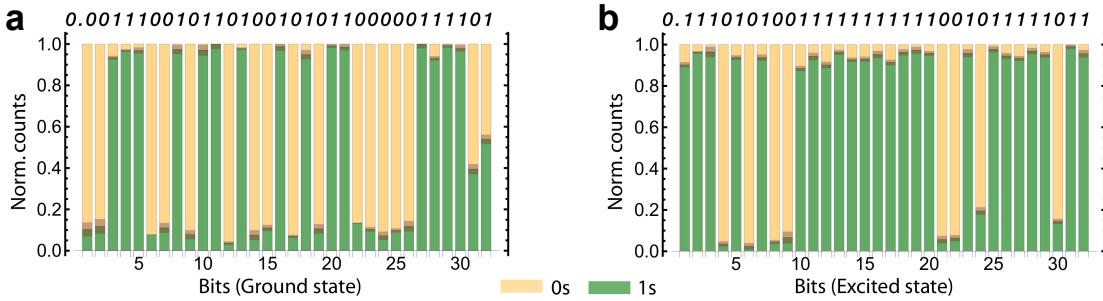


FIGURE 4.9. Experimental results for the IPEA-based refinement in WAVES. IPEA is performed on the converged final variational estimate obtained after the search outlined in Fig. 4.8 for the excitonic Hamiltonian. As a bar plot, normalised photon coincidences used to calculate the 32 IPEA-estimated bits of the eigenphase are reported for both the ground (**a**) and excited state (**b**). The theoretical bit value is reported above each bar. Uncertainties arising from Poissonian noise are reported as shaded areas on the bars.

state case we used $\mathcal{F}_{\text{obj}} \equiv -\mathcal{P}$ (recall Sect. 4.4.1). Closer to convergence, we notice fluctuations in the estimate of \mathcal{F}_{obj} , that might be interpreted as residual thermal cross-talk (see Sect. 1.4.4.2). Nevertheless, the algorithm rapidly succeeds to retrieve the correct eigenvalues, indicating robustness against this kind of noise.

After the variational search, the final step consists in improving the accuracy via IPEA, see Fig. 4.2. In our implementation we took advantage of the circuit reconfigurability, mapping each \hat{U}^{2^k} directly into the chip parameters. However, in universal quantum computers, \hat{U}^k can be efficiently achieved without classical pre-compilation by cascading k copies of \hat{U} [162]. The IPEA estimated the binary fraction expansion of the eigenphase $\varphi(\text{mod } 2\pi)$ for both the ground and excited state energies up to 32 bits (i.e. a precision of 2.9×10^{-9} eV). The normalized photon counts are reported in Figs. 4.9a-b for all the 32 bits. Such precision is higher than what is typically achievable by spectroscopic methods (Sect. 2.3).

4.6 Noise resilience of the WAVES variational search

In the light of this Thesis' point of view, a major question left unanswered so far about the WAVES approach is its resilience to various noises, or more in general, imperfect implementation. Inspired by the experimental implementation outlined in Sect. 4.5.4, we discuss two main sources of noise, expected to play a major role in affecting the estimators \mathcal{E} and \mathcal{P} :

- Poissonian noise arising from coincidence events counts;
- Infidelity in the gates implementation⁴³, affecting both the state preparation and the $C\hat{U}$

⁴³Such infidelities, for the photonic circuit considered, arise mainly from uncertainty in the phases implemented

implementation.

The effect of these noises upon QPE algorithms (and therefore the last phase of WAVES) has already been discussed in detail in Chap 2, and therefore here we are left with investigating the resilience of the variational search alone. In general, variational methods exhibit an intrinsic resilience to infidelities in the state preparation phase, such that it is customary the following [159]:

Definition 4.2. Named $\hat{A}(\vec{\theta})$ the ansatz for preparing trial states $|\Psi\rangle_T$ in a variational problem, we call *variationally suppressible* all those noise sources leading to an altered $\hat{A}'(\vec{\theta})$ such that $\exists \vec{\xi}: \|\hat{A}(\vec{\theta}) - \hat{A}'(\vec{\theta} + \vec{\xi})\| < \epsilon$, for arbitrary $\epsilon \in \mathbb{R}^+$.

Notice that Def. 4.2 does not imply any knowledge about the nature of the noise, as it relies on the fact that the variational procedure will converge to $\vec{\theta} + \vec{\xi}$ as the optimal parametrisation, compensating the infidelity in the ansatz. A counterexample to such benign cases is given by errors that break the symmetry of the ansatz [159], preventing any correction $\vec{\xi}$ to satisfy the condition in Def. 4.2. These cases are expected to exhibit a behaviour similar to what observed for ansätze that are not sufficiently flexible⁴⁴.

A different situation arises when infidelities affect the unitary evolution $\hat{U} = \exp\{-i\hat{H}t\} \rightarrow \hat{U}' = \exp\{-i\hat{H}'t\}$. In this case, a basic noise characterisation leads us to introduce an additional:

Definition 4.3. Named $\bar{\mathcal{F}}_{\text{obj}}$ the optimal objective function in a noiseless variational problem, we call *algorithmically suppressible* all those noise sources leading to a zero-mean additive white noise perturbed $\mathcal{F}_{\text{obj}}(t)$.

The reason to invoke a white noise is that this can be naturally filtered out by any optimisation schedules that do not rely on *hard choices* for the update of $\vec{\theta}$. In this way, fluctuations in $\mathcal{F}_{\text{obj}}(t)$ due to gate infidelities when implementing \hat{U}' can be statistically suppressed and rendered negligible. Such cases, can therefore be dealt with by choosing an appropriate algorithm for the optimisation, and for instance, swarm methods as those adopted here are known to be particularly well suited for compensating white noise in the estimators [331]. On the contrary, systematic non-zero-mean alterations in $\mathcal{F}_{\text{obj}}(t)$ will almost certainly lead the variational procedure to identify an incorrect $\vec{\theta}'$, which for example might correspond to the eigenstate parametrisation of a consistently different Hamiltonian \hat{H}' .

Now that this preliminary discussion is in place, we can move to two different kinds of test. A first one investigates realistic noise sources in the photonic implementation, testing to what extent they are *suppressible* in WAVES, according to the two correction processes identified by

by the thermo-optical phase shifters on chip, see Fig. 2.6. Potentially, they might also be due to imperfect fabrication of the chip, e.g. unbalanced MMI's or DCp's, however the preliminary characterisation found such imperfections negligible.

⁴⁴As discussed in Sect. 4.5.3 and more systematically below.

Defs. 4.2 and 4.3. Then, we move on to address the simulation of errors partially breaking the ansatz symmetries.

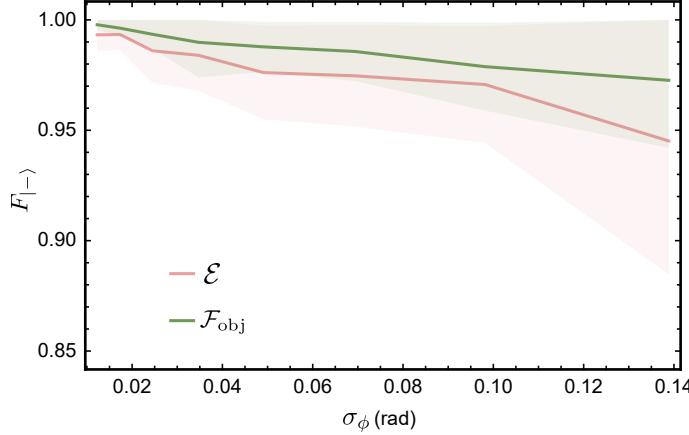


FIGURE 4.10. Numerical simulations of the variational ground-state search robustness for the bosonic 1-qubit \hat{H} against imperfect phase implementation, using \mathcal{F}_{obj} or \mathcal{E} alone. Infidelities in the gate implementation are modelled as Gaussian noise in the phase shifting components of the photonic circuit. Average fidelities with the true eigenstate $F_{|-\rangle}$ (solid lines) along with 67.5% confidence intervals (shaded areas) are reported after algorithm's convergence.

4.6.1 Resilience to experimentally realistic noises

In this section we try not only to answer the question of what degree of an imperfect photonic implementation could be tolerated by WAVES, but also understanding if the proposed objective function \mathcal{F}_{obj} (Eq. 4.25) provides any advantage (or disadvantage) in the robustness of the search. Therefore, we compare it to a variational search based upon the energy estimator \mathcal{E} alone (and thus equivalent to standard VQE, Sect. 4.3).

Data from 100 averaged numerical simulations, running the variational search for the bosonic Hamiltonian (Eq. 4.30) in both cases, are reported in Fig. 4.10 for different degrees of uncertainty in the phases implemented on the photonic chip. This is done replacing each correct phase $\bar{\varphi}_i$ – required to manipulate the target qubit⁴⁵ – with a synthetic value φ , sampled from a Gaussian distribution $\varphi_i \sim \mathcal{N}(\bar{\varphi}_i, \sigma_\varphi)$, equivalently to what already discussed in Sect. 1.4.4.3. As σ_φ increases, the behaviour of the simulated physical circuit deviates from the ideal case. In this scenario, adopting the objective function \mathcal{F}_{obj} appears a consistently more robust strategy than using \mathcal{E} alone as an estimator: a difference that increases with the level of uncertainty in the implemented phase. In particular, already for $\sigma_\varphi \simeq 0.14$ the final fidelity achieved in the

⁴⁵ Phases involved in the control register have been excluded from this simulation.

search with \mathcal{E} alone falls below $F_{|-\rangle} = 95\%$, whereas \mathcal{F}_{obj} keeps providing accurate ground state estimates.

We finally remark how this same noise model was applied to infer the expected simulated performances of the circuit, reported as dashed lines in Fig. 4.8c-f, and assuming $\sigma_\phi \simeq 0.012 \text{ rad}$, in agreement with the device characterization data. The noise model was capable of reproducing quite accurately the experimental findings within statistical error, with slight discrepancies likely to be explained by uncharacterized residual thermal cross-talk between the phase shifters (not taken into account in the model). This leads to systematic errors in the implemented phases, that cannot be suppressed naturally by WAVES.

4.6.2 Synthetically reduced ansätze

Noises breaking the symmetry of an ansatz prevent it from preparing variationally optimal states, as these would fall out of the Hilbert subspace spanned by the ansatz itself [159]. In this sense, the impact of such noises can be analysed equivalently by synthetically truncating an ansatz otherwise sufficient to learn the targeted eigenstates. At the same time, this approach addresses one of the arguments often levied against variational methods, i.e. that they strongly rely on the accuracy of the ansatz chosen for the variational state preparation. Indeed, for correlated and complex quantum systems one expects the ansätze obtained via classical approximation methods to become poor⁴⁶.

In Sect. 4.5.3 we presented and discussed simulations mostly obtained with a standard PH ansatz, expected to achieve a reasonable accuracy in describing weakly correlated systems. Here we compare such results with a truncated ansatz \hat{A}' , obtained after a gradual removal of those terms in the PH ansatz, corresponding to the smallest values $|\theta_g|$. Such terms are expected to contribute the least in the preparation of $|\bar{\Psi}\rangle$. A more detailed breakdown of the procedure is reported in Algorithm 8, and a synthetic description of the ansatz obtained can be found in Table 4.1. \hat{A}' thus simulates those cases where only limited knowledge on the physical system is available.

The results are summarised for both ground and exemplary excited state searches of the hydrogen molecules in Fig. 4.11 a&b, respectively, adopting the natural initial guess $\hat{A}(\vec{\theta}_0)|\Phi\rangle \equiv |\Phi\rangle$. For all the cases investigated, the variational search in WAVES is able to consistently identify both the ground and targeted excited states. Fidelities of approximately 95% are achieved, even for cases where the initial guess retained only small overlap with the targeted excited state and the incomplete ansatz reduces the accuracy in the state preparation.

Additional details are presented in Figs. 4.4 b&d, 4.5 b&d, for excited states searches in the H_2 and H_3^+ systems, to be compared with (a&c) in the same figures for a standard PH ansatz. The truncated ansatz leads to a consistent degradation in the fidelity, and its uncertainty, of the final estimate. Also, the poorer initial guess provided by the restricted ansatz penalises the rate

⁴⁶See Sect. 4.3.2

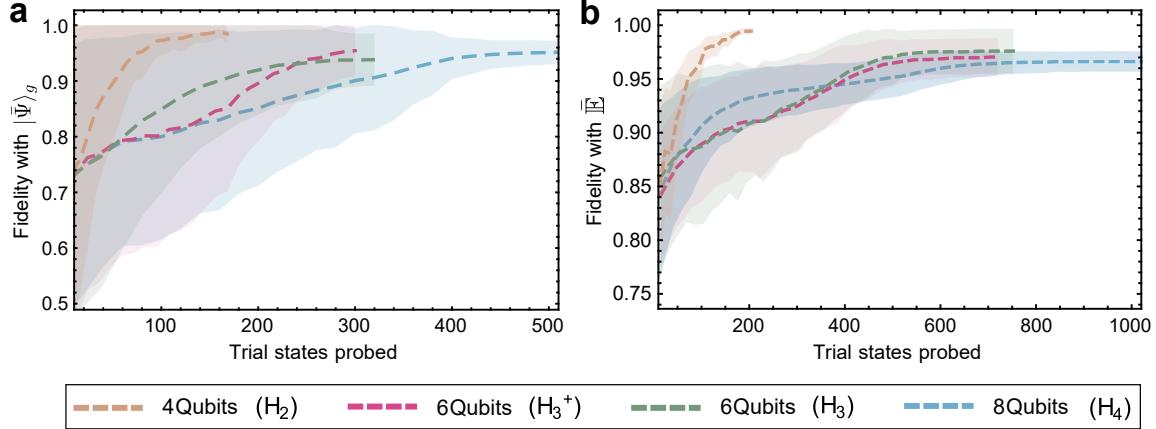


FIGURE 4.11. Numerical simulations of WAVES adopting the synthetically truncated PH ansatz, and searching eigenstates of the molecular Hydrogen systems (H_2 , H_3^+ , H_3 , H_4). **(A)** Ground state search. **(B)** Search targeting a degenerate subspace using a perturbation \hat{E}_{p_i} . The average fidelities achieved by the particle swarm optimization are calculated for 100 independent runs of WAVES. Dashed lines are average fidelities. Shaded areas indicate 67.5% confidence intervals. The x-axis reports cumulative numbers of trial states probed (number of particles times variational steps). For ease of comparison, steps have been counted only after an equivalent fidelity had been achieved. All projective measurements are here inclusive of a binomial noise model in the simulations.

of convergence. Nevertheless, the algorithm detects most excited subspaces with high fidelity, proving unaffected by slight inaccuracies in the ansätze. In addition, the reduced accuracy and overlap with the targeted eigenstate of this synthetic ansatz diminish the tendency of WAVES to collapse into the closest excited subspace, as evidenced by the different frequencies with which the swarm collapses into either one of the excited subspaces E_i (Fig. 4.5b&d). A worst-case scenario is displayed in Fig. 4.4d (corresponding to H_3^+ and ‘Exc. 2’). Here, we intentionally removed from \hat{A}' operators that are fundamental for a successful search of the Hilbert space, in proximity of the targeted E_i . This situation cannot be completely recovered by WAVES, so that an impoverished $|\Psi_0\rangle$, a slower convergence, and suboptimal fidelity $F \approx 96\%$ are observed.

4.7 Discussion and final remarks

We have introduced the concept of eigenstate witness and used it to develop a witness-assisted variational eigenspectra solver (WAVES). The tests provided show how the WAVES protocol, introduced in Sect. 4.4, is able to consistently find variationally both the ground and excited states of physical Hamiltonians with high fidelities ($\approx 99\%$ in average) in simulations as well as experimental implementations. The variational step serves as a state preparation stage for a

QPE refinement, and the enhanced approximation of the target eigenstates provided leads to a higher success probability in estimating the corresponding eigenvalue via QPE, which grows exponentially with the system's size N . This, in turn, reduces the overall cost of solving the eigenproblem. Adopting QPE is fundamental to achieve the projection onto the eigenvectors, thus addressing the shortcomings of a polynomial-sized ansatz that would not suffice to achieve accurate estimates of the eigenvalues. In WAVES experiments, eigenvalue estimates beyond spectroscopic accuracy⁴⁷ were demonstrated using IPEA, in addition to the variational search. Careful choices for the QPE procedure⁴⁸ may also provide means to learn exponentially quickly one of the eigenvalues belonging to almost-degenerate eigenstates. These might originate in complex systems of increasing size, where the energy gaps and/or the distance in the Hilbert space could be indistinguishable, within the experimental precision provided by variational methods.

WAVES offers key improvements over previous protocols. First, it proposes a genuine quantum method, in contrast to quantum-classical linear response methods (LR) [308] or the FS approach [160]. LR methods do not require non-linear optimisation schedules, but this may limit their accuracy, and both rely upon no final eigenstate projection phase⁴⁹, unlike WAVES, being ultimately bound by the accuracy of the variational ansatz. Importantly, the eigenstate witness provides *an independent test of the protocol's success*, detecting failure cases of convergence to local optima that do not represent a single eigenstate nor excited subspace, as shown in Sect. 4.6, and increasing the robustness against moderate sources of noise (Sect. 4.6.1). Also, WAVES retains the original norm, spectrum, and number of terms in the Hamiltonian $O(n^4)$, nor it depends upon an energy shift parameter, unlike FS methods⁵⁰. Finally, WAVES does not require lengthy adiabatic preparation [148, 159, 273], outlined for alternative methods in Sect. 4.2.1, and demands the availability of projectors at the end of the evolution only for the control qubit, whereas traditional VQE implementations need to measure arbitrary output qubits (Sect. 4.3.1).

These advantages come at the cost of controlling the evolution of the target register with an ancillary qubit, which is avoidable in previous VQE proposals and can pose issues to some quantum hardware architectures [332]. Also, in WAVES, the ability to find specific eigenstates relies on the quality of the excitation operators⁵¹. Further optimisation on the objective function, e.g. including the use of an energy penalty, can in principle overcome some of these limitations.

In terms of resource costs, the insets of Fig. 4.3 show empirically how the final fidelities achieved by WAVES variational searches do not decrease (within errors tolerance) when increasing the size of the Hilbert space. The discrepancy in the final fidelities, achieved for the various

⁴⁷ We name it an *accuracy* of 32-bits, because the outcomes with 32-bits of precision were checked against the *true* eigenvalues of the test \hat{H} , obtained via direct numerical diagonalisation.

⁴⁸ See Sect. 2.1.6.1.

⁴⁹ We are here excluding the availability of ancillary qubits and controlled operations that would make QPE possible, as avoiding these is the major reason supporting the adoption of LR and FS, see below.

⁵⁰ See Sect. 4.5.3, and more in detail Fig. 4.6, as well as Appendix B.3.

⁵¹ See Eq. 4.28.

cases, shall not be misinterpreted as a fundamental limit deriving from the quality of the initial guess provided to WAVES. Indeed, it is worth mentioning that in order to simulate realistic experimental conditions for near-term WAVES implementations, beside a noise model in the control-qubit tomography, we restricted the number of particles to keep execution times for the protocol within reasonable limits for state-of-art photonic implementations⁵².

Although these study cases do not imply scalability of the approach, they provide an encouraging result⁵³. This intuition has been formalised better in Thm. 4.3, which suggests that to keep constant the algorithm performances with the dimensionality of the problem, a sub-exponential increase in number of particles⁵⁴ N_ξ and iterations N_{iter} is enough, provided that a polynomial parametrisation such as those listed in Table 4.1 applies. In particular, adopting a UCC ansatz and low-order trotterisation, WAVES is expected to run in times that scale as $O(Mn^{5.5})$, with n the number of spin-orbitals and M the number of variational parameters, comparing well with the estimates reported in [329] for the Nitrogenase case, a system notoriously hard to simulate on classical hardware.

We remind the reader how WAVES, nor any other quantum method, cannot efficiently prepare eigenstates of a generic Hamiltonian, as this would imply a highly unlikely collapse in computational hierarchies [283]. As long as specific polynomial ansätze are instead of concern, the number of variational parameters, along with those describing the optimisation schedule, required to achieve chemical accuracy will likely require additional empirical studies to be precisely estimated. In full generality, it can still be stated how the use of QPE protocols will still provide a quadratic speed-up compared to standard VQE, when estimating the energy of an eigenstate within a chosen precision. Such advantages might prove relevant when targeting chemical accuracies, and beyond [161].

By introducing new objective functions for variational algorithms, WAVES opened the way to the investigation of new methods for computing full Hamiltonian spectra. Recent works have indeed proposed:

- the use of imaginary time evolution and shallow swap tests [333], further reducing circuit depth at the cost of additional measurements;
- *variational quantum deflation* algorithms, that penalise the overlap with lower-lying eigenstates to address a k -th eigenstate [332, 334], leveraging upon eigenstate orthogonality properties, at the extra cost of requiring the sequential solution of the eigenspectrum;

⁵² E.g. the longest 8 qubit simulations converged after performing only about 500 single-qubit tomographies when adopting 20 particles in the swarm optimization performed over more than 400 parameters. In comparison with ~ 100 tomographies required by our experimental proof of concept (where 8 particles in the swarm optimized over 2 parameters, see Sect. 4.5.4).

⁵³ A similar numerical approach for the empirical estimate of scaling in VQE approaches was followed by [307].

⁵⁴ As long as swarm-methods are of concern.

- hybrid quantum–classical methods, improving the outcomes obtained with a quantum device where only *active orbitals* are mapped (Sect. 4.4.2) by adding additional, low–rank contributions via perturbative methods [324];
- hybrid adiabatic and variational approaches, where the tuning of \hat{H} is performed simultaneously with a variational search [335] to reduce the length of the adiabatic preparation.

The limitations of the WAVES protocol in terms of ancillary control qubits, emphasised by part of the literature following–up our work [332, 333], are in our opinion outweighed by the benefits of WAVES. This is particularly true as long as a compiled photonic implementation is of concern, whose extremely long decoherence times grant full control over the evolution time t to the experimenter, and high–fidelity CU operations are readily available via entanglement schemes. Nevertheless, alternative approaches might be beneficial to different hardware implementations characterised by lower gate fidelity, but higher operations rate, which more than justifies leaving the quest for eigensolvers on quantum devices an open research question.

4.8 *Author's contributions to the Chapter*

Sects. 4.1–4.3 is my review of established results available in the literature, enriched by personal observations and on-purpose analyses.

The WAVES protocol in Sect. 4.4 was initially suggested by dr. Santagati and dr. Wang. I have participated to the design of the algorithm from its infancy, suggesting e.g. the objective function form in Eq.4.25. The analysis of the complexity of the protocol in Sect. 4.4.2 was mostly due to dr. Nathan Wiebe.

The cases in Sect. 4.5.1 & 4.5.2 were proposed and preliminary modelled by dr. Jarrod McClean, and prof. David Tew and Mr. Sam Morley-Short, respectively. A first simulated implementation of the protocol was performed by mr. Paesani. I have then played a leading role in modifying the original algorithm, adding also the capability to search for excited states. All the simulations and analyses reported in Sect. 4.5.3 and the whole of Sect. 4.6 were performed by me.

Contributions to design, fabrication and preliminary characterisation of the CU(2) device have already been stated at pg. 1.6. I implemented the interface of my simulation code with the experiment control, leading to the results in Sect. 4.5.4. The data collection for the variational part of WAVES was performed mostly by dr. Santagati, Wang and Paesani, whereas the data collection of the IPEA part was performed equally by all of us. The four of us contributed also equally to the data analysis.

Finally, I had a major contribution to the final preparation of the manuscript and the related review process, **that was published as [128]**.

CONCLUSIVE REMARKS AND OUTLOOK

To conclude this Thesis, we will hereby report some final remarks about the results accomplished along the research chapters. At the same time, an outlook for future developments is provided, in the belief that protocols and experiments here initiated will provide the basis for more ambitious demonstrations of the practicality of noisy intermediate quantum devices.

We start by observing how Bayesian approaches and techniques inspired by Machine Learning remarkably lowered the requirements for the implementation of routine protocols on pre-fault-tolerant quantum devices. In particular, in Chap. 2 and Sect. 3.4 we observed this respectively for QPE and CVV routines.

A thorough investigation of the advantages introduced by the RFPE protocol suggested how the impracticality of QPE in noisy quantum platforms⁵⁵ might be removed altogether (as long as the system Hilbert space can be mapped on the device's), without the need for demanding advances in the experimental setup.

The development of devices outperforming their classical counterparts introduces naturally the problem of their CVV. Our approach to the problem was based upon a potentially scalable solution, QHL, recently proposed. QHL assumes an aprioristic Hamiltonian model for the system to characterise. After demonstrating it experimentally, we considered the problem of characterising a black–box quantum device, i.e. removing the assumption of a specific model being known for it in advance. We proposed a protocol, QMLA, that can be readily implemented for many quantum systems of interest, using either classical or dedicated quantum simulators. The search for a suitable model in QMLA implemented in this Thesis is mainly deterministic, however probabilistic implementations are already being investigated⁵⁶ to extend QMLA's capabilities, particularly in terms of resilience to local optima.

In this Thesis it was also demonstrated how QMLA can be flexibly adapted to extend the characterisation from the quantum system to its environment. The approach developed in particular along Sects. 3.5.1 & 3.5.2 is again based upon a Hamiltonian picture. However, future works shall address the analysis of long-time dynamics governed by decoherence process, by describing it more in terms of its Lindbladian⁵⁷. Studying the environment of a quantum system

⁵⁵ As recently observed e.g. in [162].

⁵⁶ Results are in preparation as [259].

⁵⁷ I have been among the initiators of this improvement, that is currently being developed by collaborators at the

is of crucial importance, because any source of noise can ultimately be modelled as system–environment interactions⁵⁸, and standard strategies to cope with such noisy effects involve the design of carefully crafted control sequences to compensate for said interactions⁵⁹. Therefore, if preliminary CVV is essential to operate an ideal (but unknown) quantum device, it becomes even more essential when dealing with non–error–corrected devices, and we envisage how methods like QMLA could be embedded as a step of an optimal quantum control strategy.

CVV approaches also led to the investigation of enhanced quantum sensing protocols, MFL, developed and tested experimentally in Sect. 3.3.1 for magnetometry with an NV–centre system. Adopting MFL, record sensitivities were achieved in room–temperature conditions, operating at extremely low SNR. The latter had previously confined such sensitivities to the realm of cryogenic setups. Even if MFL was demonstrated here offline, an online implementation is already under preparation⁶⁰. Our methods would be particularly effective in applications where single-spin sensing is desired for nano-scale resolution, but cryogenic conditions are prohibitive, such as biological sensing and in new nano-MRI applications [336, 337]. Also, MFL might be used in conjunction with single-shot readout [139], faster communication, and dynamical decoupling techniques [338, 339] and adaptive measurement bases [135], to push further the possibilities in terms of achievable sensitivity, which is of particular interest when tracking dynamical processes occurring within fast timescales [224, 336, 340].

Again in the realm of applications, we investigated with Chap. 4 the adoption of quantum simulators to address the eigenproblem in Quantum Chemistry systems. We developed an algorithm, WAVES, which is in principle amenable to short circuit depths [327] and leverages upon VQE methods, known to exhibit error robustness⁶¹. In this way, we enabled near-term experiments with the eigenproblem⁶² on non-fault tolerant machines, as demonstrated in Sect. 4.5.4.

The plethora of developments following the publication of WAVES [128] and dedicated to expanding the capabilities of (variational) eigensolvers and their usability in NISQ devices demonstrated the interest and effort of the community. At the same time, it must be emphasised how most of the proposals listed in Sect. 4.7 have been only addressed theoretically and in numerical simulations. The only exceptions are represented by the linear response and quantum subspace method [149, 308], as well as its generalisations [280], and error–mitigation schemes specifically designed for VQE [341, 342]. In our interpretation, this shows how an exhaustive investigation of realistic experimental noise and limitations in NISQ devices must often rely

University of Bristol, with the outcome in preparation as [259].

⁵⁸ A formal discussion of noise in quantum devices is in Sect. 1.1.2, whereas a discussions of system–environment interactions for a specific system of our interest, NV–centres, are in Sect. 1.5 and 3.5.2.

⁵⁹ Examples of this approach quoted in this Thesis can be found e.g. in [58, 191, 212].

⁶⁰ This demonstration is led by our collaborators at the University of Ulm, and mr. Timo Joas in particular.

⁶¹ Additional robustness was added by modifying ad–hoc some of the VQE routines, e.g. adopting particle swarm methods. Details are provided in Sect. 4.4.2.

⁶² We recall how previous VQE experimental implementations only targeted the ground state problem.

upon the actual implementation of the designed algorithm, to avoid gaps between error models considered during the algorithm design and what the quantum platform of choice might offer. With WAVES⁶³ we have attempted this approach in a joint theoretical and experimental effort.

This leads us to a crucial observation. Throughout Chaps. 2–4, we have experimentally tested different noise-resilient approaches in photonic and/or NV-centre implementations. However, the implications of said approaches extend to other quantum information processing platforms. Superconducting qubits and ion traps devices might benefit as well from the enhanced performance and noise resilience of adaptive algorithms. To quote but few examples, MFL could be well applied to other sensing platforms where noise has been a limiting factor. QMLA could be extended to the characterisation and verification of systems other than single NV-centres. According to the system size, either classical simulator (via CLE) or integrated controllable quantum simulators of adequate size (via QLE) could be adopted. Possible experiments with larger-scale photonic devices may involve e.g. studying coupled NV-centres dynamics, electron spin qubits in Silicon⁶⁴, or analyse offline data from platforms accessed remotely such as the *IBM Quantum Experience*.

In conclusion, we recognise how approaches alternative to those outlined in this Thesis might be beneficial to different hardware implementations of quantum devices, with characteristics other than the photonic chips at the core of experiments reported here. This leaves the quest for optimal usage of pre-threshold quantum devices an open research question. Nevertheless, we believe that our work has suggested realistic new routes for practical and useful quantum information processing in the near future.

⁶³ As well as with MFL and QMLA.

⁶⁴ This possibility is currently being investigated using a classical supercomputer instead, by leveraging upon the collaboration with the University of New South Wales, and in particular dr. S. Knauer.

LIST OF ABBREVIATIONS

| | |
|----------------------|--|
| ϵ | error probability, if referred to a (quantum) circuit, page 49 |
| η | efficiency, if referred to a (quantum) circuit, page 49 |
| \hbar | reduced Planck constant; here $\hbar = 1$ unless otherwise stated, page 181 |
| \mathbb{N}_0 | the set of positive natural numbers, page 50 |
| \mathcal{I} | the Fisher information, page 85 |
| $\mathcal{O}(\cdot)$ | BigO notation, indicates the set of values (functions) that are (asymptotically) upper bounded by $c \cdot$, with c a real and finite constant, page 181 |
| $\mathcal{S}(\cdot)$ | von Neumann entropy, page 151 |
| $\mathcal{S}(\cdot)$ | linear entropy, page 151 |
| $\exists:$ | such that, page 190 |
| $\ \cdot\ _F$ | Frobenius norm, page 86 |
| $\Theta(\cdot)$ | Theta notation, indicates the set of values (functions) that are (asymptotically) upper and lower bounded by $c_1 \cdot$ and $c_2 \cdot$ respectively, with (c_1, c_2) real and finite constants, page 154 |
| c | velocity of light in vacuum, page 21 |
| $C\hat{U}$ | Controlled Unitary, page 48 |
| R^2 | the coefficient of determination, page 36 |
| T_2 | (homogeneous) decoherence time, page 12 |
| T_2^* | inhomogeneous decoherence time, page 44 |
| ACC | analogue classical computing, page 15 |
| AI | Artificial Intelligence, page 2 |

| | |
|--------------|--|
| APD | Avalanche Photo-Detectors, page 31 |
| ASP | adiabatic state preparation, page 142 |
| BCc | BlueCrystal (computational) cluster, page 116 |
| BCH | Baker-Campbell-Hausdorff (expansion), page 147 |
| BF | Bayes factor, page 106 |
| BPP | Bounded error Probabilistic Polynomial time, page 5 |
| CC | coupled cluster, page 146 |
| cDAG | comparative Direct Acyclic Graph, page 106 |
| CI | configuration interaction, page 147 |
| CLE | Classical Likelihood Estimation, page 78 |
| CMOS | Complementary Metal-Oxide Semiconductor, page 1 |
| CQT | Compressed quantum tomography, page 75 |
| CVV | Characterisation, Verification and Validation, page 74 |
| CW | continuous-wave, page 35 |
| d | depth, if referred to a (quantum) circuit, page 49 |
| DC | directional couplers, page 24 |
| EO | electro-optic, page 23 |
| epoch | update step in probabilistic algorithms, page 57 |
| FCI | Full Configuration Interaction, page 138 |
| FPGA | Field Programmable Gate Array, page 57 |
| FS | Folded Spectrum (method), page 149 |
| FSR | Free Spectral Range, page 27 |
| GC | Grating Coupler, page 29 |
| HOM | Hong–Ou–Mandel, page 26 |
| LOQC | linear optical quantum computer, page 28 |

| | |
|-------------|---|
| LR | Linear Response (method), page 172 |
| LSF | Least Squares Fit, page 36 |
| LWR | Liu-West resampler, page 83 |
| MFL | Magnetic Field Learning, page 88 |
| ML | Machine Learning, page 2 |
| MMI | multi-mode interferometers, page 24 |
| MW | MicroWave, page 43 |
| MZI | Mach-Zender interferometer, page 26 |
| NISQ | Noisy Intermediate-Scale Quantum Devices, page 15 |
| NN | Neural Network, page 17 |
| PE | Phase Estimation, page 47 |
| PGH | particle guess heuristics, page 57 |
| PH | Parametrised Hamiltonian, page 149 |
| PL | photoluminescence, page 41 |
| QC | Quantum Computing, page 5 |
| qFT | Quantum Fourier Transform, page 51 |
| QHL | Quantum Hamiltonian Learning, page 74 |
| QLE | Quantum Likelihood Estimation, page 80 |
| QMA | Quantum Merlin-Arthur, page 181 |
| QMLA | Quantum Model Learning Agent, page 74 |
| QPT | quantum process tomography, page 76 |
| QST | Quantum State Tomography, page 75 |
| sDAG | structural Directed Acyclic Graph, page 103 |
| SFWM | Spontaneous Four Wave Mixing, page 29 |
| SMC | Sequential Monte Carlo, page 81 |

| | |
|--------------|--|
| SNR | Signal-to-Noise Ratio, page 13 |
| SNS | Shot–Noise limited Scaling, page 6 |
| SNSPD | Superconducting Nanowire Single Photon Detectors, page 31 |
| SOI | Silicon–on–Insulator, page 21 |
| SPDC | Spontaneous Parametric DownConversion, page 29 |
| TO | thermo–optical, page 212 |
| UCCSD | unrestricted UCC, page 156 |
| UQC | Universal Quantum Computer, page 6 |
| w | width, if referred to a (quantum) circuit, page 49 |
| WAVES | Witness-Assisted Variational Eigenspectra Solver, page 150 |
| WG | waveguide, page 20 |



APPENDIX

A.1 The QFT gate and its complexity

In the main text, we have stated that the computational complexity of performing a qFT operation is $\mathcal{O}(n^2)$, with n the number of qubits in the system.

This is immediate when observing Fig. A.1, as the number of elementary one and two-qubit gates involved is: $\sum_n[(n+1)+(n)+\dots+2+1]+n = T_n + n/2$, i.e. the sum of the triangular number T_n , plus $n/2$ SWAP operations [5]. The latter ones, not displayed in the picture, reverse the order of the qubits, i.e. they perform: $|\phi_n\rangle \rightarrow |\phi_1\rangle$, $|\phi_{n-1}\rangle \rightarrow |\phi_2\rangle$, etc.

Now, $T_n = \binom{n+1}{2} \in \mathcal{O}(n^2)$, and so is the complexity of qFT , QED.

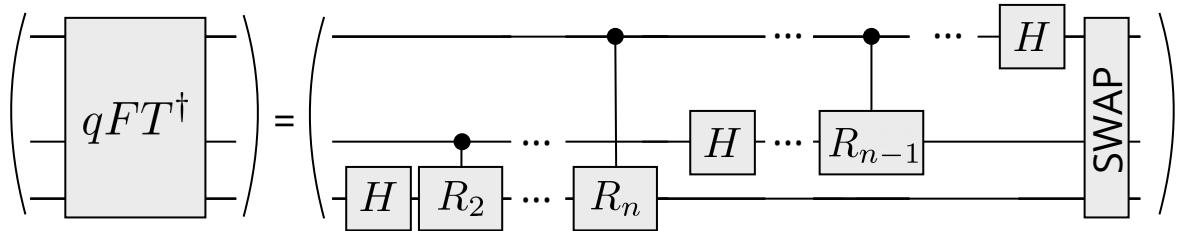


FIGURE A.1. The circuit for the Quantum Fourier Transform. H are Hadamard gates (see Eq. 1.5), and $R_k \equiv \exp(2\pi/2^{k+1})R_z(2\pi/2^{k+1})$ (see Eq. 1.7). The SWAP gate performing the final mirroring in the qubit order is not displayed for better clarity as in [5].

A.2 Pseudo-algorithms for PE applications

Algorithm 2: Bayesian update via Rejection Filtering

Input: experimental data E , mean μ and variance σ of the prior distribution, experiment parameters M and θ , number of particles N_{Part} , scale κ_E .

Output: mean μ' and variance σ' of the posterior distribution.

```

function REJFILTER ( $E, \mu, \sigma, M, \theta, N_{Part}, \kappa_E$ )
 $\mu_{acc}, \mu'_{acc}, V_{acc}, V'_{acc}, N_{acc} = 0.$        $\triangleright$  Initialize accumulators.
for  $i \in 1 \rightarrow N_{Part}$  do
     $x \sim \mathcal{N}(\mu, \sigma).$        $\triangleright$  Sample a particle according to prior distribution.
     $x = x \bmod 2\pi.$ 
     $x' = x + \pi \bmod 2\pi.$ 
     $u \sim \text{Uniform}(0, 1).$        $\triangleright$  Accept the particle with probability  $P(E|x).$ 
    if  $P(E|x) \geq \kappa_E u$  then
         $\mu_{acc} = \mu_{acc} + x.$ 
         $V_{acc} = V_{acc} + x^2.$ 
         $V'_{acc} = V_{acc} + x'^2.$ 
         $N_{acc} = N_{acc} + 1.$ 
    end if
end for
 $\mu' = \mu_{acc}/N_{acc}.$        $\triangleright$  Calculate mean and variance of the posterior distribution using
accepted particles.
 $\sigma' = \min \left( \sqrt{(V_{acc} - \mu_{acc}^2)/(N_{acc} - 1)}, \sqrt{(V'_{acc} - \mu_{acc}^2)/(N_{acc} - 1)} \right).$ 
return  $(\mu', \sigma').$ 
end function

```

Algorithm 3: Rejection Filtering Phase Estimation (RFPE)

Input: Initial prior distribution $\mathcal{N}(\mu_0, \sigma_0)$, total number of experiments N_{Steps} , number of particles N_{Part} , scale κ_E .

Output: the phase estimation μ and its uncertainty σ .

```

 $\mu = \mu_0, \sigma = \sigma_0.$        $\triangleright$  Initialize the mean  $\mu$  and variance  $\sigma$  of the distribution.
for  $i \in 1 \rightarrow N_{Steps}$  do
     $M = \lceil 1.25/\sigma \rceil, \theta \sim \mathcal{N}(\mu, \sigma).$        $\triangleright$  Get approx. optimal parameters  $M$  and  $\theta$  via PGH
    Get data  $E$  from the experiment using parameters  $M$  and  $\theta$ .
     $(\mu, \sigma) = \text{UPDATE}_{RF}(E, \mu, \sigma, M, \theta, N_{Part}, \kappa_E).$        $\triangleright$  call the Rejection Filtering
end for
return  $(\mu, \sigma).$ 

```

A.3 Algorithms for Hamiltonian Learning

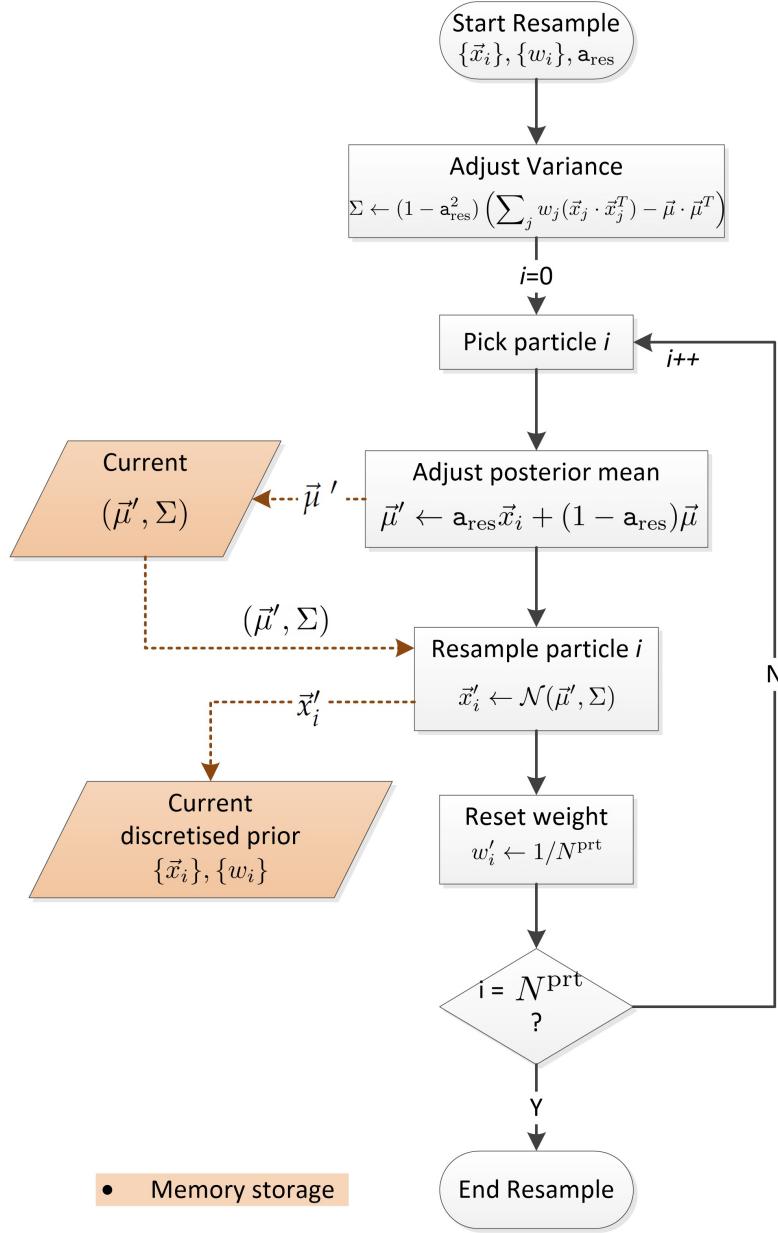


FIGURE A.2. Flowchart for the QHL resampler. Legend is in the lower box. Components invoking memory storage are colour-coded.

Algorithm 4: MFL algorithm with stepwise change detection

```

Input : An initial prior distribution  $\mathcal{F} \sim P(\vec{x})$  over models.
Input (additional) :  $p_{\text{set}}$        $\triangleright$  rate parameter, i.e. how many epochs must occur before a
RESAMPLE call
Input (additional) :  $t_{\text{set}}$        $\triangleright$  parameter adjusting the frequency of posterior-reset events
function ESTIMATEADAPTIVE (n,  $\pi_{\text{ini}}$ , N, a (the resampling parameter),  $t_{\text{resample}}$  (the
resample threshold), OPTIMIZE, UTIL, n_guesses, GUESSEXPERIMENT):
     $w_i \leftarrow 1/N^{\text{prt}}$ 
    draw each  $\omega_i$  independently from  $\pi_{\text{ini}}$ 
     $l_{\text{set}} \leftarrow 0$        $\triangleright$  initialise the reset counter
     $l_{\text{res}} \leftarrow 0$        $\triangleright$  initialise the resampling counter
    initialise  $\vec{x} = [\bar{x}_1, \dots, \bar{x}_N]$ 
    for  $e \in 1 \dots N$  do:       $\triangleright$  for each epoch
        ...
        if  $\sum_i w_i^2 < N^{\text{prt}} t_{\text{res}}$ :       $\triangleright$  if the effective sample size is below the threshold
            if  $|e - l_{\text{set}}| \geq p_{\text{set}}$  OR  $\sum_i w_i^2 < N^{\text{prt}} t_{\text{set}}$ :  $\triangleright$  if a reset occurred too recently or the
            discretised posterior has sufficient support
                 $\{w_i\}, \{x_i\} \leftarrow \text{RESAMPLE}( \{w_i\}, \{x_i\}, a )$   $\triangleright$  resample as usual
                 $l_{\text{res}} \leftarrow e$        $\triangleright$  store last resampling event
                else:       $\triangleright$  reset the procedure
                     $P'(\vec{x}) \leftarrow \mathcal{F}$ 
                     $w_i \leftarrow 1/N^{\text{prt}}$ 
                    draw each  $\omega_i$  independently from  $\mathcal{F}$ 
                     $l_{\text{set}} \leftarrow e$        $\triangleright$  store last reset event
                    continue from  $e + 1$ 
                end if
            end if
        ...
         $\bar{x}_i \leftarrow \text{MEAN} (\{w_i\}, \{x_i\})$        $\triangleright$  append the new estimate from the mean
    end for
end function

```

Output: \vec{x} , storing the instantaneous values of the unknown parameter

A.4 Algorithms for the Quantum Model Learning Agent

Algorithm 5: Routines used in Bayesian inference. See also Algorithm 1 for reference.
Subroutines QHL, Update can be found as pseudocode in [213].

```

function Converged ( $\vec{\sigma}, \epsilon, N_{\max}, N_{\min}$ ):
    if  $\dim[\vec{\sigma}] > N_{\max}$ :       $\triangleright$  stop the learning if the max number of epochs has been reached
        return True
    else if  $\dim[\vec{\sigma}] > N_{\min}$ :     $\triangleright$  as soon as a reasonable number of epochs has been processed
         $\vec{e} \leftarrow (1, 2, \dots, N_{\min})$        $\triangleright$  initialise a vector of progressive observations
         $\vec{\sigma}' \leftarrow \vec{\sigma}(\dim[\vec{\sigma}] - N_{\min}, \dots, \dim[\vec{\sigma}])$ 
        LSQ ( $y = \vec{\sigma}', x = \vec{e}$ , “ $ax + c$ ”)       $\triangleright$  perform a linear LSQ on the uncertainties  $\vec{\sigma}'$ 
        if  $a < \epsilon$ :       $\triangleright$  if the multi-parameter estimate is converging (within fluctuations)
            return True       $\triangleright$  stop the learning process
        else:
            return False
    else:
        return False       $\triangleright$  in all other cases keep processing new data

function BfBatchUpdate ( $\hat{H}_i(\vec{x}_i), \hat{H}_j(\vec{x}_j), \mathbb{D}_i, \mathbb{D}_j, P'_i(\vec{x}_i), P'_j(\vec{x}_j), f$ ):
     $\triangleright$  merge system's observations  $\mathbb{D}$  retrieved at experimental times  $\{\tau\}$ 
     $\mathbb{D}_{ij} \leftarrow \mathbb{D}_i \cup \mathbb{D}_j := \{D_e(\tau_e)\}_i \cup \{D_e(\tau_e)\}_j$ 

    for  $k$  in  $(i, j)$ :       $\triangleright$  for each model in  $(i, j)$ , batch update using the learnt parametrization
         $N^{\text{prt}} \leftarrow f(\dim[\vec{x}_k])$        $\triangleright$  assign a proper number of particles  $N^{\text{prt}}$ 
        do sample  $N^{\text{prt}}$  particles  $\{\vec{x}_\alpha\}$  and weights  $\{\vec{w}_\alpha\}$  from  $P'_k$ 
         $\ell(\mathbb{D}_{ij}|\hat{H}_k) \leftarrow 0$        $\triangleright$  reset the cumulative log-likelihood

        for  $D_e(t_e)$  in  $\mathbb{D}_{ij}$ :
             $\triangleright$  perform a QHL update of  $P'_k$  using datum  $D_e(t_e)$ 
             $P'_k(\vec{x}_k) \leftarrow \text{Update}(\{\vec{w}_\alpha\}, \{\vec{x}_\alpha\}, D_e, \tau_e)$ 
             $\vec{x}_k^* \leftarrow \mathbb{E}[P'_k(\vec{x}_k)]$ 
             $\triangleright$  update the log-likelihood  $\ell(\mathbb{D}_{ij}|\hat{H}_k)$ 
             $\ell(\mathbb{D}_{ij}|\hat{H}_k) \leftarrow \ell(\mathbb{D}_{ij}|\hat{H}_k) + \log(P(D_e|\hat{H}_k(\vec{x}_k^*, \tau_e)))$ 
        end for
    end for

     $\mathcal{B}_{ij} \leftarrow \exp[\ell(\mathbb{D}_{ij}|\hat{H}_i) - \ell(\mathbb{D}_{ij}|\hat{H}_j)]$        $\triangleright$  compute the Bayes Factor  $\mathcal{B}_{ij}$ 
    return  $\mathcal{B}_{ij}$ 
end function

```

Algorithm 6: Routines used in graphical modelling. See also Algorithm 1 for reference. The subroutine FindChampion is provided in full generality, to cover such cases where interlayer comparisons are provided not only among their respective *layer champions* - not reported in main text.

```

function DAGupdate (cDAG,  $\mathcal{B}_{ij}$ ):
    if  $\mathcal{B}_{ij} > 1$ :       $\triangleright$  edges point towards the statistically favored model
        add  $(\hat{H}_j, \hat{H}_i)$  with  $w_{ji} := w[(\hat{H}_j, \hat{H}_i)] = \mathcal{B}_{ij}$  to cDAG
    else:
        add  $(\hat{H}_i, \hat{H}_j)$  with  $w_{ij} := w[(\hat{H}_i, \hat{H}_j)] = 1/\mathcal{B}_{ij}$  to cDAG
         $\triangleright$  the edge weights  $w$  are assigned according to the value of  $\mathcal{B}$ 

    return cDAG       $\triangleright$  returns the input cDAG with the additional directed edge
end function

function FindLayerChampion (cDAG,  $b_c$ ):
    foreach  $\hat{H}_i$  in cDAG:       $\triangleright$  for each model included in the input cDAG
         $S_i \leftarrow 0$        $\triangleright$  initialize the model score  $S$ 
    foreach  $(\hat{H}_i, \hat{H}_j)$  in cDAG:       $\triangleright$  for each edge included in the input cDAG
        if  $w_{ij} > b_c$ :
             $S_j \leftarrow S_j + 1$ 
     $\hat{H} \equiv \hat{H}_i \ni: S_i = \max_k(\{S_k\})$ 

    foreach  $\hat{H}_i$  in cDAG:
        if  $\exists \mathcal{P} := ((\hat{H}_i, \hat{H}_j), (\hat{H}_j, \hat{H}_k), \dots, (\hat{H}_n, \hat{H}_i)) \in \text{cDAG}$        $\triangleright$  a loop  $\mathcal{P}$  within cDAG
             $\triangleright$  the loop is cut at its least statistically significative edge
            remove from cDAG:  $(\hat{H}_j, \hat{H}_k) \ni: \mathcal{B}_{jk} = \min_{(n,m) \in \mathcal{P}} \{\mathcal{B}_{nm}\}$ 

    return  $\hat{H}$        $\triangleright$  return the champion model given the evidence stored in cDAG
end function

```

```

function CollapseLayer (sTree, cDAG,  $b_e$ ):
    foreach  $\lambda \in$  sTree:       $\triangleright$  check for any layer  $\lambda$ 
         $B_{\text{up}} \leftarrow \min_{\text{cDAG}}(w_{kj} \mid \hat{H}_l \in \lambda)$ 
         $B_{\text{down}} \leftarrow \min_{\text{cDAG}}(w_{kl} \mid \hat{H}_k \in \lambda + 1)$ 
        if  $B_{\text{up}} > b_e$  and  $B_{\text{down}} > b_e$ :       $\triangleright$  apply collapse rules
            foreach  $\hat{H}_k \in \lambda + 1$ :
                remove  $(\hat{H}^{\lambda}, \hat{H}_k)$  from sTree       $\triangleright$  remove edges to child layer
                add  $(\hat{H}^{\lambda-1}, \hat{H}_k)$  to sTree       $\triangleright$  reassign child layer to parent champion

            foreach  $\hat{H}_l \in \lambda$ :
                remove  $(\hat{H}^{\lambda-1}, \hat{H}_l)$  from sTree       $\triangleright$  remove edges with parent's champion
                remove  $(\hat{H}^{\lambda}, \hat{H}_k)$  or  $(\hat{H}_k, \hat{H}^{\lambda})$  from cDAG       $\triangleright$  clean the cDAG

                 $\hat{H}_j \leftarrow \hat{H}^{\lambda-1}$  and  $\hat{H}_k \leftarrow \hat{H}^{\lambda+1}$ 
                 $\mathcal{B}_{jk} \leftarrow \text{BfBatchUpdate}(\hat{H}_j(\vec{x}_j), \hat{H}_k(\vec{x}_k), \mathbb{D}_j, \mathbb{D}_k, P'_j, P'_k, f)$        $\triangleright$  compute BFs
                cDAG  $\leftarrow \text{DAGupdate}(\text{cDAG}, \mathcal{B}_{jk})$        $\triangleright$  populate cDAG with reassigned edges
    end function

function FindChampion (sTree, cDAG):
    foreach  $\hat{H}_i$  in sTree:       $\triangleright$  for each model within the (residual) sTree
         $S_i \leftarrow 0$        $\triangleright$  initialise the model score  $S$ 

    foreach  $\hat{H}_i$  in sTree:
        foreach  $\hat{H}_j$  in sTree, with  $j > i$ :       $\triangleright$  for each other layer champion
            if  $(\hat{H}_i, \hat{H}_j) \notin$  cDAG and  $(\hat{H}_j, \hat{H}_i) \notin$  cDAG:       $\triangleright$  avoid recomputing BFs
                 $\mathcal{B}_{ij} \leftarrow \text{BfBatchUpdate}(\hat{H}_j(\vec{x}_i), \hat{H}_k(\vec{x}_j), D_i, D_j, \mathcal{F}'_i, \mathcal{F}'_j, f)$        $\triangleright$  compute BFs
                cDAG  $\leftarrow \text{DAGupdate}(\text{cDAG}, \mathcal{B}_{ij})$        $\triangleright$  populate the cDAG with new edges
            if  $(\hat{H}_i, \hat{H}_j) \in$  cDAG:
                else:
                     $S_i \leftarrow S_i + 1$ 

     $\hat{H}^0 \equiv \hat{H}_i \ni S_i = \max_k(\{S_k\})$        $\triangleright$   $\hat{H}^0$  is the model with highest indegree
    return  $\hat{H}$        $\triangleright$  return the champion model according to evidence stored in cDAG
end function

```

A.5 The Swarm Optimisation Algorithm for variational searches

Algorithm 7: Swarm optimization algorithm

Data: Random set of N particles, characterized by the set of parameters $\Xi := \{\vec{\theta}_i\}$, sampled uniformly from the parameters' space
initialization of parameters: $a, b, \text{threshold}(s), S$;

while convergence **not** achieved **do**

- foreach** $\vec{\theta}_i \in \Xi$ **do**
 - (Q) evaluate control qubit *purity* (\mathcal{P});
 - (Q) evaluate *energy estimator* (\mathcal{E});
 - calculate $\mathcal{F}_{\text{obj}}(\vec{\theta}') = -a\mathcal{P} + b\mathcal{E}$;
- end**
- calculate $\mu'_{F_{\text{obj}}} = \langle \mathcal{F}_{\text{obj}}(\vec{\theta}_i) \rangle_{\Xi}$;
- if** $|\mu'_{F_{\text{obj}}} - \mu_{F_{\text{obj}}}| < \text{threshold for objective function}$ **then**
 - convergence achieved (via \mathcal{F}_{obj}) ;
- else**
 - $\mu_{F_{\text{obj}}} \leftarrow \mu'_{F_{\text{obj}}}$;
- end**
- find the subset $\Xi' := \{\vec{\theta}'_i\} \subset \{\vec{\theta}_i\}$ of the S particles with lowest \mathcal{F}_{obj} ;
- if** \mathcal{F}_{obj} is adaptive **then**
 - adjust $a \propto |\langle \mathcal{P} \rangle_{\Xi'} - \langle \mathcal{P} \rangle_{\Xi}|$ and $b \propto |\langle \mathcal{E} \rangle_{\Xi'} - \langle \mathcal{E} \rangle_{\Xi}|$;
 - normalize $\{a, b\}$;
- end**
- calculate (weighted) mean $\mu_{\theta'} := \langle \vec{\theta}'_i \rangle_{\Xi'}$ for the Θ' subset parameters;
- calculate (weighted) variance $\sigma_{\theta'}^2 := \langle (\vec{\theta}'_i)^2 \rangle_{\Xi'} - \langle \vec{\theta}'_i \rangle_{\Xi'}^2$ for the Ξ' subset parameters;
- if** $\sigma_{\theta'} < \text{threshold for particle dispersion}$ **then**
 - convergence achieved (via $\sigma_{\theta'}$) ;
- end**
- generate a set Ω of $N - S$ new particles, Gaussian distributed around $\mu_{\Xi'}$ with standard deviation $\hat{\sigma}_{\Xi'}$;
- $\Xi \leftarrow \Xi' \cup \Omega$;

end

Result: $\mu_{\theta'}$ is the output best estimate for the optimal parameters, and $\sigma_{\theta'}$ the error estimate.

Algorithm 8: Protocol for the synthetic truncation of the ansatz

Data: Parametrized ansatz $\hat{A}(\vec{\theta}) \equiv \sum_{i \in I} (\theta_i \hat{A}_i)$, excitation operator \hat{E}_{p_τ} and targeted excited eigenstate $|\Psi_\tau\rangle$, threshold \tilde{F}_0 for the initial guess
obtain from WAVES optimal parameters for the ground state $\hat{A}_g = \hat{A}(\vec{\theta}_g)$, minimizing
 $\mathcal{F}_{\text{obj}} = -a\mathcal{P} + b\mathcal{E}$;
calculate $|\Psi_0\rangle = \hat{E}_{p_\tau} \hat{A}_g |\Phi\rangle$;
initialize $I' = I$;
while $|\langle \Psi_0 | \Psi_\tau \rangle|^2 \geq \tilde{F}_0$ **do**
 | find $\iota: \theta_{g_\iota} = \min_i (|\theta_{g_i}|)$;
 | $I' \leftarrow I' - \iota$;
 | $\hat{A}'_g = \sum_{i \in I} (\theta_{g_i} \hat{A}_i)$;
 | calculate $|\Psi_0\rangle = \hat{E}_{p_\tau} \hat{A}'_g |\Phi\rangle$;
end
Result: truncated ansatz $\hat{A}'(\vec{\theta}) \equiv \sum_{i \in I} (\theta_i \hat{A}_i)$, with $\dim(I') < \dim(I)$



APPENDIX

B.1 Extracting binary outcomes from a likelihood signal

In cases when $M > 1$, the datum E extracted is obtained by majority voting the single outcomes from M sequences. I.e. E is chosen according to the most frequent outcome, knowing in advance the average and maximum PL counts (\bar{n}, n_{max}), and comparing with the photon number readout after an experiment, n . If $n > \bar{n}$, the outcome is set to $|1\rangle$, and $|0\rangle$ otherwise. A probabilistic extraction, which is more robust to noisy readouts (particularly when decoherence effects are important), samples the outcome from the set $\{|0\rangle, |1\rangle\}$, with probabilities $P \propto \{1 - n/n_{max}, n/n_{max}\}$. As each extraction is affected by Poissonian noise, repeating the procedure at the same datapoint is a natural way to reproduce the Poissonian noise in the measurement.

B.2 Qualitative reasoning supporting the Heisenberg-limit

The HL can be retrieved from the phase-number uncertainty principle, for an eigenphase ϕ being estimated using a single mode optical field with number operator \hat{n} [56]:

$$\Delta\phi\Delta n \geq 1/2 \implies \Delta\phi \geq 1/N \quad (\text{B.1})$$

as it can be inferred from the $\Delta n \leq N/2$ relationship, with N the upper bound for \hat{n} , i.e. N represents the number of identical physical systems available. A more detailed analysis shows how the bound from Eq. B.1 is actually loose, and $\Delta\phi \geq \pi/N$, which is known as the *Heisenberg limit* (HL) for the achievable uncertainty $\Delta\phi$. From this bound descends immediately the concept of *HL-scaling*, which quantifies instead how $\Delta\phi$ scales with an increase in the available N , and this is expected to be $\propto 1/N$.

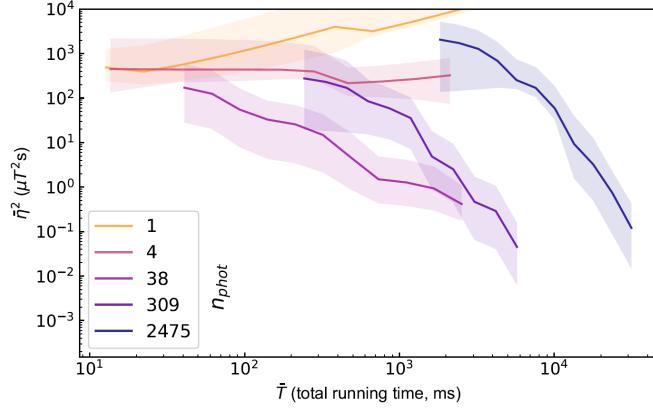


FIGURE B.1. A simulation of the non-adaptive QPE protocol (demonstrated experimentally in [135]) absolute precision, achievable with an NV-centre setup having the same characteristics as observed in Sect. 3.3.1. Performances are averaged over 1000 independent runs of the QPE protocol.

However, in Chap. 3, we made use instead of a different metric, the total evolution time T . The equivalence with the bound in Eq. B.1 is due to the following reasoning. Adopting a phase gate as in Eq. 1.7, each gate application accrues a phase given by $e^{i\phi}$. Hence, if provided with N maximum photons, the difference between the minimum and maximum number of gates application (and powers of $e^{i\phi}$) is N . When estimating the phase as the single parameter for a Hamiltonian $\hat{H}_\phi = \phi \hat{H}$, a corresponding boundary for the powers of $e^{i\phi}$ is provided by $[-\lambda_{\max} T, -\lambda_{\min} T]$, whose width is $|\Delta\lambda|T$, with T the total evolution time. In [56], it was noted how this equivalence leads to the formulation:

$$\frac{\pi}{N} \sim \frac{\pi}{|\Delta\lambda|T} \geq \Delta\phi \quad (\text{B.2})$$

which leads also to the HL-scaling $\propto 1/T$ invoked in the main text.

B.3 Comparison of WAVES and the FS method

The folded spectrum (FS) method performs variational searches after folding the energy landscape by a *shift* ϵ , so that the modified eigenvalue equation reads [319]:

$$(\hat{H} - \epsilon)^2 |\Psi\rangle = (\mathcal{E} - \epsilon)^2 |\Psi\rangle \quad (\text{B.3})$$

and an energy minimisation leads to the eigenstate with the smallest eigenvalue satisfying $\lambda_i \geq \epsilon$.

In this paragraph we comment the results of applying FS to find excited states when applying the excitation operator \hat{E}_{p_3} to the reference state of the H₂ systems, adopting a PH ansatz. We recall that \hat{E}_{p_3} provides an initial guess state with good overlaps on both the 1st and 4th excited

subspaces: $F_2 \simeq 0.39$ and $F_4 \simeq 0.61$ respectively. For a WAVES variational search, a majority of times the final state collapses on the 4th excited subspace, and in the remaining on the 1st, always with high fidelity. This two-fold convergence averaged for 100 runs is reported in blue in Fig.4.6.

In order to implement FS, a value for the energy shift ϵ is required. A natural choice is to employ the energy of the initial guess state provided by \hat{E}_{p_3} applied to the PH parametrized ground state: $|\Psi_3\rangle \equiv \hat{E}_{p_3}\hat{A}(\vec{\theta}_g)|\Phi\rangle$. Now, the energy of this trial state lies in between the two eigenvalues, corresponding to the excited subspaces with which we have consistent initial overlap. This makes them close in the folded spectrum and may become accidentally degenerate because of the initial uncertainty assigned to the set of parameters σ_{θ_g} . If it is chosen $\epsilon \simeq \langle |\Psi_3| \hat{\mathcal{H}} | \Psi_3 \rangle$ in this way, the FS search in a minority of cases converges with high fidelity to the 4th excited subspace, but most of the time converges to local minima in the folded spectrum. We remind that to have a fair comparison with the simulated runs of WAVES, here we adopted the same swarm Algorithm 7. These correspond to superpositions of eigenstates belonging to the different excited subspaces involved. In conclusion, the final excited states estimates are poor, though still resembling the two-fold behavior exhibited by WAVES (the results are shown in Fig. 4.6, in red). These limitations in the adoption of FS are known from the literature, including the original work [319].

In order to check the correct performance of the Folded Spectrum, we artificially reduced the value of ϵ to be slightly bigger than the eigenvalue of the 1st excited state. In this case, the FS search collapses consistently on the first excited subspace all the times, with a fidelity equivalent to that obtained with WAVES (see Fig. 4.6, in yellow). It is still possible to observe a slowdown in the convergence, requiring about twice the number of trial states with an equivalent choice of parameters. This poorer convergence rate is an expected effect, due to the squaring of the relevant operator $(\hat{H} - \epsilon)^2$.

B.4 Finite-size effects of the nuclear spins bath in experimental QMLA

A well-established modelling of decoherence effects in open quantum systems involves the introduction of *indirect measurements* occurring on the quantum system, when operations occur that effectively trace out environmental degrees of freedom, which have become correlated with the system state [28]. In the main paper we have systematically considered models inclusive of Hamiltonian terms of the form $\sum_i \hat{S}_i \otimes \hat{B}_i$, with \hat{S}, \hat{B} system and bath operators respectively, which is a standard interaction term leading to system-bath correlations [28, 343].

In the theoretical framework of open quantum systems, it is customary to model decoherence phenomena introducing a bath of n_{env} spins [343], or harmonic oscillators [344]. The reason to adopt a bath of interacting systems is to ensure an irreversible decoherence, as usually observed

in real physical systems. The most direct way to see this is to invoke Poincaré recurrence theorem. Assuming the state of the global quantum system starts as a separable state $|\Psi(0)\rangle = |\psi\rangle_{\text{sys}} \otimes |\phi\rangle_{\text{env}}$, the theorem states that a time $\exists \tilde{t}$, such that the evolved state $|\langle\Psi(0)|\Psi(\tilde{t})\rangle|^2 \sim 1$, provided that the global system has a finite eigenspectrum of energies $\{E_n\}$. An immediate consequence is that at times $\mathcal{O}(\tilde{t})$ (known as Poincaré recurrence times [28]), the global state can become separable again, thus leading to *revivals* in the coherence of the system state. If we consider the simplest possible interaction term:

$$\sum_b |b\rangle_{\text{sys}} \langle b|_{\text{sys}} \otimes \hat{B}_b, \quad (\text{B.4})$$

with \hat{B} a generic operator acting on the bath and $\{|b\rangle\}$ an orthonormal basis for the system, then the global evolved state can be represented as:

$$|\Psi(t)\rangle = \sum_b p_b |b\rangle_{\text{sys}} \otimes |\phi_b(t)\rangle_{\text{env}} \quad (\text{B.5})$$

and therefore the decoherence can be characterised in terms of the overlaps $M_{b,b'}(t) = |\langle\phi_b(t)|\phi_{b'}(t)\rangle|$ [28]. Now the decay of these overlaps, as well as their recurrence times, depends on the size of the bath, as a richer dynamics corresponding to a higher-dimensional bath will tend to increase the recurrence times at which $M_{b,b'}(t)$ returns non-negligible. In real systems, it is empirically observed a decay $M_{b,b'}(t) \propto \exp[-\Gamma_{bb'}t]$, in agreement with what can be predicted under Markovian assumptions. More generally, it can be proven that the recurrence times \tilde{t} for the global system scale combinatorially with the bath size n_{env} [343]. Hence the necessity in simulations to increase the bath size for a system with a discrete spectrum, in order not to observe revivals as an artefact of the finite size of the bath. This is equivalent to the known behaviour in artificial quantum simulators characterised by *finite-size effects* in the bath degrees of freedom (e.g. ion trap simulators).

In Sect. 3.5.2, we have shown how the size of the bath can be inferred from *Hahn signal* experiments. In the main paper, instead, we have investigated open quantum systems dynamics making use of a single additional qubit treated as an environmental qubit, in order to match the maximum bath dimension attainable with our quantum photonic simulator. However, this leads inevitably to finite-size effects in the system dynamics, as shown in Fig. B.2. The learning procedure can parameterize the model \hat{H} opportunely to reproduce the initial decay, but being interaction terms in the Hamiltonian time-independent, a single environmental spin will produce Poincaré recurrence times short enough to be observed in the simulations of the reduced system dynamics. The standard way to deal with these effects is to introduce phenomenological decay terms $\exp[-\Gamma(t)]$ in the model Hamiltonian [213, 265], to reproduce the observed decay. This is not done here, as it would invalidate the attempt to learn the open system dynamics as an interaction with environmental qubits, preserving a time-independent model \hat{H} .

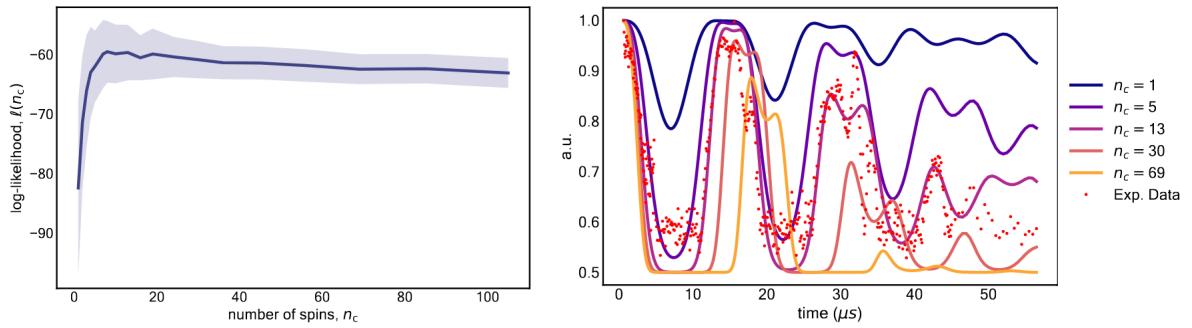


FIGURE B.2. On the left, we display the median log-likelihood obtained upon convergence of hyperparametrised QHL runs (Sect. 3.5.2), for different numbers n_c of environmental spins. On the right, predicted normalised PL for experimental data (red dots) as well as hyperparametrised models provided in the legend. Compare with Fig. 3.19 for further insight.

APPENDIX B. APPENDIX

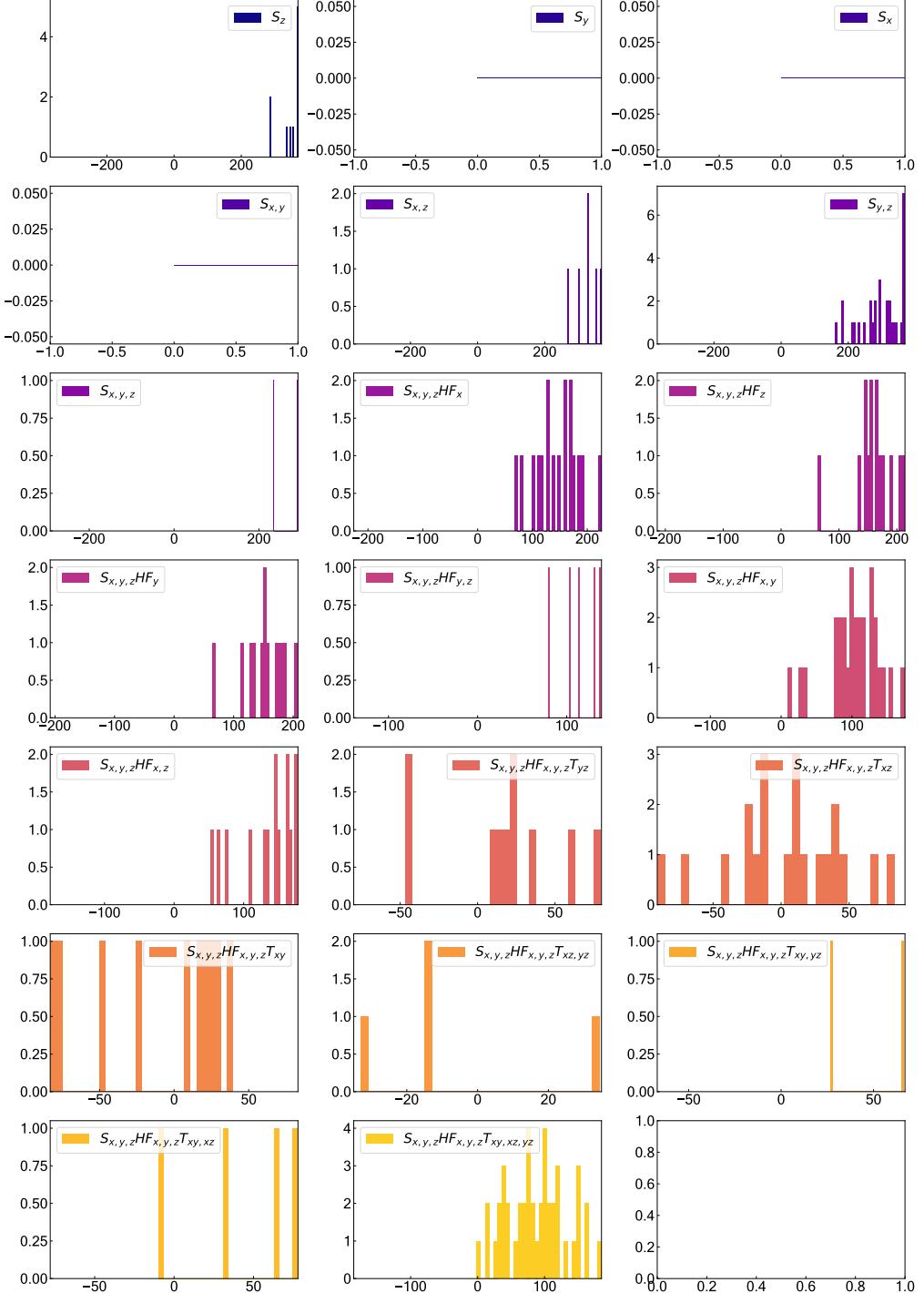


FIGURE B.3. Distributions of the computed Bayes Factor for each model in the sTree (displayed in the legend of each frame), compared with the reference model $S_{xyz}Yxyz$: positive BFs are in favour of the latter. The results are cumulative across the 100 independent QMLA runs. Models not displayed did not have any direct edges connecting them to $S_{xyz}Yxyz$ in any of the sTree's. The probe used is $|rr\rangle$.

B.4. FINITE-SIZE EFFECTS OF THE NUCLEAR SPINS BATH IN EXPERIMENTAL QMLA

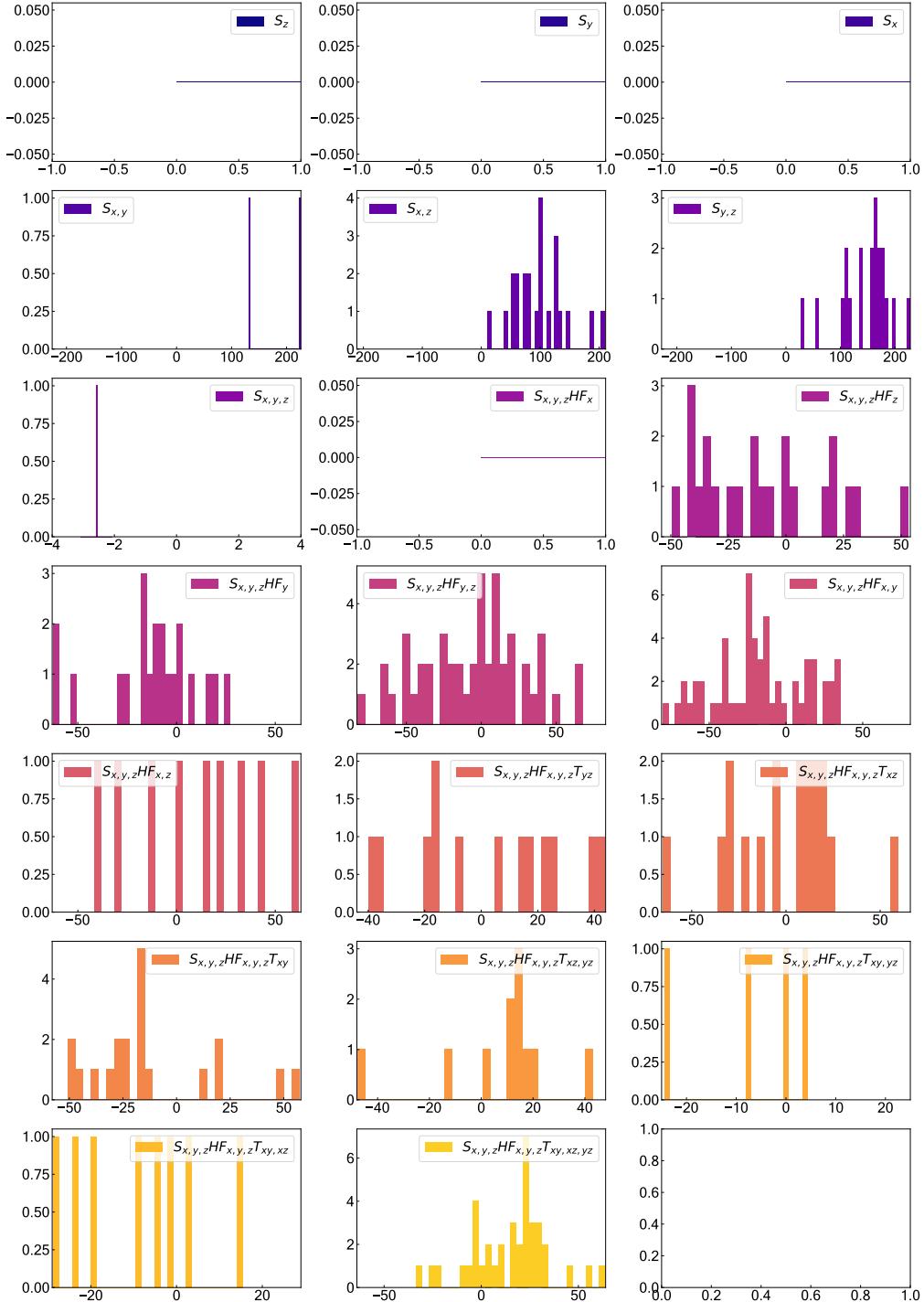
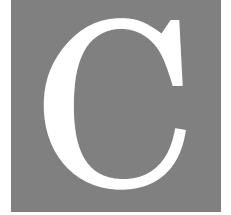


FIGURE B.4. Distributions of the computed Bayes Factor for each model in the sTree (displayed in the legend of each frame), compared with the reference model $S_{xyz}Yxyz$: positive BFs are in favour of the latter. The results are cumulative across the 100 independent QMLA runs. Models not displayed did not have any direct edges connecting them to $S_{xyz}Yxyz$ in any of the sTree's. The probe used is $|++\rangle$.



APPENDIX

C.1 Purity as an eigenstate witness

This paragraph intends to justify the adoption of \mathcal{F}_{obj} introduced for WAVES, and demonstrate that it can be actually adopted as an eigenstate witness.

Proof. We start with the evolved state after application of the $c\hat{U}$ operation:

$$\frac{|0\rangle|\psi\rangle + |1\rangle e^{-i\hat{H}t}|\psi\rangle}{\sqrt{2}}. \quad (\text{C.1})$$

A partial trace over the target qubit leads to:

$$\rho = \begin{pmatrix} \frac{1}{2} & \frac{1}{2}(\sum_j |\alpha_j|^2 e^{i\lambda_j t}) \\ \frac{1}{2}(\sum_j |\alpha_j|^2 e^{-i\lambda_j t}) & \frac{1}{2} \end{pmatrix} \quad (\text{C.2})$$

and with a few algebraic steps:

$$\text{Tr}(\rho^2) = \frac{1}{2} \left(\sum_j |\alpha_j|^4 \right) + \frac{1}{2} \left(1 + \sum_{j \neq k} |\alpha_j|^2 |\alpha_k|^2 \cos([\lambda_j - \lambda_k]t) \right) \geq \sum_j |\alpha_j|^4. \quad (\text{C.3})$$

This expression is closely related to the *participation ratio*. In particular, is $\text{Tr}(\rho^2) = 1 \iff \alpha_j = 0$, as long as $(\lambda_j - \lambda_k)t \neq 0 \pmod{2\pi}$. Most times t will grant this condition, and for each such $t > 0$, a purity $\mathcal{P} = 1 \implies |\psi\rangle$ is an eigenstate. ■

In order to provide a lower bound, if we assume $\max_{j,k} |\lambda_j - \lambda_k|t \leq \pi/2$, it is possible to invoke Jensen's inequality, obtaining:

$$\text{Tr}(\rho^2) \leq \frac{(1 + \cos(\sum_{j,k} |\alpha_j|^2 |\alpha_k|^2 |\lambda_j - \lambda_k|t))}{2}, \quad (\text{C.4})$$

Now, $\text{Tr}(\rho^2)$ is thus upper bounded by a function which decreases with the mean eigenvalue difference $|\lambda_j - \lambda_k|$. From Eqs. C.3 & C.4 follows that a purity maximisation over sufficiently long t maximises also the participation ratio. Thus, \mathcal{P} is upper and lower bounded by sensible measures of the eigenstate support.

C.2 Details about the excitonic unitary

In many quantum simulation applications the natural choice of zero-energy results in Hamiltonians where the total energy is orders of magnitude larger than the energy differences relevant to the phenomena under investigation. This is particularly true, for example, for reaction energies in quantum chemistry and is also the case for the excitonic Hamiltonian in Eq. 4.30, where we are interested in the difference between the ground and excited state $2\beta \ll \alpha$. Because IPEA requires a bitwise readout of the eigenvalue of each state of interest, any shift ℓ that reduces the magnitude of the corresponding energy increases the precision that can be obtained with a given length in the IPEA binary expansion. In practice, a reasonable choice for ℓ may be obtained, for example, from a mean-field calculation, which can be performed efficiently on a classical computer. In other words, such an algorithm directly estimates the *correlation energy* rather than the ground state energy. In order to mimic a realistic problem where mean-field theory provides a rather poor guess for the exact eigenvalues, we selected $\ell \approx 1.24$ eV, as mentioned in Sect. 4.5.2, for displacing the eigenvalues $\lambda' = \lambda - \ell$.

C.2.1 Appropriate choice of evolution t for WAVES

We are now interested in choosing an evolution time t , leading to a lower bound far from 1 in Eq. C.3, for the purity \mathcal{P} spanned by all possible trial states. This is important for the WAVES protocol, considering that eigenstate searches are (jointly) driven by \mathcal{P} . The global range of purity values will thus depend upon $\Lambda_M t = \max_{j,k} |\lambda_j - \lambda_k|t$, whereas the minimum purity observed when searching the Hilbert space between two generic eigenstates will depend upon their energy difference $\Delta_M t = \min_{j,k} |\lambda_j - \lambda_k|t$, i.e. on the degeneracy of the eigenspectrum.

If purity were to span a small range of values in the Hilbert subspace searched, a higher accuracy in the measurements would be required in turn, in order to make eigenstate witnesses sufficiently robust against experimental noise. This could be achieved, for example, via an increase in the number of measurements to obtain each data-point. Therefore, t shall be chosen big enough to keep the number of measurements as small as possible. At the same time, it is important to satisfy the requirement to invoke Jensen's inequality for \mathcal{P} to be a sensible measure of the eigenstates' support.

Considering the additional boundaries for t in WAVES, in Fig. C.1a, we plot the minimum value that \mathcal{P} assumes within the Hilbert space searched in the experiment, against t within the interval $[0, \Lambda_M]$. Considering such boundaries, we decided to pick an arbitrary integer value of

$t = 26$. Having an integer value of t simplifies the interpretation of results. This choice provides a considerable span in purity (≈ 0.35), see Fig. C.1b. At the same time, it is reasonably far from the boundary value Λ_M that occurs for \hat{H}_{exc} at $t \approx 41$, thus mimicking a situation where the span in purity for the system investigated was not pushed to an optimum: a poor initial estimate for the eigenvalues may overestimate Λ_M . It is also easy to verify that $26(\lambda_g - \lambda_e) \neq 0 \pmod{2\pi}$, therefore our choice satisfies all the conditions stated for t .

We can also easily check that $t = 26$ does not pose issues in the analysis of the eigenspectrum. Adopting again \hat{H}_{exc} , in Fig. C.1b we show that the estimators $\arg(e^{-i\lambda't})$ for both eigenvalues lie well within the interval $[-\pi, \pi]$ (even assuming approximate estimates of ± 0.2 eV).

We emphasise how the arbitrariness in preliminary choices for the parameters involved in the WAVES algorithm, within the boundaries allowed by most general considerations, are specifically intended to illustrate the robustness of the algorithm itself against poor estimates available a priori for the eigenvalues of the system investigated, which play a major role in choosing sensible parameters. The eigenvalues of \hat{H}_{exc} are readily extracted from the eigenvalues $\tilde{\lambda}_i$ provided by IPEA through back transformation as $\lambda_i = \ell + (2n\pi + \tilde{\lambda}_i)/t$, where $n \in \mathbb{N}_0$ depends upon the value chosen for t and $\tilde{\lambda}_i = 2p\tilde{\phi}_i = -2\pi[\arg(e^{-o\lambda'_it/(2\pi)})]$.

In Sect. 4.5.4 we obtained with IPEA the eigenphases $\phi_{\{e,g\}}$ leading to: $\tilde{\lambda}_g = 1.409090274$ eV and $\tilde{\lambda}_e = 0.527694267$ eV. In our case, $n = 1$ and $\lambda_g = 1.423375969$ eV, $\lambda_e = 1.497867682$ eV. Both estimates are in perfect agreement with direct diagonalisation, beyond the targeted precision (10^{-9} eV).

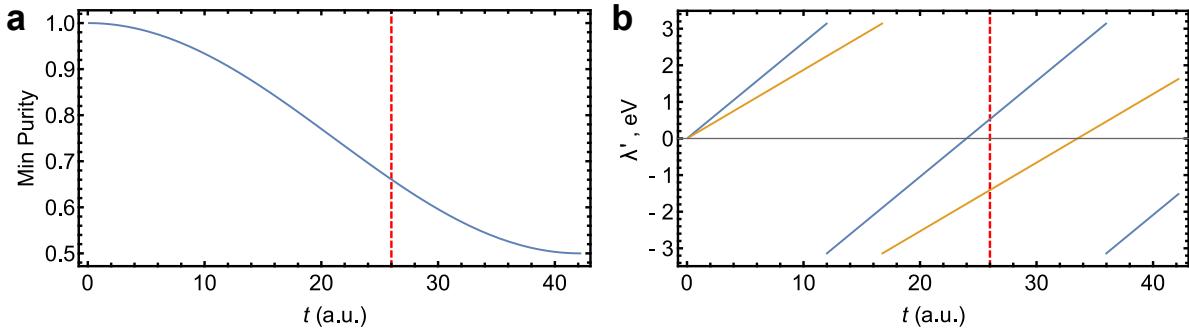


FIGURE C.1. Choice of evolution t to determine the experimental \hat{U}_{exc} **a** Span in the purity \mathcal{P} , within the interval $t \in [0, \Lambda_M]$ for \hat{H}_{exc} (solid blue line). **b** Plot of the scaled eigenphases for the ground (orange) and excited (blue) states of \hat{H}_{exc} , computed from the estimator $\tilde{\lambda}' = -\arg(e^{-i\lambda't})$. In all cases, the dashed red line indicates the choice made for both QPEA and WAVES experiment: $t = 26$.

C.3 Adiabatic State Preparation

In order to visualise the role of time in adiabatic ground state preparation, we report in Fig. C.2 the fidelity of the final state prepared when using an exemplary problem introduced by [302]:

$$\hat{H}_0 = \frac{1}{2} (\hat{I} - \hat{\sigma}_z) + \epsilon \hat{\sigma}_x \quad (C.5)$$

$$\hat{H}_1 = \frac{1}{2} (\hat{I} - \hat{\sigma}_z) \quad (C.6)$$

with ϵ capturing the intensity of the perturbation over the target Hamiltonian \hat{H}_1 . To preserve the simplicity of the example, we adopt a linear adiabatic schedule like in Eq. 4.8, with $A(s) = t/t_{\max}$. When varying t_{\max} , it is evident observing Fig. C.2 how, as soon as the condition in Eq. 4.9 is not satisfied, the fidelity F of the final state prepared¹ with the target ground state degrades quickly.

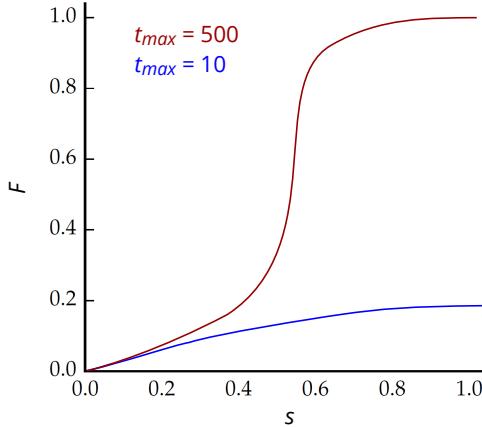


FIGURE C.2. Simulation of adiabatic ground state preparation with a linear schedule. s and t are defined in Eq. 4.8 and 4.9. The final fidelity $F = |\langle \Psi \rangle \langle \Psi_{g,1} \rangle|^2$ is reported as averaged out of 1000 independent simulations, for the two different chosen t_{\max} colour-coded in the legend.

C.4 Degeneracies in Hamiltonian models and dynamics

When discussing QHL and even more crucially QMLA strategies in the main text, we raised the issue of degeneracies in the chosen Hamiltonian model. This wording was adopted for two different situations, that we make more clear in the following.

Some degeneracies are intrinsic in the chosen model. A trivial example is provided by defining a model e.g. as $\hat{H}(a, b) = a \hat{\sigma}_x + b \hat{\sigma}_x$ to learn a true model $\hat{H}^*(\alpha) = \alpha \hat{\sigma}_x$. The model is fully legitimate, however when attempting to learn it via QHL, the expected shrinking of the covariance matrix

¹Here intended as its overlap with $|\Psi_{g,1}\rangle$.

Σ will not occur because of highly correlated (a, b) : as long as $(a + b) \simeq \alpha$ the predicted evolution $|\langle \psi | e^{-i\hat{H}^*\tau} |\psi \rangle|^2$ will satisfy the measurement for any given probe state $|\psi\rangle$.

More subtle degeneracies might instead occur because of the choice in the probe state $|\psi\rangle$, leading to undistinguishable $\mathcal{L}(D|\vec{x}; \tau)$. A first example is given by choosing an eigenstate of \hat{H}^* as the probe state. In this case, $\mathcal{L}(D|\vec{x}; \tau) = 1 \forall (\vec{x}, \tau)$ and experiments are non informative in learning a specific *parameter* value.

A less evident case might instead lead to experiments not informative in learning a specific model. We give a case of immediate applicability to the analyses of Sect. 3.4. Let us choose two models given by:

$$\hat{H}_1^* = a\hat{\sigma}_x \otimes \hat{1} + b\hat{\sigma}_y \otimes \hat{1} + c\hat{\sigma}_z \otimes \hat{1} + d\hat{\sigma}_y \otimes \hat{\sigma}_y \quad (\text{C.7})$$

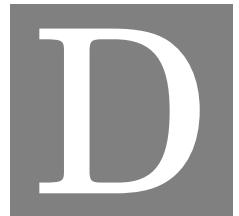
$$\hat{H}_2^* = a\hat{\sigma}_x \otimes \hat{1} + b\hat{\sigma}_y \otimes \hat{1} + c\hat{\sigma}_z \otimes \hat{1} + d\hat{\sigma}_z \otimes \hat{\sigma}_z \quad (\text{C.8})$$

(we have omitted pedices identifying the subspaces for the 2-qubit). It can be verified that the two models are distinguishable per se. However, if we choose as a probe state $|++\rangle$, the probabilities in the two cases read (we omit τ for simplicity):

$$P_1(1|a, b, c, d; \tau = 1) = \frac{1}{2} \left(\cos \left[\sqrt{a^2 + c^2 + (b - d)^2} \right] + \cos \left[\sqrt{a^2 + c^2 + (b + d)^2} \right] \right) + \\ - \frac{ia}{2} \left(\frac{\sin \left[\sqrt{a^2 + c^2 + (b - d)^2} \right]}{\sqrt{a^2 + c^2 + (b - d)^2}} + \frac{\sin \left[\sqrt{a^2 + c^2 + (b + d)^2} \right]}{\sqrt{a^2 + c^2 + (b + d)^2}} \right) \quad (\text{C.9})$$

$$P_2(1|a, b, c, d; \tau = 1) = \frac{1}{2} \left(\cos \left[\sqrt{a^2 + b^2 + (c - d)^2} \right] + \cos \left[\sqrt{a^2 + b^2 + (c + d)^2} \right] \right) + \\ - \frac{ia}{2} \left(\frac{\sin \left[\sqrt{a^2 + b^2 + (c - d)^2} \right]}{\sqrt{a^2 + b^2 + (c - d)^2}} + \frac{\sin \left[\sqrt{a^2 + b^2 + (c + d)^2} \right]}{\sqrt{a^2 + b^2 + (c + d)^2}} \right) \quad (\text{C.10})$$

Observing the two Equations, it is evident that as long as $b = c$, experiments performed on the system with the probe state $|++\rangle$ will not be able to distinguish between the models \hat{H}_1^*, \hat{H}_2^* , even if the two are not intrinsically degenerate nor indistinguishable.



APPENDIX

D.1 Simulation of an MZI

We report typical simulations and experimental results from the fabrication¹ of a fundamental building block in photonic circuits: a Mach-Zender interferometer (MZI). We investigated a set of different lengths and unbalances (ΔL) of the interferometer arms' lengths, as reported in Table D.1, along with the various Free Spectral Ranges (FSR) expected from the group index n_g , as obtained via Lumerical MODE.

| ΔL (μm) | FSR (nm) (numerical) | FSR (nm) (experimental) |
|------------------------|----------------------------|-------------------------------|
| 30 | 19.05385042 | NA |
| 100 | 5.716155127 | 5.80 |
| 180 | 3.175641737 | 3.16 height |

Table D.1: List of different ΔL investigated for the MZIs in Sect. D.1.

Using Lumerical INTERCONNECT and the SiEPIC library we also simulate the MZIs designed in detail, as reported in Fig. D.1. Simulations of the FSR compare well with what can be obtained from the spectral response. In particular, we adopt the steps below:

- We plot the MZI spectrum, and correct for any baseline with a low-order polynomial (usually a 5th order polynomial)
- We then adopt the *findpeaks* function embedded in Matlab to identify the wavelengths at which destructive interference occurs and thys the FSR, as highlighted in Fig. D.1a.

¹ In this section, as performed by Applied Nanotools Inc. which is based on direct-write 100 keV electron beam lithography technology. The results were obtained from an automated characterisation of the devices, as described in [94].

- Using the known value of ΔL for the MZI, we can use the FSR estimate (i.e. the spacing between peaks) to obtain a first guess of the group index, and hence of one fit parameter in the analytical MZI response function.
- A Taylor expansion in the proximity of one of the peaks is then adopted to find the initial guess for remaining fit parameters of the MZI response function [94].

A comparison of the final fit with the experimental data is shown in Fig. D.1b, showing a good agreement with raw experimental data.

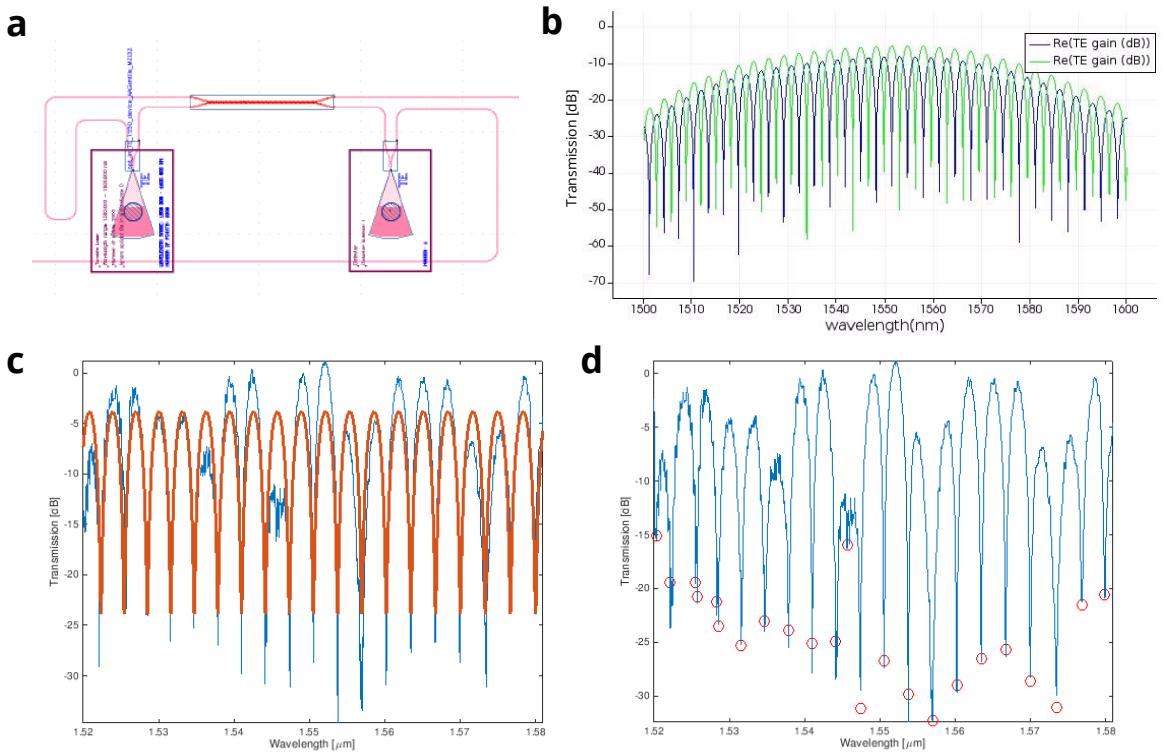


FIGURE D.1. **a** Layout of the MZI fabricated and analysed in Sect. D.1, as obtained from the Klayout software, with $\Delta L \approx 180\mu\text{m}$. **b** Spectral response simulated via SiEPIC and Lumerical Interconnect for the MZI in (a). **c** The peak finding phase of the protocol for the analysis of experimental data collected from the MZI, with in **d** the final fit of the MZI spectrum. Further details in the main text.

D.2 Qudit tomographies with the CU(4) chip

We here verify experimentally the reliability of the CU(4) chip, in dialling the appropriate states and evolution unitaries to act as a reliable quantum simulator, for the QMLA implementation described in Sect. 3.5.3. To do so, we report here the outcome of qudit state tomographies

performed with the CU(4) chip, reconstructing the output state ρ_{out} from the target registry at the end of the evolution. We performed this test with 50 different test unitaries², however for brevity we report in Fig. D.2c only the case for:

$$\hat{U}_{\text{test}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (\text{D.1})$$

In particular, in Fig. D.2a&b we report detailed tomography outcomes of a random chosen input state: $|\theta\rangle \equiv |0, 1/\sqrt{2}, 0, i/\sqrt{2}\rangle$, obtaining a final fidelity $F(|\theta\rangle) > 0.99$. In Fig. D.2 we report instead the obtained fidelities for all the tested 15 probe states (again labelled by Greek letters), after evolving them with the same \hat{U}_{test} . The average fidelity $F > 0.92$, as stated in the main text, with $F(|l\rangle) = \langle l| \hat{U}_{\text{test}}^\dagger \rho_{out} \hat{U}_{\text{test}} |l\rangle$.

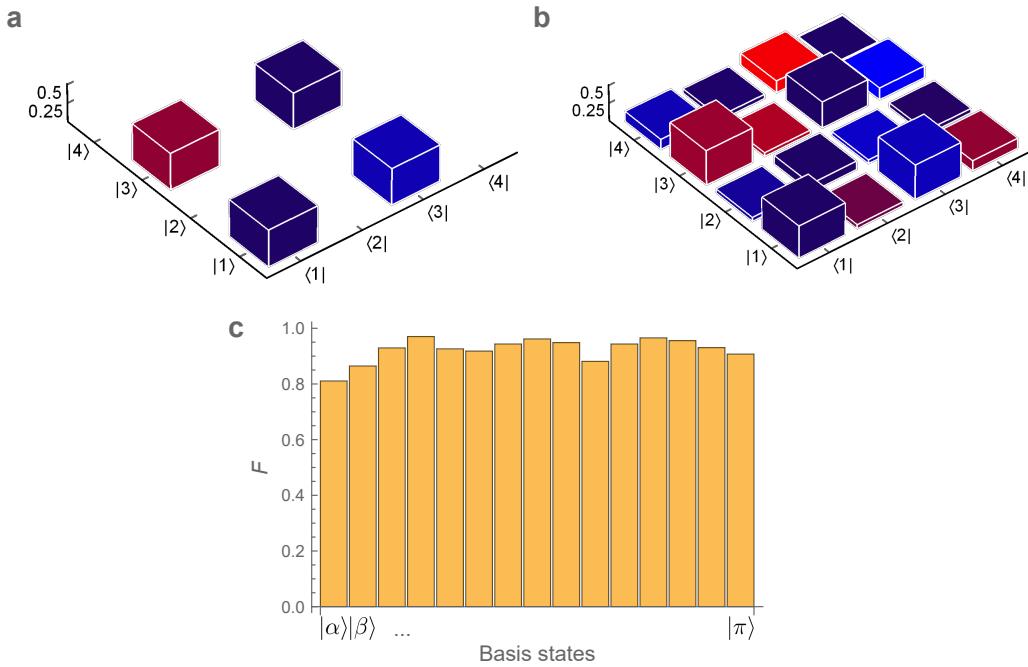


FIGURE D.2. State tomographies for selected probe states, adopting the CU(4) photonic chip introduced in Sect. 3.5.3. In particular, in: **a** simulated and **b** experimentally reconstructed ρ_{out} , when implementing \hat{U}_{test} (Eq. D.1) to evolve the probe state $|\theta\rangle$ (see Appendix D.2). **c** Fidelity for 15 test states

² The final scope was to perform QPT by performing state tomography over a set of 50 different test \hat{U} and 15 different probes $\{|l\rangle\}$.

D.3 Low-power reconfigurability for cryogenic photonics

The practicality of the quantum devices employed in delivering successful protocols has been mostly based upon designing and testing noise–resilient approaches in the data processing. In this way, it was possible to avoid quantum circuits with width and depth too high to be handled within current photonic technological limits, or the demand for a cryogenic environment when using solid–state quantum sensors.

However, software developments alone are likely not enough to push the application boundaries of quantum technologies in many cases of interest. In this Chapter, we pick a specific technological challenge: the reconfigurability of photonic devices in a cryogenic environment, discussing the investigation of possible routes to address it. As we have described in Sect. 1.4, and then observed in practice for the experiments described adopting both the CU(2) and CU(4) chips in Chaps. 2–3, the possibility to reconfigure the photonic circuit is key, in deploying a scalable photonic architecture for LOQC as well as digital quantum simulation. A brief review of technologies investigated to date, and details about the thermo–optical (TO) phase–shifters used for all photonic experiments so far, were provided in Sect. 1.4.1.

TO phase–shifters are limited to operate in the range of MHz [92, 345]. This can be a severe limitation in several applications. In LOQC, the feedforward scheme requires high–speed reconfiguration, to avoid incurring in unacceptable losses when employing on–purpose delay lines[92, 95]. High–speed reconfigurability is also a necessary requirement for classical photonic processors, recently pursued as a promising route towards achieving lower power and higher speed data processing, when compared to their electronic counterparts (e.g. [346, 347]).

A perhaps even stronger limitation appears when considering cryogenic operation scenarios, required either by efficient superconducting classical processors [345], or by the monolithic integration of SNSPDs in the quantum photonics realm³. It is known that the TO coefficient rapidly drops below 10^{-5} K^{-1} at temperatures below 50 K [348]. This might be (at least partially) compensated by the corresponding decrease in the specific heat coefficient [100], however, even if allowing for room–temperature power consumptions, TO heaters are no scalable approach in a cryogenic environment. We make this claim more evident, by considering the power consumption of the smaller photonic chip in this Thesis, the CU(2) in Fig. 1.5. As a study case, we pick the ground state search operated in Sect. 4.5.4, and keep track of the chip *active time*, i.e. the time during which at least one of the heaters on–chip was activated to perform a certain phase shift. The plots visualising this are in Fig. D.3. In order to have a reference for the chip power consumption, we consider the ICEoxford™ DryIce®, a commercial closed–loop cryostat cooled via a ^4He gas circuit⁴. At the time of writing, this system is among the commercially available cryostats with highest cooling power in continuous–mode operation⁵ ($W_{\text{cnt}} \sim 300 \text{ mW}$) at 1 K

³ A discussion is provided in 1.4.2.

⁴ I have personally contributed to the installation and preliminary characterisation of this setup for the University of Bristol

⁵ In brief, the cryostat has a limited ^4He reservoir, that can undergo a complete expansion and compression cycle

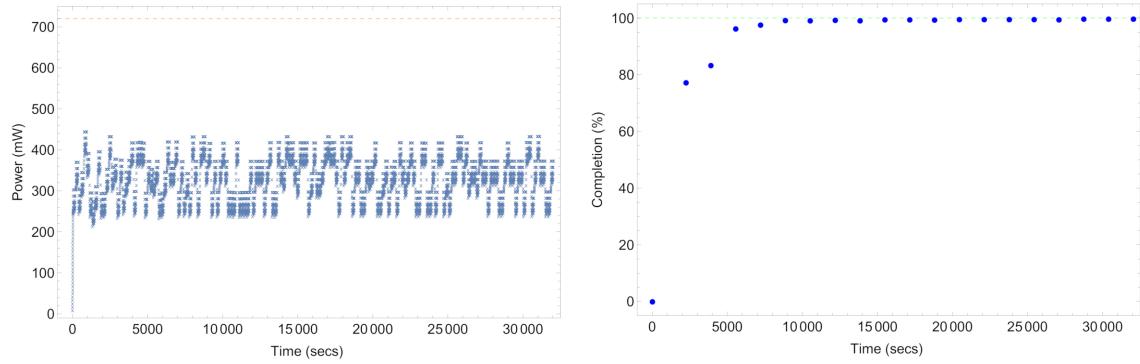


FIGURE D.3. **a** Estimated instantaneous power consumption of the CU(2) chip, during the variational search experiment shown in Fig. 4.8a. **b** Completion of the same experiment as in (a), estimated from $|\langle \Psi | \bar{\Psi} \rangle|^2$, with $|\Psi_i\rangle$ the best estimate of the target eigenstate $|\bar{\Psi}\rangle$, at the i -th step, rescaled to the interval [0%, 100%].

temperature. Observing Fig. D.3a, it is immediate to conclude that, by relying on TO effects that require $\mathcal{O}(\text{mW})$ power for a full 2π -shift, the number of heaters on a photonic chip cannot be scaled much further, whilst working in continuous. In single-shot mode, heat sources up to $W_{\text{sng}} \sim 700 \text{ mW}$ can be installed, but only for experiment durations below about 30 mins, which does not satisfy experimental requirements (see Fig. D.3b). Now, quantum photonic chips with $\sim 8\times$ the heaters in the CU(2) are being designed and operated [349], so that replicating those experiments in a cryostat is already challenging, even allowing for incremental optimisations in the TO power efficiency.

Therefore, as already mentioned in Sect. 1.4.1, electro-optic (EO) modulators offer a much more attractive solution to the problem of modulating the output signal from a photonic circuit at high-speed [94]. In photonic platforms providing strong $\chi^{(2)}$ nonlinearities (Sect. 1.4.1), this approach is rather straightforward, as it is possible to rely on the strong *Pockels effect*, whereby an applied electric field \vec{E} changes the refractive index of the material by $\Delta n = r_{\text{eff}} |\vec{E}|$. More in detail, as the Pockels effect usually occurs in anisotropic materials, it is described by a change in

within a single “clock”. The maximum heat that can be absorbed by the ${}^4\text{He}$ circulated within this cycle gives the cooling power in continuous-mode. When this limit is overcome, the temperature at the sample stage T_{sample} will not increase immediately. However, after a $\Delta\tau$, enough non-condensed ${}^4\text{He}$ (that has extracted the additional heat) will accumulate in the reservoir, heat will start dissipating within the cryostat, and T_{sample} will rise sharply. The *single-shot* mode of operation is within the $\Delta\tau$ window, with heat sources exceeding the cryostat’s W_{cnt} .

the ellipsoid of refractive indices, governed by an applied electric field \vec{E} [350]:

$$\begin{pmatrix} \Delta\left(\frac{1}{n^2}\right)_1 \\ \Delta\left(\frac{1}{n^2}\right)_2 \\ \Delta\left(\frac{1}{n^2}\right)_3 \\ \Delta\left(\frac{1}{n^2}\right)_4 \\ \Delta\left(\frac{1}{n^2}\right)_5 \\ \Delta\left(\frac{1}{n^2}\right)_6 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}, \quad (\text{D.2})$$

with r_{ij} the 18 significant Pockels coefficients⁶, and each $\Delta(1/n^2)_i$ refers to a change in the corresponding ellipsoid index n_i . We can write Eq. D.2 in compact notation as:

$$\Delta\left(\frac{\vec{1}}{n^2}\right) = R\vec{E} \quad (\text{D.3})$$

where R is known as the *Pockels tensor*. Therefore, a different EO response is expected according to the mutual geometry of the applied field \vec{E} and the crystallographic axes of the material [350, 351].

EO modulation in SOI platforms poses more intricacies, as here electro-optic phenomena are dominated by the $\chi^{(3)}$ -related *Kerr effect*, which unfortunately is weak in silicon [129]. Additional challenges arise when considering SOI quantum photonics, as for a satisfactory modulation several performances should be targeted simultaneously:

- i CMOS-compatibility, or compatibility with industrial-scale fabrication technologies for integrated photonic devices;
- ii device footprint;
- iii losses and noise introduced by the modulator;
- iv power consumption;
- v speed / bandwidth;
- vi low temperature (1 – 4 K) operability.

Standard EO modulators in silicon photonics leverage upon the *plasma dispersion effect*: altering the concentration⁷ of free charges in Si it is possible to modulate the material refractive index n [94, 129]. Unfortunately, already at room temperature these EO modulators are known to introduce important additional noise and losses in the device [92, 129]. This latter drawback is made worse when moving to lower operating temperatures, because the carrier freeze-out imposes higher concentrations of dopants in the Si WG, which in turns lead to higher losses and lower bandwidth [352, 353]. All-optical modulation, if a promising candidate for classical

⁶ From purely geometric considerations, one could introduce 81 coefficients r_{ijk} , however from symmetry considerations for the impermeability tensor, it is immediate to conclude that only 18 are non-zero [350].

⁷ Either via carrier *accumulation*, *injection* or *depletion* [129].

photonic devices [101, 354], is impractical in the realm of quantum photonics, where filtering out on-demand signals from an additional bright pump is possible in principle, but technologically difficult (refer also to Sect. 1.4.2).

The failure of $\chi^{(3)}$ -based approaches to deliver a suitable modulation of n in Si has pushed for solutions that introduce a $\chi^{(2)}$ effect in SOI platforms, which can be achieved in several different ways. As $\chi^{(2)} = 0$ is due to the centro-symmetric structure of Si crystals, a possibility is to break such symmetry, by either strain or strong electrical fields [92]. If the CMOS compatibility of the first approach is yet to be proven, such an implementation has already been demonstrated with satisfactory bandwidth, limited to few GHz [355]; however, the latter still requires improvements, as it introduces additional losses as high as ~ 0.2 dB per modulator.

Finally, a more radical and direct approach is to hybridise the SOI platform⁸ by introducing materials that natively exhibit strong non-linearities (e.g. they have $\chi^{(2)} \neq 0$) [92]. To achieve this, it is not necessary to completely replace the material of the WG: as emphasised by the simulations in Fig. 1.2, the confinement of the propagated light mode within the WG is high, but not perfect. Therefore, by depositing other materials in the proximity of the Silicon WG, where the evanescent field is not yet negligible, modulation of the effective refractive index is still possible. This is the approach pursued in the next sections of this Chapter.

A natural choice for such a hybrid material was offered by LiNbO₃. However, even if already widely adopted in the field of classical communication, its widespread adoption has been hindered by the lack of a credible CMOS-compatible process to integrate it in SOI, a problem affecting also lead zirconate titanate (PZT) modulators, that exhibit even stronger Pockels effect [356]. Finally, polymeric materials have also been considered a promising candidate, because of their strong Pockels effect, however none of such candidates is cryo-compatible [353, 356].

D.4 Characterising vibration levels in a cryogenic setup

This section is dedicated to projects and experiments ongoing, that involve setups working at cryogenic temperatures. In particular, I report data concerning the characterisation of vibrations for a recently closed-loop He-cooled commercial cryostat from ICEOxford®, and a Lakeshore® open-loop cryogenic probe station.

The quantification and optimisation of vibrations on the setup sample holder is a crucial task. The role of vibration in photonic experiments essentially resides in the vertical coupling of coherent light from external sources, into (and out of) the chip, via fibre-arrays as described in Sect. 1.4.2. This can be an important source of noise in cryogenic setups, due to mechanical vibration from pumping and cooling systems potentially coupling to the setup. As on a single chip it is often possible to couple to several structures of interest, it is a preferred option to leave

⁸ Of course, the most obvious approach is to simply replace Si WGs with another material. Even if this is an active field of research (see e.g. [94]), we will skip discussing such solution for brevity, as the focus of this Thesis are SOI platforms.

the fibre-array *floating*, instead of glueing it upon a specific structure. In this way, it is possible to reconfigure the fibre-array location with cryogenic-compatible positioners. We report here two successfully tested strategies to check vibration levels in such equipment.

A preliminary analysis, can be performed via commercially-available MEMS accelerometers, mounted directly upon the sample stage. In our case, the accelerometer is capable of providing reliable data for temperatures as low as $-50\text{ }^{\circ}\text{C}$, on all 3-axes and down to frequencies of $\sim 200\text{ Hz}$. I installed and interfaced the device with *Arduino*, and wrote from scratch the python code for control and datalogging, as well as *Mathematica* code for noise-filtering and data analysis. In Fig. D.4a we report raw data collected from the accelerometer, shown in Subfig. (b). The z axis is well distinguishable because of the gravity \vec{g} detection, with all the instrumentation required for sub-K operation connected⁹. It is immediately evident how the low-frequency component from the compressors is actually providing the spikes with biggest magnitude in (de)accelerations of the sample stage.

According to this evidence, we processed acceleration data by filtering out high frequency noise and double-integrating the residual components, allowing a quantitative estimation of the maximum relative displacements experienced by the accelerometer. The results are reported as frequency histograms (2D and 1D) in Fig. D.4c-d. The last one refers to vertical axis displacements, with values that are not expected to alter anyhow the coupling of light in and off chip. On the contrary, an un-optimised mounting for the cryo-cooler-heads leads evidently in Fig. D.4c to in-plane vibrations that are comparable with the GC spot size, which has a diameter of approximately $5\text{ }\mu\text{m}$, e.g. in the case of the CU(4) chip. This emphasises the importance of adopting appropriate vibration decoupling when working with cryogenic setups.

⁹ Acceleration was tracked also with all pumps and compressors in the setup off, to record the background (omitted), which was barely detectable by the accelerometer.

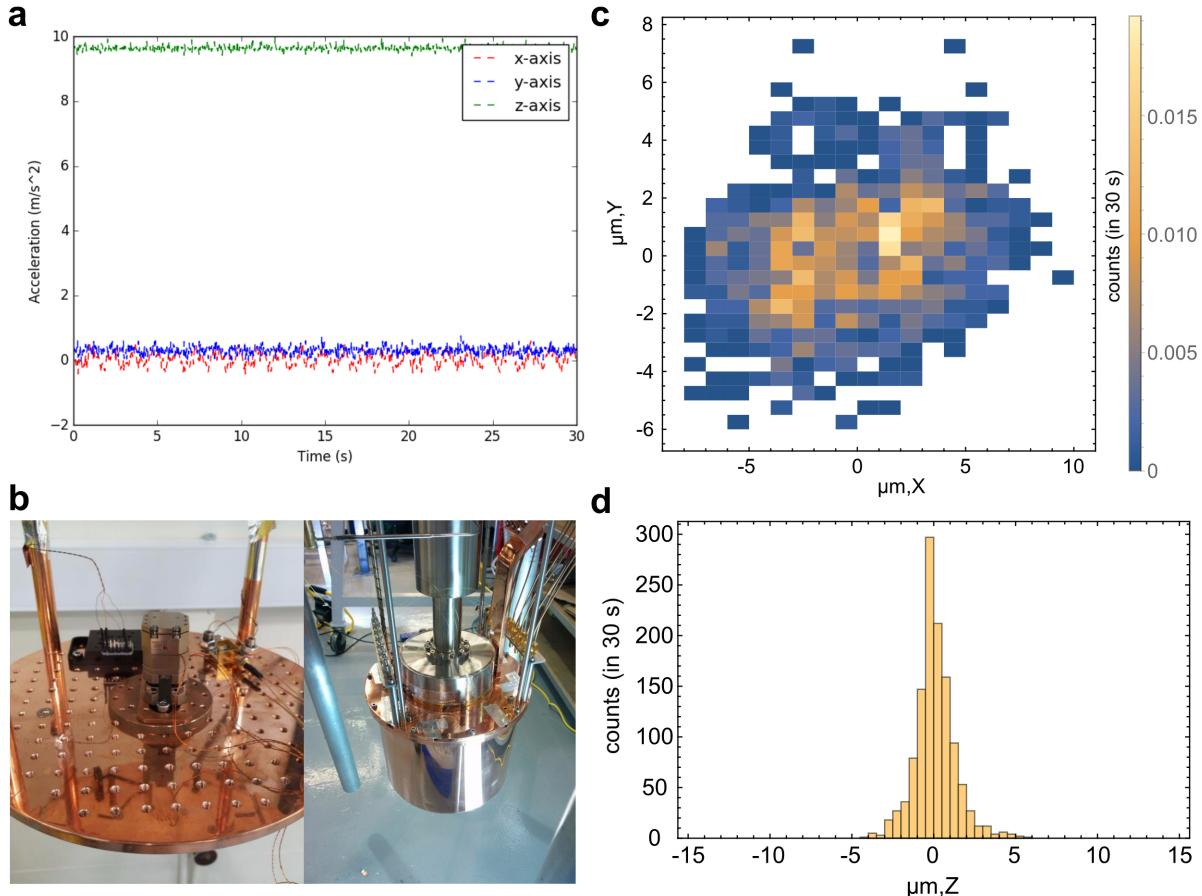


FIGURE D.4. Characterisation of vibrations in a cryogenic setup. **a** Raw acceleration data at -50°C , with the cryostat setup fully operations, from the MEMS accelerometer, shown in **b** mounted on the sample plate, hosted in the stage at temperature 4 K (on the right). **c** Post-processed data about the absolute displacement of the sample stage, with colour-coded the number of displacement events per each value on the X-Y plane. **d** Same as in (c) for the vertical displacement, with event frequencies reported on the vertical axis of the histogram.

BIBLIOGRAPHY

- [1] R. R. Schaller, “Moore’s law: past, present and future,” *IEEE spectrum* **34**, 52–59 (1997).
- [2] M. A. Kastner, “The single-electron transistor,” *Reviews of modern physics* **64**, 849 (1992).
- [3] M. Lundstrom, “Moore’s law forever?” *Science* **299**, 210–211 (2003).
- [4] A. M. Ionescu and H. Riel, “Tunnel field-effect transistors as energy-efficient electronic switches,” *Nature* **479**, 329 (2011).
- [5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2001).
- [6] W.-K. Chen, *The VLSI handbook* (CRC press, 2010).
- [7] E. Sicard, “Introducing 7-nm finfet technology in microwind,” Technical Report (2017).
- [8] H. Wasserman, M. Simmons, and A. Hayes, *An assessment of worldwide supercomputer usage*, Tech. Rep. (Los Alamos National Lab, 1995).
- [9] N. Attig, P. Gibon, and T. Lippert, “Trends in supercomputing: The european path to exascale,” *Computer Physics Communications* **182**, 2041–2046 (2011).
- [10] X. Li and C. A. Wang, “The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin,” *Decision Support Systems* **95**, 49–60 (2017).
- [11] R. S. Sutton, A. G. Barto, et al., *Introduction to reinforcement learning*.
- [12] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (Pearson Education Ltd., 2016).
- [13] M. E. Cuffaro and S. C. Fletcher, *Physical perspectives on computation, computational perspectives on physics* (Cambridge University Press, 2018).
- [14] E. H. Allen, C. S. Calude, et al., “Quassical computing,” arXiv preprint 1805.03306 (2018).
- [15] D. Deutsch, “Quantum theory, the Church–Turing principle and the universal quantum computer,” *Proceedings of the Royal Society of London A* **400**, 97–117 (1985).

BIBLIOGRAPHY

- [16] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM review* **41**, 303–332 (1999).
- [17] A. K. Lenstra, W. Hendrik Jr, et al., *The development of the number field sieve* (Springer Science & Business Media, 1993).
- [18] S. Aaronson and A. Arkhipov, “The computational complexity of linear optics,” in *Proceedings of the forty-third annual ACM symposium on Theory of computing* (ACM, 2011) pp. 333–342.
- [19] S. Aaronson, “Shtetl-optimized,” <https://www.scottaaronson.com/blog/?p=4317> (2019).
- [20] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al., “Quantum supremacy using a programmable superconducting processor,” *Nature* **574**, 505–510 (2019).
- [21] V. Giovannetti, S. Lloyd, and L. Maccone, “Advances in quantum metrology,” *Nature Photonics* **5**, 222–229 (2011).
- [22] V. Giovannetti, S. Lloyd, and L. Maccone, “Quantum-enhanced measurements: Beating the standard quantum limit,” *Science* **306**, 1330–1336 (2004).
- [23] C. L. Degen, F. Reinhard, and P. Cappellaro, “Quantum sensing,” *Reviews of Modern Physics* **89**, 035002 (2017).
- [24] D. F. Walls, “Squeezed states of light,” *Nature* **306**, 141 (1983).
- [25] V. Giovannetti, S. Lloyd, and L. Maccone, “Quantum metrology,” *Physical Review letters* **96**, 010401 (2006).
- [26] J. Aasi, J. Abadie, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, C. Adams, T. Adams, P. Addesso, R. Adhikari, et al., “Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light,” *Nature Photonics* **7**, 613 (2013).
- [27] J. P. Dowling and G. J. Milburn, “Quantum technology: the second quantum revolution,” *Philos. Trans. R. Soc. Lond. A: Mathematical, Physical and Engineering Sciences* **361**, 1655–1674 (2003).
- [28] H.-P. Breuer, F. Petruccione, et al., *The theory of open quantum systems* (Oxford University Press on Demand, 2002).
- [29] C. M. Dawson and M. A. Nielsen, “The solovay-kitaev algorithm,” *arXiv preprint quant-ph/0505030* (2005).

- [30] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, “Elementary gates for quantum computation,” *Physical Review A* **52**, 3457 (1995).
- [31] D. W. Smith, “N-representability problem for fermion density matrices. ii. the first-order density matrix with n even,” *Physical Review* **147**, 896 (1966).
- [32] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, “Simulating chemistry using quantum computers,” *Annual review of physical chemistry* **62**, 185–207 (2011).
- [33] N. Wiebe, D. Berry, P. Høyer, and B. C. Sanders, “Higher order decompositions of ordered operator exponentials,” *Journal of Physics A: Mathematical and Theoretical* **43**, 065203 (2010).
- [34] V. Kliuchnikov, D. Maslov, and M. Mosca, “Fast and efficient exact synthesis of single qubit unitaries generated by clifford and t gates,” *Quantum Information & Computation* **13**, 0607–0630 (2012).
- [35] R. Jozsa, “Fidelity for mixed quantum states,” *Journal of Modern Optics* **41**, 2315–2323 (1994).
- [36] M. A. Nielsen, “A simple formula for the average gate fidelity of a quantum dynamical operation,” *Physics Letters A* **303**, 249–252 (2002).
- [37] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, “Bounding quantum gate error rate based on reported average fidelity,” *New Journal of Physics* **18**, 012002 (2015).
- [38] D. P. DiVincenzo, “The physical implementation of quantum computation,” *Fortschritte der Physik: Progress of Physics* **48**, 771–783 (2000).
- [39] X. Hu, R. de Sousa, and S. D. Sarma, “Decoherence and dephasing in spin-based solid state quantum computers,” in *Foundations Of Quantum Mechanics In The Light Of New Technology: ISQM—Tokyo’01* (World Scientific, 2002) pp. 3–11.
- [40] W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. Gilligan, D. J. Heinzen, F. Moore, M. Raizen, and D. J. Wineland, “Quantum projection noise: Population fluctuations in two-level systems,” *Physical Review A* **47**, 3554 (1993).
- [41] T. D. Wickens, *Elementary signal detection theory* (Oxford University Press, USA, 2002).
- [42] E. Knill, “Quantum computing with realistically noisy devices,” *Nature* **434**, 39–44 (2005).
- [43] D. Gottesman, “Quantum fault tolerance in small experiments,” arXiv preprint 1610.03507 (2016).

BIBLIOGRAPHY

- [44] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, “Fault-tolerant quantum error detection,” *Science Advances* **3**, e1701074 (2017).
- [45] R. Harper and S. T. Flammia, “Fault-Tolerant Logical Gates in the IBM Quantum Experience,” *Physical Review Letters* **122**, 080504 (2019).
- [46] B. W. Reichardt, “Improved ancilla preparation scheme increases fault-tolerant threshold,” arXiv preprint quant-ph/0406025 (2004).
- [47] R. Raussendorf and J. Harrington, “Fault-Tolerant Quantum Computation with High Threshold in Two Dimensions,” *Physical Review Letters* **98**, 190504 (2007).
- [48] J. Preskill, “Quantum Computing in the NISQ era and beyond,” arXiv preprint 1801.00862 (2018).
- [49] H. T. Siegelmann, “Computation beyond the turing limit,” *Science* **268**, 545–548 (1995).
- [50] G. Kalai, “The quantum computer puzzle,” *Notices of the AMS* **63**, 508–516 (2016).
- [51] M. Dyakonov, “When will useful quantum computers be constructed? not in the foreseeable future, this physicist argues. here’s why: The case against: Quantum computing,” *IEEE Spectrum* **56**, 24–29 (2019).
- [52] A. W. Harrow and A. Montanaro, “Quantum computational supremacy,” *Nature* **549**, 203–209 (2017).
- [53] J. Preskill, “Quantum computing and the entanglement frontier,” arXiv preprint 1203.5813 (2012).
- [54] A. Neville, C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, “Classical boson sampling algorithms with superior performance to near-term experiments,” *Nature Physics* **13**, 1153 (2017).
- [55] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, “What is the computational value of finite-range tunneling?” *Physical Review X* **6**, 031015 (2016).
- [56] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. Pryde, and H. M. Wiseman, “How to perform the most accurate possible phase measurements,” *Physical Review A* **80**, 3–22 (2009).
- [57] R. S. Said, D. W. Berry, and J. Twamley, “Nanoscale magnetometry using a single-spin system in diamond,” *Physical Review B* **83**, 125410 (2011).

- [58] F. Poggiali, P. Cappellaro, and N. Fabbri, “Optimal control for one-qubit quantum sensing,” *Physical Review X* **8**, 21059 (2017).
- [59] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature* **549**, 195 (2017).
- [60] M. Schuld, “Machine learning in quantum spaces,” *Nature* **567**, 179–181 (2019).
- [61] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Physical Review letters* **113**, 130503 (2014).
- [62] S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum principal component analysis,” *Nature Physics* **10**, 631 (2014).
- [63] N. Wiebe, D. Braun, and S. Lloyd, “Quantum algorithm for data fitting,” *Physical Review letters* **109**, 050505 (2012).
- [64] A. W. Harrow, A. Hassidim, and S. Lloyd, “Quantum algorithm for linear systems of equations,” *Physical Review letters* **103**, 150502 (2009).
- [65] S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum algorithms for supervised and unsupervised machine learning,” arXiv preprint arXiv:1307.0411 (2013).
- [66] N. Wiebe, A. Kapoor, and K. M. Svore, “Quantum nearest-neighbor algorithms for machine learning,” *Quantum Information and Computation* **15** (2018).
- [67] I. Kerenidis and A. Prakash, “Quantum recommendation systems,” arXiv preprint 1603.08675 (2016).
- [68] W. Zeng and B. Coecke, “Quantum Algorithms for Compositional Natural Language Processing,” *Electronic Proceedings in Theoretical Computer Science* **221**, 67–75 (2016).
- [69] M. Schuld, I. Sinayskiy, and F. Petruccione, “The quest for a quantum neural network,” *Quantum Information Processing* **13**, 2567–2586 (2014).
- [70] N. Wiebe, A. Kapoor, and K. M. Svore, “Quantum deep learning,” *Quantum Information & Computation* **16**, 541–587 (2016).
- [71] I. Cong, S. Choi, and M. D. Lukin, “Quantum Convolutional Neural Networks,” arXiv preprint 1810.03787 (2018).
- [72] J. Adcock, E. Allen, M. Day, S. Frick, J. Hinchliff, M. Johnson, S. Morley-Short, S. Pallister, A. Price, and S. Stanisic, “Advances in quantum machine learning,” arXiv preprint 1512.02900 (2015).

BIBLIOGRAPHY

- [73] A. Narayanan and M. Moore, “Quantum-inspired genetic algorithms,” in *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on* (IEEE, 1996) pp. 61–66.
- [74] L. Chakhmakhchyan, N. J. Cerf, and R. Garcia-Patron, “Quantum-inspired algorithm for estimating the permanent of positive semidefinite matrices,” *Physical Review A* **96**, 1–8 (2017).
- [75] N. Wiebe, A. Kapoor, C. Granade, and K. M. Svore, “Quantum inspired training for boltzmann machines,” arXiv preprint 1507.02642 (2015).
- [76] E. Tang, “A quantum-inspired classical algorithm for recommendation systems,” arXiv preprint 1807.04271 (2018).
- [77] A. Gilyén, S. Lloyd, and E. Tang, “Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension,” arXiv preprint 1811.04909 (2018).
- [78] E. Zahedinejad, J. Ghosh, and B. C. Sanders, “Designing high-fidelity single-shot three-qubit gates: a machine-learning approach,” *Physical Review Applied* **6**, 054005 (2016).
- [79] L. Banchi, N. Pancotti, and S. Bose, “Quantum gate learning in engineered qubit networks: Toffoli gate with always-on interactions,” *npj Quantum Information* (2015), 10.1038/npjqi.2016.19.
- [80] M. August and X. Ni, “Using recurrent neural networks to optimize dynamical decoupling for quantum memory,” *Physical Review A* **95**, 012335 (2017).
- [81] A. Hentschel and B. C. Sanders, “Machine learning for precise quantum measurement,” *Physical Review letters* **104**, 063603 (2010).
- [82] N. Wiebe and C. Granade, “Efficient bayesian phase estimation,” *Physical Review Letters* **117**, 010503 (2016).
- [83] Z. Li, X. Liu, N. Xu, and J. Du, “Experimental realization of a quantum support vector machine,” *Physical Review letters* **114**, 140504 (2015).
- [84] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum enhanced feature spaces,” *Nature* , 7–10 (2018).
- [85] L. S. Bishop, S. Bravyi, A. Cross, J. M. Gambetta, and J. Smolin, “Quantum volume,” Technical Report (2017).
- [86] G. Kurizki, P. Bertet, Y. Kubo, K. Mølmer, D. Petrosyan, P. Rabl, and J. Schmiedmayer, “Quantum technologies with hybrid systems.” *PNAS* **112**, 3866–73 (2015).

- [87] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, “Trapped-ion quantum computing: Progress and challenges,” *Applied Physics Reviews* **6**, 021314 (2019).
- [88] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, “Experimental Comparison of Two Quantum Computing Architectures,” *PNAS* **13**, 3305 (2017).
- [89] C. Gross and I. Bloch, “Quantum simulations with ultracold atoms in optical lattices.” *Science* **357**, 995–1001 (2017).
- [90] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, et al., “A programmable two-qubit quantum processor in silicon,” *Nature* **555**, 633–637 (2018).
- [91] J. Kelly, Z. Chen, B. Chiaro, B. Foxen, J. Martinis, and Q. H. T. Team, “Operating and characterizing of a 72 superconducting qubit processor“bristlecone”: Part 1,” in *APS Meeting Abstracts* (2019).
- [92] J. Silverstone, D. Bonneau, J. L. O’Brien, and M. G. Thompson, “Silicon quantum photonics,” *IEEE* **22**, 6700113 (2016).
- [93] G. Lifante, *Integrated photonics: fundamentals* (Wiley Online Library, 2003).
- [94] L. Chrostowski and M. Hochberg, *Silicon photonics design: from devices to systems* (Cambridge University Press, 2015).
- [95] J. L. O’Brien, “Optical quantum computing,” *Science* **318**, 390–397 (2007).
- [96] A. Politi, M. J. Cryan, J. G. Rarity, S. Yu, and J. L. O’Brien, “Silicon-on-silicon waveguide quantum circuits,” *Science* **320**, 646–649 (2008), <http://science.sciencemag.org/content/320/5876/646.full.pdf>.
- [97] M. A. Ciampini, N. Spagnolo, C. Vitelli, L. Pezzè, A. Smerzi, and F. Sciarrino, “Quantum-enhanced multiparameter estimation in multiarm interferometers,” *Scientific Reports* **6**, 222 (2016).
- [98] C. Gerry, P. Knight, and P. L. Knight, *Introductory quantum optics* (Cambridge university press, 2005).
- [99] J. Wang, D. Bonneau, M. Villa, J. W. Silverstone, R. Santagati, S. Miki, T. Yamashita, M. Fujiwara, M. Sasaki, H. Terai, et al., “Chip-to-chip quantum photonic interconnect by path-polarization interconversion,” *Optica* **3**, 407–413 (2016).
- [100] G. Villareal-Garcia, *Thermo-optic phase shifters for integrated photonics at low temperatures*, Ph.D. thesis, University of Bristol (2018).

BIBLIOGRAPHY

- [101] G. P. Agrawal, *Lightwave technology: components and devices* (John Wiley & Sons, 2004).
- [102] A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, “Fabrication-constrained nanophotonic inverse design,” *Scientific Reports* **7**, 1786 (2017).
- [103] J. M. Heaton and R. M. Jenkins, “General matrix theory of self-imaging in multimode interference (mmi) couplers,” *IEEE Photonics Technology Letters* **11**, 212–214 (1999).
- [104] R. Santagati, *Towards quantum information processing with spins in silicon*, Ph.D. thesis, University of Bristol (2015).
- [105] J. Rarity, P. Tapster, E. Jakeman, T. Larchuk, R. Campos, M. Teich, and B. Saleh, “Two-photon interference in a Mach-Zehnder interferometer,” *Physical Review Letters* **65**, 1348–1351 (1990).
- [106] A. Migdall, S. V. Polyakov, J. Fan, and J. C. Bienfang, *Single-photon generation and detection: physics and applications*, Vol. 45 (Academic Press, 2013).
- [107] C. K. Hong, Z. Y. Ou, and L. Mandel, “Measurement of subpicosecond time intervals between two photons by interference,” *Physical Review Letters* **59**, 2044–2046 (1987).
- [108] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, “Experimental quantum teleportation,” *Nature* **390**, 575–579 (1997).
- [109] G. Vallone, D. Bacco, D. Dequal, S. Gaiarin, V. Luceri, G. Bianco, and P. Villoresi, “Experimental Satellite Quantum Communications,” *Physical Review Letters* **115**, 040502 (2015).
- [110] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, et al., “Satellite-to-ground quantum key distribution,” *Nature* **549**, 43–47 (2017).
- [111] B. Hensen, H. Bernien, A. E. Dreaú, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, et al., “Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres,” *Nature* **526**, 682–686 (2015).
- [112] N. Gisin and R. Thew, “Quantum communication,” *Nature Photonics* **1**, 165–171 (2007).
- [113] C. Simon, “Towards a global quantum network,” *Nature Photonics* **11**, 678–680 (2017).
- [114] J. W. Silverstone, *Entangled light in Silicon waveguides*, Ph.D. thesis, University of Bristol (2015).
- [115] S. Paesani, Y. Ding, R. Santagati, L. Chakhmakhchyan, C. Vigliar, K. Rottwitt, L. K. Oxenløwe, J. Wang, M. G. Thompson, and A. Laing, “Generation and sampling of quantum states of light in a silicon chip,” *Nature Physics* **15**, 925–929 (2019).

- [116] R. W. Boyd, *Nonlinear optics* (Elsevier, 2003).
- [117] J. W. Silverstone, R. Santagati, D. Bonneau, M. J. Strain, M. Sorel, J. L. O'Brien, and M. G. Thompson, “Qubit entanglement between ring-resonator photon-pair sources on a silicon chip,” *Nature Communications* **6**, 7948 (2015).
- [118] R. H. Hadfield, “Single-photon detectors for optical quantum information applications,” *Nature Photonics* **3**, 696–705 (2009).
- [119] C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, “Superconducting nanowire single-photon detectors: physics and applications,” *Superconductor Science and Technology* **25**, 063001 (2012).
- [120] G. N. Gol’tsman, O. Okunev, G. Chulkova, A. Lipatov, A. Semenov, K. Smirnov, B. Voronov, A. Dzardanov, C. Williams, and R. Sobolewski, “Picosecond superconducting single-photon optical detector,” *Applied Physics Letters* **79**, 705–707 (2001).
- [121] E. A. Dauler, M. E. Grein, A. J. Kerman, F. Marsili, S. Miki, S. W. Nam, M. D. Shaw, H. Terai, V. B. Verma, and T. Yamashita, “Review of superconducting nanowire single-photon detector system design options and demonstrated performance,” *Optical Engineering* **53**, 081907 (2014).
- [122] C. M. Wilkes, X. Qiang, J. Wang, R. Santagati, S. Paesani, X. Zhou, D. A. Miller, G. D. Marshall, M. G. Thompson, and J. L. O'Brien, “60 db high-extinction auto-configured mach-zehnder interferometer,” *Optics letters* **41**, 5318–5321 (2016).
- [123] A. Peruzzo, M. Lobino, J. C. F. Matthews, N. Matsuda, A. Politi, K. Poulios, X.-Q. Zhou, Y. Lahini, N. Ismail, K. Wörhoff, et al., “Quantum walks of correlated particles,” *Science* **329**, 1500–1503 (2010).
- [124] E. Knill, R. Laflamme, and G. J. Milburn, “A scheme for efficient quantum computation with linear optics,” *Nature* **409**, 46–52 (2001).
- [125] T. C. Ralph, N. K. Langford, T. B. Bell, and A. G. White, “Linear optical controlled-NOT gate in the coincidence basis,” *Physical Review A* **65**, 062324 (2002).
- [126] X. Zhou, T. C. Ralph, P. Kalasuwani, M. Zhang, A. Peruzzo, B. P. Lanyon, and J. L. O'Brien, “Adding control to arbitrary unknown quantum operations,” *Nature Communications* **2**, 413–8 (2011).
- [127] X. Zhou, P. Kalasuwani, T. C. Ralph, and J. L. O'Brien, “Calculating unknown eigenvalues with a quantum algorithm,” *Nature Photonics* **7**, 223–228 (2013).

BIBLIOGRAPHY

- [128] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, et al., “Witnessing eigenstates for quantum simulation of hamiltonian spectra,” *Science Advances* **4**, eaap9646 (2018).
- [129] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson, “Silicon optical modulators,” *Nature Photonics* **4**, 518–526 (2010).
- [130] I. Hincks, C. Granade, and D. G. Cory, “Statistical inference with quantum measurements: methodologies for nitrogen vacancy centers in diamond,” *New Journal of Physics* **20**, 013022 (2018).
- [131] B. Smeltzer, L. Childress, and A. Gali, “¹³C hyperfine interactions in the nitrogen-vacancy centre in diamond,” *New Journal of Physics* **13**, 025021 (2011).
- [132] M. W. Doherty, N. B. Manson, P. Delaney, F. Jelezko, J. Wrachtrup, and L. C. L. Hollenberg, “The nitrogen-vacancy colour centre in diamond,” *Physics reports* **528**, 1–45 (2013).
- [133] S. Knauer, *Photonic Structure Coupling and Strain Sensing with Single Photon Emitters*, Ph.D. thesis, University of Bristol (2016).
- [134] L. Childress, M. V. Gurudev Dutt, J. M. Taylor, A. S. Zibrov, F. Jelezko, J. Wrachtrup, P. R. Hemmer, and M. D. Lukin, “Coherent dynamics of coupled electron and nuclear spin qubits in diamond,” *Science* **314**, 281–285 (2006).
- [135] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen, and R. Hanson, “Optimized quantum sensing with a single electron spin using real-time adaptive measurements,” *Nature Nanotechnology* **11**, 247–252 (2015).
- [136] H. T. Dinani, D. W. Berry, R. Gonzalez, J. R. Maze, and C. Bonato, “Bayesian estimation for quantum sensing in the absence of single-shot detection,” arXiv preprint 1806.01249 (2018).
- [137] F. M. Hrubesch, G. Braunbeck, M. Stutzmann, F. Reinhard, and M. S. Brandt, “Efficient electrical spin readout of nv- centers in diamond,” *Physical Review letters* **118**, 037601 (2017).
- [138] P. Siyushev, M. Nesladek, E. Bourgeois, M. Gulka, J. Hraby, T. Yamamoto, M. Trupke, T. Teraji, J. Isoya, and F. Jelezko, “Photoelectrical imaging and coherent spin-state readout of single nitrogen-vacancy centers in diamond,” *Science* **363**, 728–731 (2019).
- [139] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, “High-fidelity projective read-out of a solid-state spin quantum register,” *Nature* **477**, 574–578 (2011).

- [140] V. V. Dobrovitski, A. E. Feiguin, D. D. Awschalom, and R. Hanson, “Decoherence dynamics of a single spin versus spin ensemble,” *Physical Review B - Condensed Matter and Materials Physics* **77**, 1–6 (2008).
- [141] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson, “Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip,” *Physical Review Letters* **118**, 100503 (2017).
- [142] D. S. Abrams and S. Lloyd, “Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors,” *Physical Review Letters* **83**, 5162 (1999).
- [143] P. Shor, “Algorithms for quantum computation: Discrete logarithms and factoring,” *Proc. 35nd Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press (1994).
- [144] S. Hallgren, “Polynomial-time quantum algorithms for pell’s equation and the principal ideal problem,” *Journal of the ACM (JACM)* **54**, 4 (2007).
- [145] P. Wocjan, C.-F. Chiang, D. Nagaj, and A. Abeyesinghe, “Quantum algorithm for approximating partition functions,” *Physical Review A* **80**, 022340 (2009).
- [146] M. Szegedy, “Quantum speed-up of markov chain based algorithms,” in *45th Annual IEEE symposium on foundations of computer science* (IEEE, 2004) pp. 32–41.
- [147] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, “Quantum metropolis sampling,” *Nature* **471**, 87 (2011).
- [148] A. Aspuru-Guzik, A. Dutoi, M. Head-Gordon, P. J. Love, and P. Love, “Simulated quantum computation of molecular energies,” *Science* **309**, 1704–1707 (2005).
- [149] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, “Implementing a Variational Quantum Eigensolver using Superconducting Qubits,” *Quantum Information and Measurement* , QF6A.2 (2017).
- [150] T. E. O’Brien, B. Tarasinski, and B. M. Terhal, “Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments,” *New Journal of Physics* **21**, 023022 (2019).
- [151] K. M. Svore, M. B. Hastings, and M. Freedman, “Faster phase estimation,” *Quantum Information & Computation* **14**, 306 (2014).
- [152] A. Kitaev, “Quantum measurements and the abelian stabilizer problem,” *Electronic Colloquium on Computational Complexity* **3** (1996).

BIBLIOGRAPHY

- [153] R. B. Griffiths and C.-S. Niu, “Semiclassical fourier transform for quantum computation,” *Physical Review Letters* **76**, 3228–3231 (1996).
- [154] M. Dobsicek, G. Johansson, V. Shumeiko, and G. Wendin, “Arbitrary accuracy iterative quantum phase estimation algorithm using a single ancillary qubit: A two-qubit benchmark,” *Physical Review A* **76**, 030306 (2007).
- [155] C. Granade, C. Ferrie, I. Hincks, S. Casagrande, T. Alexander, J. Gross, M. Kononenko, and Y. Sanders, “Qinfer: Statistical inference software for quantum applications,” *Quantum* **1**, 5 (2017).
- [156] C. Ferrie, C. E. Granade, and D. G. Cory, “How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies,” *Quantum Information Processing* **12**, 611–623 (2012).
- [157] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, “Hamiltonian learning and certification using quantum resources,” *Physical Review Letters* **112**, 190501 (2014).
- [158] E. R. Davidson and W. J. Thompson, “Monster matrices: their eigenvalues and eigenvectors,” *Computers in Physics* **7**, 519–522 (1993).
- [159] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, “The theory of variational hybrid quantum-classical algorithms,” *New Journal of Physics* **18**, 023023 (2016).
- [160] 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 photonic quantum processor,” *Nature Communications* **5** (2014).
- [161] D. Wecker, M. B. Hastings, and M. Troyer, “Progress towards practical quantum variational algorithms,” *Physical Review A* **92**, 042303 (2015).
- [162] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barrends, J. Kelly, P. Roushan, A. Tranter, N. Ding, et al., “Scalable quantum simulation of molecular energies,” *Physical Review X* **6**, 031007 (2016).
- [163] K. A. Peterson, D. Feller, and D. A. Dixon, “Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges,” *Theoretical Chemistry Accounts* **131**, 1079 (2012).
- [164] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin, and X. Yuan, “Quantum computational chemistry,” arXiv preprint 1808.10402 (2018).
- [165] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, et al., “Universal digital quantum simulation with trapped ions,” *Science* **334**, 57–61 (2011).

- [166] B. P. Lanyon, J. Whitfield, J. D. Whitfield, G. Gillett, and M. Goggin, “Towards quantum chemistry on a quantum computer,” *Nature Chemistry* **2**, 106–111 (2010).
- [167] R. Babbush, D. W. Berry, I. D. Kivlichan, A. Y. Wei, P. J. Love, and A. Aspuru-Guzik, “Exponentially more precise quantum simulation of fermions in second quantization,” *New Journal of Physics* **18**, 033032 (2016).
- [168] M. Veldhorst et al., “A two-qubit logic gate in silicon,” *Nature* **526**, 410 (2015).
- [169] D. Clery and D. Voss, “All for One and One for All,” *Science* **308**, 809–809 (2005).
- [170] R. Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences* (Sage, 2014).
- [171] C. Adam-Bourdarios, G. Cowan, C. Germain, I. Guyon, B. Kégl, and D. Rousseau, “The higgs boson machine learning challenge,” in *NIPS 2014 Workshop on High-energy Physics and Machine Learning* (2015) pp. 19–55.
- [172] D. George and E. Huerta, “Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data,” *Physics Letters B* **778**, 64–70 (2018).
- [173] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, “Functional genomic hypothesis generation and experimentation by a robot scientist,” *Nature* **427**, 247 (2004).
- [174] S. Murray, V. Bahyl, G. Cancio, E. Cano, V. Kotlyar, D. Kruse, and J. Leduc, “An efficient, modular and simple tape archiving solution for lhc run-3,” in *Journal of Physics: Conference Series*.
- [175] Z. Ghahramani, “Probabilistic machine learning and artificial intelligence,” *Nature* **521**, 452–459 (2015).
- [176] D. Waltz and B. G. Buchanan, “Automating science,” *Science* **324**, 43–44 (2009).
- [177] J. P. Crutchfield, “The dreams of theory,” *Wiley Interdisciplinary Reviews: Computational Statistics* **6**, 75–79 (2014).
- [178] A. Kirsch, *An introduction to the mathematical theory of inverse problems* (Springer Science & Business Media, 2011).
- [179] J. P. Crutchfield and B. S. McNamara, “Equation of motion from a data series,” *Complex systems* **1**, 121 (1987).
- [180] M. Schmidt and H. Lipson, “Distilling Free-Form Natural Laws,” *Science* **324**, 81–85 (2009).

BIBLIOGRAPHY

- [181] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, “Discovering physical concepts with neural networks,” arXiv preprint 1807.10300 (2018).
- [182] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, “Characterizing quantum supremacy in near-term devices,” *Nature Physics* **14**, 595 (2018).
- [183] M. R. Coudron, *Entangled protocols and non-local games for testing quantum systems.*, Ph.D. thesis, Massachusetts Institute of Technology (2017).
- [184] D. R. Wallace and R. U. Fujii, “Software verification and validation: an overview,” *IEEE Software* **6**, 10–17 (1989).
- [185] T. Häner and D. S. Steiger, “0.5 petabyte simulation of a 45-qubit quantum circuit,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2017) p. 33.
- [186] Z.-Y. Chen, Q. Zhou, C. Xue, X. Yang, G.-C. Guo, and G.-P. Guo, “64-qubit quantum circuit simulation,” *Science Bulletin* **63**, 964–971 (2018).
- [187] G. Carleo and M. Troyer, “Solving the quantum many-body problem with artificial neural networks,” *Science* **355**, 602–606 (2017).
- [188] V. Dunjko and H. J. Briegel, “Machine learning & artificial intelligence in the quantum domain: a review of recent progress,” *Reports on Progress in Physics* **81**, 074001 (2018).
- [189] J. Carolan, M. Mosheni, J. P. Olson, M. Prabhu, C. Chen, D. Bunandar, N. C. Harris, F. N. Wong, M. Hochberg, S. Lloyd, et al., “Variational quantum unsampling on a programmable nanophotonic processor,” in *CLEO: QELS_Fundamental Science* (Optical Society of America, 2019) pp. FTh3A–3.
- [190] N. Ares, “Efficiently measuring and tuning quantum devices using machine learning,” in *APS Meeting Abstracts* (2019).
- [191] L. Banchi, N. Pancotti, and S. Bose, “Quantum gate learning in qubit networks: Toffoli gate without time-dependent control,” *npj Quantum Information* **2**, 16019 (2016).
- [192] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, “Automated search for new quantum experiments,” *Physical Review letters* **116**, 090405 (2016).
- [193] M. Mohseni, A. Rezakhani, and D. Lidar, “Quantum-process tomography: Resource analysis of different strategies,” *Physical Review A* **77**, 032322 (2008).
- [194] K. Banaszek, M. Cramer, and D. Gross, “Focus on quantum tomography,” *New Journal of Physics* **15**, 125020 (2013).

- [195] M. Paris and J. Rehacek, *Quantum state estimation* (Springer Science & Business Media, 2004).
- [196] P. Alquier, C. Butucea, M. Hebiri, K. Meziani, and T. Morimae, “Rank-penalized estimation of a quantum system,” *Physical Review A - Atomic, Molecular, and Optical Physics* **88**, 1–9 (2013).
- [197] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, “Quantum state tomography via compressed sensing,” *Physical Review letters* **105**, 150401 (2010).
- [198] S. T. Flammia, D. Gross, Y. K. Liu, and J. Eisert, “Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators,” *New Journal of Physics* **14**, 095022 (2012).
- [199] W.-T. Liu, T. Zhang, J.-Y. Liu, P.-X. Chen, and J.-M. Yuan, “Experimental quantum state tomography via compressed sampling,” *Physical Review letters* **108**, 170403 (2012).
- [200] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y. K. Liu, “Efficient quantum state tomography,” *Nature Communications* **1**, 147–149 (2010).
- [201] B. Lanyon, C. Maier, M. Holzapfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. Buyskikh, A. Daley, M. Cramer, et al., “Efficient tomography of a quantum many-body system,” *Nature Physics* **13**, 1158 (2017).
- [202] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, “Permutationally invariant quantum tomography,” *Physical Review letters* **105**, 250403 (2010).
- [203] C. Schwemmer, G. Tóth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, “Experimental comparison of efficient tomography schemes for a six-qubit State,” *Physical Review Letters* **113**, 1–5 (2014).
- [204] Z. Hradil, “Quantum-state estimation,” *Physical Review A* **55**, R1561 (1997).
- [205] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, “Neural-network quantum state tomography,” *Nature Physics* **14**, 447 (2018).
- [206] J. Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, and X.-M. Jin, “Experimental Machine Learning of Quantum States,” *Physical Review Letters* **120**, 240501 (2017).
- [207] A. Seif, K. A. Landsman, N. M. Linke, C. Figgatt, C. Monroe, and M. Hafezi, “Machine learning assisted readout of trapped-ion qubits,” *arXiv preprint 1804.07718* (2018).

BIBLIOGRAPHY

- [208] M. Kieferová and N. Wiebe, “Tomography and generative training with quantum Boltzmann machines,” *Physical Review A* **96** (2017).
- [209] J. Fiurášek and Z. Hradil, “Maximum-likelihood estimation of quantum processes,” *Physical Review A* **63**, 020101 (2001).
- [210] M. Mohseni and D. Lidar, “Direct characterization of quantum dynamics,” *Physical Review Letters* **97**, 170501 (2006).
- [211] M. Ziman, M. Plesch, V. Bužek, and P. Štelmachovič, “Process reconstruction: From unphysical to physical maps via maximum likelihood,” *Physical Review A* **72**, 022106 (2005).
- [212] C. E. Granade, *Characterization, Verification and Control for Large Quantum Systems*, Ph.D. thesis, University of Waterloo (2014).
- [213] C. E. Granade, C. Ferrie, N. Wiebe, and D. G. Cory, “Robust online hamiltonian learning,” *New Journal of Physics* **14**, 103013 (2012).
- [214] N. Wiebe, C. Granade, C. Ferrie, and D. Cory, “Quantum hamiltonian learning using imperfect quantum resources,” *Physical Review A* **89**, 042314 (2014).
- [215] J. Wang, S. Paesani, R. Santagati, S. Knauer, A. A. Gentile, N. Wiebe, M. Petruzzella, J. L. O’Brien, J. G. Rarity, A. Laing, et al., “Experimental quantum hamiltonian learning,” *Nature Physics* **13**, 551 (2017).
- [216] C. Granade and N. Wiebe, “Structured filtering,” *New Journal of Physics* **19**, 1–22 (2017).
- [217] A. Beskos, D. Crisan, A. Jasra, et al., “On the stability of sequential monte carlo methods in high dimensions,” *The Annals of Applied Probability* **24**, 1396–1445 (2014).
- [218] N. Moshtagh, “Minimum Volume Enclosing Ellipsoids,” *Convex Optimization* **111**, 112–118 (2005).
- [219] S. Schmitt, T. Gefen, F. M. Stürner, T. Unden, G. Wolff, C. Müller, J. Scheuer, B. Naydenov, M. Markham, S. Pezzagna, et al., “Submillihertz magnetic spectroscopy performed with a nanoscale quantum sensor,” *Science* **356**, 832–837 (2017).
- [220] I. Jakobi, P. Neumann, Y. Wang, D. B. R. Dasari, F. El Hallak, M. A. Bashir, M. Markham, A. Edmonds, D. Twitchen, and J. Wrachtrup, “Measuring broadband magnetic fields on the nanoscale using a hybrid quantum register,” *Nature Nanotechnology* **12**, 67–72 (2017).

- [221] K. Arai, C. Belthangady, H. Zhang, N. Bar-Gill, S. J. DeVience, P. Cappellaro, A. Yacoby, and R. L. Walsworth, “Fourier magnetic imaging with nanoscale resolution and compressed sensing speed-up using electronic spins in diamond,” *Nature Nanotechnology* **10**, 859–864 (2015).
- [222] G. Waldherr, J. Beck, P. Neumann, R. S. Said, M. Nitsche, M. L. Markham, D. J. Twitchen, J. Twamley, F. Jelezko, and J. Wrachtrup, “High-dynamic-range magnetometry with a single nuclear spin in diamond,” *Nature Nanotechnology* **7**, 105–108 (2012).
- [223] G. Puentes, G. Waldherr, P. Neumann, G. Balasubramanian, and J. Wrachtrup, “Efficient route to high-bandwidth nanoscale magnetometry using single spins in diamond,” *Scientific Reports* **4**, 4677 (2014).
- [224] G. Balasubramanian, I. Y. Chan, R. Kolesov, M. Al-Hmoud, J. Tisler, C. Shin, C. Kim, A. Wojcik, P. R. Hemmer, A. Krueger, et al., “Nanoscale imaging magnetometry with diamond spins under ambient conditions,” *Nature* **455**, 648 (2008).
- [225] L. P. McGuinness, Y. Yan, A. Stacey, D. A. Simpson, L. T. Hall, D. Maclaurin, S. Prawer, P. Mulvaney, J. Wrachtrup, F. Caruso, et al., “Quantum measurement and orientation tracking of fluorescent nanodiamonds inside living cells,” *Nature Nanotechnology* **6**, 358–363 (2011).
- [226] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. Pryde, “Entanglement-free heisenberg-limited phase estimation,” *Nature* **450**, 393–396 (2007).
- [227] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. Pryde, “Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements,” *New Journal of Physics* **11**, 073023 (2009).
- [228] N. M. Nusran, M. U. Momeen, and M. V. G. Dutt, “High-dynamic-range magnetometry with a single electronic spin in diamond,” *Nature Nanotechnology* **7**, 109–113 (2012).
- [229] N. Aharon, A. Rotem, L. P. McGuinness, F. Jelezko, A. Retzker, and Z. Ringel, “NV center based nano-NMR enhanced by deep learning,” arXiv preprint 1809.02583 (2018).
- [230] I. Schwartz, J. Rosskopf, S. Schmitt, B. Tratzmiller, Q. Chen, L. P. McGuinness, F. Jelezko, and M. B. Plenio, “Blueprint for nanoscale nmr,” *Scientific Reports* **9**, 1–11 (2019).
- [231] T. Ruster, H. Kaufmann, M. A. Luda, V. Kaushal, C. T. Schmiegelow, F. Schmidt-Kaler, and U. G. Poschinger, “Entanglement-based dc magnetometry with separated ions,” *Physical Review X* **7**, 031050 (2017).
- [232] Python Packages adopted for all code implementing QHL, QMLA & MFL instances: *Matplotlib* 2.0.2, *Numpy* 1.11.1, *Pandas* 0.20.0, *Qinfer* 1.0, *Qutip* 4.1.0, *Redis* 2.10.6, *Scipy* 0.18.1.

BIBLIOGRAPHY

- [233] J. F. Barry, J. M. Schloss, E. Bauch, M. J. Turner, C. A. Hart, L. M. Pham, and R. L. Walsworth, “Sensitivity optimization for nv-diamond magnetometry,” *Reviews of Modern Physics* **92**, 015004 (2020).
- [234] C. Bonato and D. W. Berry, “Adaptive tracking of a time-varying field with a quantum sensor,” *Physical Review A* **95**, 052348 (2017).
- [235] J. F. Ralph, S. Maskell, and K. Jacobs, “Multiparameter estimation along quantum trajectories with sequential monte carlo methods,” *Physical Review A* **96**, 052306 (2017).
- [236] E. Bairey, I. Arad, and N. H. Lindner, “Learning a local hamiltonian from local measurements,” *Physical Review letters* **122**, 020504 (2019).
- [237] M. A. Nielsen and I. L. Chuang, “Programmable quantum gate arrays,” *Physical Review Letters* **79**, 321 (1997).
- [238] D. Hayes, S. T. Flammia, and M. J. Biercuk, “Programmable quantum simulation by dynamic hamiltonian engineering,” *New Journal of Physics* **16**, 083027 (2014).
- [239] C. M. Bishop, *Pattern recognition and machine learning* (Springer, 2006).
- [240] W. Buntine, “Learning classification trees,” *Statistics and computing* **2**, 63–73 (1992).
- [241] R. E. Kass and A. E. Raftery, “Bayes factors,” *Journal of the american statistical association* **90**, 773–795 (1995).
- [242] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, “Testing the dimension of hilbert spaces,” *Physical Review letters* **100**, 210503 (2008).
- [243] M. Hutson, “Artificial intelligence faces reproducibility crisis,” (2018).
- [244] H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics: an introduction* (Oxford university press, 2004).
- [245] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, “Efficient quantum algorithms for simulating sparse hamiltonians,” *Communications in Mathematical Physics* **270**, 359–371 (2007).
- [246] A. M. Childs and R. Kothari, “Simulating sparse hamiltonians with star decompositions,” in *Conference on Quantum Computation, Communication, and Cryptography* (Springer, 2010) pp. 94–103.
- [247] K. P. Murphy, *Machine Learning: a Probabilistic Perspective* (MIT Press, 2012).
- [248] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for machine learning* (Mit Press, 2012).

- [249] B. Bollobas, *Graph theory: an introductory course*, Vol. 63 (Springer Science & Business Media, 2012).
- [250] M. S. Blok, C. Bonato, M. L. Markham, D. J. Twitchen, V. V. Dobrovitski, and R. Hanson, “Manipulating a qubit through the backaction of sequential partial measurements and real-time feedback,” *Nature Physics* **10**, 189–193 (2014).
- [251] P. Y. Hou, L. He, F. Wang, X. Z. Huang, W. G. Zhang, X. L. Ouyang, X. Wang, W. Q. Lian, X. Y. Chang, and L. M. Duan, “Experimental Hamiltonian Learning of An 11-qubit Solid-State Quantum Spin Register,” arXiv preprint 1905.12248 (2019).
- [252] F. T. Charnock and T. Kennedy, “Combined optical and microwave approach for performing quantum spin operations on the nitrogen-vacancy center in diamond,” *Physical Review B* **64**, 041201 (2001).
- [253] A. Gali, M. Fyta, and E. Kaxiras, “Ab initio supercell calculations on nitrogen-vacancy center in diamond: Electronic structure and hyperfine tensors,” *Physical Review B* **77**, 1–12 (2008).
- [254] M. G. Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. Zibrov, P. Hemmer, and M. Lukin, “Quantum register based on individual electronic and nuclear spin qubits in diamond,” *Science* **316**, 1312–1316 (2007).
- [255] “Acrc: Bluecrystal user guide,” <https://www.acrc.bris.ac.uk/acrc/pdf/bc-user-guide.pdf>, accessed: 2019-08-19.
- [256] W. Pfaff, T. H. Taminiau, L. Robledo, H. Bernien, M. Markham, D. J. Twitchen, and R. Hanson, “Demonstration of entanglement-by-measurement of solid-state qubits,” *Nature Physics* **9**, 29 (2013).
- [257] L. Rowan, E. Hahn, and W. Mims, “Electron-spin-echo envelope modulation,” *Physical Review* **137**, A61 (1965).
- [258] G. A. Álvarez, C. O. Bretschneider, R. Fischer, P. London, H. Kanda, S. Onoda, J. Isoya, D. Gershoni, and L. Frydman, “Local and bulk ^{13}C hyperpolarization in nitrogen-vacancy-centred diamonds at variable fields and orientations,” *Nature communications* **6**, 8456 (2015).
- [259] B. Flynn, A. A. Gentile, and al., “Automated characterisation of quantum processes via machine learning,” to be published .
- [260] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, “An Optimal Design for Universal Multiport Interferometers,” *Optica* **3** (2016).

BIBLIOGRAPHY

- [261] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf, “Quantum fingerprinting,” *Physical Review Letters* **87**, 167902 (2001).
- [262] S. Felton, A. Edmonds, M. Newton, P. Martineau, D. Fisher, D. Twitchen, and J. Baker, “Hyperfine interaction in the ground state of the negatively charged nitrogen vacancy center in diamond,” *Physical Review B* **79**, 075203 (2009).
- [263] M. Szczykulska, T. Baumgratz, and A. Datta, “Multi-parameter quantum metrology,” *Advances in Physics: X* **1**, 621–639 (2016).
- [264] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, “Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning,” arXiv preprint 1806.00069 (2018).
- [265] R. Santagati, A. A. Gentile, S. Knauer, S. Schmitt, S. Paesani, C. Granade, N. Wiebe, C. Osterkamp, L. P. McGuinness, J. Wang, et al., “Magnetic-field learning using a single electronic spin in diamond with one-photon readout at room temperature,” *Physical Review X* **9**, 021019 (2019).
- [266] A. A. Gentile, B. Flynn, S. Knauer, N. Wiebe, S. Paesani, C. E. Granade, J. G. Rarity, R. Santagati, and A. Laing, “Learning models of quantum systems from experiments,” arXiv preprint 2002.06169 (2020).
- [267] R. P. Feynman, “Simulating physics with computers,” *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- [268] S. Lloyd, “Universal quantum simulators,” *Science* **273**, 1073–1078 (1996).
- [269] I. M. Georgescu, S. Ashhab, and F. Nori, “Quantum simulation,” *Reviews of Modern Physics* **86**, 153 (2014).
- [270] I. Bloch, J. Dalibard, and S. Nascimbène, “Quantum simulations with ultracold quantum gases,” *Nature Physics* **8**, 267 (2012).
- [271] R. Blatt and C. F. Roos, “Quantum simulations with trapped ions,” *Nature Physics* **8**, 277 (2012).
- [272] K. R. Brown, R. J. Clark, and I. L. Chuang, “Limitations of quantum simulation examined by simulating a pairing Hamiltonian using nuclear magnetic resonance,” *Physical Review Letters* **97**, 050504 (2006).
- [273] J. Du, N. Xu, X. Peng, P. Wang, S. Wu, and D. Lu, “NMR implementation of a molecular hydrogen quantum simulation with adiabatic state preparation,” *Physical Review Letters* **104**, 030502 (2010).

- [274] I. Pitsios, L. Banchi, A. S. Rab, M. Bentivegna, D. Caprara, A. Crespi, N. Spagnolo, S. Bose, P. Mataloni, R. Osellame, et al., “Photonic simulation of entanglement growth after a spin chain quench,” *Nature Communications* **8**, 1569 (2017).
- [275] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature* **549**, 242 (2017).
- [276] R. Babbush, P. J. Love, and A. Aspuru-Guzik, “Adiabatic quantum simulation of quantum chemistry,” *Scientific Reports* **4**, 6603 (2014).
- [277] E. Baake, M. Baake, and H. Wagner, “Ising quantum chain is equivalent to a model of biological evolution,” *Physical Review Letters* **78**, 559 (1997).
- [278] J. R. Lopez-Blanco, R. Reyes, J. I. Aliaga, R. M. Badia, P. Chacon, and E. S. Quintana-Orti, “Exploring large macromolecular functional motions on clusters of multicore processors,” *Journal of Computational Physics* **246**, 275–288 (2013).
- [279] A. N. Langville and C. D. Meyer, “A survey of eigenvector methods for web information retrieval,” *SIAM review* **47**, 135–161 (2005).
- [280] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, “Quantum computation of electronic transitions using a variational quantum eigensolver,” *Physical Review Letters* **122**, 230401 (2019).
- [281] M. H. M. Olsson, J. Mavri, and A. Warshel, “Transition state theory can be used in studies of enzyme catalysis: lessons from simulations of tunnelling and dynamical effects in lipoxygenase and other systems,” *Philos. Trans. R. Soc. Lond. B: Biological Sciences* **361**, 1417 (2006).
- [282] D. Lidar and H. Wang, “Calculating the thermal rate constant with exponential speedup on a quantum computer,” *Physical Review E* **59**, 2429 (1999).
- [283] J. Kempe, A. Kitaev, and O. Regev, “The complexity of the local hamiltonian problem,” in *International Conference on Foundations of Software Technology and Theoretical Computer Science* (Springer, 2004) pp. 372–383.
- [284] N. Schuch and F. Verstraete, “Computational complexity of interacting electrons and fundamental limitations of density functional theory,” *Nature Physics* **5**, 732 (2009).
- [285] A. Tranter, S. Sofia, J. Seeley, M. Kaicher, J. McClean, R. Babbush, P. V. Coveney, F. Mintert, F. Wilhelm, and P. J. Love, “The Bravyi–Kitaev transformation: Properties and applications,” *International Journal of Quantum Chemistry* **115**, 1431–1441 (2015).

BIBLIOGRAPHY

- [286] L. Serrano-Andrés and M. Merchán, “Quantum chemistry of the excited state: 2005 overview,” *Journal of Molecular Structure: THEOCHEM* **729**, 99 (2005).
- [287] T. D. Crawford and J. F. Stanton, “Some surprising failures of brueckner coupled cluster theory,” *The Journal of Chemical Physics* **112**, 7873 (2000).
- [288] N. J. Ward, I. Kassal, and A. Aspuru-Guzik, “Preparation of many-body states for quantum simulation,” *The Journal of Chemical Physics* **130**, 194105 (2009).
- [289] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Corporation, 2012).
- [290] H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics: an introduction* (Oxford university press, 2004).
- [291] W. J. Hehre, R. F. Stewart, and J. A. Pople, “Self consistent molecular orbital methods i. use of gaussian expansions of slater type atomic orbitals,” *The Journal of Chemical Physics* **51**, 2657–2664 (1969).
- [292] J. R. McClean, *Algorithms Bridging Quantum Computation and Chemistry*, Ph.D. thesis, Harvard University (2015).
- [293] M. Steudtner and S. Wehner, “Fermion-to-qubit mappings with varying resource requirements for quantum simulation,” *New Journal of Physics* **20**, 063010 (2018).
- [294] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, “A comparison of the bravyi–kitaev and jordan–wigner transformations for the quantum simulation of quantum chemistry,” *Journal of chemical theory and computation* **14**, 5617–5630 (2018).
- [295] E. P. Wigner and P. Jordan, “Über das paulische äquivalenzverbot,” *Z. Phys* **47**, 631 (1928).
- [296] S. B. Bravyi and A. Y. Kitaev, “Fermionic quantum computation,” *Annals of Physics* **298**, 210–226 (2002).
- [297] J. T. Seeley, M. J. Richard, and P. J. Love, “The Bravyi–Kitaev transformation for quantum computation of electronic structure,” *The Journal of Chemical Physics* **137**, 224109 (2012).
- [298] H. Büchler, A. Micheli, and P. Zoller, “Three-body interactions with cold polar molecules,” *Nature Physics* **3**, 726 (2007).
- [299] H. Primakoff and T. Holstein, “Many-body interactions in atomic and nuclear systems,” *Physical Review* **55**, 1218 (1939).

- [300] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, “Quantum simulation of time-dependent Hamiltonians and the convenient illusion of Hilbert space,” *Physical Review Letters* **106**, 1–4 (2011).
- [301] L. Veis and J. Pittner, “Adiabatic state preparation study of methylene,” *The Journal of Chemical Physics* **140**, 214111 (2014).
- [302] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, “Quantum computation by adiabatic evolution,” arXiv preprint quant-ph/0001106 (2000).
- [303] J. Romero, R. Babbush, J. R. McClean, and C. Hempel, “Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz,” arXiv preprint 1701.02691 (2017).
- [304] G. G. Guerreschi and M. Smelyanskiy, “Practical optimization for hybrid quantum-classical algorithms,” arXiv preprint 1701.01450 (2017).
- [305] Y. Li and S. C. Benjamin, “Efficient Variational Quantum Simulator Incorporating Active Error Minimization,” *Physical Review X* **7**, 021050 (2017).
- [306] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, “Quantum-classical computation of schwinger model dynamics using quantum computers,” *Physical Review A* **98**, 032331 (2018).
- [307] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. Joshi, P. Jurcevic, C. Muschik, P. Silvi, R. Blatt, C. Roos, et al., “Self-verifying variational quantum simulation of lattice models,” *Nature* **569**, 355 (2019).
- [308] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, “Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states,” *Physical Review A* **95**, 042308 (2017).
- [309] X.-D. Cai, D. Wu, Z.-E. Su, M.-C. Chen, X.-L. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, “Entanglement-based machine learning on a quantum computer,” *Physical Review letters* **114**, 110504 (2015).
- [310] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature* **567**, 209 (2019).
- [311] A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, “Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators?” *Chemical science* **10**, 3746–3755 (2019).

BIBLIOGRAPHY

- [312] W. Kohn, “Nobel Lecture: Electronic structure of matte-wave functions and density functionals,” *Reviews of Modern Physics* **71**, 1253 (1999).
- [313] C. E. Mora and H. J. Briegel, “Algorithmic complexity and entanglement of quantum states,” *Physical Review letters* **95**, 200503 (2005).
- [314] A. G. Taube and R. J. Bartlett, “New perspectives on unitary coupled-cluster theory,” *International journal of quantum chemistry* **106**, 3393–3401 (2006).
- [315] B. Cooper and P. J. Knowles, “Benchmark studies of variational, unitary and extended coupled cluster methods,” *The Journal of Chemical Physics* **133**, 234102 (2010).
- [316] J. B. Robinson and P. J. Knowles, “Approximate variational coupled cluster theory,” *The Journal of Chemical Physics* **135**, 044113 (2011).
- [317] C. Sparrow, E. Martin-Lopez, N. Maraviglia, A. Neville, C. Harrold, J. Carolan, Y. N. Joglekar, T. Hashimoto, N. Matsuda, J. L. O’Brien, et al., “Simulating the vibrational quantum dynamics of molecules using photonics,” *Nature* **557**, 660 (2018).
- [318] R. Babbush, J. McClean, D. Wecker, A. Aspuru-Guzik, and N. Wiebe, “Chemical basis of trotter-suzuki errors in quantum chemistry simulation,” *Physical Review A* **91**, 022311 (2015).
- [319] L.-W. Wang and A. Zunger, “Solving schrödinger’s equation around a desired energy: Application to silicon quantum dots,” *The Journal of Chemical Physics* **100**, 2394–2397 (1994).
- [320] A. R. Tackett and M. D. Ventra, “Targeting specific eigenvectors and eigenvalues of a given hamiltonian using arbitrary selection criteria,” *Physical Review B* **66**, 245104 (2002).
- [321] Z. Li, M.-H. Yung, H. Chen, D. Lu, J. D. Whitfield, X. Peng, A. Aspuru-Guzik, and J. Du, “Solving quantum ground-state problems with nuclear magnetic resonance,” *Scientific Reports* **1**, 88 (2011).
- [322] R. B. Patel, J. Ho, F. Ferreyrol, T. C. Ralph, and G. J. Pryde, “A quantum Fredkin gate,” *Science Advances* **2**, 1501531 (2016).
- [323] X. Li and X. Yao, “Cooperatively coevolving particle swarms for large scale optimization,” *IEEE Transactions on Evolutionary Computation* **16**, 210–224 (2011).
- [324] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, “Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources,” arXiv preprint 1902.10679 (2019).

- [325] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal, “Quantum simulation of many-body hamiltonians using perturbation theory with bounded-strength interactions,” *Physical Review letters* **101**, 070503 (2008).
- [326] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, 2014).
- [327] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, “Low-depth quantum simulation of materials,” *Physical Review X* **8**, 011044 (2018).
- [328] D. Poulin, M. B. Hastings, D. Wecker, N. Wiebe, A. C. Doberty, and M. Troyer, “The trotter step size required for accurate quantum simulation of quantum chemistry,” *Quantum Information & Computation* **15**, 361–384 (2015).
- [329] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, “Elucidating reaction mechanisms on quantum computers,” *PNAS* **114**, 7555 (2017).
- [330] F. Fassioli, R. Dinshaw, P. C. Arpin, and G. D. Scholes, “Photosynthetic light harvesting: excitons and coherence.” *Journal of The Royal Society - Interface* **11**, 20130901 (2014).
- [331] K. E. Parsopoulos and M. N. Vrahatis, *Particle swarm optimization and intelligence: advances and applications* (Information Science Reference New York, 2010).
- [332] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, “Generalized unitary coupled cluster wave functions for quantum computation,” *Journal of chemical theory and computation* **15**, 311–324 (2018).
- [333] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, “Variational quantum algorithms for discovering hamiltonian spectra,” *Physical Review A* **99**, 062304 (2019).
- [334] O. Higgott, D. Wang, and S. Brierley, “Variational quantum computation of excited states,” *Quantum* **3**, 156 (2019).
- [335] S. Matsuura, T. Yamazaki, V. Senicourt, and A. Zaribafyan, “Vanqver: The variational and adiabatically navigated quantum eigensolver,” arXiv preprint 1810.11511 (2018).
- [336] J. F. Barry, M. J. Turner, J. M. Schloss, D. R. Glenn, Y. Song, M. D. Lukin, H. Park, and R. L. Walsworth, “Optical magnetic detection of single-neuron action potentials using quantum defects in diamond,” *PNAS* **113**, 14133–14138 (2016).
- [337] A. Boretti and S. Castelletto, “Nanometric resolution magnetic resonance imaging methods for mapping functional activity in neuronal networks,” *MethodsX* **3**, 297 – 306 (2016).

BIBLIOGRAPHY

- [338] E. R. MacQuarrie, T. A. Gosavi, S. A. Bhave, and G. D. Fuchs, “Continuous dynamical decoupling of a single diamond nitrogen-vacancy center spin with a mechanical resonator,” *Physical Review B* **92**, 224419 (2015).
- [339] D. Farfurnik, A. Jarmola, D. Budker, and N. Bar-Gill, “Spin ensemble-based ac magnetometry using concatenated dynamical decoupling at low temperatures,” *Journal of Optics* **20**, 024008 (2018).
- [340] T. Staudacher, F. Shi, S. Pezzagna, J. Meijer, J. Du, C. A. Meriles, F. Reinhard, and J. Wrachtrup, “Nuclear Magnetic Resonance Spectroscopy on a (5-Nanometer)³ Sample Volume,” *Science* **339**, 561–563 (2013).
- [341] O. Shehab, I. H. Kim, N. H. Nguyen, K. Landsman, C. H. Alderete, D. Zhu, C. Monroe, and N. M. Linke, “Noise reduction using past causal cones in variational quantum algorithms,” arXiv preprint 1906.00476 (2019).
- [342] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. Rol, C. Bultink, X. Fu, C. Price, V. Ostroukh, N. Muthusubramanian, A. Bruno, et al., “Experimental error mitigation via symmetry verification in a variational quantum eigensolver,” *Physical Review A* **100**, 010302 (2019).
- [343] M. Schlosshauer, “Decoherence, the measurement problem, and interpretations of quantum mechanics,” *Reviews of Modern Physics* **76**, 1267 (2005).
- [344] U. Weiss, *Quantum dissipative systems* (World scientific, 2012).
- [345] F. Eltes, J. Barreto, D. Caimi, S. Karg, A. A. Gentile, A. Hart, P. Stark, N. Meier, M. G. Thompson, J. Fompeyrine, et al., “First cryogenic electro-optic switch on silicon with high bandwidth and low power tunability,” in *2018 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2018) pp. 23.1.1–23.1.4.
- [346] W. Liu, M. Li, R. S. Guzzon, E. J. Norberg, J. S. Parker, M. Lu, L. A. Coldren, and J. Yao, “A fully reconfigurable photonic integrated signal processor,” *Nature Photonics* **10**, 190–195 (2016).
- [347] N. C. Harris, J. Carolan, D. Bunandar, M. Prabhu, M. Hochberg, T. Baehr-Jones, M. L. Fanto, A. M. Smith, C. C. Tison, P. M. Alsing, et al., “Linear programmable nanophotonic processors,” *Optica* **5**, 1623 (2018).
- [348] J. Komma, C. Schwarz, G. Hofmann, D. Heinert, and R. Nawrodt, “Thermo-optic coefficient of silicon at 1550nm and cryogenic temperatures,” *Applied Physics Letters* **101**, 041905 (2012).

- [349] J. Wang, S. Paesani, Y. Ding, R. Santagati, P. Skrzypczyk, A. Salavrakos, J. Tura, R. Augusiak, L. Mančinska, D. Bacco, et al., “Multidimensional quantum entanglement with large-scale integrated optics.” *Science* **360**, 285–291 (2018).
- [350] R. A. Norwood, C. Derose, C. Greenlee, and A. Yeniay, “Organic waveguides, ultra-low loss demultiplexers and electro-optic (eo) polymer devices,” in *Handbook of Organic Materials for Optical and (Opto) Electronic Devices: Properties and Applications* (Elsevier Inc., 2013) pp. 709–785.
- [351] P. Bernasconi, M. Zgonik, and P. Günter, “Temperature dependence and dispersion of electro-optic and elasto-optic effect in perovskite crystals,” *Journal of Applied Physics* **78**, 2651–2658 (1995).
- [352] M. Gehl, C. Long, D. Trotter, A. Starbuck, A. Pomerene, J. B. Wright, S. Melgaard, J. Siirola, A. L. Lentine, and C. DeRose, “Operation of high-speed silicon photonic micro-disk modulators at cryogenic temperatures,” *Optica* **4**, 374 (2017).
- [353] F. Eltes, G. E. Villarreal-Garcia, D. Caimi, H. Siegwart, A. A. Gentile, A. Hart, P. Stark, G. D. Marshall, M. G. Thompson, J. Barreto, et al., “An integrated cryogenic optical modulator,” arXiv preprint 1904.10902 (2019).
- [354] R. Bruck, K. Vynck, P. Lalanne, B. Mills, D. J. Thomson, G. Z. Mashanovich, G. T. Reed, and O. L. Muskens, “All-optical spatial light modulator for reconfigurable silicon photonic circuits,” *Optica* **3**, 396 (2016).
- [355] E. Timurdogan, C. V. Poulton, M. J. Byrd, and M. R. Watts, “Electric field-induced second-order nonlinear optical effects in silicon waveguides,” *Nature Photonics* **11**, 200–206 (2017).
- [356] S. Abel, F. Eltes, J. E. Ortmann, A. Messner, P. Castera, T. Wagner, D. Urbonas, A. Rosa, A. M. Gutierrez, D. Tulli, et al., “Large Pockels effect in micro- and nanostructured barium titanate integrated on silicon,” *Nature Materials* **18**, 42–47 (2018).