Variational Quantum Algorithms: a Noisy Landscape

Enrico Fontana

A thesis presented for the degree of

Doctor of Philosophy

Academic supervisor: Clemens Kupke

Industrial supervisors: Ivan Rungger (NPL) and Ross Duncan (Quantinuum)

Department of Computer and Information Sciences

University of Strathclyde, Glasgow

September 17, 2025

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Abstract

Quantum computing is an alternative paradigm of computation to the classical Boolean logic based systems, leveraging the laws of quantum mechanics. There is ample evidence that a large-scale, error-corrected quantum computer would be able to outperform classical computers for specific problems. However, current technology is not yet capable of realising such systems, meaning that the computers of present day are too small and too prone to errors (also known as *quantum noise*). Furthermore, the problems for which we have convincing proof of *quantum advantage* are mostly of academic interest. Thus over the past two decades the scientific community has been actively searching for novel algorithms that can be run on current devices, and that may be of commercial interest. A class of quantum algorithms that has received much attention in recent times is *variational quantum algorithms*, which are the focus on this Thesis. These algorithms include a feedback loop between a parameterised quantum circuit and a classical optimiser, the latter receiving input in the form of a cost value.

The overarching topic of investigation in this Thesis is whether variational quantum algorithms are inherently resilient to noise, a question of paramount importance since these algorithms are intended to be applicable to quantum devices that lack error correction. The problem is approached by examining, analytically and numerically, how noise affects optimisation, with a particular focus on how the cost function (or energy) landscape is modified by various types and rates of errors.

On the way to answering this guiding question, new perspectives and theoretical tools are developed. These include a linearised model for the impact of noise on optimisation, symmetries in the cost function landscape that are sensitive to certain types of noise, a generic theory of the exponential accumulation of errors, and a variety of

methods based on Fourier analysis to distinguish the disparate effects of hardware noise on the cost function. These tools in turn reveal new phenomena including noise thresholds for optimisation and simulatability, noise-induced barren plateaus, and different impact on optimisation of different error models.

The results overall paint a mixed picture. While there appears to be some inherent noise resilience in the optimisation process, and in some cases noise may give rise to larger gradients, under mild assumptions on the noise and the quantum circuit it can be shown that noise frequently acts as an impediment to optimisation. From the energy landscape point of view, noise generically appears to have a smoothing effect leading to vanishing gradients for deep circuits. This phenomenon also manifests as a dampening of high-frequency Fourier modes, making the noisy landscape classical simulatable on average. The evidence suggests that further advances are needed before such algorithms can deliver advantage over classical methods.

Acknowledgements

A PhD is notoriously not an easy task, though I found that it has not been hard in the way I originally thought. Looking back it is not the pressures of academia that stand out, but everything that went on around it: being independent, relations with supervisors and with colleagues, the wild mess that are conferences, time management, and dealing with the clash against reality when plans do not work out. If I succeeded in getting through it, it is in large part because of the many great people that stood by my side along the way.

First I would like to thank my PhD supervisors. Ivan, you have been an excellent guide and mentor, especially at the beginning when you pushed me to pursue my instincts and chart my own way, one should not underestimate the impact of early encouragement. You are one of the nicest people I know, and I wish you the best with your research team, that I am sure you are going to lead with your trademark kindness and enthusiasm. Thank you Ross for teaching me that scientific writing should be approached with precision and method, that form has a substance all of its own. The more I go on, the more I appreciate your distrust of Overleaf and love of Github. Thank you Clemens for putting up with my sometimes erratic communication schedule and for handling the bureoucratic matters, you have been very patient and I appreciate that.

Next for my colleagues and collaborators. Alessandra, it feels like a lifetime ago when you and I were the only quantum computing PhDs at NPL, and I have very fond memories of those times spent figuring everything out. Carla, Abhishek, Nikos, Weixi, Deep, François, you all are talented scientists that also know how to have a fun time after work, I wish you all the best in your future research. Thank you Cristina and Nathan for having shown me for the first time what real mathematical talent looks

like. Marco, you have made a massive impact to my career, first and foremost by being the best example for a scientist that truly loves their research. Thank you Manuel for giving me a masterclass in German efficiency and productivity, as well as showing me the crazy side of Germans too.

I must now thank all the people outside work that stood by my side during these hectic years. First, those friends that made the miscalculation of being my housemates: thank you Nunzio and Laura, I cannot think of a nicest couple out there, and being with you made lockdown so much easier than it could've been. Thank you Peter and Dario (and Säde), we really had a good couple years. Yao and Matti, having you as neighbours was almost too good to be true, I live with the hope that it may one day happen again. Andrea, Nicolò and Zeyu, after Cambridge had forged us into brothers it was wonderful to have you around in London. I must also mention some people that were my companions during that strange time that was 2021, in that magical place that is the Canary archipelago. Thank you Emanuele for having been my housemate and having shared some of the most memorable moments. Dave, thank you for showing me that you can be a nerd and an amazing surfer at the same time. Thank you Robby, Renny, Sara, Giulia, Federica, Ico, Alex, Saria, Leonardo, Davide, Luce, Alessandra, you are an amazing bunch of people. Churri, Edoardo, Nicola, Riccardo - you guys are just mad.

Finally, my love and appreciation goes to my family, though we may fight and disagree on almost everything, our bond is stronger. Last but most definitely not least, thank you Malina for having been by my side, you are truly a rare person, I am glad and lucky to have you in my life.

Contents

Introduction

History of Quantum Computing

It is a natural tendency of narrators of history to assign the birth of a new idea into a unique moment, a singular flash of genius, often accompanied by a colorful story of how that insight came to be and how all of a sudden everything changed. In reality, everything is rarely that simple. However in the field of quantum computing there is a popular contender for such a "let there be light" moment. In 1981 Richard Feynman gave a lecture at MIT titled "Simulating physics with computers" [Fey82]. There, the legendary physicist suggested using a quantum mechanical computer to simulate quantum mechanical systems like molecules and atomic nuclei. Like many of his ideas, his reasoning was simultaneously brilliant and simple: the universe, he argued, is fundamentally quantum mechanical in nature. These laws make it so that fully simulating a quantum mechanical system quickly becomes unmanageable for classical computers based on the Turing machine paradigm. Therefore, why instead not use another naturally occurring quantum mechanical system? He correctly pointed out that it is entirely possible for such a "quantum computer" to exist and successfully simulate another different quantum mechanical system, provided that we have precise control over the interactions within the system. Feynman concludes his brief lecture with the following quote:

> ...nature isn't classical, dammit, and if you want to make a simulation of nature, you better make it quantum mechanical, and by golly it's a beautiful problem, because it doesn't look so easy.

In reality, I have committed a minor historical fallacy, since Feynman was not the first person to conceive of a computer based on quantum mechanics: Paul Benioff [Ben80] and Yuri Manin [Man80] predated him by about one year. Still, he was in making these claims he was right on all accounts: quantum computing is the best way to simulate large quantum mechanical systems, it is a beautiful problem and, most of all, it has proven to be awfully hard to solve.

After this glove of challenge was metaphorically thrown, it wasn't too long before somebody picked it up. In 1985 David Deutsch published a work titled "Quantum theory, the Church-Turing principle and the universal quantum computer" where the principles of a quantum computer capable of realising any natural quantum phenomenon were first described [Deu85]. Thus he introduced many of the basic tools for quantum computing, including a 2-state system as its unit (what came to be known as a qubit), sets of unitary operations to modify the state (including single-qubit rotations, quantum analogues of reversible classical operations, and phase tagging operations), and measurement to extract the result. He also recognised the limitations of real-world quantum computers to perfectly simulate zero-temperature, ideal systems and discussed how the density matrix formalism can be used to describe imperfect, "noisy" quantum states. Almost equally important, Deutsch identified the first problem where a quantum algorithm had a clear advantage over a classical computer. This was generalised in 1992 into the Deutsch-Josza algorithm [DJ92]. Other notable quantum algorithms discovered around this time and exhibiting similar separations were the Bernstein-Vazirani [BV93] algorithm from 1993 and Simon's [Sim97] algorithm first presented in 1994.

These algorithms represented a significant revolution in theoretical computer science, as they demonstrated that there existed tasks for which quantum computers could outperform classical computers, in some cases with seemingly exponential advan-

tage. Indeed Ref. [BV93] introduced the Bounded Quantum Polynomial time (BQP) complexity class, as well as the (still unanswered) question of whether an exponential separation can be conclusively shown to exist between it and BPP. However none of these algorithms solved a problem with any practical relevance. That all changed in 1994 with the discovery by Peter Shor of a polynomial-time quantum algorithm for prime factorisation [Sho99], a problem for which no polynomial-time classical algorithms were (and are) known. This achievement represented a watershed in the field, since the hardness of prime factorisation is the foundation for public-key cryptographic systems such as RSA [RSA78], which in turn underlie many systems requiring secure communication. Thus, although hard to measure, it can be argued that Shor's algorithm was the single biggest factor responsible for elevating quantum computing from a relatively obscure, purely theoretical field to an endeavour with society-changing potential, and put it squarely in the public's eye.

To conclude this "golden decade" of quantum computing, Grover's algorithm was introduced in 1996, providing quadratic speedups for unstructured search problems [Gro96]. Notable are also the discoveries in the field of quantum error correction, again spearheaded by Shor [Sho95]. These provided convincing justification that such algorithms could in principle be run even on noisy devices, provided that systems of large enough size and good enough quality are available [DMN13]. I would also be remiss not to mention the result on quantum simulators by Seth Lloyd, which proved Feynman's hunch right: quantum computers really can simulate physical systems efficiently [Llo96].

Perhaps just as important, the same decade saw the very first quantum computers, which mostly consisted of molecules in solution, the qubits being the nuclear spins of particular atoms accessed via magnetic resonance techniques. The scale was far too small and the noise far too great for any computation of value, but sufficed as a proof of principle [Hol24]. A notable experiment, at the turn of the millennium, was the demonstration by a team at IBM of Shor's algorithm applied to the number 15 (which as the authors helpfully remind us, equals 3×5) [VSB$^+$01].

From this point on up to the present day, the advances have been simultaneously too many and too few to tell. Too many, because the progress has been astounding,

a veritable explosion of papers, experiments, and funding. Just to name a figure: in the year 2000, 500 papers were released on the ArXiv containing the words "quantum computer" or "quantum computing". In 2023, that number had grown more than tenfold to over 5500. The quantum computing economics have also evolved to the point that there are in the range of 400 different quantum computing companies, with dozens of hardware companies, developing several architectures including superconducting, trapped ions, neutral atoms, and photonics [BDG+24]. Too few, because in many senses we are not far from where we were: the basic architecture of quantum algorithms is unchanged, the flagship quantum algorithms are still the same, and the computers are still too small and noisy to run said algorithms.

However, a hint that things could soon change emerged in the mid 2010s, with the suggestion that maybe these small, noisy computer could still perform useful tasks. In 2018 Prof. John Preskill of CalTech captured this climate of hopefulness by introducing the notion of Noisy Intermediate-Scale Quantum or NISQ devices and suggesting that there may be opportunities in this *NISQ era* [Pre18]. Amonst the contenders for the first algorithms to deliver quantum advantage with noisy computers, was a type of algorithm that had emerged a few years prior and had quickly gathered the attention of a large part of the research community: variational quantum algorithms. The variational subfield adopted the acronym NISQ to the point that the two almost became synonymous.

The concept and origins of variational quantum algorithms will be explored in detail in later sections. For the purpose of this Introduction, it suffices to say that variational quantum algorithms employ parameterised quantum circuits as a flexible template to solve a range of different problems. In contrast to previous algorithms requiring fault-tolerance, these algorithms are not motivated by rigorous proofs but are instead heuristic. Their motivation lies in the successes of artificial neural networks in machine learning and of the variational method in quantum chemistry, two quite distinct methods that inspired the creation of variational quantum algorithms. Furthermore, these algorithms are from the start conceived as applicable to non-error corrected machines, due to them potentially giving useful algorithms with much lower gate count and depth requirements, and due to a vague intuition according to which

tunable parameters may help mitigate for noise. However, since the start questions abounded on the validity of these assumption and therefore on the actual usefulness of these algorithms.

When I first encountered variational quantum algorithms at the beginning of my Ph.D. in 2019, the interest of the community had matured enough to the point that it justified making them the central topic of my study. Specifically, the goal was to investigate whether the assumptions of usefulness at small scale and resilience to noise really held. In the course of the doctorate the focus was gradually sharpen to look specifically at the cost function landscape of noisy variational algorithms, as little was known of the mathematical properties of this object.

This Thesis will recount some key messages that my analysis has brought to light. These messages, alongside the numerous discoveries by other researchers, largely point to the reality being much different from what was initially hoped. However, the difference is not always in the negative direction and the conclusions are often considerably nuanced. In addition, as can be glimpsed in this Thesis, the proofs recruit a wide variety of mathematical techniques sometimes leading to the development of useful and novel algorithms. Personally this has made researching variational quantum algorithms a very rewarding experience.

Overview of Thesis

The structuring of this Thesis is as follows. Broadly speaking, it revolves around the topic of noise and its effect on the cost function landscape of variational quantum algorithms. A necessary disclaimer is that even this subfield is enormous, therefore I am going to focus only on those issues that lie at the intersection of what was of interest to me, what I worked on and where most progress has been made.

Following the necessary introduction to quantum computing in general in Chapter 1, and variational quantum algorithms in Chapter 2, Part I will focus on the critical evaluation of noise resilience of these algorithms. Chapter 3 presents a mostly numerical study of small variational algorithms, that still manage to reveal interesting and novel phenomena upon the addition of different noise models. Some of these observa-

tions served as inspiration for the work presented in Chapter 4, which aims to build analytical frameworks for explaining the phenomena. The ultimate goal is generalising to variational algorithms of any size.

Part II of the Thesis broadly explores the same questions but takes a different point of view: that of the Fourier expansion of the landscape. In Chapter 5 this spectral approach is mathematically motivated and used to formulate measures of noise specifically tailored to variational algorithms. An attempt is also made at developing specific noise mitigation strategies. In Chapter 6, the same approach is repurposed for proving that any level of decoherent noise makes a variational algorithm simulatable in polynomial time (with many caveats). The proof is constructive as an algorithm is given that performs these simulations, and shown to work in practice.

Note: for the rest of the Thesis I will use plural tenses as I will be discussing work that was produced in collaboration with other researchers.

List of papers

The following is a list of the papers written during the course of my Ph.D. that are presented in the Thesis, in order of appearance.

1. **Fontana, E.**, Fitzpatrick, N., Ramo, D. M., Duncan, R., & Rungger, I. (2021). *Evaluating the noise resilience of variational quantum algorithms.* Physical Review A, 104(2), 022403.
   Discussed in Chapter 3.

2. **Fontana, E.**, Cerezo, M., Arrasmith, A., Rungger, I., & Coles, P. J. (2022). *Non-trivial symmetries in quantum landscapes and their resilience to quantum noise.* Quantum, 6, 804.
   Discussed in Chapter 4.

3. Wang, S., **Fontana, E.**, Cerezo, M., Sharma, K., Sone, A., Cincio, L., & Coles, P. J. (2021). *Noise-induced barren plateaus in variational quantum algorithms.* Nature communications, 12(1), 6961.
   Discussed in Chapter 4.

4. **Fontana, E.**, Rungger, I., Duncan, R., & Cîrstoiu, C. (2022). *Spectral analysis for noise diagnostics and filter-based digital error mitigation.* arXiv preprint arXiv:2206.08811.
   Discussed in Chapter 5.

5. **Fontana, E.**, Rudolph, M. S., Duncan, R., Rungger, I., & Cîrstoiu, C. (2023). *Classical simulations of noisy variational quantum circuits.* arXiv preprint

arXiv:2306.05400.

Discussed in Chapter 6.

List of itemised contributions to the above papers:

1. Designed and ran numerical experiments and performed the data analysis.
2. Identified phenomenon of parameter symmetries, developed most of the theory.
3. Developed the theory on barren plateaus induced by measurement noise, contributed to general discussions on the rest of the theory.
4. Assisted in the theoretical proofs. Devised experimental procedure, including benchmarking and noise mitigation. Collected experimental data using device APIs.
5. Developed most of the theory, including the algorithm and the proof of efficient simulation.

The following papers were also written in my Ph.D. with myself as first author, but are only mentioned briefly in this Thesis for reasons of space and clarity.

- **Fontana, E.**, Rungger, I., Duncan, R., & Cîrstoiu, C. (2022). *Efficient recovery of variational quantum algorithms landscapes using classical signal processing.* arXiv preprint arXiv:2208.05958.
- **Fontana, E.**, Herman, D., Chakrabarti, S., Kumar, N., Yalovetzky, R., Heredge, J., Sureshbabu, S. H., & Pistoia, M. (2024). *Characterizing barren plateaus in quantum ansätze with the adjoint representation.* Nature Communications 15(1), 7171. [1]

[1]The research for this paper was carried out during an internship at JPMorgan Chase, for the duration of which the PhD was formally suspended.

Chapter 1

Background

Here we provide some background on quantum computing. Since any exposition of quantum computing cannot ignore its foundations in quantum mechanics, first we attempt to give a brief summary of this monumental field. Then we discuss the basic components of any quantum algorithm, and finally explain how errors and external noise can be modelled in the density matrix picture.

This Chapter is based on material from [NC11], to which we refer the Reader for a more in-depth and less rushed explanation.

1.1 Quantum Mechanics

To brutally summarise what is arguably one of the most complex and deep achievements of human intellect, quantum mechanics describes components of reality as belonging to a Hilbert space, with an additional axiom for probabilistic measurement outcomes. Formally a Hilbert space is a vector space equipped with an inner product and satisfying certain technical requirements. We will concern ourselves exclusively with finite dimensional Hilbert spaces, so we can use the definition:

Definition 1 (Finite dimensional Hilbert space)**.** *A finite dimensional Hilbert space is a finite dimensional vector space $\mathcal{H}$ equipped with an inner product $\langle\cdot|\cdot\rangle$, which is a*

function $\mathcal{H} \times \mathcal{H} \mapsto \mathbb{C}$ *with the following conditions:*

- $\langle x|y\rangle = \langle y|x\rangle^*$*;*
- $\langle z|ax + by\rangle = a\langle z|x\rangle + b\langle z|y\rangle$*;*
- $\langle x|x\rangle > 0$ *if* $x \neq 0$ *otherwise* $\langle x|x\rangle = 0$*;*

with $x, y, z \in \mathcal{H}$ *and* $a, b \in \mathbb{C}$*.*

From now on we will assume finite dimensionality and just say Hilbert space. In practice we will almost always use $\mathcal{H} = \mathbb{C}^d$ for some $d \in \mathbb{N}$ and the Euclidean inner product, this can be assumed unless otherwise stated.

We will employ what is known as the *Dirac bra-ket notation*:

Definition 2 (Dirac bra-ket notation)**.** *We assign the following meanings:*

- *The* ***ket*** $|x\rangle$ *denotes a vector in* $\mathcal{H}$*;*
- *The* ***bra*** $\langle x|$ *denotes the conjugate transpose of a vector in* $\mathcal{H}$*:* $\langle x| = |x\rangle^\dagger$*;*
- *Matrix multiplication is implied whenever objects are placed next to each other.*

This has the intended consequence that a bra-ket $\langle x|y\rangle$ is indeed the inner product between $|x\rangle$ and $|y\rangle$. In general the notation makes it possible to understand at a glance what the type of any composite object, by looking at the number and order of bras and kets. For instance, when A is a linear operator, $A|x\rangle$ is a vector and $\langle x|A|y\rangle$ a complex number.

For now it was just mathematical definitions. The innovation of quantum mechanics is mapping these definitions to real objects and experimental procedures. This provides them with an operational nature. The correspondence is as follows:

- Normalised vectors in a complex Hilbert space $\mathcal{H}$ correspond to *quantum states*;
- Unitary operators on $\mathcal{H}$ correspond to the allowed transformations that can be performed on the states;
- Hermitian operators on $\mathcal{H}$ correspond to *observables*, special operations that yield a *probabilistic* outcome.

Let us be more precise. Suppose the state of a quantum system is $|\psi\rangle$, then a unitary transformation U applied on the system yields an outcome $U|\psi\rangle$, another valid quantum state. If we instead measure the quantum state with a Hermitian measurement operator M representing some observable quantity, we would get as outcome a real number. This will be a random variable with expected value $\langle\psi|M|\psi\rangle$ (real because of Hermiticity), which is therefore called the *expectation value* of M on $|\psi\rangle$. The post-measurement state will be:

$$\frac{\Pi_i|\psi\rangle}{\|\Pi_i|\psi\rangle\|_2}, \tag{1.1}$$

where Π_i is an operation that depends on the outcome i that was measured, more specifically a projector onto the eigenspace of M consistent with the measurement outcome. This means that measurement in general changes the state being measured.

The strange rescaling to the state is due to a further postulate of quantum mechanics: *that all allowed quantum states must be normalised to 1 in ℓ^2 norm.* This is due to the fact that the quantum state has a probabilistic interpretation. Take an orthonormal basis of the Hilbert space $\{|i\rangle\}_{i=1}^d$. The operator $M_i = |i\rangle\langle i|$ is a proper measurement operator, in fact it is a *projective measurement* since $M_i^2 = M_i$. The expectation value of this measurement is:

$$p_i = \langle\psi|M_i|\psi\rangle = |\langle i|\psi\rangle|^2 = |[|\psi\rangle]_i|^2. \tag{1.2}$$

This has the physical interpretation of the *probability* of measuring the state $|i\rangle$ given the system is in the state $|\psi\rangle$. We can check that indeed $0 \leq p_i \leq 1$, and also that $\sum_i M_i = \mathbb{1}$ implying $\sum_i p_i = 1$, both derived from state normalisation. A set of bounded linear operators $\{M_i\}$ obeying the conditions $M_i \geq 0$ and $\sum_i M_i = \mathbb{1}$ is called a *positive operator-valued measure* or POVM. POVMs have a special significance in quantum information because they furnish measurement outcomes with a probability measure.

One notices that the global phase of the state has no impact on the probabilities of measurement outcomes, and therefore, no physical significance. This means that in fact quantum states must be identified with *rays* in a Hilbert space, i.e. we are really

dealing with a *projective* Hilbert space, and the transformations are in turn *projective* unitary operators [Mor16, SN20]. However this distinction will in practice not matter in this Thesis as long as we ensure proper normalisation of states.

1.2 Quantum Computing

In the next Chapter we will present the basic structure of the most popular family of near-term quantum algorithms, the variational quantum algorithms. For now we give a very quick overview of the common attributes of qubit-based digital quantum algorithms.

The most basic building block of a quantum computer is the quantum bit, or *qubit*, a two-level quantum system. In analogy with classical bits we label these two states as $|0\rangle$ and $|1\rangle$, which form the *computational basis*. In real implementations these may be distinct energy levels of an atom or of a superconducting circuit. As vectors, these states have the following representation:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.3}$$

The single qubit Hilbert space therefore is $\mathbb{C}^2$.

A single qubit is a relatively useless system, qubits truly become powerful when they are coupled together. In quantum theory, given two quantum systems in Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ the combined quantum system is described by the tensor product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The tensor product of two Hilbert spaces is defined as the Hilbert space spanned by all $u \otimes v$ with $u \in \mathcal{H}_A$ and $v \in \mathcal{H}_B$, where $u \otimes v$ is identified with the tuple (u, v) modulo the following linearity constraints: $(au) \otimes v = u \otimes (av) = a(u \otimes v)$ $\forall a \in \mathbb{C}$, $(u_1 + u_2) \otimes v = (u_1 \otimes v) + (u_2 \otimes v)$, $u \otimes (v_1 + v_2) = (u \otimes v_1) + (u \otimes v_2)$.

Thus, the Hilbert space for n qubits is the tensor product of the single qubit Hilbert spaces $(\mathbb{C}^2)^{\otimes n} = \mathbb{C}^{2^n}$, meaning that the size of the space grows exponentially fast with the addition of new qubits which ultimately underlies the power of quantum computers. By virtue of the exponential growth, a simulation of a quantum computater based solely

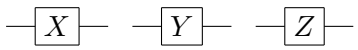
on the statevector becomes impractical beyond about a few dozen qubits even on the largest supercomputers available today.

1.2.1 Quantum gates

To fully explore the Hilbert space we must be able to perform transformations on the state vector in a controllable way, which as mentioned must be unitary. These also interchangeably referred to as *quantum gates*. The most basic ones are single qubit gates, for example the *Pauli operators*:

$$\mathbb{1}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.4}$$

which are both Hermitian and unitary, and therefore square to $\mathbb{1}_2$. By $\mathbb{1}_d$ we indicate the d-dimensional identity matrix. The three nonidentity Pauli operators are given the following circuit diagrams:



Exponentiating these matrices gives the *single qubit Pauli rotations*, conventionally defined as:

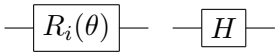
$$R_x(\theta) = e^{-\frac{i}{2}\sigma_x\theta},\ R_y(\theta) = e^{-\frac{i}{2}\sigma_y\theta},\ R_z(\theta) = e^{-\frac{i}{2}\sigma_z\theta}. \tag{1.5}$$

Another significant transformation is the Hadamard gate:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{1.6}$$

also Hermitian, that can be used to switch between the computational basis and the basis spanned by $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$. Note that $|+\rangle^{\otimes n}$ is the uniform superposition of all n qubit computational basis states. The single qubit Pauli rotation gates and the Hadamard gate are indicated as:

Among the possible two-qubit operations the most frequently used is the controlled-



NOT (CNOT) gate, also known as controlled-X (CX):

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.7}$$

with circuit diagram

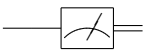
The CNOT belongs to a class of two-qubit gates that can generate entanglement between qubits. For instance, the entangled Bell state $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$ can be created from the unentangled $|00\rangle$ state using a Hadamard followed by a CNOT gate. Such *entangling gates* are fundamental components of all quantum algorithms, as without entanglement they would be easy to simulate classically. Indeed the gate set $\{H, R_z\left(\frac{\pi}{4}\right), CNOT\}$ is *universal* since it can express any unitary. Thus for most intents and purposes, including this Thesis, the gates introduced will be sufficient to build up quantum algorithms.

1.2.2 Measurements

Measurement, or observation, is a fundamental component of any quantum computing algorithm, as it extracts classical information from the quantum state and as such signals the termination of the quantum routine. Observation of a quantum system is truly a bridge between the quantum and classical worlds, and as such the interpretations on what "really" happens during measurement vary wildly. Operationally however the act of measuring a quantum system is quite simple, and it simply involves performing some physical operation that reveals the state of the system, such as shining a laser or

a particular microwave signal and recording the outcome.

For qubit-based quantum computers, measurement projects the quantum state on the computational basis and returns a string of 0s and 1s with the probabilities dictated by the quantum state. However this is only one kind of measurement, known as Z basis measurement, as it projects onto the eigenstates of the Pauli-Z operator σ_z on each qubit. Measurement in other bases is possible by suitably "rotating" the state before measurement using single-qubit gates, for instance a Pauli-X operator can be measured by applying a Hadamard gate. This method can be used to calculate *expectation values* of any multi-qubit Pauli operator, by repeatedly measuring in the suitable basis and averaging the resulting eigenvalues. For instance, take the operator $M = \sigma_z \otimes \sigma_z$, then if by repeatedly preparing the two-qubit state $|\psi\rangle$ and measuring in the computational basis one obtains the bitstrings $\{00, 01, 11\}$, one would assign the eigenvalues $\{+1, -1, +1\}$ and obtain $+\frac{2}{3}$ as an approximation to $\langle\psi|M|\psi\rangle$. Increasingly accurate approximations may be obtained by using a larger number of repetitions, or *shots*. Computational basis measurements appear in circuit diagrams as:



where the double lines signify a wire in a classical state (0 or 1).

Finally, it should be noted that the class of physical measurements is broader than these examples, and consists of the previously mentioned positive operator-valued measures (POVMs). These can in practice be constructed from regular measurements in conjunction with additional unitary gates, ancillary systems and postprocessing of the measurement outcomes.

1.2.3 Notation and structure of Pauli basis

From now on we will use the shorthand I, X, Y, Z for $\mathbb{1}_2, \sigma_x, \sigma_y, \sigma_z$ and imply the tensor product sign, so that $\sigma_z \otimes \sigma_z$ becomes ZZ. The fact that multi-qubit Pauli operators may be written in that form leads to them being occasionally referred to as *Pauli strings*.

Since the Pauli operators form the most natural operator basis for measurement, it is common to express arbitrary measurement operators in this basis. Indeed the 4^n n-qubit Pauli operators form a basis for Hermitian operators on $\mathbb{C}^{2^n}$: writing a generic Pauli operator as P_i (e.g. $P_1 = II\cdots I$ and $P_{4^n} = ZZ\cdots Z$) we can express any measurement operator as:

$$M = \sum_{i=1}^{4^n} c_i P_i. \tag{1.8}$$

Sometimes, we may utilise a different indexing of Pauli operators that more directly reveals their content. First note that since $XZ = iY$ we can decompose a given Pauli string into an X and a Z part, for instance on 3 qubits $XYZ = i \cdot XXI \cdot IZZ$. We then use the notation:

$$X^{\boldsymbol{a}} = \bigotimes_{i=1}^{n} X^{a_i}, \quad Z^{\boldsymbol{b}} = \bigotimes_{i=1}^{n} Z^{b_i}, \quad \boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^n \tag{1.9}$$

such that any Pauli string may be written in the form $cX^{\boldsymbol{a}}Z^{\boldsymbol{b}}$ with $c \in \{\pm 1, \pm i\}$, for instance $XYZ = iX^{(1,1,0)}Z^{(0,1,1)}$.

The efficiency of many quantum algorithms, including most variational ones, depends on M being expressible as a sum of polynomially many Paulis. Finally we note that the Pauli basis is orthogonal under the Hilbert-Schmidt inner product:

$$\langle P_i, P_j \rangle_{HS} = \mathrm{Tr}\{P_i^\dagger P_j\} = 2^n \delta_{ij}. \tag{1.10}$$

All elements of the n-qubit Pauli basis are Hermitian and unitary and retain the property $P_i^2 = P_1 = \mathbb{1}_{2^n}$. We denote the ordered set of all n-qubit Pauli operators by $\mathbb{P}^n$:

$$\mathbb{P}^n = (P_1, ..., P_{4^n}) = (I\cdots I, ..., Z\cdots Z). \tag{1.11}$$

Note that we can form a group under matrix multiplication, the *Pauli group*, by taking the product with $\{\pm 1, \pm i\}$:

$$G_{\mathbb{P}}^n = \{\pm 1, \pm i\} \times \mathbb{P}^n. \tag{1.12}$$

1.3 Quantum noise

The central theme of this Thesis is the impact of errors during the quantum computation, also known as quantum noise. It turns out that treating errors in a formal way requires going beyond the simple picture of vector quantum states and unitary transformations. This picture is only accurate for *closed* quantum systems, which can be pictured as completely independent quantum entities. In contrast most errors in quantum computation occur because of interactions between the quantum system and the external environment, and so we are dealing with *open* quantum systems. These are most properly described via the density matrix formalism, which we will now describe.

1.3.1 Density matrix formalism

Starting from a state vector $|\psi\rangle$, we define the corresponding density matrix as the projector $|\psi\rangle\langle\psi|$. In general:

Definition 3 (Density matrix)**.** *A density matrix in a Hilbert space $\mathcal{H}$ is a Hermitian operator on $\mathcal{H}$ that is positive semidefinite and has trace 1.*

A generic density matrix is usually given the notation ρ. The power of this formalism is that it extends easily from *pure* states to noisy or *mixed* quantum states. Suppose we have a process that outputs state $|0\rangle$ or state $|1\rangle$ with equal *classical* probability. Then the overall quantum mixed state is $\frac{|0\rangle\langle 0|+|1\rangle\langle 1|}{2}$. Contrasting this with the pure state corresponding to their superposition $\frac{|0\rangle\langle 0|+|0\rangle\langle 1|+|1\rangle\langle 0|+|1\rangle\langle 1|}{2}$ one sees that the mixed state lacks the cross terms. Mathematically, we can distinguish between pure and mixed state using the *purity*:

Definition 4 (Purity)**.** *The purity of a quantum state ρ is defined as* $\mathrm{Tr}\{\rho^2\}$.

The purity of a state ρ has its maximum value 1 if and only if the state is pure, and has minimum value $\frac{1}{2^n}$ for n qubit states, attained by the *maximally mixed* state $\frac{\mathbb{1}}{2^n}$.

A density matrix, being an Hermitian operator, can be decomposed in the Pauli basis, yielding a *Pauli vector*, with the following notation:

$$\rho \mapsto |\rho\rangle\!\rangle, \quad [|\rho\rangle\!\rangle]_i = \frac{1}{\sqrt{d}}\langle \rho, P_i\rangle_{HS}. \tag{1.13}$$

Notice that the purity of the state is related to the 2-norm of its Pauli vector:

$$\mathrm{Tr}(\rho^2) = \langle\langle \rho | \rho \rangle\rangle = \| |\rho\rangle\rangle \|_2^2. \tag{1.14}$$

A similar process may be done for measurement operators. In general, writing an operator as a vector in some basis of operators is called *vectorisation*.

1.3.2 Quantum operations

The most general type of transformations on density matrices are called *quantum operations*. It is easy to see that a unitary transformation on a state $|\psi\rangle \mapsto U|\psi\rangle$ takes the form $\rho \mapsto U\rho U^\dagger$ on the corresponding density matrix: unitary transformations yield *unitary operations*. The key characteristic of these operations is that they preserve purity:

$$\mathrm{Tr}\{(U\rho U^\dagger)^2\} = \mathrm{Tr}\{U\rho^2 U^\dagger\} = \mathrm{Tr}\{\rho^2\}. \tag{1.15}$$

They are only a subset of all quantum operations, however. The most general definition is the following:

Definition 5 (Quantum operation)**.** *A quantum operation $\mathcal{E}$ is a superoperator, i.e. a linear map between density matrices, that obeys the following conditions:*

- *Trace preservation (TP):* $\mathrm{Tr}\{\mathcal{E}(\rho)\} = \mathrm{Tr}\{\rho\}$*;*
- *Complete positivity (CP):* $\mathcal{E} \otimes \mathcal{I}_d$ *is a positive map for all* $d \in \mathbb{N}^+$*,*

where $\mathcal{I}_d$ is the identity channel of dimension d.

Note that quantum operations are maps. Therefore we will use the symbol "$\circ$" to refer to concatenation of quantum operations and "$\bigcirc_{i=1}^{m}$" for repeated concatenation. These conditions ensure that the action of a quantum operation on any subsystem of a state results in the global state still being a valid density matrix. In reference to this axiomatic definition quantum operations are also called *CPTP maps*. Sometimes the name *quantum channel* is also used. We shall use all of them interchangeably.

Describing quantum operations is somewhat more complicated than describing unitary transformations, and many different formalisms exist. We present two: the *Kraus* and the *Pauli Transfer Matrix (PTM)* formalism.

Definition 6 (Kraus formalism)**.** *We associate to a quantum channel $\mathcal{E}$ a set of operators $\{K_i\}$ called* ***Kraus operators****, such that:*

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{1.16}$$

which is called the ***Kraus decomposition****.*

To ensure TP we further require that $\sum_i K_i^\dagger K_i = \mathbb{1}$ while CP follows automatically from the Kraus decomposition. Notice that the Kraus decomposition is not unique.

The Pauli Transfer Matrix (PTM) or Pauli-Liouville formalism originates from the previous observation that density matrices may be written in the Pauli basis. This allows us to express a channel as a linear operator acting on Pauli vectors:

Definition 7 (Pauli Transfer Matrix formalism)**.** *The Pauli Transfer Matrix (PTM) formalism associates to a quantum channel $\mathcal{E}$ on $\mathcal{H} = \mathbb{C}^d$ a Pauli Transfer Matrix $\boldsymbol{\mathcal{E}}$, a matrix in $\mathbb{R}^{d^2 \times d^2}$ defined as:*

$$[\boldsymbol{\mathcal{E}}]_{ij} = \frac{1}{d} \mathrm{Tr}\{P_i \mathcal{E}(P_j)\}. \tag{1.17}$$

The TP condition is enforced via the requirement that $[\boldsymbol{\mathcal{E}}]_{1j} = \delta_{1j}$, while there appears to be no straight-forward rule for ensuring CP. A PTM acting on a Pauli vector takes the form of matrix multiplication:

$$|\mathcal{E}(\rho)\rangle\!\rangle = \boldsymbol{\mathcal{E}} |\rho\rangle\!\rangle. \tag{1.18}$$

Notice that the PTM is real by the following argument. Quantum states are Hermitian and therefore they can only have real coefficients when expressed in the Pauli basis. If a channel's PTM had complex coefficients, on some input states it would output invalid states with complex Pauli coefficients. For instance if the element $[\boldsymbol{\mathcal{E}}]_{ij}$ was complex,

one could pass the mixed state $I + P_j$ and the output would contain $[\boldsymbol{\mathcal{E}}]_{ij} P_i$ making it an invalid non-Hermitian state.

We can also very naturally vectorise the Pauli operators themselves, which allows us to write the Pauli normalisation and PTM definition very naturally as

$$\langle\!\langle P_i | P_j \rangle\!\rangle = d\delta_{ij}, \quad [\boldsymbol{\mathcal{E}}]_{ij} = \frac{1}{d} \langle\!\langle P_i | \mathcal{E} | P_j \rangle\!\rangle. \tag{1.19}$$

For example, the unitary channels corresponding to the three types of single qubit Pauli rotations in PTM form are:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad R_y(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 & \sin(\theta) \\ 0 & 0 & 1 & 0 \\ 0 & -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}, \tag{1.20}$$

$$R_z(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.21}$$

One may use the fact that purity is conserved by unitary channels and the relation between purity and 2-norm of Pauli vectors to show that the PTM of a unitary channel must be a matrix from the orthogonal group, as can be verified in the matrices above. Owing to the TP and CP conditions however, PTMs have additional structure and so do not cover the entire orthogonal group on d^2 dimensions.

Finally another common characterisation of quantum channels is the χ-matrix representation. This can be derived from the Kraus decomposition by selecting any basis of operators $\{A_i\}$ and decomposing $K_i = \sum_j a_{ij} A_j$, giving:

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger = \sum_{jk} A_j \rho A_k^\dagger \left(\sum_i a_{ij} a_{ik}^* \right) = \sum_{jk} A_j \rho A_k^\dagger \chi_{jk}, \tag{1.22}$$

where by definition χ is Hermitian and positive [NC11]. A common choice for the

basis of operators is the Pauli basis. If instead the basis of projectors $|i\rangle\langle j|$ is used the resulting χ-matrix is equivalent to the Choi matrix:

$$C_{\mathcal{E}} = \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|). \tag{1.23}$$

For any basis, the CP condition corresponds to χ being positive, while TP corresponds to:

$$\sum_{jk} \chi_{jk} A_k^\dagger A_j = \mathbb{1}. \tag{1.24}$$

1.3.3 Decoherent channels

While unitary quantum channels preserve purity, generic quantum channels may decrease or even increase it (although never beyond 1). If when acting on a pure state a channel decreases its purity then it is called a *decoherent* quantum channel. These are the type of channel most commonly associated to quantum noise (however one may also have coherent errors). We now examine some of single-qubit channels most commonly used in literature.

We begin with the simplest decoherent channel, the *symmetric depolarising* channel. This channel can be expressed simply as:

$$\mathcal{E}_{SD}(\rho) = (1-p)\rho + p\frac{\mathbb{1}}{d}, \tag{1.25}$$

where $p \in [0,1]$ is a free parameter that determines how strongly decoherent the channel is. Clearly, it pushes the state closer to the maximimally mixed state. Its Kraus decomposition is:

$$K_1 = \sqrt{1 - \frac{3p}{4}}\,\mathbb{1}_2, \quad K_2 = \sqrt{\frac{p}{4}}\,\sigma_x, \quad K_3 = \sqrt{\frac{p}{4}}\,\sigma_y, \quad K_4 = \sqrt{\frac{p}{4}}\,\sigma_z, \tag{1.26}$$

while the PTM is:

$$\boldsymbol{\mathcal{E}}_{SD} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1-p \end{pmatrix}. \tag{1.27}$$

Next we examine the *phase damping* channel. This channel the following Kraus operators:

$$K_1 = \sqrt{1-p}\,\mathbb{1}_2, \;\; K_2 = \sqrt{p}\,\sigma_z, \tag{1.28}$$

and the PTM is:

$$\boldsymbol{\mathcal{E}}_{PD} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.29}$$

The channel has the effect of shrinking the off-diagonal components of ρ, when expressed in the computational basis. Both these channels are examples of *Pauli channels*, which have Pauli Kraus operators and are diagonal in the PTM form. Another channel of this family is the *bit-flip* channel that has:

$$K_1 = \sqrt{1-p}\,\mathbb{1}_2, \;\; K_2 = \sqrt{p}\,\sigma_x, \tag{1.30}$$

and

$$\boldsymbol{\mathcal{E}}_{BF} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1-p \end{pmatrix}. \tag{1.31}$$

and, as the name implies, is equivalent to the classical bit-flip channel where a bit is inverted with probability p.

Finally we consider a channel that is not Pauli but is very important for modelling

noise on real devices. This is the *amplitude damping* channel with Kraus operators:

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}, \tag{1.32}$$

and PTM:

$$\boldsymbol{\mathcal{E}}_{AD} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p} & 0 & 0 \\ 0 & 0 & \sqrt{1-p} & 0 \\ p & 0 & 0 & 1-p \end{pmatrix}. \tag{1.33}$$

Notice the nonzero element in the first column, meaning that acting on the completely mixed state will return a less mixed state. This channel models a spontaneous emission causing a decay from the $|1\rangle$ to the $|0\rangle$ state, which is a common issue of many real implementations of qubits.

1.3.4 Coherent errors

Finally, we define *coherent errors* as those errors that preserve purity, which means that they must be unitary operations and so theoretically reversible. In practice however, these errors may be impossible to identify and therefore to correct. Furthermore, since they can be caused from miscalibrations of gates they are very common as any experimental system suffers from some amount of systematic error. Coherent errors can be severely destructive and protocols like randomised compiling have been designed to convert them into decoherent errors [WE16].

Chapter 2

Variational Quantum Algorithms

In this Chapter, we go over the main object of focus of the Thesis, variational quantum algorithms. While the study of these algorithms is recent and in constant flux, it is helpful to still attempt a historical overview of the topic. This will allow us to review the original motivations behind variational quantum algorithms, a critical analysis of which is the recurring theme in this Thesis. We will then outline the general architecture and features and present some of the most widely studied implementations. Finally we conclude with a discussion on the opportunities and challenges that variational quantum algorithms face in finding commercially relevant applications.

The algorithms and theory discussed in this Chapter draw upon several papers that have been duly cited in the text.

2.1 The road to VQAs

2.1.1 NISQ

In the field it is considered good practice to start by introducing the concept of Noisy Intermediate-Scale Quantum (NISQ) [Pre18]. The success of the term, wildly used in both scientific and popular press alike, lies in placing the two largest problems that face modern quantum computing research in one package (with the added advantage

of having a memorable sound). These problems are the presence of noise and the small scale of current quantum computers, and together conspire to make running the most powerful quantum algorithms all but impossible. At the same time, the term NISQ has since its inception been paired with the suggestion that some useful applications may be possible even with our modest quantum resources. By and large, efforts in this direction have been focused on variational quantum algorithms (VQAs).

2.1.2 General structure

The distinctive trait of VQAs is a hybrid classical-quantum optimisation loop, where the quantum computer is paired to a classical device that repeatedly updates the circuit based on the measurement outcomes. To be more precise, the update process typically consists in modifying a set of continuous parameters $\boldsymbol{\theta}$ which determine the specific unitary that is being expressed. Without much originality, this kind of circuit is called a *parameterised quantum circuit* (PQC). Typically the circuit structure is fixed and the parameters that are varied correspond to the angles of rotation gates, such as single-qubit rotations. The fixed aspect of the circuit such as unparameterised gates and the type of rotation gates is called the *ansatz*. Generally we will refer with ansatz to the general structure or architecture of a VQA or a class of VQAs, and we will use PQC to indicate a specific parameterised circuit.

In general we write the parameterised unitary as $U(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, ...)$ is a vector of parameters. Again typically, one takes an initial state $|\psi_0\rangle$ and an Hermitian measurement operator M, and estimates the expectation value which will be a function of the parameters:

$$f(\boldsymbol{\theta}) = \langle\psi_0|U^\dagger(\boldsymbol{\theta})MU(\boldsymbol{\theta})|\psi_0\rangle. \tag{2.1}$$

In many algorithms, including VQE and QAOA, this becomes the *cost function* of the variational algorithm, which the classical optimiser seeks to minimise (conventionally). Sometimes a nonlinear transformation g may be applied such that the cost function is $g \circ f(\boldsymbol{\theta})$.

When discussing noise, it is necessary to replace the unitary with a parameterised

channel $\mathcal{U}_{\boldsymbol{\theta}}$, in which case the cost function becomes:

$$f(\boldsymbol{\theta}) = \mathrm{Tr}\{\mathcal{U}_{\boldsymbol{\theta}}(\rho_0)M\}, \tag{2.2}$$

where ρ_0 is a potentially mixed initial state.

2.1.3 VQE

A near-term algorithm involving a hybrid classical-quantum loop was proposed in 2005 by Aspuru-Guzik *et al.* [AGDLHG05] as a way to reduce the number of qubits needed for a quantum phase estimation (QPE) subroutine, for the purpose of estimating molecular ground states. However the first "true" VQA is arguably the Variational Quantum Eigensolver (VQE), proposed by Peruzzo *et al.* in 2013 [PMS$^+$14]. Similarly introduced to circumvent the limitations of QPE for finding ground states of chemical systems, the main contribution of Peruzzo and collaborators was to bring into quantum computing ideas previously reserved to classical techniques, chiefly that of using the classical outer loop to optimise a quantity measurable on the output state that certifies the solution quality.

More specifically, a common problem in quantum chemistry is finding the ground state and low-lying excited states (or their eigenvalues), for a molecule or material, given knowledge of its Hamiltonian H. For these problems the Hamiltonian can be relatively simple, consisting of polynomially many Pauli terms in the system size, and yet finding the ground state exceedingly hard. It is known that the ground state problem of even 2-local Hamiltonians may be QMA-hard, which is considered unsolvable even for quantum computers [KKR05]. A k-local Hamiltonian is one where the Pauli terms act on at most k qubits.

Yet, in practice many classical methods exist that can approximate the ground state to relatively good accuracy. VQE is inspired by one such method, the variational method, where one classically prepares a trial wavefunction $|\Psi\rangle$ (not necessarily

normalised) and seeks to minimise the Rayleigh-Ritz functional [AW05]:

$$\frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{2.3}$$

The solution is the ground state, if unique, or if degenerate a superposition of orthogonal ground states. This method is limited to those wavefunctions that can be treated classically. The motivation of VQE is that a quantum computer can prepare a vastly bigger range of quantum state, with normalisation coming for free. An example is the unitary coupled cluster (UCC) approach, which uses the ansatz:

$$U_{CC}(\boldsymbol{\theta}) = \sum_i e^{i\theta_i T_i}. \tag{2.4}$$

Here T_i are formed from the excitation operators of the fermionic system. The initial state is the Hartree-Fock state $|\Psi_{HF}\rangle$. The cost function that is minimised is then:

$$f(\boldsymbol{\theta}) = \langle\Psi_{HF}|U_{CC}^{\dagger}(\boldsymbol{\theta})HU_{CC}(\boldsymbol{\theta})|\Psi_{HF}\rangle. \tag{2.5}$$

When the system is mapped to qubits via Jordan-Wigner transform, both the unitary and the initial state can be prepared easily, since the T_i's can be expanded in polynomially many Pauli terms and $|\Psi_{HF}\rangle$ is a single bitstring [RBM$^+$19]. The UCC is a promising candidate as it is a natural extension of the coupled cluster approach which is considered the "gold standard" in quantum chemistry, and furthermore it is not approachable for classical computers [PMS$^+$14].

Another significant ansatz for VQE suggested by Ref. [PMS$^+$14] and developed in Ref. [KMT$^+$17] is the hardware-efficient ansatz (HEA). The idea is to arrange the one- and two-qubit gates native to the hardware in such a way to maximise entanglement, while concurrently optimising for the hardware geometry in order to reduce the gate count. This takes the form of alternating layers of fixed entangling gates and parameterised rotation gates applied to all qubits, making use of the parallelism in gate execution:

$$U_{HEA}(\boldsymbol{\theta}) = \prod_{i=1}^{L} U_{rot}(\boldsymbol{\theta})U_{ent}. \tag{2.6}$$

Typically, $U_{rot}(\boldsymbol{\theta}) = \bigotimes_{i=1}^{n} R(\theta_i)$ with R a Pauli rotation and U_{ent} is a staircase of CNOT gates. The objective is to make the circuit as classically intractable as possible while reducing execution time to reduce the detrimental effects of decoherence. However, the HEA suffers from the grave issue of barren plateaus, which is the vanishing of the variance of the gradient with system size, preventing efficient optimization. This will be explored in detail later.

The final generic ansatz that is noteworthy is the Hamiltonian Variational Ansatz (HVA), introduced in Ref. [WHT15]. In this ansatz, the state is time-evolved in stages, each one corresponding to a term of the problem Hamiltonian. For example, assuming the Hamiltonian can be decomposed as $H = H_1 + H_2$, then the ansatz takes the form:

$$U_{HVA}(\boldsymbol{\theta}) = \prod_{i=1}^{L} e^{-iH_1\theta_{2i}} e^{-iH_2\theta_{2i+1}}. \tag{2.7}$$

The interested Reader may find more information about VQE in Ref. [TCC+22].

2.1.4 QAOA

Not long after the invention of VQE, another key variational quantum algorithm was introduced. In 2014 Farhi, Goldstone and Gutmann released a preprint on the Quantum Approximate Optimisation Algorithm, or QAOA [FGG14]. The problem addressed by this algorithm is combinatorial optimisation, focusing on hard MaxSat problems such as MaxCut. While this problem is radically different from that solved by VQE, the solution is similar since QAOA also maps the problem to ground state estimation. Specifically, the MaxCut objective function can be mapped to a Ising Hamiltonian:

$$H = \sum_{\langle ij \rangle} \frac{Z^{(i)} Z^{(j)} - 1}{2}, \tag{2.8}$$

where the graph vertices are mapped to qubits. Notice that the Hamiltonian is diagonal and thus the eigenvectors are computational basis states. Therefore the ground state of this Hamiltonian will be the solution of the MaxCut problem. In order to find the

ground state, the authors suggest an ansatz of the form:

$$U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{i=1}^{p} U_M(\beta_i) U_P(\gamma_i), \tag{2.9}$$

applied to the uniform superposition state $|+^n\rangle = |+\rangle^{\otimes n}$, where the mixer and problem unitaries are defined as:

$$U_M(\beta) = \prod_{i=1}^{n} e^{-iX^{(i)}\beta}, \;\; U_P(\gamma) = e^{-iH\gamma}. \tag{2.10}$$

As in VQE, the cost function:

$$f(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle +^n | U^\dagger(\boldsymbol{\gamma}, \boldsymbol{\beta}) H U(\boldsymbol{\gamma}, \boldsymbol{\beta}) | +^n \rangle \tag{2.11}$$

is then repeatedly evaluated and the parameters are updated using a classical optimisation algorithm to minimise it. Given a set of converged angle parameters $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$, the solution is extracted by sampling from the output state $U(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)|+^n\rangle$ repeatedly and selecting the bitstring giving the lowest cost function value.

The QAOA ansatz is inspired from adiabatic quantum computation, which it approximates in the large p and small angle limit. The authors of Ref. [FGG14] however suggest keeping p small and letting the angles take the full range of values. They show that this may still yield good solutions to the problem, while being at the same time inefficient to simulate for a classical computer and relatively easy for a near-term quantum computer.

Indeed due to strong theoretical guarantees the QAOA has emerged as a promising candidate of near-term quantum supremacy [FH16]. The variational character of QAOA has been observed to circumvent some of the issues of adiabatic algorithms [ZWC+20], and notably for this Thesis it has been found to offer some resilience to noise when implemented on real devices [OMA+17]. At the time of writing, an up-to-date review of the merits and problems of QAOA can be found at Ref. [BBC+24].

2.1.5 Other algorithms

Besides VQE and QAOA, other algorithms that exploit a classical-quantum optimisation loop have been developed in recent years. Some notable examples are variational quantum factoring [AOAGC19], quantum neural networks [BLSF19], quantum generative algorithms [GZD17, LW18], VQAs for dynamics [LYPS17], the variational quantum linear solver [BPLC$^+$23], and many others. For a review see Ref. [CAB$^+$21] or Ref. [BCLK$^+$22]. As a note, variational quantum algorithms for machine learning like quantum neural networks or quantum GANs fall within the broader category of Quantum Machine Learning (QML) [SSP15, BWP$^+$17], a term that is sometimes applied to all VQAs due to the similarity between the optimisation loop of VQA and the training of neural networks used in classical ML, and which we will avoid due to the possible confusion.

2.2 Challenges and opportunities

The original justification for VQE in the paper by Peruzzo *et al.* was threefold:

- **Simplicity of circuits**: smaller system sizes, no control or ancilla qubits usually required.
- **Adaptability**: the optimisation loop allows for flexibility in the ansatz used, such that circuits tailored to the hardware may be used effectively.
- **Classical intractability**: quantum computers can realise classically intractable ansätze.

Similar arguments are featured in most proposals for other VQAs. For instance in the case of QAOA for MaxCut, the original motivation for QAOA, the ansatz is efficiently realised on most quantum computer architectures by using single-qubit rotations and CNOTs. Conversely sampling from the QAOA output state is believed to be classically intractable even for depth $p = 1$ [FH16].

2.2.1 Noise resilience

The original VQE paper makes only a reference in passing to quantum noise, while the original QAOA paper does not mention it all. The first proper error analysis of VQAs is the 2015 Ref. [MRBAG16], which suggested that VQAs were capable of *variational error suppression*. The idea behind this was that the variational ansatz possessed additional degrees of freedom that may be exploited to compensate for errors in the calculation. Since those errors would cause the expectation value to be higher, the optimiser would naturally adjust the parameters to yield such compensation. More precisely, they consider *coherent errors* which can be corrected by a shift of parameters, such that if U is the noiseless parameterised unitary and $\widetilde{U}$ the noisy one, then at a particular value of $\boldsymbol{\theta}$:

$$U(\boldsymbol{\theta}) \approx \widetilde{U}(\boldsymbol{\theta} + \boldsymbol{\gamma}), \tag{2.12}$$

more exactly their difference is small in some matrix norm. Provided that the noisy unitary yields a higher value for the cost function, the optimiser will then be able to correct for the presence of noise by shifting the parameters accordingly. The authors acknowledge the limitations of this model in the case where the coherent noise violates the symmetries of the ansatz and therefore cannot be corrected.

Arguably a larger limitation is that the effect of decoherent noise cannot be captured by this model. Nonetheless VQE was soon found experimentally to be partly resilient to errors, at least systematic ones and in a small system, confirming the prediction [OBK$^+$16]. Some evidence for resilience to decoherent errors also emerged, for a different variational algorithm [LB17], along with promising simulations for VQE [SSMAG16]. Also of note are the findings of Ref [MKSCD17] that see the variational loop being able to compensate for the effect of noise by identifying decoherence-free subspaces, and that also presents a variant of VQE (quantum subspace expansion) that offers enhanced resilience.

Exploring the truthfulness behind the hypothesis of VQA resilience to decoherence was the main guiding priciple for the work that is presented in this Thesis. For now, it suffices to say that as things stood in 2019, VQAs seemed promising for achieving

"quantum supremacy" in the NISQ area, as they were designed for Intermediate-Scale and potentially could deal with Noisy quantum computers, too. Part of the reason was the development of error mitigation strategies which we will summarise later.

2.2.2 Barren plateaus

Historically barren plateaus (BPs) were amongst the first issues specific to VQAs to be found, and they still remain highly influential in the field. BPs were identified in 2018 by McClean *et al.* [MBS+18] and can be seen as a manifestation of the phenomenon of concentration of measure. Simply put, concentration of measure refers to a collection of many related mathematical results, in which a measure on a space of large dimensionality concentrates on an exponentially small region. The most relevant manifestation is in probability theory, and specifically in the fact that, as 2024 Abel prize winner Michel Talagrand puts it [Tal96]:

> A random variable that depends (in a "smooth" way) on the influence of many independent variables (but not too much on any of them) is essentially constant.

For BPs we consider as random variables the parameters of a VQA under independent random initialisation, and we look at the VQA expectation gradient as a function of the variables. The reason is that most implementations of VQAs assume that the optimisation is started from a set of parameters picked uniformly at random, and thus it is reasonable to ask how large of a gradient we should expect in this setting. The answer by the authors of [MBS+18] is that, for sufficiently expressive ansatze, the variance of the gradient of any parameter under uniform random initialisation decays exponentially with the number of qubits n:

$$\operatorname*{Var}_{\boldsymbol{\theta}\sim\text{Unif}} \partial_i f(\boldsymbol{\theta}) \in \mathcal{O}(2^{-bn}), \quad \forall i,\ b > 0, \tag{2.13}$$

in other words, all gradient components concentrate around the mean (typically equal to zero). For large systems, the VQA cost function landscape then would appear approximately flat with tiny fluctuations, hence the name "barren plateaus". An alternative

definition uses the cost function instead of its partial derivatives:

$$\operatorname*{Var}_{\boldsymbol{\theta}\sim\mathrm{Unif}} f(\boldsymbol{\theta}) \in \mathcal{O}(2^{-bn}), \quad b > 0. \tag{2.14}$$

These two definitions have been shown to be equivalent for most intents and purposes [AHCC22, MB24]. Intuitively, for periodic functions the gradient mean is 0, and therefore concentration in the gradient implies concentration of the cost function along any path by integration. The reverse direction holds by application of the parameter shift rule [MNKF18] that connects gradient to cost function value for typical PQCs.

In many ways, the BP phenomenon is related to the vanishing gradients in recurrent neural networks [BSF94], and the problem it poses can be partially understood from a purely classical standpoint. Like for neural networks, the optimisation loop of VQAs relies either directly or indirectly on the gradient information to optimise the cost function towards its minimum, and therefore one would need to determine the gradients with exponential precision in n in order to successfully run the algorithm. On this point, the quantum vanishing gradients become a much tougher challenge than their classical counterpart: since the cost is calculated from an expectation value, to get a precision of order $\epsilon \in \mathcal{O}(2^{-bn})$ would require running the quantum circuit $\mathcal{O}(\epsilon^{-2}) = \mathcal{O}(2^{2bn})$ times. Contrast this with classical neural networks where the cost of estimating gradients scales with $\mathcal{O}(\log(\epsilon^{-1}))$ [MBS$^+$18]. Overall this means that ansatze exhibiting BPs have a fundamental limit to their system size, at least for uniform parameter initialisation.

Mathematically BPs emerge when the ansatz is an approximate 2-design, i.e. averaging over parameter initialisations approximately returns the same first and second moments as averaging over the Haar measure of the unitary group [RBKSC04]. Put simply, the ansatz produces circuits that are almost as random as possible. Indeed there exists a well-established link between BPs and ansatz expressiveness: the larger the fraction of the unitary group expressible by the ansatz, the smaller the gradient or cost function variance [HSCC22]. This hold also when restricting to subgroups [LCS$^+$22, FHC$^+$23, RBS$^+$23]. Ultimately BPs can be seen a manifestation of the "curse of dimensionality" that often appears across mathematics and computer science [Bel57]: the larger the space explored by the ansatz, the harder it is to optimise

it.

For an recent review on BPs see Ref. [LTW+24].

2.3 Benchmarking noise

Noise is a distinguishing and thus unavoidable feature of the NISQ era. Understanding the quantities and kinds of noise present during the computation is therefore of paramount importance, informing everything from device selection to error mitigation strategies. Characterising an unknown quantum channel is however complicated, not least because the dimension of the space of channels scales quadratically in the Hilbert space dimension, and thus exact characterisation quickly becomes unfeasible. For this reason multiple benchmarking techniques have been developed that attempt to give operationally significant measures of noise that can also be efficiently determined.

2.3.1 Quantum process tomography

We first present Quantum Process Tomography (QPT) an inefficient but simple protocol that historically was the first one to be introduced [CN97]. Briefly the protocol aims to determine any representation of a channel from measurements. In Ref. [CN97] they rewrite the Kraus decomposition of a channel as:

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger = \sum_{ij} A'_i \rho A'^\dagger_j \chi_{ij}, \tag{2.15}$$

where $\{A'_i\}$ is an orthonormal basis of operators and χ_{ij} is a Hermitian matrix encoding the channel. The χ-matrix is determined by performing state tomography on $\{\mathcal{E}(\rho_i)\}$, where $\{\rho_i\}$ is a basis of states. Thus the method assumes that one can exactly produce the intended states and measurements. Experimental realisations often yield χs corresponding to nonphysical channels, to remedy this a maximum likelihood method can be used with the constraint that the channel is CPTP [OPG+04].

The Kraus representation is just a particular choice, in practice any other representation may be used. While this method uniquely determines the channel, it scales poorly with the number of qubits and in practice has been limited to small systems

like single gates or few qubit circuits [OPG+04, WHE+04]. However, in recent times the performance of QPT has been extended to up to 10 qubits using machine learning and tensor networks [TWA+23].

2.3.2 Gate set tomography

Gate set tomography (GST), introduced in Refs. [MGS+13, BKGN+13], is an extension of QPT that seeks to remedy some of its pitfalls, namely the fact that the states and the measurements required for tomography may themselves be noisy or inaccurate [Gre15, NGR+21]. In other words, unlike QPT, GST does not require calibrating the states and measurements to some ideal, perfectly known quantities, but utilises device-native states and measurements. One other difference with QPT is that, like the name suggests, the tomography is done for the entire set of gates that can be enacted on the device, all at once [NGR+21].

GST begins with a gateset $\mathcal{G}$ of initial states, gates, and measurements, represented in vectorised form:

$$\mathcal{G} = \{\{|\rho_i\rangle\!\rangle\}, \{\boldsymbol{\mathcal{E}}_j\}, \{\langle\!\langle M_k|\}\}. \tag{2.16}$$

These operations are device-native and initially unknown, and is what GST seeks to characterise. Then one selects a set of fiducial (i.e. ideal) states and measurements $\{\{|\rho'_l\rangle\!\rangle\}, \{\langle\!\langle M'_m|\}\}$, not necessarily orthogonal, but required to form complete sets for the corresponding vector spaces. It turns out that using just the native operations it is possible to estimate $\langle\!\langle \rho'_l|\boldsymbol{\mathcal{E}}_j|M'_m\rangle\!\rangle$, $\langle\!\langle \rho'_l|\rho_i\rangle\!\rangle$ and $\langle\!\langle M'_m|M_k\rangle\!\rangle$, thus a complete characterisation of the native gateset in terms of the fiducial states. This is achieved using sequences of gates and classical postprocessing techniques like maximum likelihood estimation which also allow to enforce CPTP conditions. Notably, the characterisation is not unique but is always up to a *gauge transformation* given by an invertible A:

$$|\rho_i\rangle\!\rangle \to A|\rho_i\rangle\!\rangle\}, \ \boldsymbol{\mathcal{E}}_j \to A\boldsymbol{\mathcal{E}}_j A^\dagger, \ \langle\!\langle M_k| \to \langle\!\langle M_k|A^\dagger, \tag{2.17}$$

which leaves any physically observable quantity invariant [Gre15, NGR+21].

2.3.3 Randomised benchmarking

Tomographic techniques like QPT and GST can comprehensively characterise noise, however they suffer from an unavoidable exponential scaling with number of qubits n which limits their applicability to very small systems. This is because a superoperator between operators in a Hilbert space of dimension 2^n has dimensionality $\mathcal{O}(2^{4n})$. Randomised benchmarking (RB), in contrast, does not aim to completely characterise the channels but instead to return a compact representation of noise in a circuit, and as such has found a wide applicability in the NISQ era.

To be specific, for a single quantum gate, the RB protocol returns its *average gate fidelity*, that is the gate fidelity averaged over all input states. If the gate channel is $\mathcal{E}$, the average gate fidelity is measured as:

$$\bar{F}_{\mathcal{E}} = \int_U \mathrm{Tr}\{U|0\rangle\langle 0|U^\dagger \mathcal{E}(U|0\rangle\langle 0|U^\dagger)\}\, dU, \tag{2.18}$$

where the average is over the Haar measure for the unitary group $U(d)$. The conjugation of a channel by a random unitary is sometimes called *twirling*. This is the original approach to RB as introduced in Ref. [EAŻ05]. For $\mathcal{E}$ an arbitrary depolarising channel, twirling has the effect of converting $\mathcal{E}$ to a symmetric depolarising noise channel [EAŻ05]:

$$\bar{\mathcal{E}}(\rho) = \int_U U^\dagger \mathcal{E}(U\rho U^\dagger)U\, dU = (1-p)\rho + p\frac{\mathbb{1}}{d}, \tag{2.19}$$

with depolarisation probability p related to the average gate fidelity:

$$p = \frac{1-\bar{F}_{\mathcal{E}}}{1-\frac{1}{d}}. \tag{2.20}$$

Averaging over the unitary group however is impractical. Luckily there's a solution: since the fidelity depends twice on U, a unitary 2-design may be used, the most well-known example being the Clifford group [DCEL09,KLR$^+$08]. Small (1-2 qubits) Clifford gates are easy to implement natively on most devices and also have the advantage of being classically simulatable.

The typical RB protocol thus selects a Clifford gate $G \in U(d)$ to be benchmarked, and starting from a fixed initial state implements sequences composed of m repeated applications of G conjugated by random Clifford gates $R_1, ..., R_m$:

$$S_m = R_m^c G R_m \cdots R_2^c G R_2 R_1^c G R_1, \tag{2.21}$$

$R_i^c = G_i R_i G_i^\dagger$ is a Clifford gate that ensures that the random gates add up to no net unitary transformation, and would have no effect if G is noiseless. However in real devices G is a noisy quantum channel that gets twirled by the random gates and therefore can be written as a noiseless unitary followed by a symmetric depolarising channel.

In a single benchmark several such random sequences of varying length m are applied and the final state is measured in such a way to estimate the probability p_m that the sequence yields the correct state (from classical simulation), also called survival probability. The different estimates as a function of m are then used for an exponential regression, the decay rate being related to the depolarising coefficient p and therefore the average gate fidelity [KLR$^+$08]. The RB protocol is very robust to state preparation and measurement (SPAM) noise [MGE12], and deviations from exponential decay for the survival probability signal breakdown of the Markovian assumptions of noise [Wal18].

RB is much less demanding than full tomography, however in practice it is limited to small Clifford gates, since realising n-qubit Clifford gates requires $O(n^2/\log n)$ primitive gates and so the fidelity decreases rapidly with n [PSR$^+$22]. Still, RB of 1- and 2-qubit gates is the de-facto industry standard for assessing and reporting the quality of quantum computing devices [PCDR$^+$19].

2.4 Error mitigation

Arguably quantum error mitigation (QEM) for NISQ is the natural counterpart to error correction for long-term devices. As the nomenclature suggests, while error correction seeks to eliminate errors (or at least exponentially suppress them), error mitigation is any practice that can reduce the impact of errors on the calculation. The underlying

implication is that NISQ devices are far too noisy and small scale for fault tolerance, therefore we must settle for methods that reduce but do not eliminate the effects of quantum noise. This was seen as sufficient because of the aforementioned narrative on the noise resilience of VQAs, according to which these algorithms may still work even with some noise present. Indeed QEM can be applied very naturally to VQAs, as most techniques seek to reduce the effect of noise on quantum expectation values. One point of note is that such techniques must be applied to a fixed quantum state and therefore in a variational algorithm they have to be repeated for every parameter value.

2.4.1 Zero Noise Extrapolation

To review the most widely used QEM algorithms, we begin with Zero Noise Extrapolation (ZNE), proposed independently in Refs. [TBG17, LB17]. Here we let the quantum state ρ depend on a noise parameter λ, such that $\rho_{\lambda=1}$ corresponds to the state output by the real (noisy) device and $\rho_{\lambda=0}$ is the ideal (noiseless) state. Then the expectation value of an operator M may be Taylor expanded in powers of λ:

$$\mathrm{Tr}\{M\rho_\lambda\} = \mathrm{Tr}\{M\rho_0\} + \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} \frac{d^i}{d\lambda^i}\bigg|_0 \mathrm{Tr}\{M\rho_\lambda\}. \tag{2.22}$$

Now the linear approach assumes that the series can be approximately truncated at the first order:

$$\mathrm{Tr}\{M\rho_\lambda\} \approx \mathrm{Tr}\{M\rho_0\} + \lambda \frac{d}{d\lambda}\bigg|_0 \mathrm{Tr}\{M\rho_\lambda\}, \tag{2.23}$$

and the slope may be estimated by artificially increasing the noise in a controlled manner to $\lambda = r > 1$, which may be done for instance by extending gate times or by adding filler operations that add up to unitary. This is done multiple times obtaining different expectation values, which are then linearly extrapolated to find $\mathrm{Tr}\{M\rho_0\}$. This is the approach suggested by Ref. [LB17], Ref. [TBG17] instead uses a different linear extrapolation technique based on Richardson extrapolation. Some variations of ZNE with improved performance include ZNE with exponential extrapolation [EBL18] and digital ZNE [GTHL$^+$20] which can be implemented at circuit level. The success of exponential ZNE in practice is notable because it shows that the linear approximation

may not hold on NISQ devices.

2.4.2 Probabilistic Error Cancellation

Probabilistic Error Cancellation (PEC) was also introduced in Ref. [TBG17]. The idea is to express a clean circuit using a (quasi)probabilistic mixture of noisy ones. More precisely, it is assumed that we can decompose an intended unitary channel as:

$$\mathcal{U} = \sum_i \eta_i \widetilde{\mathcal{U}}_i, \tag{2.24}$$

where $\eta_i \in \mathbb{R}$ are expansion coefficients. This is true whenever the noisy channels $\widetilde{\mathcal{U}}_i$ form a sufficiently large basis in the space of quantum operations. Then one can rewrite the noiseless expectation value in the *quasiprobability representation* (QPR):

$$\mathrm{Tr}\{M\mathcal{U}(\rho)\} = \gamma \sum_i s_i p_i \mathrm{Tr}\{M\widetilde{\mathcal{U}}_i(\rho)\}, \tag{2.25}$$

where $s_i = \mathrm{sign}(\eta_i)$, $\gamma = \sum_i |\eta_i| \geq 1$ is a scaling factor and $p_i = \frac{|\eta_i|}{\gamma}$ is a probability. Assuming we can produce the noisy expectation value with the circuit and the decomposition coefficients are known, the noiseless expectation value may be approximated.

A more practical point of view on the QPR comes from assuming that for any unitary channel $\mathcal{U}$ the noisy quantum device implements $\widetilde{\mathcal{U}} = \mathcal{D}\mathcal{U}$ where the noise $\mathcal{D}$ is independent of the unitary. Then one selects correction unitary channels $\mathcal{C}_i$ and coefficients η_i to invert the noise:

$$\sum_i \eta_i \mathcal{C}_i = \mathcal{D}^{-1}. \tag{2.26}$$

By setting $\widetilde{\mathcal{U}}_i = \widetilde{\mathcal{C}_i\mathcal{U}}$ one obtains an appropriate QPR [EBL18]. Note that however there exist channels that cannot be inverted, such as reset channels.

Ref. [TBG17] provide decompositions for the case of symmetric depolarising and amplitude damping noise on up to two qubits. PEC was subsequently updated in Ref. [VDBMKT23] to cover arbitrarily circuits, by constructing, via benchmarking a

sparse representation of noise in the Pauli basis adapted to hardware connectivity. The resulting compact Pauli noise model can be efficiently inverted to yield a suitable QPR. More recently the method was successfully applied to a large 127 qubit time evolution experiment by IBM, alongside ZNE [KEA+23].

2.4.3 Clifford Data Regression

Another technique that like ZNE utilises a linear approximation to invert the effect of noise is Clifford Data Regression (CDR). Introduced in Ref. [CACC21], given a quantum circuit the idea is to create a linear model that takes noisy expectation values as input:

$$f(\mathrm{Tr}\{M\widetilde{\mathcal{U}}(\rho)\}) = a_1\mathrm{Tr}\{M\widetilde{\mathcal{U}}(\rho)\} + a_2. \tag{2.27}$$

and aims to approximate $\mathrm{Tr}\{M\mathcal{U}(\rho)\}$ on unseen samples, therefore serving as a denoising model. The model is trained by least-squares regression on data gathered from circuits similar to the target one, chosen such that the noiseless circuit can be simulated classically. Specifically the suggestion is to substitute some non-Clifford gates by the closest Clifford gate, since Pauli expectation values of circuits with t non-Clifford gates can be simulated in time $2^{\Omega(t)}$ [BBC+19]. Given a set of such circuits $\{\mathcal{C}_i\}_i$, the dataset for regression is:

$$\mathcal{D} = \{(\mathrm{Tr}\{M\widetilde{\mathcal{C}}_i(\rho)\}, \mathrm{Tr}\{M\mathcal{C}_i(\rho)\})\}_i, \tag{2.28}$$

with the noisy expectation values sampled from the device and the noiseless ones obtained from the classical simulator. The regression then minimizes the loss

$$\mathcal{L} = \sum_{(\mathcal{X},\mathcal{Y})\in\mathcal{D}} (\mathcal{Y} - f(\mathcal{X}))^2 \tag{2.29}$$

by finding the best-fit parameters (a_1, a_2).

The rationale for CDR is that some channels like the symmetric depolarising one lead precisely to a linear relationship between noisy and noiseless expectation values. Specifically, as shown in Ref. [CACC21], m layers of symmetric depolarising noise with

depolarising probability p lead to coefficients

$$a_1 = \frac{1}{(1-p)^m}, \quad a_2 = -\frac{1-(1-p)^m \mathrm{Tr}M}{d(1-p)^m} \tag{2.30}$$

where d is the size of the Hilbert space.

2.4.4 Randomised compiling

Randomised compiling takes the core idea of randomised benchmarking, using twirling to convert complex noise channels into simpler ones, and applies it to noise mitigation. The reason is that in many situations it is preferable to be able to treat noise as simple decoherent channels as opposed to having arbitrary noise. One example for the fault-tolerant era is that most quantum error correction codes have been designed to correct Pauli errors. Furthermore, as mentioned previously, in the fault-tolerant setting coherent errors are believed to be more destructive than decoherent errors, since they accumulate quadratically instead of linearly and therefore require longer code distances [Got19]. These are the motivations behind randomised compiling [WE16], which is thus a method that finds more use in fault-tolerant algorithms than for VQAs [JIBE23], also because there are good theoretical arguments for the latter being resilient to coherent errors since the optimization can partially correct them (see Chapter on VQAs). Yet it is useful to briefly review the basics of randomised compiling, both because some NISQ algorithms like quantum simulation may benefit [HNM+21], and to get further intuition on quantum noise.

Randomised compiling utilises twirling to obtain simpler effective noise channels. However, since reduction to Pauli channels is sufficient for most purposes, instead of Clifford gates the much simpler Pauli gates are used, leading to Pauli twirling:

$$\tilde{\mathcal{E}}(\rho) = \mathbb{E}_{P\sim\mathbb{P}^n} P\mathcal{E}(P\rho P)P. \tag{2.31}$$

It can be shown that $\tilde{\mathcal{E}}$ is a Pauli channel: there exists a probability vector $\boldsymbol{p}$ such that:

$$\tilde{\mathcal{E}}(\rho) = \sum_{i=1}^{4^n} p_i P_i \rho P_i. \tag{2.32}$$

In practice this is done by compiling the quantum circuit into alternating layers of less noisy "easy" gates C_i and more noisy "hard" gates G_i; the latter are error-mitigated while the former are used for mitigation. In Ref. [WE16] the hard gates are Cliffords (such as Hadamard and CNOT) and the easy gates arbitrary products of single-qubit rotations on all qubits. This gateset is universal and so any circuit may be compiled with it. At each round the easy gates C_i are modified such that they enact a random conjugation by $P_i \in \mathbb{P}^n$ on the hard gates. The algorithm's gate sequence is modified to:

$$C_m G_m \cdots C_1 G_1 C_0 \to \tilde{C}_m G_m \cdots \tilde{C}_1 G_1 \tilde{C}_0, \tag{2.33}$$

where $\tilde{C}_i = P_i C_i P_i^c$ is the new easy gate and like before $P_i^c = G_i P_{i-1}^\dagger G_i^\dagger$ compensates for the net unitary rotation, but this time is Pauli since the Clifford group normalises $\mathbb{P}^n$. Selecting different random Paulis for each circuit run leads to the intended Pauli twirling of the noise channels on the hard gates G_i [WE16].

2.4.5 Other techniques

Symmetry verification refers to QEM techniques, predominantly for algorithms in quantum chemistry like VQE, where errors are mitigated by knowledge of the symmetries that the noiseless output state should obey. The mitigation occurs either by projecting onto the symmetric subspace throughout the quantum circuit, which requires mid-circuit measurements, or more simply via postselection of the final state [BMSSO18, SBMS$^+$19]. The method has also been applied to QAOA which also offers symmetries [KLGS22].

Another method that, while not specifically created for this purpose, can improve the resilience of VQE to noise is quantum subspace expansion [MKSCD17]. In this method the final state $|\Psi\rangle$ found via VQE is expanded into a subspace by applying

fermionic operators:

$$\mathcal{S} = \{|\Psi\rangle\} \cup \{a_j^\dagger a_i|\Psi\rangle\}_{i\neq j} \cup \{a_k^\dagger a_j^\dagger a_l a_i|\Psi\rangle\}_{i\neq j\neq k\neq l} \cup \cdots, \tag{2.34}$$

up to a predetermined order. The Hamiltonian is that projected onto this subspace giving a generalised eigenvalue problem (since the vectors may not be orthogonal), which can be solved classically for low orders. This may be used to find excited states or, in the presence of noise, to attempt to find a new improved solution for the original ground state problem. QSE may also naturally be combined with symmetry verification techniques [MKSCD17].

Virtual distillation, introduced simultaneously in Refs. [Koc21, HMO$^+$21], refers to protocols that reduce errors by calculating expectation values, instead on a noisy ρ, on the state:

$$\rho^{(m)} = \frac{\rho^m}{\mathrm{Tr}\rho^m}, \tag{2.35}$$

where all eigenvalues except the dominant one are exponentially suppressed. With the assumption that the dominant eigenvalue is the noiseless state, this improves the solution quality. This is achieved by preparing m copies of the state and performing various techniques for estimating $\mathrm{Tr}\{M\rho^{(m)}\}$.

Part I

Noise and optimization

Chapter 3

Numerically evaluating noise resilience

This Chapter explores the impact of noise on variational quantum algorithms, specifically looking at changes in the landscape and at convergence under different noise models. An attempt is made at creating simple models for understanding these effects. Some new interesting phenomena are seen with noise: notably the presence of multiple local minima with a range of cost function values, which can cause sharp transitions in the converged parameters beyond some noise thresholds, and an inherent resilience to noise when the circuit is overparameterised.

This Chapter is based on Ref. [FFR+21b].

3.1 Methods

3.1.1 Noise model

Recall from Chapter 1 that the ansatz of a VQA without noise is a parameterised unitary operator $U(\boldsymbol{\theta})$, with the vector $\boldsymbol{\theta}$ representing the used parameters, that is

applied to a fixed initial state ρ_0 to yield a desired output state:

$$\rho(\boldsymbol{\theta}) = U(\boldsymbol{\theta})\rho_0 U(\boldsymbol{\theta})^\dagger. \tag{3.1}$$

Equivalently, we can associate to the unitary a superoperator $\mathcal{U}_{\boldsymbol{\theta}}$. Decomposing the PQC in its constituent gates, the overall quantum operation can be written as a composition of superoperators:

$$\rho(\boldsymbol{\theta}) = \mathcal{U}(\boldsymbol{\theta})\,(\rho_0) = \mathcal{U}^L_{\boldsymbol{\theta}_L}\mathcal{U}^{L-1}_{\boldsymbol{\theta}_{L-1}}\cdots\mathcal{U}^1_{\boldsymbol{\theta}_1}\,(\rho_0), \tag{3.2}$$

where each unitary is independently parameterised. In a slight abuse of notation we omit the map composition symbol.

The cost function, $C(\boldsymbol{\theta})$, typically corresponds to the expectation value of an operator O for the prepared state. The classical optimiser then attempts to find the minimum of the real-valued cost function:

$$C(\boldsymbol{\theta}) = \mathrm{Tr}[O\,\rho(\boldsymbol{\theta})]. \tag{3.3}$$

Eqs. (3.2) and (3.3) encapsulate the unitary dynamics of a closed system representing a noiseless quantum computer. Here we simulate this on a classical machine, which can be achieved by representing the operators in the equation as matrices and performing the operations numerically.

A real quantum computer is an open quantum system; therefore, our basic unitary evolution needs to be expanded with a noise model. Here we construct the model by interleaving noiseless operations and noisy quantum channels that aim to replicate decoherent processes in quantum computers. Indicating the effect of noisy quantum channels as Λ_l, with the integer l indexing the specific channel, Eq. (3.2) is modified to:

$$\rho(\boldsymbol{\theta}) = \Lambda_L\mathcal{U}^L_{\boldsymbol{\theta}_L}\Lambda_{L-1}\mathcal{U}^{L-1}_{\boldsymbol{\theta}_{L-1}}\cdots\Lambda_1\mathcal{U}^1_{\boldsymbol{\theta}_1}\Lambda_0\,(\rho_0). \tag{3.4}$$

For simplicity we use identical noise channels throughout the circuit evaluation. This

also makes sense because we will apply these channels only after 2-qubit gates in our experiments, which will always be the same type of gate (CNOT), and therefore we can assume a similar noise profile. Each of these spans all the qubits in the system, and is defined as the application of an identical one-qubit noise channel to every qubit:

$$\Lambda(\rho) = \left(\bigotimes_{i=1}^{N} \Lambda^{(i)} \right) (\rho), \tag{3.5}$$

where $\Lambda^{(i)}$ is the one-qubit channel acting on the i^{th} qubit, ρ is the input state and N is the number of qubits. This is termed a *product channel* [ZYZP19]. The same approximation is used in Ref. [BD08], and is valid when qubits are sufficiently separated physically and there is only small cross-talk. Note that, again for simplicity, we have also assumed that the noise is identical on every qubit, however in general the qubits of a real quantum computer have noise characteristics that can differ significantly from one another [AAB+19]. In our approach we neglect cross-talk and coherent errors, as our main focus is on decoherent noise channels. Similarly, we do not consider readout noise, even though this is an important feature of real quantum devices [AAB+19]. The justification for this choice is that readout noise is independent from ansatz design, while the main question of this Chapter is to address the effect of a noisy ansatz itself. We also neglect finite sampling (shot) noise, which allows us to employ exact statevector simulations.

We consider three types of noise channels: amplitude damping, phase damping, and symmetric depolarising channels. Two common metrics of qubit quality, the longitudinal relaxation ($T1$) and dephasing ($T2$) times, can be related directly to amplitude and phase damping [HdSS02, SSMAG16, RJV+20]. Symmetric depolarising noise describes a decay to a completely mixed state, and hence is useful as a prototypical decoherent channel [NC10].

We make use of the Kraus operator formalism (Sec. 1.3.2) to apply these channels onto the quantum state. In our implementation, the operators are parameterised by $\gamma \in [0, 1]$, representing the strength of the noise. A value of 0 represents no noise (and hence an identity channel), while a value of 1 is maximal in the sense that the output of

the noise channel corresponds to the fixed point of the channel. For amplitude damping, this is the state $|0\rangle\langle 0|$, for dephasing it is any linear combination of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, and for symmetric depolarising it is the completely mixed state.

As a simplification, we apply noise only on two-qubit gates. This is justified by physical considerations valid for most hardware systems, where 2-qubit operations are considerably slower than the 1-qubit ones, and have a much higher noise rate, usually by an order of magnitude in their gate fidelity figures [TQ18]. Consistently with previous work [DH21,BD08], we place a noisy channel after the gate, ensuring that at maximum noise the state output by the circuit will be unentangled even for non-depolarising noise channels like phase and amplitude damping.

3.1.2 VQE

The first system we will examine is VQE. The results are presented in Sec. 3.2.1. The VQE algorithm seeks to identify the ground state of a given Hamiltonian H, by setting $O = H$ in Eq. (3.3) and minimising the cost function $C(\boldsymbol{\theta})$. The value of $C(\boldsymbol{\theta})$ itself is thus a good measure of the quality of the solution.

To assess the quality of the output after convergence we also consider the fidelity with respect to the exact ground state of H. The fidelity is a measure of closeness between states, given two states ρ and σ it is defined as [Joz94]:

$$F(\sigma,\rho) \coloneqq \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}\right)^2 = \mathrm{Tr}\left[\sigma\rho\right],$$

where the last equality is valid whenever one of the two states is pure. For ρ and σ both pure, it has the minimum value of 0 when they are orthogonal and 1 when $\rho = \sigma$. Assuming the Hamiltonian is non-degenerate, if we denote the exact ground state as $|\psi_{\mathrm{gs}}\rangle$, the fidelity of the output state corresponds to:

$$F(\boldsymbol{\theta}) = \langle\psi_{\mathrm{gs}}|\,\rho(\boldsymbol{\theta})\,|\psi_{\mathrm{gs}}\rangle\,. \tag{3.6}$$

Finally, since in VQE we are almost exclusively interested in Hamiltonians with entangled ground states, we use entanglement as a further important test of the quality of the

output state. If we have a 2-qubit system, there exists a broad selection of measures of bipartite entanglement. Since the noisy circuit produces mixed states, we choose one that is valid in this regime, namely the concurrence, $Q(\rho)$ [HW97]. The concurrence is frequently used in literature, as it is monotonically related to entanglement of formation, a meaningful measure of entanglement, while being easier to calculate in practice [CKW00, Woo98, LKL$^+$17]. It has the closed form:

$$Q(\rho) \coloneqq \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4). \tag{3.7}$$

Here λ_i are the eigenvalues, in decreasing order, of the Hermitian matrix $\sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$, where $\tilde{\rho}$ is the spin-flipped density matrix $(\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$, with $*$ indicating complex conjugation. $Q(\rho) = 0$ if and only if ρ is a linear combination of product states, and $Q(\rho) = 1$ if and only if ρ is a Bell state [Woo01].

When working with more than 2 qubits we cannot directly apply concurrence since it is specifically valid for 2-qubit systems. Out of the several possible measures of multi-qubit entanglement [LVS$^+$07], we choose the maximum concurrence taken over all pairs of qubits in the system [YYE07, MM19] as a natural extension of concurrence that is simple to calculate and to interpret. In particular if the maximum pairwise concurrence is zero then the state has no bipartite entanglement across any partition (however it may still have multipartite entanglement).

Initially we consider the following Hamiltonian on two qubits:

$$H_2 = \hat{\sigma}_z^1 \hat{\sigma}_z^2 + \hat{\sigma}_x^1 + \hat{\sigma}_x^2. \tag{3.8}$$

This is an example of a transverse-field Ising Hamiltonian [HPK13], and appears in this form in dynamical mean field theory (DMFT) simulations of the single-impurity Anderson model (SIAM) for its 2-electron ground state [RFC$^+$19]. Since this minimal Hamiltonian has an entangled ground state, it forms an ideal starting point for the investigation of the effect of noise. As will be shown in the subsequent sections, the conclusions found here are applicable also to wavefunctions obtained with more complicated Hamiltonians. Furthermore, as the Hamiltonian is real and nondegenerate, the

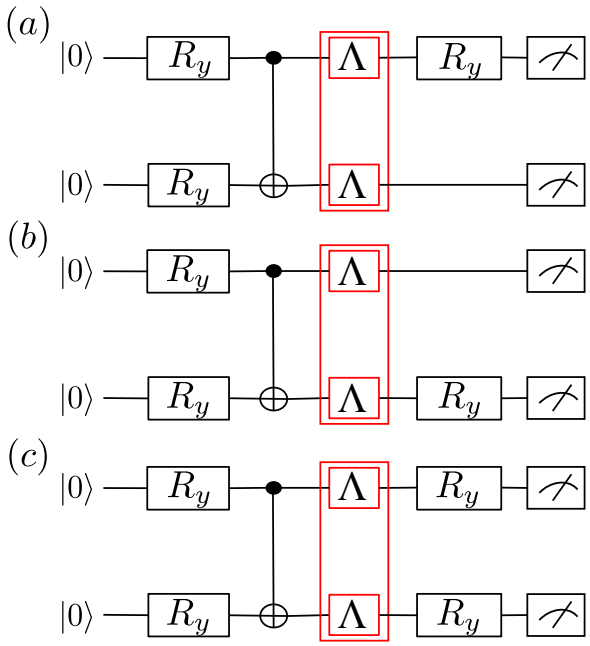


Figure 3.1: Circuit ansätze used for state preparation on two qubits. Circuits (a) and (b) have 3 rotation parameters, and the circuit in (c) has 4 rotation parameters. In the noiseless case they all allow to cover the full real-states space of two qubits, and hence allow to construct any real 2-qubit state. There is one noise channel applied after the CNOT gate, as indicated by the Λ blocks on each qubit.

ground state will be a real vector too. This is because:

$$\begin{aligned} H|\psi_{gs}\rangle = E_{gs}|\psi_{gs}\rangle &\xrightarrow{*} H^*|\psi_{gs}\rangle^* = E^*_{gs}|\psi_{gs}\rangle^* \\ &\xrightarrow{=} H|\psi_{gs}\rangle^* = E_{gs}|\psi_{gs}\rangle^* \\ &\xrightarrow{\text{nondeg.}} |\psi_{gs}\rangle^* = |\psi_{gs}\rangle. \end{aligned}$$

Thus we can restrict our choice to just those circuits that always output a real wavefunction, enabling us to significantly reduce the number of parameters.

As ansätze we choose the three circuits shown in Fig. 3.1, each of which can be shown analytically to be able to prepare any possible real 2-qubit state. These ansätze include two inequivalent 3-parameter circuits, termed circuit (a) and (b), which differ on the position of the final rotation gate. We also consider a 4-parameter circuit that has rotations on both qubits before measurement. This latter circuit is over-parameterised, since it has one extra parameter compared to the previous two, and therefore allows us to explore the impact of redundant parameters.

The VQE algorithm is implemented using an exact density matrix simulator, which allows the use of a gradient-based classical optimiser, specifically the Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimiser in our case [Bro70,Fle70,Gol70, Sha70]. The algorithm is evaluated on a range of $\gamma \in [0, 1]$, for phase, amplitude damping and symmetric depolarising noise, for all the ansatz circuits.

We also perform VQE on a larger 4-qubit, 2-local Hamiltonian:

$$H_4 = \hat{\sigma}^1_z\hat{\sigma}^3_z + \frac{1}{2}(\hat{\sigma}^1_x\hat{\sigma}^2_x + \hat{\sigma}^1_y\hat{\sigma}^2_y + \hat{\sigma}^3_x\hat{\sigma}^4_x + \hat{\sigma}^3_y\hat{\sigma}^4_y). \tag{3.9}$$

This Hamiltonian describes a similar physical system as H_2 [RFC+19] and is similarly real. The chosen state preparation ansatz circuit is shown in Fig. 3.2, and gives the exact energy in the noiseless case [RFC+19].

3.1.3 Random target state

Next we will investigate how noise affects a variational algorithm for solving a more general task. We consider random state fidelity optimisation, where, rather than choosing

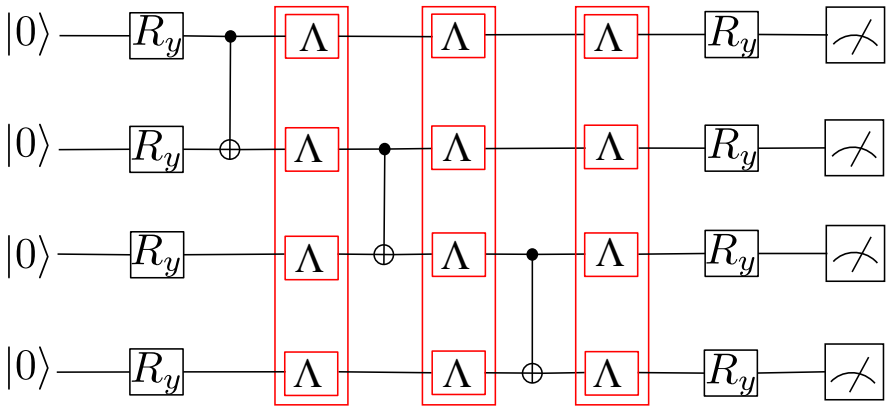


Figure 3.2: Circuit used for the 4-qubit Hamiltonian VQE simulations. We apply a noise channel on all qubits after each CNOT gate.

a specific Hamiltonian and evaluating its ground state, we select at random a pure target state ρ_{T}. This approach allows to estimate how closely ground state wavefunctions of arbitrary Hamiltonians can be reproduced with a given circuit ansatz in presence of noise. The results are presented in Sec. 3.2.2.

The optimisation procedure is modified to maximising the fidelity $F(\rho_{\mathrm{T}}, \rho(\boldsymbol{\theta}))$. Equivalently, the problem can be formulated as a minimisation of the infidelity, defined as:

$$R \coloneqq 1 - F, \tag{3.10}$$

and hence the cost function is:

$$C(\boldsymbol{\theta}) = R(\rho_{\mathrm{T}}, \rho(\boldsymbol{\theta})) \coloneqq 1 - F(\rho_{\mathrm{T}}, \rho(\boldsymbol{\theta})). \tag{3.11}$$

We consider pure target states, so a VQE cost function can be used by choosing $O = \mathbb{1} - \rho_{\mathrm{T}}$.

We then extend this to the case where one is provided with a set of n_{T} pure target states sampled from a uniform distribution. As figure of merit we use the *average optimal infidelity* over the set, which we define as:

$$\bar{R} = \frac{1}{n_{\mathrm{T}}} \sum_{n=1}^{n_{\mathrm{T}}} \min_{\boldsymbol{\theta}} R(\rho_{\mathrm{T},n}, \rho(\boldsymbol{\theta})), , \tag{3.12}$$

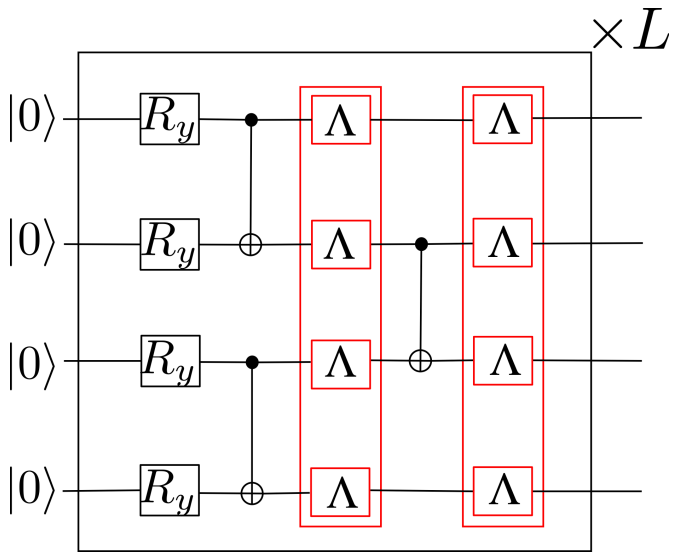


Figure 3.3: Circuit block used for the random target state fidelity maximisation algorithm; shown is one layer of the ansatz, which is repeated a number L times in the full ansatz. The noise channels are added to each qubit after the CNOT gates.

where $\rho_{\mathrm{T},n}$ is the target state with index n. The same measure has been used in Ref. [GZB+20]. We use a set of target states instead of a single target state since we are investigating the capability of the ansatz to prepare a variety of states, as it is intended as a general purpose ansatz in the style of the HEA. For each target, the optimisation procedure is run in presence of noise. $\bar{R} = 0$ would imply that the quantum circuit can represent any state in the ensemble exactly. The addition of noise is expected to increase $\bar{R}$, as mixed states cannot have perfect overlap with pure states. As the distribution of target states, we consider the Haar distribution, which is the unique uniform distribution over a space of pure quantum states [DCEL09]. More precisely, the ensemble consists of 1000 real states generated by sampling a random orthogonal matrix from the circular real matrix distribution, the Haar distribution over real orthogonal matrices [Res], and picking its first column.

We choose a hardware-efficient ansatz [KMT+17] consisting of an identical layer of 4 rotation gates and 3 CX gates, repeated L times, as illustrated in Fig. 3.3. The choice of this specific structure is motivated by its high expressibility as demonstrated in Ref. [SJAG19], and by its compactness, which reduces the number of noise channels per layer. Indeed, the first two CX gates can be executed in parallel, and hence according

to our noise model we insert only two noise channels per layer. Hardware-efficient circuits dense in parameterised operations are well-suited for preparing general quantum states [MEAG$^+$20, RBM$^+$19, M$^+$18]. In our experiments, the circuit depth L ranges from 1 to 5 to explore ansätze with different expressiveness.

We consider again phase damping, amplitude damping and symmetric depolarising noise, and use the same density matrix simulator and local gradient-based minimiser that we employed in the VQE simulations. We perform two types of numerical simulations: in the first, we optimise the circuit without noise to obtain the optimal parameters for the ideal case but evaluate the infidelity using the noisy circuit ("non-reoptimised"); in the second type we start from the same noiseless optimum and then reoptimise the circuit by performing gradient descent with the noise channels in place ("reoptimised"). The non-reoptimised cost function provides an upper bound to the reoptimised cost function, and the two will be equal only if the location of the minimum is unaffected by noise. In order to isolate the effect of noise from other contributions to the infidelity, we consider the average optimal relative infidelity $\bar{R}_{\text{rel}}$, which we define as:

$$\bar{R}_{\text{rel}} \coloneqq \bar{R} - \bar{R}_{\text{id}}, \tag{3.13}$$

where $\bar{R}_{\text{id}}$ indicates the infidelity evaluated in the ideal noiseless case.

3.2 Results

3.2.1 VQE

Analytically we calculate the exact ground state energy to be $E_{\text{gs}} = -\sqrt{5} \approx -2.236$, and the concurrence to be $Q_{\text{gs}} = \frac{1}{\sqrt{5}} \approx 0.447$. We verified that the numerical simulations with our used ansätze reproduce these analytical results exactly for $\gamma = 0$. The outcome of the noisy VQE simulation is shown in Fig. 3.4 for all state preparation circuits of Fig. 3.1. Straight away, we notice clear differences between the noise channels. In all measures of state quality, phase and amplitude damping channels are the least destructive, while the symmetric depolarising channel has a much more dramatic effect.

We perform a comprehensive search of the parameter space, allowing us to identify

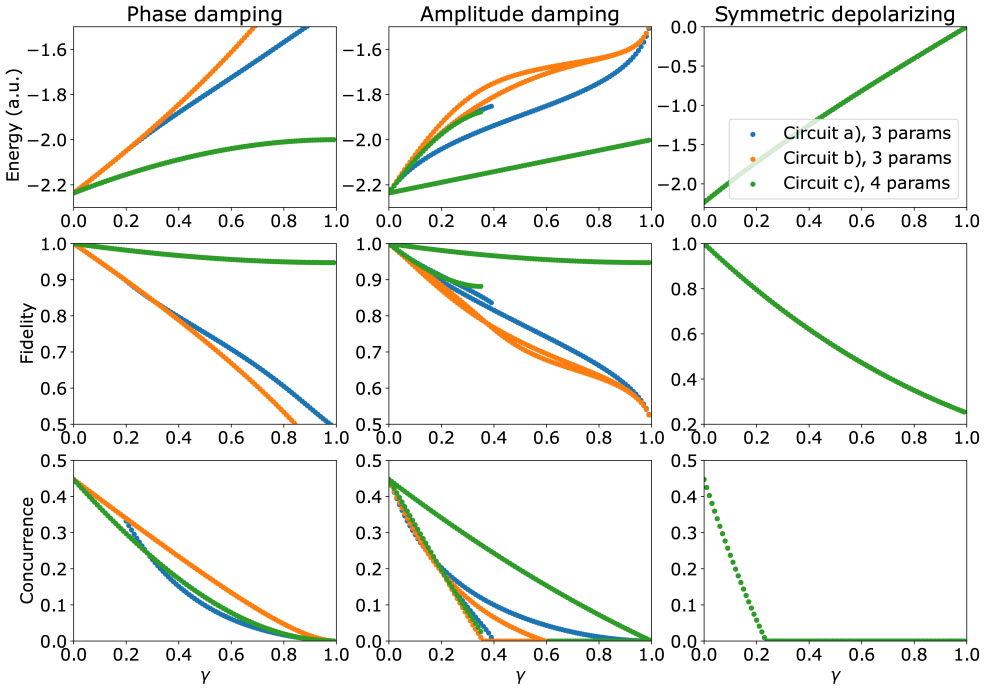


Figure 3.4: Energy, fidelity and concurrence as function of the noise parameter γ for the all energy minima found for the 2-qubit circuits with 3 and 4 gates (Fig. 3.1), obtained applying the three different indicated noise types. We plot all the obtained local minima in the energy landscape, so that at a given value of γ there can be multiple points for the same circuit. For symmetric depolarising noise there is no circuit-dependent difference, and hence only one curve is visible as all curves overlap.

all the local minima in the energy landscape, which we represent by plotting multiple points for the same value of γ. The number of local minima also shows a clear dependence on the noise type. Amplitude damping noise reveals local minima that branch out at low noise levels, resulting in multiple points for the same noise value. Additionally, the number of minima depends on the circuit: although difficult to deduce from the figure, by numerically analyzing the results we find two solutions for circuits (a) and (b), and three solutions for the 4-parameter circuit. Symmetric depolarising and phase damping noise instead present a single global energy minimum.

While for symmetric depolarising noise there is no dependence on the circuit, for phase and amplitude damping noise the quality of the prepared state depends strongly on the used circuits. The fact that circuits (a) and (b) give different results shows that in presence of noise, shifting a rotation gate from one qubit to another can improve the quality. Furthermore, we consistently see that one of the solutions of the 4-parameter circuit is significantly better than any solution in the 3-parameter circuits, in all measures of state quality. This indicates that for this system the over-parameterised circuit with one redundant angle of rotation exhibits improved capabilities of noise mitigation, or equivalently a higher noise resilience.

To determine how these findings generalise to higher qubit counts we perform the same analysis on the four-qubit system given by H_4, which also has ground state energy $E_{gs} = -\sqrt{5}$. As explained in Sec. 3.1.2, since we are dealing with a system larger than 2 qubits we employ the maximum pairwise concurrence as a measure of entanglement. We find that the maximum pairwise concurrence of the new ground state is $Q_{\text{gs}} = \frac{2}{\sqrt{5}} \approx 0.894$.

The results for maximum concurrence are shown alongside the results for energy and fidelity in Fig. 3.5. We see that the effect of the three types of noise channels on the state optimisation is very similar to the 2-qubit case. In particular, amplitude and phase damping have a less destructive impact than symmetric depolarising. Furthermore, the 4-qubit system exhibits multiple local minima at low noise for amplitude damping that are absent for phase and symmetric depolarising noise. However, compared to the 2-qubit system, the number of local minima is now larger and their appearance and

disappearance more irregular.

3.2.2 Random target states

Initially we consider the noise levels $\gamma \in \{0.1\%, 1\%, 5\%, 10\%\}$, which span the range of noise found in current devices [DH21]. The results are shown in Fig. 3.6. In the figure we present the average optimal fidelity for each γ for the three types of noise channels, where the vertical bars indicate the standard deviation over the ensemble. For comparison, we also plot the result for the noiseless case ($\gamma = 0$). We consider the two separate cases of noiseless training with noisy evaluation (non-reoptimised), and noisy training with noisy evaluation (reoptimised). By construction the reoptimised results are always better or equal to the non-reoptimised ones, as the former will take into account any noise-induced change in the cost function landscape. Comparing both results thus gives an insight on the degree to which noise affects the landscape, and on to what extent a variational algorithm can compensate for it. From Fig. 3.6 we can see that the reoptimised results in general improve significantly on the non-reoptimised ones.

We can observe that for $L \geqslant 4$ the noiseless fidelity is maximised for all the target states and equal to one, with zero standard deviation, showing that for such overparameterised circuits any target state can be essentially exactly prepared. For $L < 4$ on the other hand the circuit does not reach all target states even without noise, and the standard deviation increases as L decreases. As expected, the addition of noise further reduces the average fidelity in all plots, with a larger noise level resulting in bigger deviation from the noiseless fidelity. For the non-reoptimised case, with all types of noise the fidelity reaches a peak in L, corresponding to the optimal circuit depth in presence of noise. Interestingly, in the case of phase and amplitude damping noise with reoptimisation, the fidelity continues to increase with the number of layers, even for high noise levels.

We also explore noise down to $\gamma = 10^{-4}$, which is representative of the higher quality end of current quantum devices [DH21]. Here we directly compare the relative infidelity as a measure of the effects of noise only, as a function of layers. The results for phase

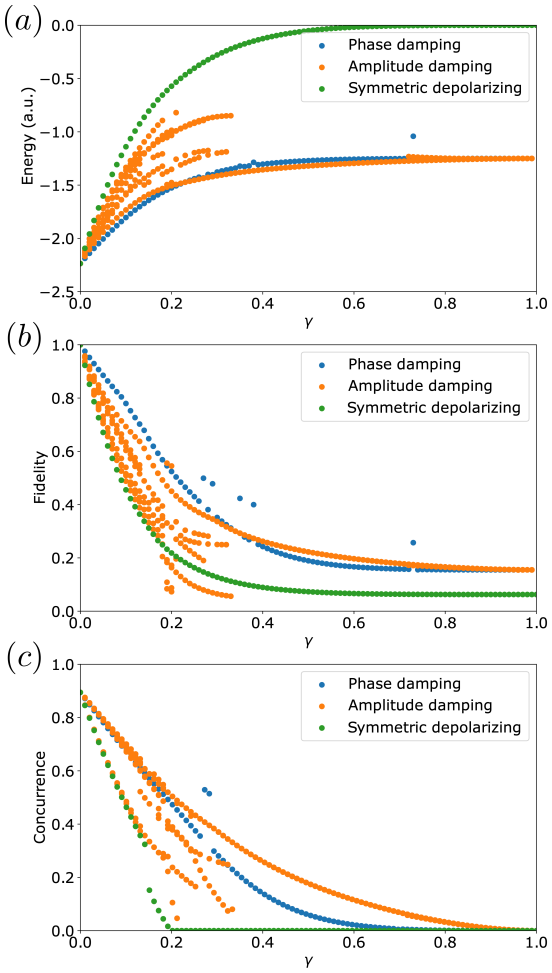


Figure 3.5: Energy, fidelity and concurrence as function of the noise level γ for all the energy minima found for the 4-qubit VQE simulations with the circuit shown in Fig. 3.2, obtained applying the three different indicated noise types. We plot all the obtained local minima in the energy landscape, so that at a given value of γ there can be multiple points for the same circuit. Some points are missing due to imperfect optimisation, and due to the fact that local minima can appear and disappear for increasing noise levels.

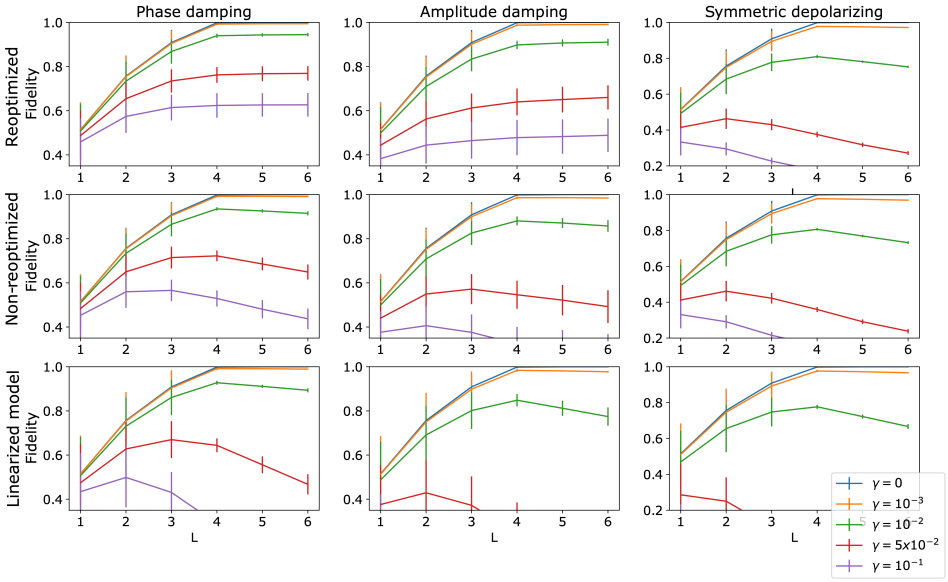


Figure 3.6: Fidelity vs number of layers, L, at realistic noise levels, for the circuit ansätze illustrated in Fig. 3.3 ($\gamma = 0\%$ (blue), 0.1% (orange), 1% (green), 5% (red), 10% (purple)). The plots in the first column are for phase damping noise, the second column plots are for amplitude damping noise, and the third column plots are for symmetric depolarising noise. The top row of plots are for non-reoptimised parameters, in the second row of plots the rotation parameters are reoptimised at each noise level, and the in the third row of plots the linear noise model results are presented. Each point shows the average over 1000 target states, and the vertical bars at each point indicate the standard deviation.

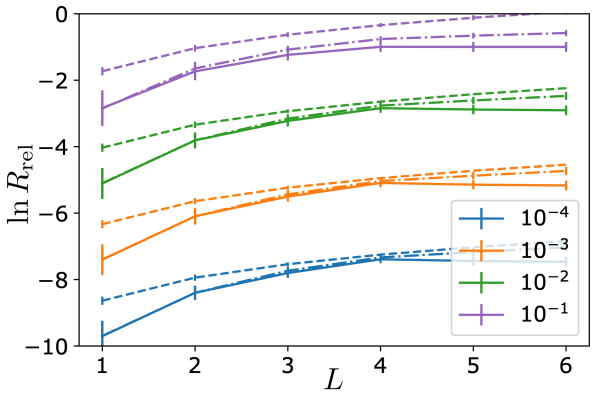


Figure 3.7: Relative infidelity as function of number of layers, L, for different levels γ of phase damping noise. The solid curves are for noise-aware reoptimised parameters, the dash-dotted curves are for parameters fixed at their values optimised in absence of noise (non-reoptimised), and the dashed curves indicate the results of the linear noise model. Each point shows the average over 1000 target states, and the vertical bars at each point indicate the standard deviation.

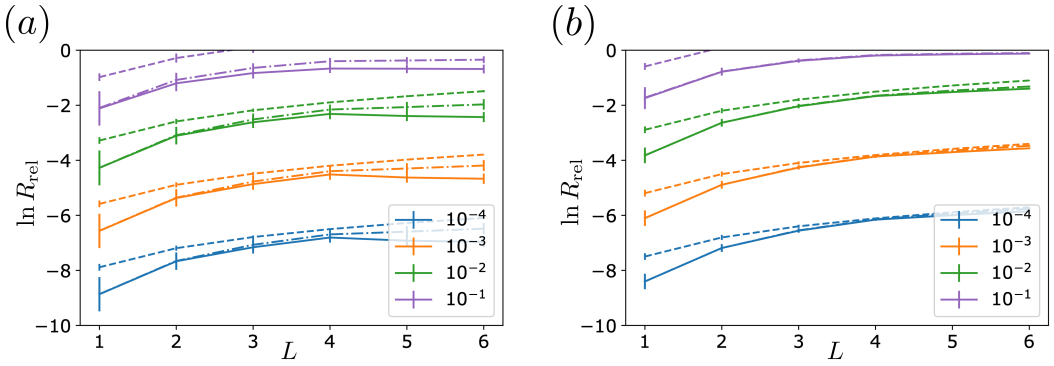


Figure 3.8: Relative infidelity vs layers, for a) amplitude damping noise and b) symmetric depolarising noise, for different values of the noise level γ ($\gamma = 10^{-4}$ (blue, lowermost), 10^{-3} (orange, third from top), 10^{-2} (green, second from top), 10^{-1} (purple, uppermost)). The solid curves are for noise-aware reoptimised parameters, the dash-dotted curves are for parameters optimised in the absence of noise (non-reoptimised), and the dashed curves indicate the results for the linear noise model. Each point shows the average over 1000 target states, and the verical bars at each point indicate the standard deviation.

damping noise are shown in Fig. 3.7. The relative infidelity reaches a peak at $L = 4$ for the reoptimised case, while for the non-reoptimised case it increases monotonically.

3.2.2.1 Noise-induced state transitions

In the energy-based optimisation (Sec. 3.2.1) we observed that under symmetric depolarising and amplitude damping noise, there exists a threshold $\gamma < 1$, past which the algorithm converges to a non-entangled state, corresponding to an undesired noise-induced transition. It is important to verify whether a similar phenomenon appears for the fidelity maximisation with general target states, as to determine that it is a general phenomenon in VQA optimisation and not specific to VQE. We therefore choose a target state at random from the distribution, and plot the fidelity of the output state after optimisation as a function of noise. The results are shown in Fig. 3.9 for phase damping noise, $\gamma \in [0, 0.1]$ and $L = 3$. In Fig. 3.9a we show fidelity and concurrence, and in Fig. 3.9c we show a representative subset of the parameters optimised at each γ. In Fig. 3.9b we show fidelity and concurrence evaluated without noise, but with the circuit rotation parameters optimised with noise (Fig. 3.9c).

We indeed observe a noise-induced transition in the example considered as a discontinuity in the concurrence values (red line in Fig. 3.9a), which appears at a much lower noise level threshold ($\gamma \approx 0.04$) than in the VQE simulations. This transition is also visible in the converged parameter values (Fig. 3.9c) and the resulting quality measures evaluated for those parameters without noise (Fig. 3.9c). However, the state after the transition is still entangled, and indeed the concurrence behaves in a nontrivial way, highlighting that the transition is more complex in the general case. We note that the detailed behaviour depends on the specific target state, and other examples of such transitions display different behaviours, ranging from sharp thresholds to smoother transitions (see Appendix A.2). There are also target states for which no well-defined transition can be observed.

From the point of view of the cost function landscape, is likely that such transitions are caused by the increased noise shifting the position and the energies of the local minima, or creating new local minima, such that optimiser converges to a completely different set of parameters.

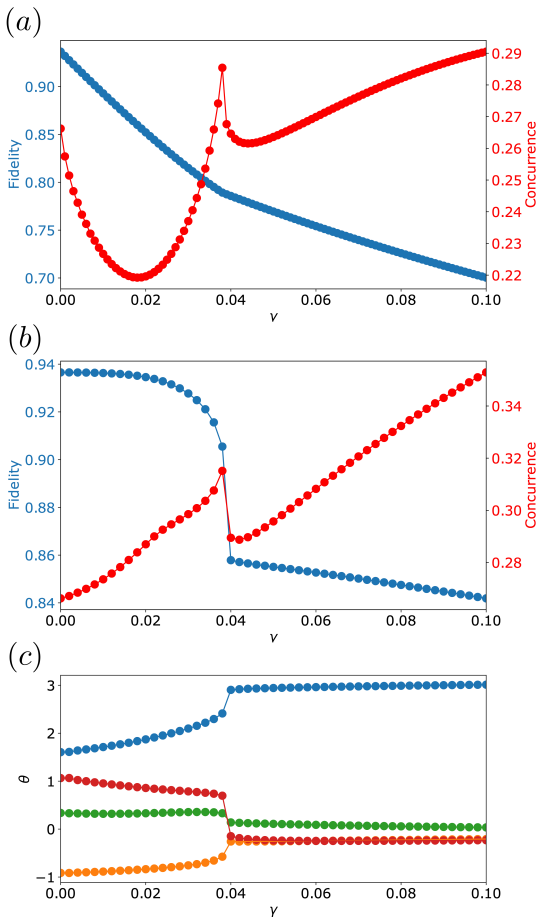


Figure 3.9: Evolution of fidelity and concurrence at convergence for fidelity optimisation at different strengths γ of phase damping noise, for a single random target state at $L = 3$. Four indicative optimised rotation angles are shown in the bottom panel (c) to illustrate the evolution of the optimal parameters with increasing noise. The resulting fidelity and maximum pairwise concurrence are shown in the top panel (a). The central panel (b) shows the measures evaluated without noise for the angles optimised at each γ value. A discontinuity in the slope of both state quality measures and converged angles is found at $\gamma \approx 0.04$, which therefore corresponds to the threshold γ value for this state and circuit, above which linear extrapolation of the properties to zero noise is not possible.

3.2.2.2 Linearised model

To provide further insight in our numerical results for optimisation with random target states, here we present a model for the infidelity at small γ. In Appendix A.1 we formulate a model that approximates noise propagation as linear, in the sense that each noise channel contributes an additive factor to the final relative infidelity. This comes from taking a first-order expansion of a noise channel's output in the noise parameter, and gives a formula for infidelity that is consistent with the assumption that fidelity is multiplicative, common in the literature [CDWE19, AAB+19, ZSW20]. The resulting formula for the fidelity depends however on the precise states acted by each noise channel; furthermore, we are interested in the infidelity averaged over an ensemble of circuits, each optimised towards a random target state. We thus make the simplification that the state at every layer can be described with the same distribution of target states. This is clearly inaccurate, however as we will see it leads to an effective model at low noise and large depth.

Overall the linearised model gives an estimate for the average relative infidelity and its variance as:

$$\bar{R}_{\text{rel}} \approx \alpha\,\gamma\,d, \tag{3.14}$$

$$\Delta^2_{\text{rel}} \approx \beta\,\gamma^2 d^2, \tag{3.15}$$

where α and β are constants obtained from the target state distribution and the noise channel, and d is the number of noise channels, which for our ansatz is $d = 2L$. In Appendix A.1 we provide a detailed description including the procedure for the calculation of α and β, together with their numerical values obtained for our systems.

We plot the expectations of the model for the fidelity in Fig. 3.6 (bottom row of panels), and for the relative infidelity in Fig. 3.7. Overall the model captures the numerical trends rather well. As expected from the model being a linear approximation, it describes better the behaviour of the fidelity at low noise levels, while for higher noise levels the deviations compared to the numerical results become larger. Since the model is formulated under the assumption of no noisy optimisation, it matches the

non-reoptimised results much better than the reoptimised ones. In the latter case, the trend is similar until $L < 4$, however it diverges significantly for larger L, since the model does not include the improvement of fidelity due to parameter reoptimisation with the number of layers for phase and amplitude damping noise.

A further effect of the noisy optimisation that the linear model cannot capture is the sudden transition to more noise-resilient set of parameters above a noise threshold, as observed in Fig. 3.9. This suggests the existence of an upper bound to the noise level that such simplified models can adequately describe: past the transition, the noisy state fidelity and energy cannot be extrapolated back to the noiseless state, as it can be seen from the plots. However, as these transitions happen at different noise values depending on the circuit, the noise and the loss function, we expect that the precise upper bound will be highly dependent on the setting. For noise mitigation techniques that extrapolate finite noise data down to the zero noise level like ZNE [TBG17, LB17], this implies that noise data needs to be collected below this critical threshold to avoid extrapolating to potentially undesired zero noise states. This is because these techniques rely on similar assumptions to our linear model such as smoothness of the energy curves with changing levels of noise, which is clearly violated at the transitions.

It should be noted that the model consistently overestimates the infidelity, especially for amplitude damping, at larger noise levels and for low depth circuits. For large noise levels the discrepancy is likely due in part to higher-order contributions. In the case of amplitude damping, it is possible that the linearised approximation is not appropriate since the channel is more complex. However it is certainly the case that the assumption that the state at every layer is identically distributed does not align with reality, and this may play a major role in the discrepancy. Also note that the estimate for the variance is clearly much lower than what is observed, which is also likely affected by this assumption. The model could certainly be made more complete by considering different states in different layers, but we leave this to future work as the model is sufficient for the present discussion.

3.3 Overview of results

3.3.1 VQE

Overall, the 2-qubit system and the 4-qubit system show a similar behaviour with respect to the effect of the three types of noise channels, in all the measures of state quality. Consistently, the most destructive noise channel is the symmetric depolarising channel, where energy, fidelity and concurrence rapidly move away from the exact values. In both simulations the concurrence falls to zero at $\gamma \approx 0.2 - 0.3$. In contrast, the other two noise channels concede a degree of robustness, yielding better performance even at high noise. We note that this is likely not due to the ground state being close to a computational basis state, since the states are chosen at random. Nevertheless, as noise rises towards $\gamma = 1$ all circuits eventually tend towards unentangled states. We note that for this system the concurrence is a much more stringent quality criterion than the energy and fidelity, since it inevitably goes towards zero for all circuits at $\gamma = 1$, while in some cases energy and fidelity only deviate by 10-20% from the exact value at maximum noise. For example, for 2-qubits and complete dephasing, the energy is only about 10% higher than the noiseless value. This shows that the energy alone can be a deceptively poor quality measure for quantum algorithms.

Another common feature of the VQE experiments is the presence of multiple local minima when amplitude damping noise is present in the circuit, which are visualised as multiple lines branching off from the same point at zero noise as the noise parameter is increased. For some of these solutions the measures of quality decrease less with noise, and hence appear to be more resilient than others. The damping noise models bring to light differences between otherwise equivalent circuits at zero noise. All circuits considered for the 2-qubit case (3.1) perform indistinguishably with respect to symmetric depolarising noise. Under the two damping noise channels, however, the three circuits are affected differently, with circuit (c) appearing to perform considerably better than circuits (a) and (b). Importantly, the different resilience to noise found for circuits (a) and (b) for the 2-qubit case shows that the circuit configuration needs to be optimised for maximisation of noise resilience. We expect that this optimisation of gate placement

is even more important when the noise level varies across qubits. Our results also show that the inclusion of redundant parameters can further improve the quality of the final state.

3.3.2 Random state optimisation

For the non-reoptimised case, the fidelity reaches a peak in L, which depends on the noise level, signaling the point where the noise from the increased number of noise channels overcomes the improvement in the accuracy of the circuit with the additional parameters. In contrast, in the case of phase and amplitude damping noise with reoptimisation the fidelity continues to increase with the number of layers even for high noise levels. Past $L = 4$ the reoptimised simulations show a noticeable improvement in noisy fidelity compared to the non-reoptimised simulations. This depth threshold is significant, as it marks the point past which the circuit can perfectly reproduce all target states at zero noise. Any additional layer beyond $L = 4$ therefore does not contribute to the noiseless fidelity and only introduces redundancy. In the noisy case this overparameterisation leads to improved resilience for amplitude and phase damping noise, while for symmetric depolarising noise no improvement is found. This is analogous to what found for the VQE simulations and points to a general phenomenon where redundancy enables the optimiser to find better solutions under asymmetric noise channels. Note that we do not expect the improvement to continue to arbitrarily large number of layers, as presumably the capability of overparameterisation to minimise the effect of noise is limited and eventually the latter dominates. More research on real hardware is needed to understand the practical limits of noise-aware optimisation.

Improved state quality upon parameter reoptimisation in quantum algorithms has previously been reported in Refs. [SSMAG16] and [MKSCD17], where the authors find that phase and amplitude damping noise generally impact state preparation in VQE less than symmetric depolarising noise, with phase damping noise being the least impactful. In particular, Ref. [MKSCD17] studies how reoptimisation under noise significantly improves the results of a VQE for a chemical problem. However, their model applies noise after the state has been prepared, as opposed to the interleaved noise model

proposed here.

3.4 Conclusions

We studied the effects of different types and levels of quantum noise on VQAs, by simulating small-scale VQE and random target state fidelity maximisation. For the VQE simulations we used three measures of state quality, namely energy, fidelity and entanglement, and showed that the energy alone can be a deceptively poor quality measure for quantum algorithms if the goal is not merely to produce a low-energy state but an entangled state with correlation to the true ground state of interest. When maximising the fidelity of the state produced with a given ansatz with a target state, we found that in presence of symmetric depolarising noise there is a circuit depth at which fidelity is maximised. For amplitude and phase damping noise, and for the considered circuit depths, noise-aware parameter reoptimisation allows to progressively improve the fidelity as the circuit depth is increased.

Overall we found symmetric depolarising noise to be the most detrimental, while for amplitude and phase damping noise it is possible to mitigate the effects of noise by optimised gate placement, overparameterisation and noise-aware reoptimisation. We obtained these results consistently across our considered systems, a 2- and 4-qubit simulation for a specific Hamiltonian, and 4-qubit simulations for general target states.

We uncovered new, previously unobserved phenomena. We noticed that amplitude damping noise, but not symmetric depolarising or phase damping noise, led to multiple minima with different characteristics, which is not observed in the noiseless case. Furthermore it appears that redundancy in the circuit, for instance by including parameterised rotation gates which do not improve the expressivity, in the presence of noise can be exploited by the optimiser to reach better solutions. Interestingly this is observed to occur for amplitude and phase damping, but not symmetric depolarising noise. The explanation of these phenomenona will be the focus of the next Chapter.

Finally, for a number of target states and circuits there seems to be a noise threshold above which the states produced by the circuit have largely different physical properties from the true target state, especially entanglement as measured by concurrence. For

practical applications it is critical to ensure that noise levels are below this threshold to prevent convergence to undesired solutions. As such it would be valuable to be able to predict the noise range at which these transitions happen. However at least with our models there appears to be too much variability across different circuits, and even the same circuit under a different cost function, to be able to formulate a conclusive theory.

Chapter 4

Theoretical analysis of noisy optimisation

Starting from some observations made in the previous Chapter on the effects of noise on the converged optima of VQAs, in the first part we formulate a theory of parameter symmetries and how they are affected by noise. Specifically we construct an algorithm to enumerate all discrete symmetries of certain VQAs, and prove that while unital channels preserve these symmetries, nonunital channels do not. In the second part, we discuss a more general phenomenon caused by decoherence, the vanishing of variations in the cost function, or noise-induced barren plateaus. This phenomenon applies to many NISQ algorithms and therefore carries large practical relevance for non-error-corrected quantum computers.

The first part of this Chapter is based on Ref. [FCA+22]. The second part is based on Ref. [WFC+21], focusing especially on the result for measurement noise where the Author made his largest contribution to the paper.

4.1 Overview

In the previous Chapter, upon introducing a noise model in two VQA models we experimentally witnessed some phenomena which may have consequences for the practical implementation of these algorithms, and therefore deserve further study. To briefly summarise, we saw that functionally equivalent ansätze (in the sense of which unitaries they can express) under amplitude and phase damping noise can display different behaviour under optimisation. In particular overparameterised ansätze seemed to be more noise resilient that would otherwise be expected. Interestingly this behaviour was not observed for symmetric depolarising noise. In addition, focusing on a single ansatz, under amplitude damping noise only, many new local minima of different cost function values appeared, signaling that different output states were being expressed. Both these phenomenona appeared with minimal noise, become more apparent with increased noise, while they were absent in the noiseless case.

A natural hypothesis from these observations is that there are parameter configurations that yield equivalent states in the absence of noise, but that give different states when certain kinds of noisy channels are introduced. We refer to this phenomenon as *parameter symmetry breaking*. This can lead to spurious local minima if the original state was a local minimum of the cost function, or to improved noise resilience if they give better paths through parameter space that the optimiser is able to exploit. In this Chapter we shall demonstrate that these phenomena do in fact occur and most likely underlie the observations.

Finally, we also give a mathematical explanation of an observation which is simple but has powerful implications for the scalability of VQAs in noisy devices. This is the evident fact that with increased noise, in the form of more decoherent channels or more layers of decoherent channels, comes a suppression of the fluctuations of the cost function and a decay towards a constant value. This phenomenon is known as *noise-induced barren plateaus* (NIBP). We report the results from the original paper Ref. [WFC$^+$21] in detail and go through the proof for measurement noise, which was the author's main contribution.

4.2 Noise and parameter symmetries

In this section we show that the noise-induced phenomena discussed above, namely the presence of multiple local minima for amplitude damping and the differing performance of functionally equivalent ansätze with amplitude and phase damping, arise due to symmetries of parameterised quantum circuits.

Let us first define parameters for a PQC:

Definition 8 (Symmetric parameters)**.** *Let $U(\boldsymbol{\theta})$ be a PQC. We say that two distinct sets of parameters $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ are symmetric if $U(\boldsymbol{\theta})$ is equal to $U(\widetilde{\boldsymbol{\theta}})$ (up to a global phase).*

Now we can define parameter symmetries as maps that relate symmetric parameters:

Definition 9 (Parameter symmetries)**.** *A parameter symmetry of a PQC is any map $f : \Theta \mapsto \Theta$ such that $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}} = f(\boldsymbol{\theta})$ are symmetric for all $\boldsymbol{\theta} \in \Theta$.*

For a given circuit we may have sets of parameter symmetries. In case a set is countable, we say that the circuit has *discrete parameter symmetries*. Conversely, if the members of the set vary continuously across all the set, then the circuit has *continuous parameter symmetries*.

A circuit with parameter symmetries will feature symmetries of the cost function in the parameter space, since symmetric sets of parameters must yield the same value of the cost function. Therefore, the presence of parameter symmetries implies the existence of multiple identical minima in the noiseless cost function. In the case of discrete symmetries the minima are distinct and separated in the cost function landscape, while for a continuous symmetry the minima are connected and may be visualised as a valley in the landscape. If one introduces a small state-dependent disturbance in the circuit, which breaks the symmetry between parameter symmetric states, such as for specific noise channels, this leads to different states being produced by the circuit. The symmetric global energy minima for the noiseless case then splits in local minima with different energies.

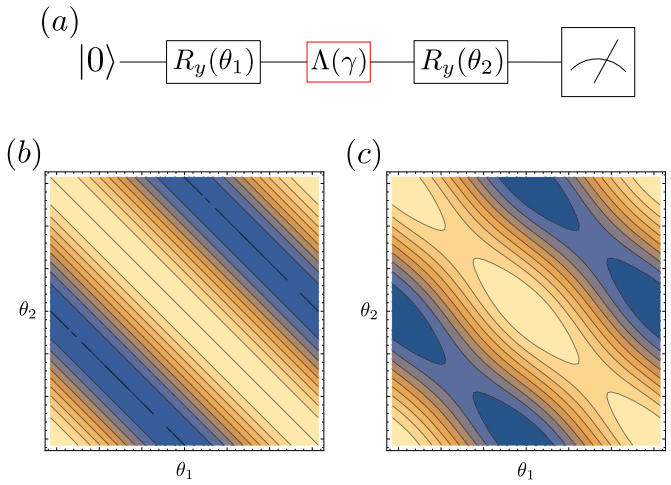


Figure 4.1: Illustrative example for a continuous parameter symmetry for an overparameterised circuit, where the second rotation by θ_2 is redundant in the absence of noise; the plots in (a) and (b) show the parameter space landscape of the cost function $C(\theta_1, \theta_2) = \mathrm{Tr}[\rho(\theta_1, \theta_2)\,|0\rangle\langle 0|]$. In (a) the noiseless results are shown, and in (b) a phase damping channel is added between the Ry gates, with $\gamma = 0.4$. When noise is added, the continuous parameter symmetry is broken, so that the valley in (a) is replaced by a set of minima in (b).

4.2.1 Continuous symmetries

Continuous parameter symmetries provide a useful framework for explaining the observed improved resilience to noise when including redundant rotation gates. Overparameterised quantum circuits automatically have continuous parameter symmetries, since any variation in the redundant parameter can be compensated by modifying the remaining parameters accordingly in order not to change the final state. As illustrative example circuit we consider the placement of two identical single-qubit rotation gates next to each other, shown in Fig. 4.1. As cost function we consider the overlap with the $|0\rangle$ state: $C(\theta_1, \theta_2) = \mathrm{Tr}[\rho(\theta_1, \theta_2)\,|0\rangle\langle 0|]$. The computed cost function exhibits a 1D family of continuous parameter symmetries $\{f_\alpha\}$ in the noiseless case (Fig. 4.1(a)), given by:

$$f_\alpha((\theta_1, \theta_2)) = (\theta_1 + \alpha, \theta_2 - \alpha) \mod 2\pi, \quad \alpha \in [0, 2\pi]. \tag{4.1}$$

When a phase damping noise channel is added between the two rotations, the continuous parameter symmetries are broken, and there now exists a discrete set of minima

(Fig. 4.1(b)). For such cases re-optimisation of the parameters obtained for $\gamma = 0$ for the noisy case will generally lead to improved energies. Since at zero noise the minima exist as a continuous valley, the addition of noise introduces a gradient in this valley, which the optimiser can exploit to arrive to a better solution.

The improved noise resilience due to a redundant parameter can therefore be explained as resulting from the ability of the circuit to explore more state preparation paths in parameter space when compared to a circuit with fewer parameters. It can therefore reach additional paths that are potentially less affected by the noise. Notice that in our toy model above, the parameter symmetries are preserved by symmetric depolarising noise. In general, for n qubits any global symmetric depolarising channel with action:

$$\mathcal{N}(\rho) = p\frac{\mathbb{1}}{2^n} + (1-p)\rho, \tag{4.2}$$

can always be commuted to the end of the circuit and therefore it cannot break continuous symmetries. However, *local* symmetric depolarising channels like the ones used in the experiments are able to break the symmetries. The fact that we did not observe significant noise resilience in that case may be related to the uniform flattening of the cost function with noise, which is the NIBP phenomenon that we will explore later.

4.2.1.1 Relation to Quantum Fisher Information

So far we gave a toy example of such symmetries. since a circuit may be overparameterised in multiple different ways it is difficult to find an overarching mathematical treatment. However there exists an interesting approach to understand the interaction of quantum noise and overparameterisation that goes via the Quantum Fisher Information Matrix (QFIM) [LYLW20, ASZ+21, HBK21].

Given a parameterised (mixed) state $\rho_{\boldsymbol{\theta}} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$ such as those output by (noisy) PQCs, the QFIM is defined as [LYLW20]:

$$F_{ab}(\rho_{\boldsymbol{\theta}}) := \sum_{\lambda_i+\lambda_j\neq 0} \frac{2\mathrm{Re}(\langle\lambda_i|\frac{\partial\rho}{\partial\theta_a}|\lambda_j\rangle\langle\lambda_j|\frac{\partial\rho}{\partial\theta_b}|\lambda_i\rangle)}{\lambda_i+\lambda_j}, \tag{4.3}$$

where the sum is over pairs of eigenvectors of $\rho_{\boldsymbol{\theta}}$ in which at least one has a nonzero

eigenvalue. Intuitively, the QFIM measures how an infinitesimal change in the parameters leads to a change in the corresponding quantum state. In particular, the rank of the QFIM gives the dimensionality of the state space that can be explored at that particular parameter setting. Thus, an overparameterised PQC would lead to the QFIM being rank-deficient (rank less than maximum, i.e. its dimension) for any value of the parameters. Importantly for our analysis, in Ref. [GMLC24] it has been observed that the addition of noise within a circuit can increase the rank of the QFIM, opening up new directions in state space.

This can be seen as a local perspective of the phenomenon of continuous parameter symmetry breaking. A somewhat imprecise argument goes as follows. Assume that we have a continuous symmetry given by the continuous set $\{f_\alpha\}$ with parameter α determining the symmetry, and assume that $f_0 = \mathrm{id}$. If we fix a parameter vector $\boldsymbol{\theta}$ then $U(\boldsymbol{\theta}) \approx U(\boldsymbol{\theta} + d\boldsymbol{\theta})$ (up to global phase) with:

$$d\boldsymbol{\theta} = \frac{\partial f_\alpha(\boldsymbol{\theta})}{\partial \alpha}\Big|_{\alpha=0} d\alpha, \tag{4.4}$$

with $d\alpha \in \mathbb{R}$ infinitesimally small. The PQC output state will approximately (to first order in dx) be unchanged upon shifting the parameters by $d\boldsymbol{\theta}$, and thus the QFIM must be rank-deficient. Then if noise makes the QFIM full rank at $\boldsymbol{\theta}$ (as observed in Ref. [GMLC24]) it must be that the continuous symmetry is broken.

By the same argument, the presence of s independent continuous 1D parameter symmetries requires the QFIM rank everywhere to be at most $m - s$, where m is the number of parameters. We remark that due to the locality of the QFIM, it is possible that in some locations the rank is smaller than the bound even in the absence of symmetries: the QFIM is an imprecise measure of symmetries in the landscape. However the fact that it is observed to become full rank with noise provides a theoretical explanation for continuous symmetry breaking.

4.2.2 Discrete symmetries

By analysing the parameters for the equivalent minima for the noiseless case in our 2- and 4-qubit VQE simulations, we find that they obey fixed relations consisting of shifts of the angles by π and inversions. Furthermore, we find numerically that phase damping and symmetric depolarising channels preserve these symmetries, while amplitude damping channels can break them. This is thus consistent with the multiple noise induced minima appearing for amplitude damping noise. We will show that this is a common feature in PQCs, especially a type of PQC which we call *buffered*, by constructing a method to move between symmetric parameters.

4.2.2.1 Buffered PQCs

Given a PQC $U(\boldsymbol{\theta})$, we define its buffered version as follows.

Definition 10 (Buffered PQC)**.** *We define the buffered version of a PQC $U(\boldsymbol{\theta})$ as a gate sequence:*

$$U_B(\boldsymbol{\theta}, \boldsymbol{\gamma}) = V_B(\boldsymbol{\gamma}) U(\boldsymbol{\theta}) \,, \tag{4.5}$$

where $V_B(\gamma)$ is the so-called buffer unitary given by:

$$V_B(\boldsymbol{\gamma}) = \bigotimes_{j=1}^{n} e^{-iX_j \gamma_{2j}} e^{-iY_j \gamma_{2j-1}} \,. \tag{4.6}$$

As shown in Fig. 4.2, the buffer unitary is simply given by a tensor product of single qubit rotations around the x and y axes which are parameterised by the vector $\boldsymbol{\gamma}$ of length $2n$.

We note that our main results, stated below, are valid for buffered PQCs. However, since $U(\boldsymbol{\theta}) = U_B(\boldsymbol{\theta}, \mathbf{0})$ one can always trivially extend any PQC to its buffered version. That is, any PQC can be considered as a buffered PQC with trivial rotation angles in the buffer unitary. In addition, we remark that our results will also hold for any PQC where $U(\boldsymbol{\theta})$ contains (at least) two single-qubit rotations about different axes in every qubit (not necessarily sequentially or in parallel). However, for the sake of simplicity in introducing the method, we consider the case where one appends a buffer unitary to

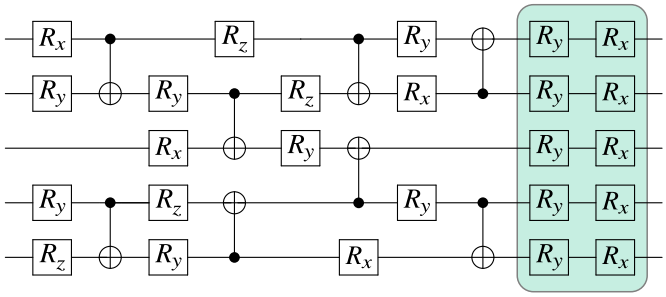


Figure 4.2: Schematic representation of a buffered PQC. The buffered version of any parameterised quantum circuit $U(\boldsymbol{\theta})$ can be obtained by appending to the circuit a buffer unitary as described in (4.6). Here we have highlighted this buffer unitary.

the PQC.

Finally, we remark that in some practical settings the buffer unitary can be implemented virtually without any additional computational overhead. Whenever $U(\boldsymbol{\theta})$ acts before measurement, the single-qubit rotations in the buffer layer can be absorbed into the measurement operator and executed classically by post-processing the measurement statistics. However, when the buffered unitary does not act prior to the measurements, such as when only a portion of the circuit is buffered, then $U_B(\boldsymbol{\gamma})$ must be included.

4.2.2.2 The σ-Pulse method

Recall from Definition (9) that symmetries occur when different sets of parameters $\boldsymbol{\theta}$ in a PQC can leads to the same unitary being produced. Let us here make two important remarks. First, note that the definition implies that the structure of the circuit remains unchanged between $U(\boldsymbol{\theta})$ and $U(\widetilde{\boldsymbol{\theta}})$, as no gates in the circuit are being added or replaced; only their parameter values differ. Second, we remark that these parameter symmetries naturally translate into cost function landscape symmetries. That is, given two symmetric sets of parameters $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ we have $C(\boldsymbol{\theta}) = C(\widetilde{\boldsymbol{\theta}})$.

Note that there are many mechanics which can lead to symmetries in $\boldsymbol{\theta}$. For instance, they can arise from the wrapping symmetry in a rotation, i.e., from the fact that for any single qubit rotation we have $R_\mu(\theta_\mu) = R_\mu(\theta_\mu + 2\pi)$. Similarly, parameter symmetries can also be obtained from other types of mechanisms, such as commutation symmetries. Consider for example a two-qubit PQC composed of a CNOT preceded

and followed by single qubit rotations about the z axis on the first qubit. That is, $V(\theta_1, \theta_2) = (R_z(\theta_1) \otimes \mathbb{1}_2) . C_X^{12} . (R_z(\theta_2) \otimes \mathbb{1}_2)$. Noting that $[(R_z(\theta_1) \otimes \mathbb{1}_2), C_X^{12}] = 0$, then it follows that $V(\theta_1, \theta_2) = V(p\theta_1 + (1-q)\theta_2, (1-p)\theta_1 + q\theta_2)$ for any θ_1 and θ_2, and for any $p, q \in (0, 1)$.

Here we provide a general theory to analyse a class of non-trivial discrete parameter symmetries in buffered circuits. Specifically we introduce a method for finding and characterizing the following symmetries:

Definition 11 (σ-Pulse symmetries)**.** *Let $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ be a buffered PQC as in Definition 10. $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ possesses a σ-Pulse symmetry whenever, for any $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$, there exists a set of $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ related to $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ via:*

$$\widetilde{\theta}_j = (-1)^{p_j}\theta_j + q_j\pi\,, \quad \widetilde{\gamma}_j = (-1)^{p'_j}\gamma_j + q'_j\pi\,, \tag{4.7}$$

for some $p_j, q_j, p'_j, q'_j \in \{0, 1\}$, such that $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ and $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ are symmetric.

.

We now present a method that allow us to start from a set of parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ of a buffered PQC and obtain a second set $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ which are σ-Pulse symmetric to $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ according to Definition 11. We call the procedure the *σ-Pulse method*. This method is based on three basic steps: (1) The creation of the so-called σ-Pulses, (2) The propagation of said pulses through the circuit, and (3) The absorption of the σ-Pulses in the buffer unitary. The mathematical rules for these steps (described below in detail) are schematically shown in the ZX-calculus [CD11] notation of Fig. 4.3(a).

The first step of the σ-Pulse method is based on the fact that any single qubit rotation around a principal axis satisfies the following identity:

$$\begin{aligned} R_\mu(\theta) &= e^{-i\theta\sigma_\mu/2} = e^{-i(\theta+\pi)\sigma_\mu/2} e^{i\pi\sigma_\mu/2} \\ &= R_\mu(\theta + \pi)\,(i\sigma_\mu) = R_\mu^{01}\,(i\sigma_\mu)\,, \end{aligned} \tag{4.8}$$

where we defined the shifted rotations:

$$R_\mu^{pq} = R_\mu\left((-1)^p\theta + q\pi\right)\,, \tag{4.9}$$

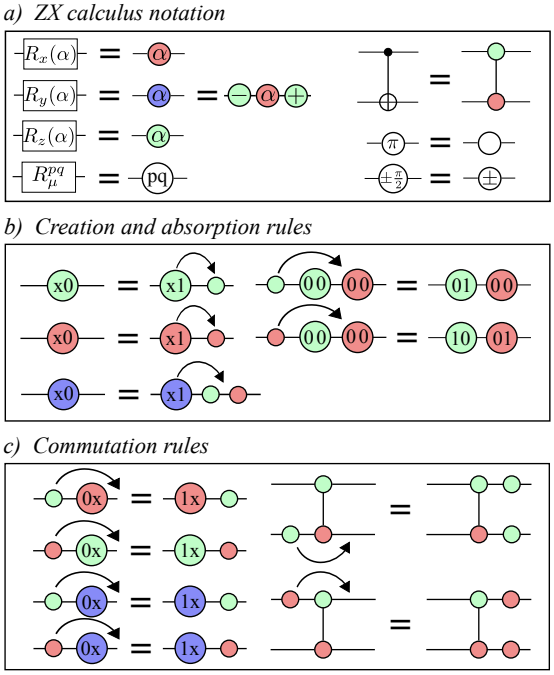


Figure 4.3: σ-Pulse method rules in the notation of ZX-calculus. (a) Notation for rotations around the principal axes, shifted rotations (of Eq. (4.9)), and for a CNOT gate. We additionally introduce notation for the σ-Pulses (rotation of angle π), and for rotations of angles $\pm\frac{\pi}{2}$. (b) Rules for creating and absorbing σ-pulses. (c) Non-trivial commutation rules for σ-Pulses. We remark that in panels (b) and (c), the rules fora $(i\sigma_y)$ pulse can be derived from those of $(i\sigma_x)$ and $(i\sigma_z)$. In addition, in those panels an equal sign indicates that the unitaries are equal up to a global phase.

for $p, q \in \{0, 1\}$ and $\mu \in \{x, y, z\}$. As shown in Fig. 4.3(b), Eq. (4.8) implies that the angle of a single-qubit rotation in $U(\boldsymbol{\theta})$ can be shifted by π at the expense of adding to the circuit a $i\sigma_\mu$ gate, i.e., at the expense of creating a σ-Pulse. When Eq. (4.8) is employed to generate a pulse, we say that the gate is a *generator* of a *primary* pulse. Note that simply employing (4.8) changes the structure of the ansatz as we have added new gates to the circuit. For the structure of $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ to be preserved, the σ-Pulses need to be propagated through the circuit towards $U_B(\boldsymbol{\gamma})$ where they be absorbed.

Propagating the σ-Pulses through the circuit implies knowing how $i\sigma_\mu$ commutes with all other gates in the ansatz. The commutation of a σ_μ-Pulse through a single qubit rotation $R_\nu(\theta)$ is given by:

$$R_\nu(\theta)\,(i\sigma_\mu) = i\sigma_\mu R_\nu\left((-1)^{\delta_{\mu\nu}}\theta\right) = i\sigma_\mu R_\nu^{\delta_{\mu\nu}0}\,. \tag{4.10}$$

Eq. (4.10) shows that if $\mu \neq \nu$, the commutation of a pulse with a rotation can lead to said rotation picking up a minus sign. Moreover, the following identities provide the commutation rules between a pulse and a CNOT:

$$\begin{aligned} C_X(i\sigma_\mu \otimes \mathbb{1}) &= -i(i\sigma_\mu \otimes iX)C_X, \quad \text{for } \sigma_\mu = Y, X \\ C_X(\mathbb{1} \otimes i\sigma_\mu) &= -i(iZ \otimes i\sigma_\mu)C_X, \quad \text{for } \sigma_\mu = Y, Z \\ [C_X, (iZ \otimes \mathbb{1})] &= [C_X, (\mathbb{1} \otimes iX)] = 0\,, \end{aligned} \tag{4.11}$$

These commutation rules, which we illustrate in Fig. 4.3(c), show that commuting a σ-Pulse on the control (target) qubit through a CNOT can lead to the creation of a *secondary* pulse on the target (control) qubit, plus a global unobservable phase. For the gate structure of $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ to remain unchanged, the secondary pulses also need to be propagated towards the measurement, which in turn means that they can create additional secondary pulses.

Once all the primary and secondary pulses have been propagated to the buffer unitary, they can be absorbed by shifting the rotation in $U_B(\boldsymbol{\gamma})$ angles as:

$$\begin{aligned} R_x(\gamma')R_y(\gamma)(iX) &= R_x(\gamma' - \pi)R_y(-\gamma) = -R_x^{01}R_y^{10}, \\ R_x(\gamma')R_y(\gamma)(iY) &= R_x(\gamma')R_y(\gamma - \pi) = -R_x^{00}R_y^{01}, \\ R_x(\gamma')R_y(\gamma)(iZ) &= R_x(\gamma' - \pi)R_y(\pi - \gamma) = -R_x^{01}R_y^{11}, \end{aligned} \tag{4.12}$$

where we use the definition of the shifted rotations R_μ^{pq} of (4.9). Here we remark that the minus signs on the right-hand side of (4.12) simply correspond to unobservable global phases. These absorption rules are shown in Fig. 4.3(b).

4.2.2.3 Parameter symmetries

Eqs. (4.10, 4.12) provide the framework for determining symmetries in $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ with the σ-Pulse method. Given a set of angles $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$, one can select any number of rotations to generate primary pulses. Once the primary and secondary pulses are propagated and absorbed in the buffer unitary, we define $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ as the ensuing new set of angles. From Eqs. (4.8) and (4.12) it is straightforward to see that $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ and $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ are symmetric

according to Eq (4.7) in Definition 11. In Fig. 4.4 we explicitly show this procedure.

Definition (11), and more specifically, Eq. (4.7), allows us to derive the following proposition:

Proposition 1 (Exponential Symmetry)**.** *Let $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma}) = V(\boldsymbol{\gamma})U(\boldsymbol{\theta})$ be a buffered PQC, where the original PQC $U(\boldsymbol{\theta})$ has M parameters. Then, for any $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ there exists 2^M sets of σ-Pulse symmetric parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ according to Definition (11). Each symmetric set can be characterised by a bitstring $\boldsymbol{\beta}$ of length M such that $\beta_j = 0$ if $\widetilde{\theta}_j \in [0, \pi)$ and $\beta_j = 1$ if $\widetilde{\theta}_j \in [\pi, 2\pi)$. We denote as $\mathcal{B}$ the set of such bitstrings.*

Proof. Let us first recall that, as mentioned in the main text, there are M parameters in the PQC $U(\boldsymbol{\theta})$. Moreover let us denote as $\mathcal{G}$ the set of all possible generator choices over the parameters in $U(\boldsymbol{\theta})$. The number of distinct symmetric sets of parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ obtained through the σ-Pulse method is given by the cardinality of $\mathcal{G}$:

$$|\mathcal{G}| = \sum_{\alpha=0}^{M} \binom{M}{\alpha} = 2^M \,. \tag{4.13}$$

Note that the shifts in the buffer layer parameters $\boldsymbol{\gamma}$ are entirely determined, via the σ-Pulse method, by the shifts in the parameters $\boldsymbol{\theta}$, and so they do not contribute to the number of symmetries. Similarly, we can count the number of symmetric sets of parameters as the number of bitstrings $\boldsymbol{\beta}$ in $\mathcal{B}$ of length M, which is precisely 2^M. □

In addition, Proposition 1 implies the next Corollary.

Corollary 1 (Effective Parameter Space Reduction)**.** *Let $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ be a buffered PQC. Then, for every $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ there always exists a set of parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ which are σ-Pulse symmetric to $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$, and which are such that $\widetilde{\theta}_j \in [0, \pi)$ for all j.*

To prove Corollary 1, we show that given any $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ one can always find a set $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ with $\boldsymbol{\beta} = \mathbf{0}$:

Proof. Let us assume that the parameters in $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ are order by layer, where a layer consists of quantum gates that can be performed in parallel and where the first layer contains the first gates in $U(\boldsymbol{\theta})$. In general we can assume without loss of

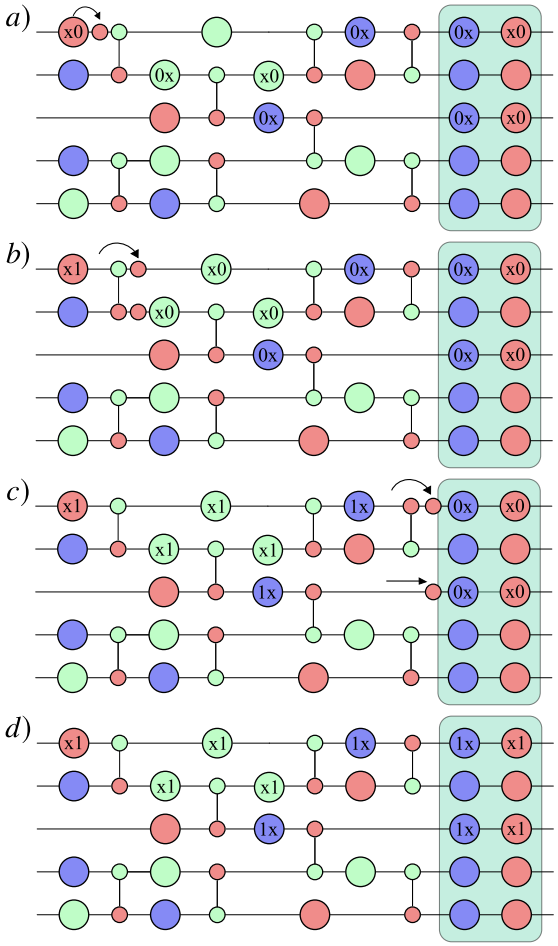


Figure 4.4: Schematic of the σ-Pulse method for finding parameter symmetries. We start with the buffered PQC of Fig. 4.2(a) with parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$. (a) A primary σ-Pulse is generated in the first R_x rotation acting on the first qubit. (b) While propagating the pulse through the CNOT gate, secondary pulses are created. (c) All the primary and secondary pulses are propagated to the buffer layer, where they can be absorbed. (d) Once the pulses are absorbed, we have a new set of parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$ which are symmetric to $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ according the Definition 11.

generality that the angles in $\boldsymbol{\theta}$ are in $[0, 2\pi)$. We now describe a sequential procedure that can be used to obtain the vector $\widetilde{\boldsymbol{\theta}}$ where every parameter $\widetilde{\theta}_j \in \widetilde{\boldsymbol{\theta}}$ not in the buffer layer are in the reduced domain $[0, \pi)$.

If $\theta_1 \in [0, \pi)$ then we do nothing, but if $\theta_1 \in [\pi, 2\pi)$ we create and forward propagate a σ-Pulse. According to Eq. (4.8), this will add π to θ_1, which maps it to the interval $[0, \pi)$. This procedure is then sequentially repeated for each parameter in $\boldsymbol{\theta}$ not in the buffer layer. We remark that since σ-Pulse propagate forward in the circuit, creating a pulse in θ_j does not affect any angle θ_k with $k < j$. Moreover, we know from (4.10) that as the σ-Pulses propagate they can add a minus sign to other angles in $\boldsymbol{\theta}$. Hence, if a given θ_j that was originally in $[\pi, 2\pi)$ picked up a minus sign then we do nothing as it will now be in $[0, \pi)$. On the other hand, if it was in $[0, \pi)$ we have to create a σ-Pulse to map it to $[0, \pi)$. Note that at the end of this procedure every parameter not in the buffer layer will be mapped to the reduced domain $[0, \pi)$. □

Proposition 1 has several important implications. First, it shows that the each point in the cost function landscape is exponentially periodic, as the unitary and therefore the cost function value is symmetric over the parameter translation $\theta_j \to (\theta_j + \pi)$ for every j. This holds up to a sign change $\theta_k \to -\theta_k$ in some other parameters with $k \neq j$ and a correction rotation in the buffer layer parameters $\boldsymbol{\gamma}$, due to the necessity to propagate and absorb the extra σ-Pulses. In particular, denoting as $\{\boldsymbol{\theta}_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ a set of parameters that minimise the cost $C(\boldsymbol{\theta}, \boldsymbol{\gamma})$ we have that the global minimum is 2^M-fold symmetric, since there exist an additional $2^M - 1$ parameter sets $\{\boldsymbol{\theta}_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ with the same exact cost function value. This Corollary implies an exponential reduction of the effective hyperarameter space by restricting the domain of all angles in $U(\boldsymbol{\theta})$ from $[0, 2\pi)^M$ to $[0, \pi)^M$. Hence, all relevant features of the cost function landscape (including the global minima) can be found in $[0, \pi)^M$. We finally remark that the domain restriction in Corollary (1) is non-trivial as it does not arise from a wrapping symmetry in the rotation parameters (i.e., it does not arise from the fact that $R_\mu(\theta) = R_\mu(\theta + 2\pi)$). Instead this domain reduction arises from the σ-Pulse symmetries.

4.2.2.4 Noise-induced lifting of the symmetries

In this section we analyse how noise affects the symmetries in $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ and hence the symmetries in the cost landscape. Our main results are presented in the form of two theorems, with Theorem 1 analyzing the effect of unital Pauli noise, and Theorem 2 the effect of non-unital Pauli noise. We recall that unital Pauli noise channels include $T2$ processes (i.e. dephasing channel), and depolarizing as special cases. On the other hand, non-unital Pauli noise channels include $T1$ processes as a special case, i.e. the amplitude damping channel is a non-unital Pauli channel.

Consider the following definition:

Definition 12 (Unital Pauli noise model)**.** *We define the unital Pauli noise model as a process in which a unital Pauli channel acts after every layer of gates acting in parallel in* $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$*.*

A unital Pauli noise channel corresponds to the action of random Pauli operators according to a given probability distribution. Here we employ the notation for Pauli operators:

$$X^{\boldsymbol{a}} = X_1^{a_1} \otimes \cdots \otimes X_n^{a_n}\,, \qquad Z^{\boldsymbol{b}} = Z_1^{b_1} \otimes \cdots \otimes Z_n^{b_n}\,. \tag{4.14}$$

where $\boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^{\otimes n}$ are bitstrings of length n. Let $\mathcal{P}_U$ denote an n-qubit Pauli channel. The action of $\mathcal{P}_U$ on any given n-qubit Pauli operator is given by:

$$\mathcal{P}_U(X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) = \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}} X^{\boldsymbol{l}} Z^{\boldsymbol{k}} (X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) (X^{\boldsymbol{l}} Z^{\boldsymbol{k}})^\dagger, \tag{4.15}$$

where $0 \leqslant p^A_{\boldsymbol{l},\boldsymbol{k}} \leqslant 1$, and $\sum_{\boldsymbol{l},\boldsymbol{k}} p^A_{\boldsymbol{l},\boldsymbol{k}} = 1$. By using the fact that:

$$X^{\boldsymbol{l}} Z^{\boldsymbol{k}} = (-1)^{\boldsymbol{l}\cdot\boldsymbol{k}} Z^{\boldsymbol{k}} X^{\boldsymbol{l}}\,, \tag{4.16}$$

where $\boldsymbol{l} \cdot \boldsymbol{k}$ is the dot product modulo 2, we find:

$$\begin{aligned}\mathcal{P}_U(X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) &= \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}} X^{\boldsymbol{l}} Z^{\boldsymbol{k}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}} Z^{\boldsymbol{k}} X^{\boldsymbol{l}} \\ &= \sum_{\boldsymbol{l},\boldsymbol{k}} (-1)^{\boldsymbol{a}\cdot\boldsymbol{k}} (-1)^{\boldsymbol{b}\cdot\boldsymbol{l}} p_{\boldsymbol{l},\boldsymbol{k}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}} \\ &= f_{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}}, \end{aligned} \tag{4.17}$$

where $f_{\boldsymbol{a},\boldsymbol{b}} = \sum_{\boldsymbol{l},\boldsymbol{k}} (-1)^{\boldsymbol{a}\cdot\boldsymbol{k}} (-1)^{\boldsymbol{b}\cdot\boldsymbol{l}} p_{\boldsymbol{l},\boldsymbol{k}}$ and $-1 \leqslant f_{\boldsymbol{a},\boldsymbol{b}} \leqslant 1$ for all $\boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^n$. The channels so defined are completely positive trace-preserving maps whose superoperator is diagonal in the Pauli basis. They are also unital since by construction

$$\mathcal{P}_U(\mathbb{1}) = \mathbb{1}. \tag{4.18}$$

Then the following theorem holds.

Theorem 1 (Symmetry preservation)**.** *Let $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ be a buffered PQC as in Definition 10. Then, the σ-Pulse parameter symmetries in $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ are preserved under the action of the unital Pauli noise model in Definition 12.*

Proof. Let us now consider a buffered circuit $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ which is implemented in the presence of a unital Pauli noise mode as presented in Definition 12. Here we show since the Pauli operators commute with unital Pauli noise, the σ-Pulse parameter symmetries are preserved in the presence of noise.

Let us now consider the channel $\mathcal{V}_B$ which implements the unitary $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$. This channel can be expressed as:

$$\mathcal{V}_B = \mathcal{U}_B \circ \mathcal{V}_L \circ \cdots \circ \mathcal{V}_1 \,, \tag{4.19}$$

where $\mathcal{V}_l$ is the channel that implements the unitaries in the l-th layer, and where $\mathcal{U}_B$ the channel that implements the buffer unitary. From the σ-Pulse method we know that we can find a symmetric set of parameters by creating a sigma pulse, propagating it, and absorbing it in the buffer layer. In the channel notation this procedure can be expressed as

- Creation of a primary σ-Pulse:

$$\mathcal{V}_B = \mathcal{U}_B \circ \mathcal{V}_L \circ \cdots \circ \Sigma_1 \circ \widetilde{\mathcal{V}}_1 \,, \tag{4.20}$$

  with Σ_1 the channel that implements the σ-Pulse, and with $\widetilde{\mathcal{V}}_1$ the parameter shifted unitary.

- Propagation of the σ-Pulses:

$$\mathcal{V}_B = \mathcal{U}_B \circ \Sigma_L \circ \widetilde{\mathcal{V}}_L \circ \cdots \circ \widetilde{\mathcal{V}}_1 \,, \tag{4.21}$$

  where now Σ_L is the channel that implements the primary and secondary σ-Pulses.

- Absorption of the σ-Pulses:

$$\mathcal{V}_B = \widetilde{\mathcal{U}_B} \circ \widetilde{\mathcal{V}}_L \circ \cdots \circ \widetilde{\mathcal{V}}_1 \,. \tag{4.22}$$

Here we can see that the channel $\mathcal{V}_B$ remains unchanged, meaning that $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma}) = V(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}})$.

Let us now analyse this procedure in the presence of noise. The noisy version of the channel that implements $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$ can be expressed as:

$$\widehat{\mathcal{V}_B} = \mathcal{P}_U^{(L+2)} \circ \mathcal{U}_B \circ \mathcal{P}_U^{(L+1)} \circ \mathcal{V}_L \circ \mathcal{P}_U^{(L)} \cdots \circ \mathcal{P}_U^{(1)} \circ \mathcal{V}_1 \circ \mathcal{P}_U^{(0)} \,, \tag{4.23}$$

with $\mathcal{P}_U^{(l)}$ the noisy channel acting after every layer of gates. Once a σ-Pulse has been created we will have:

$$\widehat{\mathcal{V}_B} = \mathcal{P}_U^{(L+2)} \circ \mathcal{U}_B \circ \mathcal{P}_U^{(L+1)} \circ \mathcal{V}_L \circ \mathcal{P}_U^{(L)} \cdots \circ \mathcal{P}_U^{(1)} \circ \Sigma_1 \circ \widetilde{\mathcal{V}}_1 \circ \mathcal{P}_U^{(0)} \,. \tag{4.24}$$

As we now show, any unital noisy channels $\mathcal{P}_U^{(i)}$ always commute with the channels Σ_k that implement σ-Pulses. Explicitly, the action of Σ_k on any given n-qubit Pauli

operator is given by:

$$\begin{aligned}\Sigma_k(X^{\boldsymbol{a}}Z^{\boldsymbol{b}}) &= (X^{\boldsymbol{p}}Z^{\boldsymbol{q}})X^{\boldsymbol{a}}Z^{\boldsymbol{b}}(X^{\boldsymbol{p}}Z^{\boldsymbol{q}})^\dagger \\ &= (-1)^{\boldsymbol{a}\cdot\boldsymbol{q}}(-1)^{\boldsymbol{b}\cdot\boldsymbol{p}}X^{\boldsymbol{a}}Z^{\boldsymbol{b}}\,. \end{aligned} \tag{4.25}$$

Hence, using Eqs. (4.16), (4.25) and (4.15) we have that the following chain of equalities always hold:

$$\begin{aligned}\mathcal{P}_U^{(i)}\circ\Sigma_k(X^{\boldsymbol{a}}Z^{\boldsymbol{b}}) &= \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}X^{\boldsymbol{l}}Z^{\boldsymbol{k}}X^{\boldsymbol{p}}Z^{\boldsymbol{q}}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}})Z^{\boldsymbol{q}}X^{\boldsymbol{p}}Z^{\boldsymbol{k}}X^{\boldsymbol{l}} \\ &= \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}X^{\boldsymbol{l}}X^{\boldsymbol{p}}Z^{\boldsymbol{k}}Z^{\boldsymbol{q}}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}})Z^{\boldsymbol{q}}Z^{\boldsymbol{k}}X^{\boldsymbol{p}}X^{\boldsymbol{l}} \\ &= \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}X^{\boldsymbol{l}+\boldsymbol{p}}Z^{\boldsymbol{k}+\boldsymbol{q}}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}})Z^{\boldsymbol{q}+\boldsymbol{k}}X^{\boldsymbol{p}+\boldsymbol{l}} \\ &= \Sigma_k\circ\mathcal{P}_U^{(i)}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}})\,. \end{aligned} \tag{4.26}$$

From 4.26 we can commute the σ-Pulses through the noisy channel so that:

$$\begin{aligned}\widehat{\mathcal{V}_B} =&\mathcal{P}_U^{(L+2)}\circ\mathcal{U}_B\circ\mathcal{P}_U^{(L+1)}\circ\mathcal{V}_L\circ\mathcal{P}_U^{(L)}\cdots\circ\mathcal{P}_U^{(1)}\circ\Sigma_1\circ\mathcal{V}_1\circ\mathcal{P}_U^{(0)} \\ =&\mathcal{P}_U^{(L+2)}\circ\mathcal{U}_B\circ\Sigma_L\circ\mathcal{P}_U^{(L+1)}\circ\widetilde{\mathcal{V}}_L\circ\mathcal{P}_U^{(L)}\cdots\circ\mathcal{P}_U^{(1)}\circ\widetilde{\mathcal{V}}_1\circ\mathcal{P}_U^{(0)} \\ =&\mathcal{P}_U^{(L+2)}\circ\widetilde{\mathcal{U}}_B\circ\mathcal{P}_U^{(L+1)}\circ\widetilde{\mathcal{V}}_L\circ\mathcal{P}_U^{(L)}\cdots\circ\mathcal{P}_U^{(1)}\circ\widetilde{\mathcal{V}}_1\circ\mathcal{P}_U^{(0)}\,. \end{aligned} \tag{4.27}$$

Eq. (4.27) proves Theorem 1 as it shows that the noisy channels that implement $V(\boldsymbol{\theta},\boldsymbol{\gamma})$ and $V(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\gamma}})$ are the same.

□

Theorem 1 shows that the parameter symmetries in $U_B(\boldsymbol{\theta},\boldsymbol{\gamma})$ arising from σ-Pulse symmetries are completely preserved by the action of unital Pauli noise channels which includes as special cases the action of local (or global) depolarizing channels, as well as dephasing channels.

In addition, Theorem 1 implies that the symmetry in the cost function landscape also remains unchanged. Particularly, we then know that the optimal parameters $\{\widehat{\theta}_{\text{opt}},\widehat{\boldsymbol{\gamma}}_{\text{opt}}\}$ leading to the global minimum of the noisy cost function will still be 2^M-fold symmetric in $[0,2\pi)$. That is, starting from $\{\widehat{\theta}_{\text{opt}},\widehat{\boldsymbol{\gamma}}_{\text{opt}}\}$, all the symmetric parameters obtained from any bitstring in $\mathcal{B}$ will have the same energy. In a practical scenario

Theorem 1 implies that the minimum cost achievable from randomly initializing the parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ will be independent of the bitstring $\boldsymbol{\beta}$ characterizing the initial point.

Note however that for general cost functions the presence of quantum noise can change the cost landscape such that the optimal parameters of the noisy cost function are different than the ones for the noiseless case [MRBAG16]. That is, one generally has $\{\widehat{\theta}_{\text{opt}}, \widehat{\boldsymbol{\gamma}}_{\text{opt}}\} \neq \{\theta_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ (the optimal parameters of the noiseless circuit). For the special cases when $\{\widehat{\theta}_{\text{opt}}, \widehat{\boldsymbol{\gamma}}_{\text{opt}}\} = \{\theta_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ we say that the cost has *optimal parameter resilience*. This phenomenon has been analysed in [SKCC20] for the problem of variational quantum compiling.

In addition, we also know that the value of the cost function evaluated in the presence of noise at the optimal parameters $\widehat{C}(\widehat{\theta}_{\text{opt}}, \widehat{\boldsymbol{\gamma}}_{\text{opt}})$ can also change due to the presence of noise. Here $\widehat{C}$ denotes the noisy cost function. In fact, let us consider a cost function of the form $C(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \text{Tr}[OU(\boldsymbol{\theta})\rho V^{\dagger}(\boldsymbol{\theta})]$, where O is a Hermitian operator, and assume that the parameters $\{\theta_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}\}$ yield the operator's ground state. Then we have trivially that

$$C(\boldsymbol{\theta}_{\text{opt}}, \boldsymbol{\gamma}_{\text{opt}}) \leqslant \widehat{C}(\widehat{\boldsymbol{\theta}}_{\text{opt}}, \widehat{\boldsymbol{\gamma}}_{\text{opt}})\,. \tag{4.28}$$

In general, the output state of the PQC converges to the fixed point of the noise model [SFGP21a], and so the cost function will be increasingly different as noise increases.

Let us here analyse the effect of non-unital Pauli noise on the σ-Pulse symmetries. Hence, consider the following definition

Definition 13 (Non-unital Pauli noise model)**.** *We define the non-unital Pauli noise model as a process in which a non-unital Pauli channel acts after every layer of gates acting in parallel in* $U_B(\boldsymbol{\theta}, \boldsymbol{\gamma})$*.*

Here we recall that non-unital Pauli noise channels are completely positive trace-preserving maps $\mathcal{P}_{NU}$ whose action on the identity operator is:

$$\mathcal{P}_{\text{NU}}(\mathbb{1}) = \mathbb{1} + \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} d_{\boldsymbol{ab}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}}\,. \tag{4.29}$$

On the other hand, its action on all other Pauli operators is given by:

$$\mathcal{P}_{\mathrm{NU}}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}}) = q_{\boldsymbol{ab}}X^{\boldsymbol{a}}Z^{\boldsymbol{b}}\,. \tag{4.30}$$

Then the following theorem holds.

Theorem 2 (Symmetry breaking)**.** *Let $U_B(\boldsymbol{\theta},\boldsymbol{\gamma})$ be a buffered PQC as in Definition 10. Then, the σ-Pulse parameter symmetries in $U_B(\boldsymbol{\theta},\boldsymbol{\gamma})$ can be broken under the action of a non-unital Pauli noise model.*

To prove Theorem 2 we show that there exist noisy channels $\mathcal{N}$ (which include as special case then non-unital Pauli noise channels $\mathcal{P}_{NU}$) that do not commute with the channels Σ that implement the σ-Pulses and which hence break the parameter symmetries.

Proof. Let us consider a general noisy channel which acts on an n-qubit bitstring as

$$\mathcal{N}(X^{\boldsymbol{a}}Z^{\boldsymbol{b}}) = \sum_{\boldsymbol{k},\boldsymbol{l}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{k}}Z^{\boldsymbol{l}}\,, \tag{4.31}$$

which has some $c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} \neq 0$ for $\boldsymbol{k},\boldsymbol{l} \neq \boldsymbol{a},\boldsymbol{b}$. A non-unital Pauli channel in particular must have $c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{0},\boldsymbol{0}} \neq 0$ for some $\boldsymbol{k},\boldsymbol{l} \neq \boldsymbol{0},\boldsymbol{0}$ (see Eq. (4.29) of the main text). Similarly to the previous section, we now can define noisy version of the channel that implements $U_B(\boldsymbol{\theta},\boldsymbol{\gamma})$ as:

$$\widehat{\mathcal{V}_B} = \mathcal{N}^{(L+2)} \circ \mathcal{U}_B \circ \mathcal{N}^{(L+1)} \circ \mathcal{V}_L \circ \mathcal{N}^{(L)} \cdots \circ \mathcal{N}^{(1)} \circ \mathcal{V}_1 \circ \mathcal{N}^{(0)}\,. \tag{4.32}$$

Let us now analyse if the channel Σ_k commutes with a noise channel $\mathcal{N}$ when acting on an n-qubit bitstring $X^{\boldsymbol{a}}Z^{\boldsymbol{b}}$.

$$
\begin{aligned}
&[\Sigma_k, \mathcal{N}](X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) = (\Sigma_k \circ \mathcal{N} - \mathcal{N} \circ \Sigma_k)(X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) \\
&= (-1)^{\boldsymbol{a}\cdot\boldsymbol{q}}(-1)^{\boldsymbol{b}\cdot\boldsymbol{p}} \sum_{\boldsymbol{k},\boldsymbol{l}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{k}} Z^{\boldsymbol{l}} - \sum_{\boldsymbol{k},\boldsymbol{l}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} (-1)^{\boldsymbol{k}\cdot\boldsymbol{q}}(-1)^{\boldsymbol{l}\cdot\boldsymbol{p}} X^{\boldsymbol{k}} Z^{\boldsymbol{l}} \\
&= \sum_{\boldsymbol{k},\boldsymbol{l}} \left[(-1)^{\boldsymbol{a}\cdot\boldsymbol{q}}(-1)^{\boldsymbol{b}\cdot\boldsymbol{p}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} - (-1)^{\boldsymbol{k}\cdot\boldsymbol{q}}(-1)^{\boldsymbol{l}\cdot\boldsymbol{p}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}}\right] X^{\boldsymbol{k}} Z^{\boldsymbol{l}} \\
&= \sum_{\boldsymbol{k},\boldsymbol{l}} \left[(-1)^{\boldsymbol{a}\cdot\boldsymbol{q}+\boldsymbol{b}\cdot\boldsymbol{p}} - (-1)^{\boldsymbol{k}\cdot\boldsymbol{q}+\boldsymbol{l}\cdot\boldsymbol{p}}\right] c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{k}} Z^{\boldsymbol{l}} \\
&= \sum_{\boldsymbol{k},\boldsymbol{l}} M_{\boldsymbol{p},\boldsymbol{q}}^{\boldsymbol{a},\boldsymbol{b},\boldsymbol{k},\boldsymbol{l}} c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{k}} Z^{\boldsymbol{l}}
\end{aligned} \tag{4.33}
$$

where:

$$
M_{\boldsymbol{p},\boldsymbol{q}}^{\boldsymbol{a},\boldsymbol{b},\boldsymbol{k},\boldsymbol{l}} = \begin{cases} 2 & \text{if } \boldsymbol{a}\cdot\boldsymbol{q} = \boldsymbol{b}\cdot\boldsymbol{p} \pmod 2 \text{ and } \boldsymbol{k}\cdot\boldsymbol{q} \neq \boldsymbol{l}\cdot\boldsymbol{p} \pmod 2, \\ -2 & \text{if } \boldsymbol{a}\cdot\boldsymbol{q} \neq \boldsymbol{b}\cdot\boldsymbol{p} \pmod 2 \text{ and } \boldsymbol{k}\cdot\boldsymbol{q} = \boldsymbol{l}\cdot\boldsymbol{p} \pmod 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.34}
$$

Thus, given $\boldsymbol{p}, \boldsymbol{q}$ and $\boldsymbol{a}, \boldsymbol{b}$, there exists noise channel defined by some $c_{\boldsymbol{k},\boldsymbol{l}}^{\boldsymbol{a},\boldsymbol{b}}$ such that the σ-Pulse does not commute with the noise channel. That any such noise channel must be non-Pauli is clear since $M_{\boldsymbol{p},\boldsymbol{q}}^{\boldsymbol{a},\boldsymbol{b},\boldsymbol{k},\boldsymbol{l}} = 0$ for all $\boldsymbol{p}, \boldsymbol{q}$ whenever $\boldsymbol{a}, \boldsymbol{b} = \boldsymbol{k}, \boldsymbol{l}$.

□

From Theorem 2 we have that given two sets of symmetric parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ and $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}}\}$, then there exists some non-unital noise such that we have $U(\boldsymbol{\theta}, \boldsymbol{\gamma}) \neq U(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\gamma}})$. This implies that the 2^M-fold symmetry of the optimal noisy parameters, and hence the symmetry in the cost landscape, is broken.

In particular we now have that some of the (previously) global minima are transformed into local minima. In practical terms, due to this symmetry breaking in the cost landscape not all randomly initialised $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ will converge to the global minima. In the next section we show how this effect can be mitigated by exploiting the knowledge of parameter symmetries to construct an optimiser.

Note that it is possible that $\mathcal{N} \circ \Sigma_k = \Sigma_k' \circ \mathcal{N}$ where Σ_k' is a different σ-Pulse, which may then lead to a different set of symmetries. However it is still true that the original σ-Pulse symmetries are broken.

We remark that the proof for Theorem 2 in the Appendix is valid for more general noisy channels that include as special case non-unital Pauli noise channels. For instance, we show that coherent error, such as qubit drift, can also break the σ-Pulse symmetries.

4.2.2.5 Example: parameter symmetries in QAOA

As a special case of our analysis we can explicitly characterise the symmetries possessed by a Quantum Alternating Operator Ansatz (QAOA) [HWO$^+$19], notably without requiring a buffer layer. To recap, the QAOA, which generalises the ansatz in the Quantum Approximate optimisation Algorithm [FGG14], has a layered structure, each layer being composed of two unitaries: a problem unitary $U_P(\beta_i) = e^{-i\beta_i H_P}$, where H_P is the problem Hamiltonian (consisting of Z-terms only) and a mixing unitary $U_M(\gamma_i) = e^{-i\gamma_i H_M}$, where typically $H_M = \sum_j X_j$. The parameters of QAOA can therefore be represented as two vectors $\{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$ each corresponding to each type of unitary. Now let us consider the effect of shifting one of these parameters by π.

If the parameter belongs to a mixing unitary, the shift will be cancelled by the creation of a σ-Pulse in the X direction on all the qubits. It can be shown that the commutation of any number of X pulse through a gate consisting of Z terms of any order generates no additional pulses, but leads to the parameter of the gate acquiring a negative sign. Therefore, the X pulses generated by such a shift can only be annihilated by a restoring shift of another mixing parameter. Therefore, a symmetry is the following:

$$\begin{aligned}
&\gamma_i \to \gamma_i + \pi \\
&\gamma_j \to \gamma_j + \pi,\ i < j \\
&\beta_k \to -\beta_k,\ \forall k \text{ s.t. } i < k \leqslant j.
\end{aligned}$$

If instead we focus on the problem unitary, we see that the parameter shift will create, on a given qubit, as many Z pulse as there are Z operators in the problem Hamiltonian acting on that qubit. Since the mixing unitary presents one X rotation for each qubit, one concludes that, in order to be able to commute all of the pulses

past a mixing unitary with only a sign modification of its parameter, there must be the same number of Z pulses on all the qubits. Therefore, the ansatz will only possess such a symmetry if the problem Hamiltonian features the same number of Zs on each qubit. This happens, for example, in MaxCut problems on n-regular graphs. In this specific case, when n is odd each shift will generate one total Z per qubit, and for any i, j with $i < j$ the symmetries are:

$$\beta_i \to \beta_i + \pi,\ \beta_j \to \beta_j + \pi,$$
$$\gamma_k \to -\gamma_k\ \forall k\, i < k \leqslant j.$$

Otherwise, for n even all the pulses will cancel and the symmetries will simply be:

$$\beta_i \to \beta_i + \pi\ \forall i.$$

Identifying these symmetries is significant, because they imply that restricting the range of some parameters will not affect the result of the algorithm.

4.3 Noise-induced barren plateaus

In this section, we analytically study the scaling of the cost function and its gradient for VQAs as a function of the number of qubits n, the circuit depth L, and the noise parameter $q < 1$. We consider a general class of local noise models that includes depolarizing noise and certain kinds of Pauli noise. Furthermore, we investigate a general, abstract ansatz that allows us to encompass many of the important ansätze in the literature, hence allowing us to make a general statement about VQAs. This includes the Quantum Alternating Operator Ansatz (QAOA) which is used for solving combinatorial optimisation problems [FGG14, WHJR18, Cro18, HWO$^+$19] and the Unitary Coupled Cluster (UCC) Ansatz which is used in the Variational Quantum Eigensolver (VQE) to solve chemistry problems [CRO$^+$19, BM07, LHHGW18]. This is also applicable for the Hardware Efficient Ansatz and the Hamiltonian Variational Ansatz (HVA) which are employed for various applications [KMT$^+$17, AA20, AAB$^+$20b,

WHT15, WZdS$^+$20]. Our results also generalise to settings that allow for multiple input states or training data, as in machine learning applications, often called quantum neural networks [SSP14, SSP15, BWP$^+$17, BBF$^+$20, ASZ$^+$21].

Our main result (Theorem 3) is an upper bound on the magnitude of the gradient that decays exponentially with L, namely as $2^{-\kappa}$ with $\kappa = -L\log_2(q)$. Hence, we find that the gradient vanishes exponentially in the circuit depth. Moreover, it is typical to consider L scaling as poly(n) (e.g., in the UCC Ansatz [LHHGW18]), for which our main result implies an exponential decay of the gradient in n. We refer to this as a Noise-Induced Barren Plateau (NIBP). We remark that NIBPs can be viewed as concomitant to the cost landscape concentrating around the value of the cost for the maximally mixed state, and we make this precise in Lemma 1. See Fig. 4.5 for a schematic diagram of the NIBP phenomenon.

To be clear, any variational algorithm with a NIBP will have exponential scaling. In this sense, NIBPs destroy quantum speedup, as the standard goal of quantum algorithms is to avoid the typical exponential scaling of classical algorithms. NIBPs are conceptually distinct from the noise-free barren plateaus of Refs. [MBS$^+$18, SCCC22, CSV$^+$21, MKW21, PNGY21, HSCC22]. Indeed, strategies to avoid noise-free barren plateaus [CSV$^+$21, UB21, VC21, VBM$^+$19, GWOB19, SMM$^+$20] do not appear to solve the NIBPs issue.

The obvious strategy to address NIBPs is to reduce circuit complexity, or more precisely, to reduce the circuit depth. Hence, our work provides quantitative guidance for how small L needs to be to potentially avoid NIBPs.

In what follows, we present our general framework followed by our main result. We also present two extensions of our main result, one involving correlated ansatz parameters and one allowing for measurement noise. The latter indicates that global cost functions exacerbate the NIBP issue. In addition, we provide numerical heuristics that illustrate our main result for MaxCut optimisation with the QAOA, and an implementation of the HVA on superconducting hardware, both showing that NIBPs significantly impact this application.

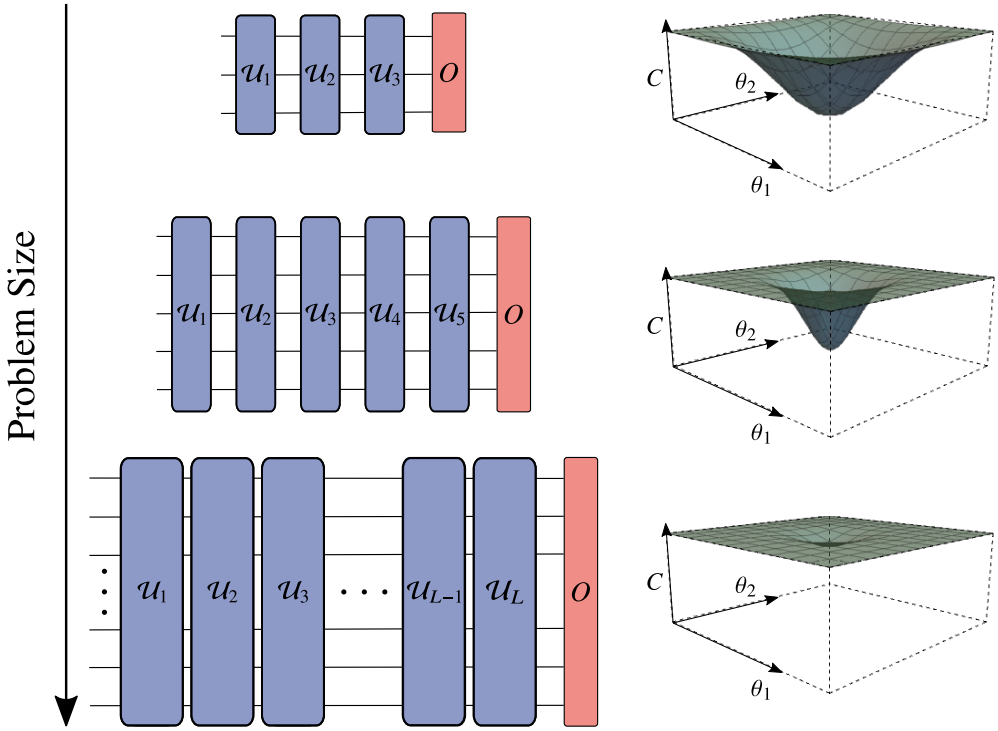


Figure 4.5: Schematic diagram of the Noise-Induced Barren Plateau (NIBP) phenomenon. For various applications such as chemistry and optimisation, increasing the problem size often requires one to increase the depth L of the variational ansatz. We show that, in the presence of local noise, the gradient vanishes exponentially in L and hence exponentially in the number of qubits n when L scales linearly in n. This can be seen in the plots on the right, which show the cost function landscapes for a simple variational problem with local noise.

4.3.1 Framework

In this section we analyse a general class of parameterised ansätze $U(\boldsymbol{\theta})$ that can be expressed as a product of L unitaries sequentially applied by layers

$$U(\boldsymbol{\theta}) = U_L(\boldsymbol{\theta}_L) \cdots U_2(\boldsymbol{\theta}_2) \cdot U_1(\boldsymbol{\theta}_1)\,. \tag{4.35}$$

Here $\boldsymbol{\theta} = \{\boldsymbol{\theta}_l\}_{l=1}^{L}$ is a set of vectors of continuous parameters that are optimised to minimise a cost function C that can be expressed as the expectation value of an operator O:

$$C = \mathrm{Tr}[OU(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta})]\,. \tag{4.36}$$

As shown in Fig. 4.6, ρ is an n-qubit input state. Without loss of generality we assume that each $U_l(\boldsymbol{\theta}_l)$ is given by:

$$U_l(\boldsymbol{\theta}_l) = \prod_m e^{-i\theta_{lm}H_{lm}} W_{lm}\,, \tag{4.37}$$

where H_{lm} are Hermitian operators, $\boldsymbol{\theta}_l = \{\theta_{lm}\}$ are continuous parameters, and W_{lm} denote unparameterised gates. We expand H_{lm} and O in the Pauli basis as:

$$H_{lm} = \boldsymbol{\eta}_{lm} \cdot \boldsymbol{\sigma}_n = \sum_i \eta^i_{lm}\sigma^i_n\,, \quad O = \boldsymbol{\omega} \cdot \boldsymbol{\sigma}_n = \sum_i \omega^i \sigma^i_n\,, \tag{4.38}$$

where $\sigma^i_n \in \{\mathbb{1}, X, Y, Z\}^{\otimes n}$ are Pauli strings, and $\boldsymbol{\eta}_{lm}$ and $\boldsymbol{\omega}$ are real-valued vectors that specify the terms present in the expansion. Defining $N_{lm} = |\boldsymbol{\eta}_{lm}|$ and $N_O = |\boldsymbol{\omega}|$ as the number of non-zero elements, i.e., the number of terms in the summations in Eq. (4.38), we say that H_{lm} and O admit an efficient Pauli decomposition if $N_{lm}, N_O \in \mathcal{O}(\text{poly}(n))$, respectively.

As shown in Fig. 4.6, we consider a noise model where local Pauli noise channels $\mathcal{N}_j$ act on each qubit j before and after each unitary $U_l(\boldsymbol{\theta}_l)$. The action of $\mathcal{N}_j$ on a local Pauli operator $\sigma \in \{X, Y, Z\}$ can be expressed as:

$$\mathcal{N}_j(\sigma) = q_\sigma \sigma\,, \tag{4.39}$$

where $-1 < q_X, q_Y, q_Z < 1$. Here, we characterise the noise strength with a single parameter $q = \sqrt{\max\{|q_X|, |q_Y|, |q_Z|\}}$. Let $\mathcal{U}_l$ denote the channel that implements the unitary $U_l(\boldsymbol{\theta}_l)$ and let $\mathcal{N} = \mathcal{N}_1 \otimes \cdots \otimes \mathcal{N}_n$ denote the n-qubit noise channel. Then, the noisy cost function is given by:

$$\widetilde{C} = \text{Tr}\left[O\left(\mathcal{N} \circ \mathcal{U}_L \circ \cdots \circ \mathcal{N} \circ \mathcal{U}_1 \circ \mathcal{N}\right)(\rho)\right]\,. \tag{4.40}$$

4.3.2 Results

There are some VQAs, such as the VQE [PMS+14] for chemistry and other physical systems, where it is important to accurately characterise the value of the cost function

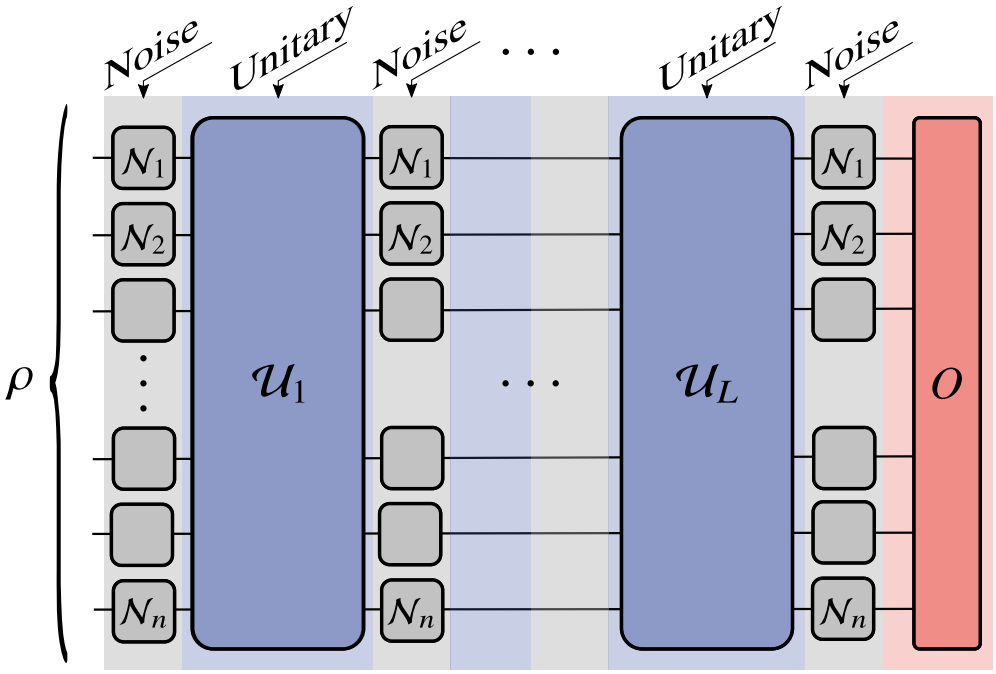


Figure 4.6: Setting for our analysis. An n-qubit input state ρ is sent through a variational ansatz $U(\boldsymbol{\theta})$ composed of L unitary layers $U_l(\boldsymbol{\theta}_l)$ sequentially acting according to Eq. (4.35). Here, $\mathcal{U}_l$ denotes the quantum channel that implements the unitary $U_l(\boldsymbol{\theta}_l)$. The parameters in the ansatz $\boldsymbol{\theta} = \{\boldsymbol{\theta}_l\}_{l=1}^{L}$ are trained to minimise a cost function that is expressed as the expectation value of an operator O as in Eq. (4.36). We consider a noise model where local Pauli noise channels $\mathcal{N}_j$ act on each qubit j before and after each unitary.

itself. We provide an important result below in Lemma 1 that quantitatively bounds the cost function itself, and we envision that this bound will be especially useful in the context of VQE. On the other hand, there are other VQAs, such as those for optimisation [FGG14, WHJR18, Cro18, HWO$^+$19], compiling [KLP$^+$19, SKCC20, JB22], and linear systems [BPLC$^+$23, XSE$^+$21], where the key goal is to learn the optimal parameters and the precise value of the cost function is either not important or can be computed classically after learning the parameters. In this case, one is primarily concerned with trainability, and hence the gradient is a key quantity of interest. These applications motivate our main result in Theorem 3, which bounds the magnitude of the gradient. We remark that trainability is of course also important for VQE, and hence Theorem 3 is also of interest for this application.

With this motivation in mind, we now present our main results. We first present our

bound on the cost function, since one can view this as a phenomenon that naturally accompanies our main theorem. Namely, in the following lemma, we show that the noisy cost function concentrates around the corresponding value for the maximally mixed state.

Lemma 1 (Concentration of the cost function)**.** *Consider an L-layered ansatz of the form in Eq.* (4.35)*. Suppose that local Pauli noise of the form of Eq.* (4.39) *with noise strength q acts before and after each layer as in Fig. 4.6. Then, for a cost function* $\widetilde{C}$ *of the form in Eq.* (4.40)*, the following bound holds:*

$$\left|\widetilde{C} - \frac{1}{2^n}\mathrm{Tr}[O]\right| \leqslant G(n)\,, \tag{4.41}$$

where:

$$G(n) = \sqrt{2\,\ln 2}\, N_O \|\boldsymbol{\omega}\|_\infty\, n^{1/2} q^{cL+1}\,. \tag{4.42}$$

Here $\|\cdot\|_\infty$ *is the infinity norm,* $\|\cdot\|_1$ *is the trace norm,* $\boldsymbol{\omega}$ *is defined in Eq.* (4.38)*, and* N_O *is the number of non-zero elements in the Pauli decomposition of* O*, and* $c = 1/(2\ln 2)$ *is a constant.*

The proof of this lemma can be found in the Supplementary Information of Ref. [WFC+21]. Briefly, it makes use of standard norm inequalities and crucially on inequalities on the relative entropy and the 2-Renyi relative entropy of a quantum state.

Lemma 1 implies the cost landscape exponentially concentrates on the value $\mathrm{Tr}[O]/2^n$ for large n, whenever the number of layers L scales linearly with the number of qubits. While this lemma has important applications on its own, particularly for VQE, it also provides intuition for the NIBP phenomenon, which we now state.

Let $\partial_{lm}\widetilde{C} = \partial\widetilde{C}/\partial\theta_{lm}$ denote the partial derivative of the noisy cost function with respect to the m-th parameter that appears in the l-th layer of the ansatz, as in Eq. (4.37). For our main technical result, we upper bound $|\partial_{lm}\widetilde{C}|$ as a function of L and n.

Theorem 3 (Upper bound on the partial derivative)**.** *Consider an L-layered ansatz as defined in Eq.* (4.35)*. Let* θ_{lm} *denote the trainable parameter corresponding to the Hamiltonian* H_{lm} *in the unitary* $U_l(\boldsymbol{\theta}_l)$ *appearing in the ansatz. Suppose that local Pauli*

noise of the form in Eq. (4.39) *with noise parameter* q *acts before and after each layer as in Fig. 4.6. Then the following bound holds for the partial derivative of the noisy cost function:*

$$|\partial_{lm}\widetilde{C}| \leqslant F(n)\,, \tag{4.43}$$

where:

$$F(n) = \sqrt{8\ln 2}\, N_O \|\boldsymbol{\omega}\|_\infty \left\|\boldsymbol{\eta}_{lm}\right\|_1 n^{1/2} q^{cL+1}\,, \tag{4.44}$$

and $\boldsymbol{\eta}_{lm}$ *and* $\boldsymbol{\omega}$ *are defined in Eq.* (4.38)*, and* N_O *is the number of non-zero Pauli terms in* O*, and* $c = 1/(2\ln 2)$ *is a constant.*

The proof of this theorem can be found in the Supplementary Information of Ref. [WFC+21].

Let us now consider the asymptotic scaling of the function $F(n)$ in Eq. (4.44). Under standard assumptions such as that O in Eq. (4.38) admits an efficient Pauli decomposition and that H_{lm} has bounded eigenvalues, we now state that $F(n)$ decays exponentially in n, if L grows linearly in n.

Corollary 2 (Noise-induced barren plateaus)**.** *Let* $N_{lm}, N_O \in \mathcal{O}(\mathrm{poly}(n))$ *and let* $\eta^i_{lm}, \omega^j \in \mathcal{O}(\mathrm{poly}(n))$ *for all* i, j*. Then the upper bound* $F(n)$ *in Eq.* (4.44) *vanishes exponentially in* n *as:*

$$F(n) \in \mathcal{O}(2^{-\alpha n})\,, \tag{4.45}$$

for some positive constant α *if we have:*

$$L \in \Omega(n)\,. \tag{4.46}$$

The asymptotic scaling in Eq. (4.45) is independent of l and m, i.e., the scaling is blind to the layer, or the parameter within the layer, for which the derivative is taken. This corollary implies that when Eq. (4.46) holds, i.e. L grows at least linearly in n, the partial derivative $|\partial_{lm}\widetilde{C}|$ exponentially vanishes in n across the entire cost landscape. In other words, one observes a Noise-Induced Barren Plateau (NIBP). We note that Eq. (4.46) is satisfied for all $q < 1$. That is, NIBPs occur regardless of the noise strength, it only changes the severity of the exponential scaling.

In addition, Corollary 2 implies that NIBPs are conceptually different from noise-free barren plateaus. First, NIBPs are independent of the parameter initialization strategy or the locality of the cost function. Second, NIBPs exhibit exponential decay of the gradient itself; not just of the variance of the gradient, which is the hallmark of noise-free barren plateaus. Noise-free barren plateaus allow the global minimum to sit inside deep, narrow valley in the landscape [CSV+21], whereas NIBPs flatten the entire landscape.

Finally, we present an extension of our main result to the case of measurement noise. First consider the following lemma:

Lemma 2 (Pauli coefficients under noise)**.** *Consider an operator* Λ *whose Pauli basis decomposition is:*

$$\Lambda = \lambda_0 \mathbb{1}^{\otimes n} + \boldsymbol{\lambda} \cdot \boldsymbol{\sigma}_n, \tag{4.47}$$

where $\lambda_0 \in \mathbb{R}$ *and* $\boldsymbol{\lambda} \in \mathbb{R}^{4^n-1}$*. Under the action of a noise channel* $\mathcal{N}$ *of the form in Eq.* (4.39) *on a single Pauli string* σ_n^i *we have:*

$$|\widetilde{\lambda}_i| \leqslant q|\lambda_i| \quad \forall\, i > 0, \tag{4.48}$$

where we define $\widetilde{\lambda}_i$ *as the coefficient such that* $\widetilde{\lambda}_i\, \sigma_n^i = \mathcal{N}(\lambda_i \sigma_n^i)$*.*

Proof. The effect of a single layer of noise can be expressed as follows:

$$\lambda_i \xrightarrow{\mathcal{N}} q_X^{x(i)} q_Y^{y(i)} q_Z^{z(i)} \lambda_i \,, \tag{4.49}$$

for all $i \in [4^n - 1]$, where $x(i) + y(i) + z(i) \leqslant n$ is the number of non-identity terms in the i-th Pauli string. Noting that $x(i) + y(i) + z(i) \geqslant 1 \ \forall i$ and using the definition of q, we obtain the desired statement. □

Now consider a model of measurement noise where each local measurement independently has some bit-flip probability given by $(1 - q_M)/2$, which we assume to be symmetric with respect to the 0 and 1 outcomes. This leads to an additional reduction

of our bounds on the cost function and its gradient that depends on the locality of the observable O.

Theorem 4 (Measurement noise)**.** *Consider expanding the observable O as a sum of Pauli strings, as in Eq.* (4.38)*. Let w denote the minimum weight of these strings, where the weight is defined as the number of non-identity elements for a given string. In addition to the noise process considered in Fig. 4.6, suppose there is also measurement noise consisting of a tensor product of local bit-flip channels with bit-flip probability $(1 - q_M)/2$. Then we have:*

$$\left|\widetilde{C} - \frac{1}{2^n}\mathrm{Tr}[O]\right| \leqslant q_M^w \, G(n)\,, \tag{4.50}$$

and:

$$|\partial_{lm}\widetilde{C}| \leqslant q_M^w F(n)\,, \tag{4.51}$$

where $G(n)$ and $F(n)$ are defined in Lemma 1 and Theorem 3, respectively.

Proof. We prove in detail the proposition about the gradient of the cost function. The proposition about the cost function is derived in an analogous manner.

As a model of measurement noise we consider a classical bit-flip channel applied to every qubit, such that the standard POVM elements get replaced by:

$$P_0 = |0\rangle\langle 0| \to \widetilde{P}_0 = p_{00}\,|0\rangle\langle 0| + p_{01}\,|1\rangle\langle 1| \tag{4.52}$$

$$P_1 = |1\rangle\langle 1| \to \widetilde{P}_1 = p_{10}\,|0\rangle\langle 0| + p_{11}\,|1\rangle\langle 1|\,, \tag{4.53}$$

where $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$. Furthermore, we take this channel to be unital, such that $\widetilde{P}_0 + \widetilde{P}_1 = (p_{00} + p_{10})P_0 + (p_{01} + p_{11})P_1 = P_0 + P_1$ giving $p_{00} + p_{10} = 1$ and $p_{01} + p_{11} = 1$. Thus, there is only one free parameter q_M, and we set $p_{00} = p_{11} = \frac{1+q_M}{2}$, $p_{01} = p_{10} = \frac{1-q_M}{2}$. Note that without loss of generality we can assume $p_{00}, p_{11} > 1/2$,

and hence $q_M \geqslant 0$. Overall:

$$P_0 = |0\rangle\langle 0| \to \widetilde{P}_0 = \frac{1+q_M}{2}|0\rangle\langle 0| + \frac{1-q_M}{2}|1\rangle\langle 1| \tag{4.54}$$

$$P_1 = |1\rangle\langle 1| \to \widetilde{P}_1 = \frac{1-q_M}{2}|0\rangle\langle 0| + \frac{1+q_M}{2}|1\rangle\langle 1| \,. \tag{4.55}$$

The equivalence between this classical channel and a quantum bit-flip channel is seen by writing $P_0 = \frac{\mathbb{1}+Z}{2}$ and $P_1 = \frac{\mathbb{1}-Z}{2}$, such that the bit-flip channel is equivalent to a transformation of the Pauli Z operator: $Z' = q_M Z$. This corresponds to the effect of a bit-flip channel $\mathcal{N}(\rho) = \frac{1+q_M}{2}\rho + \frac{1-q_M}{2}X\rho X$.

The reasoning so far only applies to measurements in the Z basis. However, in our model we do not consider a standard projective measurement, but the expectation value with respect to a general Hermitian operator. This assumes the capability of performing measurements in any basis. If we assume that the classical bit-flip acts independently of the basis we choose to measure in, then we see that the corresponding quantum channel must be a depolarizing channel $\mathcal{D}_M$ such that:

$$\mathcal{D}_M(\sigma) = q_M\sigma \,, \tag{4.56}$$

where σ is any single-qubit Pauli operator. An alternative realistic assumption that also leads to (4.56) is that the quantum computer can only measure in the computational basis, and so one implements measurements in general bases by applying an extra layer of (noisy) one-qubit rotations before measurement. We thus proceed to model measurement noise as a tensor product of such local depolarizing channels applied prior to measurement and denote the overall channel as $\mathcal{N}_M$. From Eq. (4.56) we have that :

$$\mathcal{N}_M(O) = \sum_i \omega^i \mathcal{N}_M(\sigma_n^i) = \widetilde{\boldsymbol{\omega}} \cdot \boldsymbol{\sigma}_n \,, \tag{4.57}$$

where $\widetilde{\boldsymbol{\omega}}$ is a vector of elements $\widetilde{\omega}^i = q_M^{w(i)}\omega^i$, and where $w(i) = x(i) + y(i) + z(i)$ is the weight of the Pauli string. Here we recall that we have respectively defined $x(i)$, $y(i)$ and $z(i)$ as the number of Pauli operators X, Y, and Z in the i-th Pauli string. Let $w = \min_i w(i)$ be the minimum weight of the Pauli strings of O.

Let us now write the noisy cost function partial derivative as:

$$\partial_{lm}\widetilde{C} = \mathrm{Tr}\left[\mathcal{N}_M(O)\partial_{lm}\rho_L\right] . \tag{4.58}$$

We write:

$$|\partial_{lm}\widetilde{C}| \leqslant \|\mathcal{N}\circ\mathcal{N}_M(O)\|_\infty \left\|\mathcal{W}_+\left(\partial_{lm}\,\rho_{l-}\right)\right\|_1 . \tag{4.59}$$

The first term can be bounded as:

$$\begin{aligned}
\|\mathcal{N}\circ\mathcal{N}_M(O)\|_\infty &= \|\mathcal{N}\left(\widetilde{\boldsymbol{\omega}}\cdot\boldsymbol{\sigma}_n\right)\|_\infty && (4.60)\\
&\leqslant N_O \max_i \left\|\mathcal{N}(\widetilde{\omega}^i\sigma_n^i)\right\|_\infty && (4.61)\\
&\leqslant N_O\, q \max_i |\widetilde{\omega}^i| && (4.62)\\
&\leqslant N_O\, q\, q_M^w \|\boldsymbol{\omega}\|_\infty , && (4.63)
\end{aligned}$$

where the first inequality is due to the triangle inequality, the second inequality comes from Lemma 2 and the fact that Pauli operators have sup-norm 1, and the final inequality from the definition of w. We see this result is identical to that in Theorem 3, aside from an extra factor q_M^w.

The second term in (4.59) is bounded in the proof of Theorem 3 as:

$$\left\|\mathcal{W}_+\left(\partial_{lm}\,\rho_{l-}\right)\right\|_1 \leqslant \sqrt{8\ln 2}\left\|\boldsymbol{\eta}_{lm}\right\|_1 n^{1/2} q^{cL} . \tag{4.64}$$

Putting the two parts together we obtain:

$$\begin{aligned}
\left|\partial_{lm}\widetilde{C}\right| &\leqslant \sqrt{8\ln 2}\, N_O \|\boldsymbol{\omega}\|_\infty \left\|\boldsymbol{\eta}_{lm}\right\|_1 n^{1/2}\, q_M^w\, q^{cL+1} && (4.65)\\
&= q_M^w F(n) , && (4.66)
\end{aligned}$$

as required.

Now let us prove the complimentary result for the cost function magnitude. We

write:

$$\left|\widetilde{C} - \frac{1}{2^n}\mathrm{Tr}[O]\right| = \left|\mathrm{Tr}\left[\mathcal{N} \circ \mathcal{N}_M(O)\, \boldsymbol{a}^{(L)} \cdot \boldsymbol{\sigma}_n\right]\right| \tag{4.67}$$

$$\leqslant \left\|\mathcal{N} \circ \mathcal{N}_M(O)\right\|_\infty \left\|\boldsymbol{a}^{(L)} \cdot \boldsymbol{\sigma}_n\right\|_1 \tag{4.68}$$

$$\leqslant q_M^w\, N_O \left\|\boldsymbol{\omega}\right\|_\infty \sqrt{2\ln 2 \cdot n}\, q^{cL+1}\,. \tag{4.69}$$

where the first inequality comes from Hölder's inequality and the second one from the same argument as before and the proof of Lemma 1. Thus we can write:

$$\left|\widetilde{C} - \frac{1}{2^n}\mathrm{Tr}[O]\right| \leqslant q_M^w\, G(n)\,, \tag{4.70}$$

where $G(n) = \sqrt{2\ln 2}\, N_O \left\|\boldsymbol{\omega}\right\|_\infty n^{1/2} q^{cL+1}$ is the concentration factor in Lemma 1. Hence observables with $w \in \Omega(n)$ will suffer from an exponential decay in n of the cost function and its gradient, independent of circuit depth. □

Theorem 4 goes beyond the noise model considered in Theorem 3. It shows that in the presence of measurement noise there is an additional contribution from the locality of the measurement operator. It is interesting to draw a parallel between Theorem 4 and noise-free barren plateaus, which have been shown to be cost-function dependent and in particular depend on the locality of the observable O [CSV$^+$21]. The bounds in Theorem 4 similarly depend on the locality of O. For example, when $w = n$, i.e., global observables, the factor q_M^w will hasten the exponential decay. On the other hand, when $w = 1$, i.e., local observables, the scaling is unaltered by measurement noise. In this sense, a global observable exacerbates the NIBP issue by making the decay more rapid with n.

4.4 Conclusion

In this Chapter we discussed symmetries of the cost function landscape which can be broken by quantum noise, giving rise to a more complex landscape. The first type of symmetry is a continuous one that emerges when the quantum circuit is overparam-

eterised, meaning that some parameterised gates are redundant and could be set to zero (or any fixed value) without affecting the range of quantum states that can be expressed. In general, these symmetries may broken by many types of channel with the exception of the global depolarising channel that can be commuted with any gate. We have explored the connection to the QFIM and how it supports breaking of these symmetries by noise. However, making precise statements about the phenomenon is difficult due to the wide range of symmetries and noise channels that can exist.

The second type of symmetry links discrete sets of parameters associated with single-qubit rotation gates which differ by rotations of $\pi/2$. To understand and analyse these symmetries we have introduced the σ-Pulse method. The main idea behind this method is the creation, propagation, and absorption of virtual Pauli rotation gates in the PQC, which allow us to obtain symmetric sets of parameters. These symmetries are occur in large number in circuits that we call *buffered*, but are also present in many common VQAs. Regarding noise, we rigorously showed that the parameter symmetries are preserved under the action of unital Pauli noise, implying that dephasing and depolarizing noise acting throughout the PQC have no effect on the overall symmetric structure of the cost landscape. We then proved that non-unital Pauli can break the parameter symmetries and hence the symmetries in the cost landscape. This result implies that, when training in the presence of noise, some of the previously exponentially symmetric global minima can become local minima. Hence, optimisation strategies that randomly initialise the parameters could converge to one of those local minima and not obtain an optimal solution.

As an aside, for buffered PQCs, as shown in the proof of Prop. 1, given a parameter $\boldsymbol{\theta}$ there exists an exponentially big set of symmetric parameters $\boldsymbol{\theta}'$ such that $U(\boldsymbol{\theta}) = U(\boldsymbol{\theta}')$ (up to a global phase). These symmetries translate in turn into exponential symmetries in the cost landscape, and further we can show that all the relevant features of the cost landscape can be found in a subspace of the parameter hyperspace, hence providing an exponential reduction of the search space of variational quantum algorithms. In the context of variational algorithms, the result may have numerous implications. The fact that the noiseless quantum landscape has been demonstrated to consist of an

exponentially large number of translated copies of a single landscape "unit cell" means that the informational content of the landscape is actually greatly reduced. As such, one may hope to construct quantum algorithms that are more sample-efficient by exploiting these regularities.

Ultimately any parameter symmetry that can be broken by the presence of noise channels can pose a challenge for the optimisation of the VQA. This is because previously identical points in the cost function landscape now may become different, leading to a proliferation of local minima. Spurious local minima, also called traps, are a phenomenon which already in the noiseless case is known to cause issues for gradient-based optimisers [WGK20, BK21, AK22, LJGM$^+$23, CHPZ24]. In particular, in Ref. [AK22] the authors find that underparameterised VQAs suffer from exponentially many such traps, and therefore that overparameterisation is beneficial. Our results partial contrast this view when noise is presence, as noise can induce multiple new traps especially when the PQC is overparameterised.

However, we also argued in our discussion of continuous symmetries that noise may partially aid optimisation by introducing gradients along previously flat directions in the landscape. This may underlie the partial resilience of VQAs to noise observed in the numerical experiments in the previous Chapter, as the optimiser would be able to exploit these additional gradients to improve the quality of the solution and thus mitigate the negative effects of noise. Furthermore, in Ref. [FCA$^+$22] an optimiser based on the σ-Pulse method is presented, which exploits the symmetries to identify parameters which offer improved noise resilience. Overall it can be argued that symmetry breaking complicates the cost function landscape and thus the optimisation of VQAs, but not in a universally negative way.

In contrast, the phenomenon of noise-induced barren plateaus (NIBPs) is arguably a negative result for the field of variational quantum computing. Like noise-free barren plateaus [MBS$^+$18], NIBPs require the precision and hence the algorithmic complexity to scale exponentially with the problem size. Thus, avoiding NIBPs is necessary for a VQA to have any hope of exponential quantum speedup. However, unlike noise-free barren plateaus, the cost function and the gradient vanish with increasing problem

size at every point on the cost function landscape, rather than probabilistically. As a consequence, NIBPs cannot be addressed by layer-wise training, correlating parameters and other strategies [CSV+21, UB21, VC21, VBM+19, GWOB19, SMM+20], all of which can help avoid noise-free barren plateaus. At the moment, the only strategies we are aware of for avoiding NIBPs are: (1) reducing the hardware noise level, or (2) improving the design of variational ansätze such that their circuit depth scales more weakly with n.

We emphasise that naïve mitigation strategies such as artificially increasing gradients cannot remove the exponential scaling of NIBPs as this simply increases the variance of any finite-shot evaluation of derivatives, and it does not improve the resolvability of the landscape. This argument extends simply to include any error mitigation strategy that implements an affine map to cost values [CACC21, MS21, VKG+21, RGM22, HNdJB20, Sha21, AAB+20a]. Further, most error mitigation techniques consist only of postprocessing noisy circuits. Thus, we deem it unlikely many strategies can remove exponential NIBP scaling as information about the cost landscape has fundamentally been lost (or at least been made exponentially inaccessible). This is in contrast to error correction where information is protected and recovered. However, in general it is an open question as to whether or not error mitigation strategies can mitigate NIBPs, and we leave this question for future work.

Part II

Fourier picture

Chapter 5

Fourier spectrum of noisy Variational Quantum Algorithms

In this Chapter, we take the view of spectral analysis to investigate the effects of noise on parameterised quantum circuits. Interpreting the quantum landscape as a band-limited multidimensional function in the parameters allows us to use classical signal processing tools for quantifying the impact of various error channel models. For different models, we quantify the additional, higher frequency modes in the output signal caused by device errors. We show that filtering these noise-induced modes effectively mitigates device errors. Moreover, we describe the classical and quantum resource requirements for these techniques and test their effectiveness for application motivated-circuits on quantum hardware.

This work is based on Ref. [FRDC22b].

5.1 Fourier decomposition of variational algorithms

As we have outlined previously, there are many ways in which VQAs resemble the neural networks using in classical machine learning: they consist of a parameterised computational structure, they are trained using a cost (loss) function that measures

their performance, the training loop often uses gradient-based methods like gradient descent. This is largely a tautological observation since some of the first VQAs have been designed with neural networks in mind [FN18]. However the similarities go deeper and extend to the main results and outstanding questions in both domains. For instance, both suffer from an interpretability problem [PF23], since the computational structures are largely black boxes for the analysis of which we lack the sufficient tools, although for both heuristically tuning the architecture can lead to improved performance [WZdS$^+$20]. Importantly for our discussion, the cost function landscape for both is similarly poorly understood, with indications that it is highly nonconvex [BK21] and full of local minima [AK22] and flat regions [MBS$^+$18].

There exist however several peculiarities for each domain. For instance, classical neural networks typically employ nonlinearities, which are hard to introduce in the naturally linear quantum setting (but not impossible, see for instance Ref. [TMN$^+$21, MGV$^+$23]). In particular for the cost function and optimisation, large part of the success of classical neural networks can be attributed to the existence of backpropagation, which allows for rapid and parallelised computation of gradients even for very large neural networks. In contrast this is generally not available for VQAs [BWP23], and the best one can hope for in the quantum setting is methods like the parameter shift rule [SBG$^+$19] that improve on finite differences but still require to compute each gradient component separately.

The latter method is actually indicative of a broader phenomenon: despite the cost function of VQAs being very complex, the unitary structure, the intrinsic linearity and the kind of gates typically employed in PQCs allow to demonstrate results that have no parallel in classical neural networks. In this and the next Chapter we will explore arguably the most powerful characteristic of VQA cost functions, the fact that they can be decomposed as multivariate Fourier series or *trigonometric polynomials*. We shall use these terms interchangeably as they are essentially equivalent in our case. To be more precise, a Fourier series in the variables $\boldsymbol{\theta} \in \mathbb{R}^m$ takes the form:

$$f(\boldsymbol{\theta}) = \sum_{\boldsymbol{\omega} \in \Omega} c_{\boldsymbol{\omega}} e^{i\boldsymbol{\theta}\cdot\boldsymbol{\omega}}, \tag{5.1}$$

where $\boldsymbol{\omega} \in \mathbb{R}^m$ is the frequency, Ω is the frequency spectrum, a countable set which may be infinite, and $c_{\boldsymbol{\omega}} \in \mathbb{C}$ the Fourier coefficients. A (real, multivariate) trigonometric polynomial takes the form:

$$f(\boldsymbol{\theta}) = \sum_{\boldsymbol{k} \in \mathcal{K}} [a_{\boldsymbol{k}} \cos(\boldsymbol{k} \cdot \boldsymbol{\theta}) + b_{\boldsymbol{k}} \sin(\boldsymbol{k} \cdot \boldsymbol{\theta})], \tag{5.2}$$

where $a_{\boldsymbol{k}}, b_{\boldsymbol{k}} \in \mathbb{R}$ are the trigonometric coefficients. It is easy to show that for real functions, $c_{\boldsymbol{\omega}} = c^*_{-\boldsymbol{\omega}}$ and the two formulations are can be converted into each other. In this Thesis we shall also use alternative formulations for trigonometric polynomials that more directly relate to the quantum case. In particular when $\mathcal{K} \subseteq \mathbb{Z}_d^m$,[1] we may write:

$$f(\boldsymbol{\theta}) = \sum_{\boldsymbol{k} \in \mathcal{K}} d_{\boldsymbol{k}} \Phi_{\boldsymbol{k}}(\boldsymbol{\theta}), \tag{5.3}$$

where the *trigonometric monomials* $\Phi_{\boldsymbol{k}}(\boldsymbol{\theta})$ are:

$$\Phi_{\boldsymbol{k}}(\boldsymbol{\theta}) = \prod_{i=1}^{m} t^{|k_i|}_{\mathrm{sgn}(k_i)}(\theta_i), \quad t_s(\theta) = \begin{cases} 1 \text{ if } s = 0, \\ \cos(\theta) \text{ if } s = 1, \\ \sin(\theta) \text{ if } s = -1. \end{cases} \tag{5.4}$$

(Here we assume $\mathrm{sgn}(0) = 0$). This form is again equivalent and interchangeable with the previous ones.

We will denote by $\mathcal{T}_{m,d}$ the set of all real multivariate trigonometric polynomials on m variables with degree (maximum frequency for any parameter $\max_{\mathcal{K}} \|\boldsymbol{k}\|_\infty$) at most d.

5.1.1 Canonical approach

Several authors have observed the simple relation between expectation values of observables with respect to parameterised circuits and trigonometric functions or finite Fourier series [SSM21, VT18, GVT20, NFT20, PIOM19, OGB21, WIWL22]. In particular Schuld et al. in Ref. [SSM21] gave a general analysis that is used to explore the expressivity of ansätze for data-encoding. We shall summarise it here.

[1] In an abuse of notation we define $\mathbb{Z}_d := \{-d, -d+1, ..., d-1, d\}$.

The particular circuit involved repeated applications of the same parameterised unitary of the form:

$$U := V_0 U(\theta) V_1 U(\theta) ... U(\theta) V_m, \tag{5.5}$$

with one-parameter unitaries $U(\theta) = e^{i\theta H}$ generated by a Hermitian operator H acting on the Hilbert space corresponding to n qubits. The key idea is that H can be diagonalised as:

$$H = W \Sigma W^\dagger, \tag{5.6}$$

with $\Sigma = \text{diag}(\lambda_1, ..., \lambda_{2^n})$ the diagonal matrix of eigenvectors, and :

$$U(\theta) = W e^{i\theta\Sigma} W^\dagger, \tag{5.7}$$

with $e^{i\theta\Sigma} = \text{diag}(e^{i\theta\lambda_1}, ..., e^{i\theta\lambda_{2^n}})$. Therefore the expectation value of a target observable O is given by:

$$\begin{aligned} \text{Tr}(OU(\theta)\rho U(\theta)^\dagger) &= \text{Tr}(OWe^{i\theta\Sigma}W^\dagger \rho W e^{-i\theta\Sigma} W^\dagger) \\ &= \text{Tr}(\tilde{O} e^{i\theta\Sigma} \tilde{\rho} e^{-i\theta\Sigma}) \\ &= \sum_{ij} e^{i\theta(\lambda_j - \lambda_i)} [\tilde{O}]_{ij} [\tilde{\rho}]_{ji}, \end{aligned} \tag{5.8}$$

where $\tilde{O} := WOW^\dagger$ and $\tilde{\rho} := W\rho W^\dagger$ are the observable and input state in a rotated frame. Then we conclude that the expectation value must be a 1D Fourier series with frequency spectrum $\Omega = \{\lambda_j - \lambda_i\}_{ij}$. The coefficient for a frequency ω are given by summing all $[\tilde{O}]_{ij}[\tilde{\rho}]_{ji}$ such that $\omega = \lambda_j - \lambda_i$. As such, even though there are 2^n eigenvalues of H, the size of the spectrum and thus the Fourier series may be much smaller. For instance, when H is a Pauli operator the eigenvalues are $\lambda_i \in \{\pm 1\}$ giving just 3 frequencies in the spectrum $\omega \in \{-2, 0, 2\}$. This means that 3 coefficients completely characterise the cost function of a PQC with a single parameterised Pauli gate regardless of the rest of the circuit, as observed in Ref. [OGB21]. This can be generalised to PQCs with $m > 1$ parameters and generators, obtaining a multivariate

Fourier series the frequency spectrum of which is a finite lattice:

$$\Omega = \{\boldsymbol{\lambda}^{(j)} - \boldsymbol{\lambda}^{(i)}\}_{ij} \subset \mathbb{R}^m, \tag{5.9}$$

where $[\boldsymbol{\lambda}^{(j)}]_i$ is the jth eigenvalue of the ith generator.

5.1.2 Pauli paths

We now give an alternative derivation of the trigonometric form of circuits with m independent parameters. This approach is conceptually different from the one outlined above, as we now decompose the target unitary operation into a linear combination of quantum channels that separates the dependency on the parameterisation from the specific circuit structures involved. This is related to sum-over-paths approaches to classical simulation of circuits [BBC+19, HL21]. Crucially, the channel-based perspective is particularly well-suited to analyse the effects of noise.

First consider a single-qubit system with some input state, a z-rotation $R_z(\theta) = e^{-iZ\theta/2}$ and a Pauli measurement. Shifting to the Heisenberg picture, we consider the conjugate channel $\mathcal{R}_z^\dagger(\theta) := R_z^\dagger(\theta_j)(\cdot)R_z(\theta_j)$. The result of its application onto a Pauli operator P is:

$$\mathcal{R}_z^\dagger(\theta)(P) = \begin{cases} P & \text{if } P = I, Z, \\ \cos(\theta)P \pm \sin(\theta)P' & \text{otherwise.} \end{cases}, \tag{5.10}$$

where P' is another Pauli operator. Specifically, if $P = X$ then we gain a negative sign and $P' = Y$, while if $P = Y$ then we get a positive sign and $P' = X$.

This can be used to prove the following result:

Lemma 3 (Trigonometric form of expectation values)**.** *Consider a parameterised quantum unitary of the form:*

$$U(\boldsymbol{\theta}) := \prod_{j=1}^{m} V_j U_j(\theta_j)\, V_0, \tag{5.11}$$

and the state resulting from applying the corresponding channel to a fixed initial state: $\rho(\boldsymbol{\theta}) = \mathcal{U}(\boldsymbol{\theta})(\rho_0)$. Now suppose each of the one-parameter unitaries is generated by Pauli operators $U(\theta_j) = e^{-iP_j\theta_j/2}$, and in addition the unparameterised unitaries V_i are

Clifford and the input state ρ_0 is a stabiliser state.

Then the Fourier coefficients of the trigonometric polynomial $\mathrm{Tr}(\rho(\boldsymbol{\theta})P)$ *where P is a Pauli operator are* $c_{\mathbf{k}} \in \{\pm 1/2^{||\mathbf{k}||_1}, \pm i 1/2^{||\mathbf{k}||_1}, 0\}$*, and the spectrum is* $\mathcal{K} \subseteq \{-1, 0, 1\}^m$*, so* $\mathrm{Tr}(\rho(\boldsymbol{\theta})P) \in \mathcal{T}_{m,1}$*. Furthermore, the expansion in terms of trigonometric functions has coefficients* $d_{\mathbf{k}} \in \{0, \pm 1\}$*.*

Proof. It is most direct to prove first that the expectation value $\mathrm{Tr}(\rho(\boldsymbol{\theta})P)$ is in the given trigonometric form. To do this, first consider that every parameterised operator $U(\theta_j) = e^{-iP_j\theta_j/2}$ is a Pauli gadget, and as such it can be decomposed into one single-qubit parameterised Z-rotation and Clifford gates, which can be absorbed into the fixed gates. Therefore, the circuit consists entirely of fixed Clifford gates V' and parameterised single-qubit Z-rotations:

$$U(\boldsymbol{\theta}) = V_0' Z_1(\theta_1) V_1' Z_2(\theta_2) V_2' \dots Z_m(\theta_m) V_m'. \tag{5.12}$$

Now consider the channel form of the unitary as it appears in the calculation of the expectation value $\mathrm{Tr}(P\mathcal{U}(\boldsymbol{\theta})(\rho_0))$, which in the Heisenberg picture becomes $\mathrm{Tr}(\mathcal{U}^\dagger(\boldsymbol{\theta})(P)\rho_0)$. Every step of the calculation of the expectation value then consists in the successive application of either a conjugate Z-rotation channel or a Clifford channel. Clifford channels by definition normalise the Pauli group and hence turn a Pauli operator into another Pauli operator. On the other hand Z-rotation channels may, depending on the operator, produce a linear combination of operators with trigonometric coefficients, as seen before. The coefficients are always either sin or cos of the rotation angle, with potentially a negative sign. Therefore, at the end of the procedure, one is left with a sum of Pauli operators:

$$\mathcal{U}^\dagger(\boldsymbol{\theta})(P) = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} \pm \Phi_{\mathbf{k}} P_{\mathbf{k}}, \tag{5.13}$$

where $\boldsymbol{\Lambda} = \{-1, 0, 1\}^m$ and $\Phi_{\boldsymbol{k}}$ is a trigonometric monomial. As a final step, the expectation value of this operator is taken with respect to the stabiliser ρ_0 state, which for each term yields either 0 or ± 1. One is therefore left with a function of the wanted form.

To express the trigonometric function as a Fourier series, each trigonometric mono-

mial can be expanded in complex exponentials, giving:

$$\mathrm{Tr}(P\mathcal{U}(\boldsymbol{\theta})(\rho_0) = \sum_{\boldsymbol{k}\in\boldsymbol{\Lambda}} c_{\boldsymbol{k}} e^{i\boldsymbol{\theta}\cdot\boldsymbol{k}}. \tag{5.14}$$

where the coefficients are $c_{\mathbf{k}} \in \{\pm 1/2^{||\mathbf{k}||_1}, \pm i 1/2^{||\mathbf{k}||_1}, 0\}$, where $||\boldsymbol{k}||_1 = \sum_i |\boldsymbol{k}_i|$. This is because each sine or cosine in a monomial $\Phi_{\mathbf{k}}$ contributes a factor of $1/2 \times \{1, -1, i, -i\}$, and there are $||\mathbf{k}||_1$ sines and cosines in $\Phi_{\mathbf{k}}$. As the function is real, by standard Fourier analysis $c_{-\boldsymbol{k}} = c^*_{\boldsymbol{k}}$. □

5.1.3 Process modes

In summary, we saw that we can recover the trigonometric form of the cost function by working in the Heisenberg picture and considering how a Pauli observable is converted by Pauli rotations into more Pauli operators with trigonometric coefficients. At most, every application of a rotation channel results in doubling the Pauli support of the total operator. Therefore we can view the computation as a tree, where every leaf corresponds to a Pauli operator weighted by a trigonometric monomial. Alternatively we can see the computation as branching into different *Pauli paths* for every rotation channel. This concept was explored before in different contexts, for example in Refs. [GD18, AGL$^+$22] for the simulation of random circuit sampling, where Haar-random two-qubit unitaries take the place of the parameterised rotation gates.

Here we aim to examine these computational paths from the point of view of the channel itself. Take again the rotation channel $\mathcal{R}_z(\theta)$. We can decompose it as a linear combination of Clifford unitary channels:

$$\mathcal{R}_z(\theta_j)(\rho) = p_{\mathcal{I}}(\theta_j)\rho + p_{\mathcal{Z}}(\theta_j) Z\rho Z^\dagger + p_{\mathcal{S}}(\theta_j) S\rho S^\dagger, \tag{5.15}$$

where $S = \sqrt{Z}$ and coefficients given by:

$$p_{\mathcal{I}}(\theta) = \frac{1}{2}(-\sin\theta + \cos\theta + 1) \tag{5.16}$$

$$p_{\mathcal{Z}}(\theta) = \frac{1}{2}(-\sin\theta - \cos\theta + 1) \tag{5.17}$$

$$p_{\mathcal{S}}(\theta) = \sin\theta. \tag{5.18}$$

These may be seen as quasi-probabilities as they sum to 1 (thus preserving the trace of ρ) but can be negative. Alternatively we can rearrange into a different equation that makes immediate the connection to Fourier series:

$$\mathcal{R}_z(\theta_j)(\rho) = \mathcal{C}_0 + e^{i\theta}\mathcal{C}_1 + e^{-i\theta}\mathcal{C}_{-1}, \tag{5.19}$$

with $\mathcal{C}_0, \mathcal{C}_{\pm 1}$ linear combinations of Clifford unitary channels as follows:

$$\mathcal{C}_0 = \frac{1}{2}\mathcal{I} + \frac{1}{2}\mathcal{Z} \tag{5.20}$$

$$\mathcal{C}_1 = \frac{1+i}{4}\mathcal{I} - \frac{1-i}{4}\mathcal{Z} - \frac{i}{2}\mathcal{S} \tag{5.21}$$

$$\mathcal{C}_{-1} = \frac{1-i}{4}\mathcal{I} - \frac{1+i}{4}\mathcal{Z} + \frac{i}{2}\mathcal{S}. \tag{5.22}$$

Where $\mathcal{Z}(\cdot) = Z(\cdot)Z^\dagger$ and $\mathcal{S}(\cdot) = S(\cdot)S^\dagger$. Following previous work [CJ17], we call these channels *process modes*.

We can easily generalise to parameterised unitaries taking the form $U(\theta_j) = e^{-iP_j\theta_j/2}$ for some n qubit Pauli operator P_j, as they can be decomposed into Cliffords and a single-qubit z-rotation. Therefore we can easily show that given any P_j there exist channels $\{\mathcal{C}_k^{(j)}\}_{k=0,\pm 1}$ such that:

$$\mathcal{U}(\theta_j) = \mathcal{C}_0^{(j)} + e^{i\theta_j}\mathcal{C}_1^{(j)} + e^{-i\theta_j}\mathcal{C}_{-1}^{(j)}. \tag{5.23}$$

Now we extend to circuits $U(\boldsymbol{\theta})$ with parameterised unitaries generated by arbitrary Hamiltonians, and arbitrary non-parameterised gates:

$$U(\boldsymbol{\theta}) = V_0 U_1(\theta_1) V_1 ... U_m(\theta_m) V_m, \quad U_j(\theta_j) = e^{i\theta_j H_j}. \tag{5.24}$$

In such a case the resulting expected values with respect to $\boldsymbol{\theta}$ will be described by multivariate trigonometric polynomials with possibly higher frequencies, i.e. some components of the frequency vector $\mathbf{k}$ could have absolute value larger than 1. We seek to determine the allowed values for the frequencies and the Fourier coefficients in terms of operations derived from the circuit structure.

In this general setting, by the results in Sec. 5.1.1 we know that the frequency spectrum associated with a particular unitary is discrete and dictated by the spectrum of its generating Hamiltonian. Therefore we expect that each unitary channel decomposes into a number of process modes. We will see that by simply concatenating the channels for the individual gates expressed as combinations of process modes, we can express $\mathcal{U}(\boldsymbol{\theta})$ as a linear combination of operations that isolate the parameter dependency:

$$\mathcal{U}(\boldsymbol{\theta}) = \sum_{\boldsymbol{k}} e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}}\, \mathcal{V}_0 \circ \mathcal{U}_{k_1} \circ \mathcal{V}_1 \circ ... \circ \mathcal{U}_{k_m} \circ \mathcal{V}_m, \tag{5.25}$$

where $\mathcal{U}_{k_j}$ is the k_jth process mode of channel $\mathcal{U}_j = U_j \cdot U_j^\dagger$. Now we recognise in the summands an analogue of the Pauli paths from the point of view of channels, and therefore we call this result *process path decomposition*.

In the following Theorem we show how the decomposition arises, and how it directly gives the trigonometric polynomial form taken by expectation values; to emphasise their dependency on the measured observable, we employ the notation $c_{\mathbf{k}}(O)$ for the coefficients.

Theorem 5. *Consider a parameterised quantum state on n qubits with m parameters $\rho(\boldsymbol{\theta}) = U(\boldsymbol{\theta})\rho_0 U(\boldsymbol{\theta})^\dagger$ with unitary as in Eq.* (5.24)*. The expectation value of any Hermitian operator O with respect to $\rho(\boldsymbol{\theta})$ is a generalised m-variate trigonometric polynomial of bounded degree:*

$$\langle O \rangle_{\rho(\boldsymbol{\theta})} = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} c_{\mathbf{k}}(O) e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}}, \tag{5.26}$$

where $c_{-\mathbf{k}} = c_{\mathbf{k}}^*$ and the lattice $\boldsymbol{\Lambda} \subset \mathbb{Z}_1^{r_1} \times ... \times \mathbb{Z}_1^{r_m}$ where each r_j is the rank of

the Walsh-Hadamard transform of the eigenvalue vector of H_j[2]. *Furthermore, the frequency vector* $\boldsymbol{\omega}_{\mathbf{k}} = ((\omega_{\mathbf{k}})_1, \ldots, (\omega_{\mathbf{k}})_m)$ *ranges over a discrete set with bounded degree* $\sup_{\mathbf{k}\in\boldsymbol{\Lambda}} |(\omega_{\mathbf{k}})_j| \leqslant 2||H_j||_\infty$.

Proof. From Ref. [WGMAG14] we have that any diagonal operator has a decomposition in terms of the diagonal Pauli operators acting only with Z or I:

$$\sum_z \lambda_z |z\rangle\langle z| = \sum_{\mathbf{j}=(j_1,\ldots,j_n)} a_{\mathbf{j}} Z^{j_1} \otimes \ldots \otimes Z^{j_n}. \tag{5.27}$$

Each j_i is a binary variable with $\mathbf{j} := (j_1, \ldots, j_n)$ and the vector $\mathbf{a} := (a_0, a_1, \ldots, a_{2^n-1})$ is the Walsh-Hadamard transform of the eigenvalue vector $\boldsymbol{\lambda} = (\lambda_z : z \in \{0,1\}^n)$:

$$a_{\mathbf{j}} = \frac{1}{2^n} \sum_{z=0}^{2^n-1} \lambda_z (-1)^{\mathbf{z}\cdot\mathbf{j}}, \tag{5.28}$$

with inverse given by:

$$\lambda_z = \sum_{\mathbf{j}} a_{\mathbf{j}} (-1)^{\mathbf{z}\cdot\mathbf{j}}. \tag{5.29}$$

Therefore, the action of the unitary:

$$U_i(\theta_i) = e^{iH\theta_i} = W \prod_{\mathbf{j}:a_{\mathbf{j}}\neq 0} e^{ia_{\mathbf{j}} Z^{j_1}\otimes\ldots\otimes Z^{j_n}\theta_i} W^\dagger \tag{5.30}$$

upon diagonalisation can be decomposed into a series that involves replacing the action of each *phase gadget* $Z^{\mathbf{j}}(\theta) = e^{ia_{\mathbf{j}} Z^{j_1}\otimes\ldots\otimes Z^{j_n}\theta}$ that has, as we have seen in Eq. (5.23), a process modes decomposition:

$$\mathcal{Z}^{\mathbf{j}}(\theta) = \mathcal{C}_0^{\mathbf{j}} + e^{2i\theta a_{\mathbf{j}}} \mathcal{C}_{-1}^{\mathbf{j}} + e^{-2i\theta a_{\mathbf{j}}} \mathcal{C}_1^{\mathbf{j}}. \tag{5.31}$$

Therefore:

$$\mathcal{U}_i(\theta_i) = \sum_{\mathbf{k}\in\{0,\pm 1\}^r} e^{i\theta_i \omega_{\mathbf{k}}^i} \mathcal{W} \circ \mathcal{C}_{\mathbf{k}}^i \circ \mathcal{W}^\dagger, \quad \mathcal{C}_{\mathbf{k}}^i = \mathcal{C}_{k_1}^{\mathbf{j}_1} \circ \ldots \circ \mathcal{C}_{k_r}^{\mathbf{j}_r}, \tag{5.32}$$

[2]The 2^m-dim Walsh-Hadamard transform matrix can be defined as $[H_m]_{ij} := \frac{1}{2^{m/2}}(-1)^{i\cdot j}$ where the $i \cdot j$ indicates the bitwise dot product between the binary representations of the indices.

with $r = |\{a_{\mathbf{j}} : a_{\mathbf{j}} \neq 0\}|$ is the number of phase gadgets that compose the unitary, $\omega^i_{\mathbf{k}} = 2\mathbf{a}\cdot\mathbf{k}$ and $\mathbf{a}$ is the vector of the nonzero $a_{\mathbf{j}}$. Each vector $\mathbf{k}$ has entries in $\{-1, 0, 1\}$ and labels a particular process mode. Clearly $\omega^i_{\mathbf{k}}$ is maximised when $\mathbf{k} \in \{\pm 1\}^r$, when this occurs then $\mathbf{a}\cdot\mathbf{k}$ is an element of the inverse Walsh-Hadamard transform of $\mathbf{a}$, and so is an eigenvalue of H_j. Therefore $|\omega^i_{\mathbf{k}}| \leq 2\|H_j\|_\infty$.

Now to obtain the full decomposition of the channel $\mathcal{U}(\boldsymbol{\theta})$ we simply concatenate the decomposed $\mathcal{U}_i(\theta_i)$ and the $\mathcal{V}_i$ channels, obtaining process paths that are weighted by a trigonometric coefficient $e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}}$. Applying each path to the starting state and measuring with the observable yields a complex number for each path

$$c_{\mathbf{k}}(O, \rho_0) = \mathrm{Tr}[O\mathcal{V}_0 \circ \mathcal{W} \circ \mathcal{C}^0_{\mathbf{k}} \circ \mathcal{W}^\dagger \circ ... \circ \mathcal{W} \circ \mathcal{C}^m_{\mathbf{k}} \circ \mathcal{W}^\dagger \circ \mathcal{V}_m(\rho_0)], \tag{5.33}$$

which obeys $c_{-\mathbf{k}} = c^*_{\mathbf{k}}$ since the function must be real.

□

The decomposition of $\mathcal{U}$ in the form of Eq. (5.32), under certain circumstances, may in fact naively overcount the number of frequencies $\omega_{\mathbf{k}}$. This comes from the fact that phase gadgets commute and particular compositions of process modes of the type $\mathcal{C}^{\mathbf{j}}_k$ give a null channel, so the corresponding term in Eq. (5.32) vanishes. Also note that generally, the frequencies are not uniquely defined by a given path $\mathbf{k}$: It may be the case that two different paths give $\omega_{\boldsymbol{k}} = \omega_{\boldsymbol{k}'}$, in which case the corresponding coefficient in the Fourier transform will be $c_{\boldsymbol{k}} + c_{\boldsymbol{k}'}$. Therefore in the Fourier-transform frequency spectra $\mathcal{S}$ the two paths will have indistinguishable contributions.

This analysis involving process modes may be more suitable than the expressions in Eq. (5.8) to determine the allowed frequencies and coefficients in the expectation value in the following situations: when one does not have direct access to the eigenvalues; when the parameterised unitary is given as a quantum circuit, or when the eigenvalues have higher multiplicities where it provides a more compact form for the coefficients. Our focus in this Chapter is ultimately on analysing and mitigating the effects of quantum noise on expectation values of observables viewed as trigonometric polynomials, where such a quantum channel perspective is more suitable.

5.2 Error diagnostics from Fourier spectra

5.2.1 Fourier spectra of quantum channels

In the previous section we showed how a quantum channel approach can be used to derive the Fourier spectra of parameterised unitary circuits, via a process mode decomposition. We now drop the unitary requirement, and discuss Fourier spectra of general quantum channels that may represent noisy implementations. The previous decomposition of unitary operations into process modes can also be seen as a consequence of the following general result.

Lemma 4 ([BFH$^+$17])**.** *Any trace-preserving quantum channel χ on n qubits can be expressed as a linear combination of Clifford unitary channels $\mathcal{C}_i$ and Pauli reset channels $\mathcal{R}_j$:*

$$\chi = \sum q_i \mathcal{C}_i + \sum q'_j \mathcal{R}_j, \tag{5.34}$$

for $q_i, q'_i \in \mathbb{R}$. Trace preservation is equivalent to $\sum_i q_i + \sum_j q'_j = 1$, and the S_j terms are only required for non-unital channels[3]*. The decomposition is not unique.*

A *Pauli reset channel* maps any input state to a specific fixed eigenstate of a Pauli operator, conditional to a measurement by the same operator. An example is a single qubit channel that measures in the Z basis and, if the eigenvalue -1 is measured (state is $1\rangle$), applies a X to reset to the $|0\rangle$ state. This specific channel appears, for instance, in the decomposition of the amplitude damping channel.

In general, the Fourier spectrum of an expectation value evaluated on a parameterised state produced by a noisy channel will differ from that of a noiseless one in two distinct ways: the Fourier coefficients might be different, and the frequency support might vary.

Theorem 6. *Writing $\tilde{\rho}(\boldsymbol{\theta}) = \chi(\rho(\boldsymbol{\theta}))$ for the noisy implementation of the target state*

[3]Note that there is no simple rule to determine if the resulting linear map will be completely positive, and therefore physical.

$\rho(\boldsymbol{\theta})$, and allowing the noise channel χ to depend on $\boldsymbol{\theta}$, we have:

$$\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})} = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} \sum_{j} q_j(\boldsymbol{\theta}) c_{\mathbf{k}}(\mathcal{C}_j^\dagger(O)) e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}} + q'(\boldsymbol{\theta}), \tag{5.35}$$

where $q_j(\boldsymbol{\theta})$ and $q'(\boldsymbol{\theta})$ capture the dependency on $\boldsymbol{\theta}$ of the Clifford unitary and Pauli reset channels respectively.

Proof. One may formally write $\tilde{\mathcal{U}}(\boldsymbol{\theta}) = \mathcal{E}\circ\mathcal{U}(\boldsymbol{\theta})$, where the noise channel $\mathcal{E}$ may itself be dependent on the angles $\boldsymbol{\theta}$. One may then employ decompositions as in Lemma 4 for $\mathcal{E}$ so that only the quasiprobability distribution carries an angle dependency, namely:

$$\mathcal{E} = \sum_j q_j(\boldsymbol{\theta})\, \mathcal{C}_j + \sum_j q'_j(\boldsymbol{\theta}) \mathcal{R}_j. \tag{5.36}$$

Therefore the noisy expected values take the form:

$$\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})} = \sum_j q_j(\boldsymbol{\theta}) \mathrm{Tr}[\rho(\boldsymbol{\theta})\mathcal{C}_j^\dagger(O)] + \sum_j q'_j(\boldsymbol{\theta}) \langle P_j|O|P_j\rangle, \tag{5.37}$$

where we have used that the adjoint Pauli reset channel is $\mathcal{R}_j^\dagger(O) = \mathbb{I}\,\langle P_j|O|P_j\rangle$ for some eigenstate $|P_j\rangle$ of a Pauli operator. Following the analysis in the previous section, $\mathrm{Tr}[\rho(\boldsymbol{\theta})\mathcal{C}_i^\dagger(O)] = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} c_{\mathbf{k}}(\mathcal{C}_i^\dagger(O)) e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}}$ where we explicitly add the dependency of the coefficients $c_{\mathbf{k}}$ on the observable. This implies that:

$$\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})} = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} \sum_{i} q_i(\boldsymbol{\theta}) c_{\mathbf{k}}(\mathcal{C}_i^\dagger(O)) e^{i\boldsymbol{\omega}_{\mathbf{k}}\cdot\boldsymbol{\theta}} + \sum_j q'_j(\boldsymbol{\theta}) \langle P_j|O|P_j\rangle. \tag{5.38}$$

The second summation can be absorbed into the $\omega_{\boldsymbol{k}} = 0$ term giving the intended result.

Let us denote by $\mathcal{S}$ the (maximal) frequency support for the Fourier decomposition of the unitary channel $\mathcal{U}(\boldsymbol{\theta})$. For any given observable, the frequency spectrum of $\langle O\rangle_{\rho(\boldsymbol{\theta})}$ will be included in $\mathcal{S}$, but may contain fewer modes. As we have seen in the previous section, for each independent parameter θ_i the frequency spectrum is included in a finite lattice $\mathcal{S}_i := \{\omega_{\mathbf{k}_i} = -2\mathbf{a}_i\cdot\mathbf{k}_i : \mathbf{k}_i \in \mathbb{Z}_3^{\times r_i}\}$ over $\mathbb{Z}_3$ with real basis vectors

$-2\mathbf{a}_i$ which can be determined from the circuit description of $U_i(\theta_i)$. On the other hand, there's no a priori constraint for q_i, q_i' to have discrete spectra since the noise channel may not present periodicity in the angles. □

Recall that in the noiseless case, a discrete set of frequencies suffices, corresponding to the points of the lattice $\Lambda \subset \mathbb{R}^m$. In the noisy case, the expectation value $\langle O \rangle_{\tilde{\rho}(\boldsymbol{\theta})}$ is still a function of $\boldsymbol{\theta}$, but it may no longer have a discrete Fourier series. As mentioned, this would happen if any $q_i(\boldsymbol{\theta}), q_i'(\boldsymbol{\theta})$ is not a periodic function of $\boldsymbol{\theta}$ and therefore cannot be written as a discrete Fourier series. Therefore we define the *noisy frequency spectrum* as:

$$\tilde{\mathcal{S}} := \{\tilde{\boldsymbol{\omega}} \in \mathbb{R}^m : \mathcal{F}[\langle O \rangle_{\tilde{\rho}}](\tilde{\boldsymbol{\omega}}) \neq 0\}, \tag{5.39}$$

where $\mathcal{F}$ denotes the Fourier transform, computed as:

$$\mathcal{F}[\langle O \rangle_{\tilde{\rho}}](\tilde{\boldsymbol{\omega}}) = \int \langle O \rangle_{\tilde{\rho}(\boldsymbol{\theta})} e^{i\tilde{\boldsymbol{\omega}} \cdot \boldsymbol{\theta}} d\,\boldsymbol{\theta}. \tag{5.40}$$

The difference between $\mathcal{S}$ and $\tilde{\mathcal{S}}$ depends on the specific noise parameters $q_j(\boldsymbol{\theta})$, and different channels have distinct effects on the spectra. In the following section we illustrate this for several noise models based on varying physical assumptions.

5.2.2 Effect of decoherent channels

Consider a noisy implementation where each unitary gate operation is followed by a noise channel that is independent of the parameters:

$$\tilde{\mathcal{U}} = \tilde{\mathcal{V}}_0 \circ \mathcal{U}_1(\theta_1) \circ \tilde{\mathcal{V}}_1 \circ \cdots \circ \mathcal{U}_m(\theta_m) \circ \tilde{\mathcal{V}}_m. \tag{5.41}$$

It is clear that for any fixed observable $\tilde{\mathcal{S}} \subseteq \mathcal{S}$, in other words the range of frequencies remains the same. However the magnitude of the Fourier coefficients might be different. Stronger assertions are possible if we specialise to particular cases.

For example, if each noisy gate $\mathcal{V}_i$ is affected by a (global) depolarising channel with parameter p, then one finds that the noisy Fourier coefficients $\tilde{c}_{\mathbf{k}}$ are homogeneously

contracted:

$$\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})} = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} (1-p)^{m+1} c_{\mathbf{k}}(O) e^{i\omega_{\mathbf{k}}\cdot\boldsymbol{\theta}} + \frac{(1-(1-p)^{m+1})}{2^n}\mathrm{Tr}(O). \tag{5.42}$$

This can be seen as a special case of the following result.

Lemma 5 (Pauli channel)**.** *Let $\tilde{U}$ be as in Eq.* (5.41) *and suppose each $\tilde{\mathcal{V}}_i$ is a Clifford unitary followed by a Pauli channel*[4]*, and each $\mathcal{U}_i$ is a Pauli gadget; then, for any Pauli observable O, the coefficients are contracted for all paths so that $|\tilde{c}_{\mathbf{k}}(O)| \leq |c_{\mathbf{k}}(O)|$.*

The proof can be found in Appendix B.2. For general observables O and non-Clifford $\mathcal{V}_i$, some modes that were previously zero due to cancellation of paths might find themselves with a non-zero value under a Pauli noise model.

In Appendix B.2, we provide a comprehensive analysis of noise channels and their effect on Fourier coefficients, summarised in Table 5.1. Importantly, the effect of coherent, non-unital and parameter-dependent channels is detectable from the Fourier spectrum as additional modes. Coherent channels may create additional modes by adding new process paths. For instance, consider a small single-qubit Z-rotation: this would create a two paths with an additional $\mathcal{Z}$ or $\mathcal{S}$ channel. While the channel with $\mathcal{Z}$ behaves like a Pauli-Z error and therefore does not create additional coefficients, the channel with $\mathcal{S}$ might. For the case of non-unital channels, consider amplitude damping, which adds a new path with a reset channel. This is equivalent to resetting a qubit to $|0\rangle$, which is like adding a path with a different circuit, and therefore potentially different frequencies. Parameter-dependent channels may lead to modes with frequencies outside the theoretical spectrum $\mathcal{S}$. A simple example is an error channel that behaves like a single-qubit rotation that depends on a parameter, which by being a rotation would then add frequencies to the spectrum outside the range of frequencies that one could infer from the circuit alone.

Finally, correlated noise falls under one of these categories depending on the specific

[4] A *Pauli channel* is an incoherent convex combination of Pauli errors.

Noise channel	Effect on Fourier coefficients	$\tilde{\mathcal{S}} \subseteq \mathcal{S}$?
Depolarising (Lemma 10)	Uniform contraction	✓
Pauli, SPAM (Lemma 11)	Contraction (O Pauli and $\mathcal{V}$ Clifford)	✓
Aligned Pauli (Lemma 12)	Uniform contraction (any O and $\mathcal{V}$)	✓
Coherent (corollary of Lemma 8)	Different coefficients but conserved norm of coefficient vector $\tilde{\boldsymbol{c}}$	✓
Non-unital (Lemma 13)	Different coefficients, new modes with subset of parameters	✓
Parameter-dependent (Lemma 14)	Different coefficients, new modes outside theoretical spectrum	✗

Table 5.1: Effect of different noise channels on Fourier coefficients of the landscape. Proofs can be found in Appendix B.2. Rightmost column indicates whether the noisy spectrum $\tilde{\mathcal{S}}$ is included in the theoretical noiseless spectrum $\mathcal{S}$.

type. Correlated noise across qubits that can be described as multi-qubit Pauli noise will give the same effect of any Pauli noise channel. Noise that is correlated to some rotation gate would fall under parameter-dependent noise.

5.2.3 Shot noise

Overall the analysis in previous sections confirms that noise is capable of expanding the frequency spectrum. With knowledge of the exact, noiseless spectrum, one thus has a direct way of quantifying the impact of noise on a quantum variational landscape. In practical terms, an implementation of error diagnosis from Fourier data would involve performing a discrete Fourier transform (DFT) by sampling parameter values. Since the number of points determines the resolution at which the frequencies in the noisy spectrum can be determined, the grid should be fine enough to be able to resolve frequencies beyond $\mathcal{S}$. In Appendix B.1 we discuss possible approaches to optimising the number of calls to the quantum computer.

These sampling requirements are in turn different to those needed to resolve the individual points of the landscape with sufficient accuracy. We consider this requirement in the next section.

Typically, for DFT the sampled points $\boldsymbol{\theta_i}$ are equally spaced on a grid $\boldsymbol{\Theta} = \frac{2\pi}{d}\mathbb{Z}_d^{\times m}$, where $d = 2N + 1$ for N the largest frequency resolvable. For each sampled point, the corresponding circuit $U(\boldsymbol{\theta}_i)$ is repeated n_s times to construct sample mean estimators

$\tilde{x}_i$ for the expectation value x_i. From the central limit theorem, the standard deviation due to a finite number of shots scales as $1/\sqrt{n_s}$ and using properties of the DFT a similar scaling occurs in estimating the Fourier coefficients under shot noise. The full proof is in Appendix B.2.1.

Lemma 6 (Shot noise)**.** *Given a number of shots n_s, each Fourier coefficient is normally distributed around the noiseless mean with standard deviation:*

$$\sigma(Re\,(\mathcal{F}[\tilde{x}]_{\boldsymbol{k}})) = \frac{\sqrt{1-||x||_2^2/d^m}}{\sqrt{2n_s d^m}}, \tag{5.43}$$

and similarly for the imaginary part.

Two observations can be made on this result: first, the standard deviation of each Fourier coefficient due to shot noise is independent of the coefficient's frequency vector $\boldsymbol{k}$, and second, the dependence on d^m means that increasing the resolution in the frequency spectrum or the number of parameters varied reduces the shot noise in the Fourier coefficients with a scaling of $1/O(\sqrt{d^m})$. This is because increasing the resolution and/or the parameters increases the number of grid points and therefore the total number of samples that go into estimating the Fourier coefficients.

5.2.4 Figures of merit from Fourier spectrum

5.2.4.1 Power spectra

If a frequency spectrum is known to be included into $\mathcal{S}$, one can quantify the extra noise-induced modes from an experimentally measured spectrum $\tilde{\mathcal{S}}$ using the signal power (or energy).

Definition 14 (Quality measures)**.** *The **power in the extra modes** due to noise is given by:*

$$P_N := \sum_{\boldsymbol{\omega}\in\tilde{\mathcal{S}}/\mathcal{S}} |\tilde{c}_{\boldsymbol{\omega}}|^2, \tag{5.44}$$

*while the **signal power** will be given by*

$$P_S := \sum_{\boldsymbol{\omega}\in\mathcal{S}} |\tilde{c}_{\boldsymbol{\omega}}|^2. \tag{5.45}$$

*The **total power** $P := P_N + P_S$ is proportional, via Parseval's theorem, to the power of the real-space signal $\sum_i |x_i|^2$, where the sum is over all grid points $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$ and x_i is the sample expectation value at point $\boldsymbol{\theta}_i$. One can also define the observed **signal-to-noise ratio** (SNR) as $SNR := P_S/P_N$.*

Distinguishing noise from the signal will depend on the knowledge of the spectrum $\mathcal{S}$, as well as the size of the experimentally measured spectrum $\tilde{\mathcal{S}}$. To maximise the distinction, it is beneficial to identify $\mathcal{S}$ as small as possible, which might be possible for specific circuit structures. In Sec. 5.3.2 we show how the noiseless spectrum can be refined for the case of QAOA. However, for general circuits one is confined to the results of Theorem 5.

5.2.4.2 Average fidelity and purity

A quantity of practical relevance for benchmarking that can be extracted from the Fourier representation is the average fidelity over all sampled parameter values $\langle F\rangle_{\boldsymbol{\theta}}$, equal to $\mathcal{F}[F]_{\mathbf{0}}$. Let us represent the Fourier coefficients over all Pauli operators in the vectorised notation $\boldsymbol{c}$, where $[\boldsymbol{c}]_{\boldsymbol{k},i} = c_{\boldsymbol{k}}(P_i)$. In Appendix B.3 we show the following result:

Lemma 7 (Average fidelity)**.** *The average fidelity of the output state over all sampled parameters is proportional to the inner product of the vectors of Fourier coefficients:*

$$\langle F\rangle_{\boldsymbol{\theta}} = \int F(\tilde{\rho}(\boldsymbol{\theta}), \rho(\boldsymbol{\theta}))\, d\,\boldsymbol{\theta} = \frac{1}{2^n}\boldsymbol{c}^{\,\dagger}\tilde{\boldsymbol{c}}, \tag{5.46}$$

where $F(\tilde{\rho}(\boldsymbol{\theta}), \rho(\boldsymbol{\theta}))$ is the fidelity between the target state ρ and the noisy implementation $\tilde{\rho}$.

From this it is clear that the Fourier modes in $\tilde{\boldsymbol{c}}$ that fall outside the support of $\boldsymbol{c}$ do not contribute to the average fidelity. In other words, additional modes such as those

produced by parameter-dependent errors are not captured by this averaged measure, but nonetheless will affect fidelity of individual circuits.

Of course, in a practical setting it is unrealistic to measure all Pauli observables, while direct fidelity estimation [FL11] requires at worst $\Omega(2^n)$ measurements, for many classes of states resource requirements are reduced, and one could also further employ estimation of multiple observables using classical shadows to reduce the computational burden [HKP20].

Similarly, the average purity can be related to the norm of the noisy Fourier coefficient vector:

Lemma 8 (Average purity)**.** *The average purity of the output state $\tilde{\rho}(\boldsymbol{\theta})$ over all sampled parameters is proportional to the squared norm of the vector of Fourier coefficients:*

$$\langle P \rangle_{\boldsymbol{\theta}} = \frac{1}{2^n} |\tilde{\boldsymbol{c}}|^2. \tag{5.47}$$

The proof can be found in Appendix B.3. The relation between purity and spectral power is therefore that the average power over all Pauli operators is proportional to the average purity of the state over all sampled parameters. As a corollary, coherent noise channels in the preparation circuit preserve purity of the final state and therefore the norm of $\tilde{\boldsymbol{c}}$.

Furthermore, since coherent errors preserve purity of the state, it must be that under coherent noise channels the norm of $\boldsymbol{c}$ is preserved. Vice versa, decoherent channels will decrease purity, and therefore lead to a contraction in the components of $\boldsymbol{c}$, consistent with the analysis presented in Appendix B.2.

5.2.5 Experiments

5.2.5.1 Simple circuit

In the following we apply spectral noise evaluation to example circuits run on real quantum devices. The first circuit examined is a two-qubit circuit, consisting of two parameterised R_y rotations, and the measurement operator is the parity operator ZZ. The parameterisation is the canonical one, therefore the noiseless expectation value

is $f(\theta_1, \theta_2) = \cos(\theta_1)\cos(\theta_2)$, which corresponds to 4 nonzero Fourier coefficients with maximum absolute frequency 1. Therefore, to study the presence of higher frequency modes we sample up to a maximum absolute frequency of 2, giving a 5x5 grid. The experiment was run on a superconducting device (`ibmq_lima`).

To assess the effect of noise on the frequency spectrum, after state preparation and prior to measurement we introduce a series of identity gates, composed of two successive CNOT gates. These would cause an increase in noise by extending the depth and therefore computation time, increasing decoherence, and by being noisy themselves. Furthermore, we note that error mitigation via extrapolation for digitised computations [GTHL$^+$20] involves introducing such fiducial gates to artificially boost noise levels, and so the analysis here may also be used to test noise assumptions involved in deriving fitting functions.

In Fig. 5.1 we show the total power in noise modes, defined above as modes not present in the noiseless spectrum. The latter is further divided in noise on the expected noiseless spectrum $\mathcal{S}$ ('on support') and noise outside of it ('off support'), which in this case consists of modes having any components with frequency higher than 1. The error bars correspond to one standard deviation as evaluated by a bootstrapping variance estimation method [You96]. The results reveal that, as expected, increasing the number

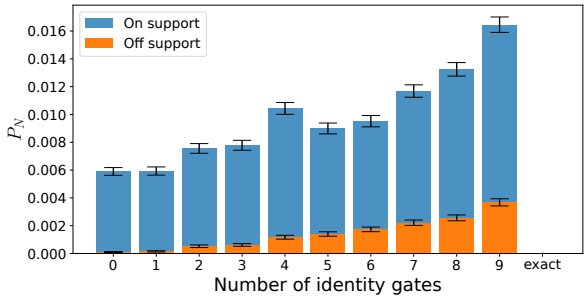


Figure 5.1: Power in noise modes for signal measuring the observable $O = ZZ$ for 2-qubit 2-rotations circuit, for increasing depth of fiducial gates corresponding to a logical identity (in this case two CNOT operations). The experiment was run on `ibmq_lima`, for each depth a grid of 5x5 points was evaluated, with 8192 shots per point.

of identity gates has a progressive detrimental effect on the quality of the output. Most of the noise appears to come from on-support modes, while off-support extra modes

remain minor, however they also appear to increase progressively with depth. Based on the previous discussion, the decay in signal power is to be interpreted as likely being caused by decoherence, while the increase in additional modes, both on- and off-support, signals the increased effect of more complex noise channels.

5.2.5.2 UCC circuit

Next, we evaluate the spectrum of a more complex 2-qubit circuit inspired by the unitary coupled cluster (UCC) approach to VQE [MEAG+20], consisting of two Pauli gadgets with $P_1 = XY$ and $P_2 = YX$. The measurement operator is ZI, which yields the expectation value $f(\theta_1, \theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$, giving 2 Fourier coefficients. This time, the experiment is run on four superconducting devices provided by IBM, as well as the ion trap computer Quantinuum H1-1, powered by Honeywell. The sampling grid is identical to the previous experiment (5x5). The results are shown in Fig. 5.2.

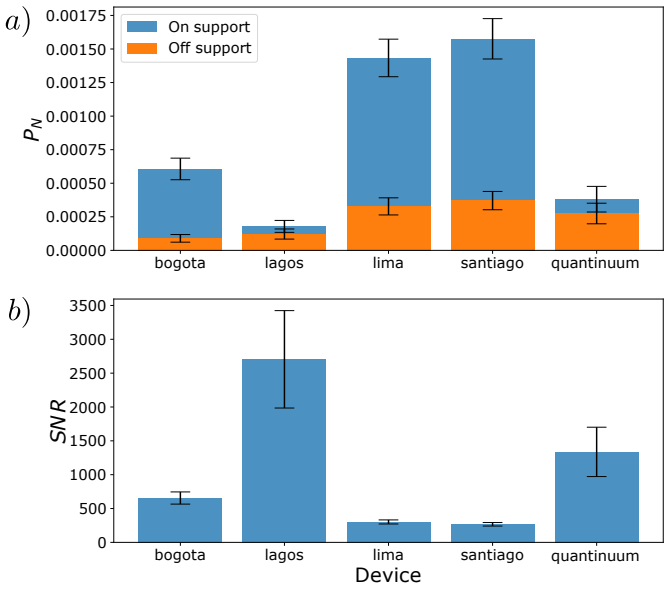


Figure 5.2: Figures of merit for 2-qubit UCC circuit, a) power in noise modes, b) signal-to-noise ratio. Measured observable $O = ZI$. Run on different systems: Quantinuum H1-1 (ion-based), IBM's `ibmq_santiago`, `ibmq_lima`, `ibmq_bogota`, `ibmq_lagos` (superconducting). Evaluated on a 5x5 parameter grid with 8192 shots per grid point.

Overall, the devices with the worst performance are `ibmq_santiago` and `ibmq_lima`,

while the best performing on the task are HQ1 and `ibmq_lagos`. This might be ascribed to their greater number of qubits (20 and 7), giving a larger parameter space for the compiler to optimise noise, as well as to higher quantum volume in the case of H1-1 (which, at the time of writing, holds the world record for this measure [Hon21]). On the other hand, the other machines have all a lower number of qubits (5). Interestingly, for the most accurate devices the majority of the noise modes fall outside the exact support, hinting that these devices might benefit the most from spectral analysis and noise mitigation. The noise characteristics of the quantum computing devices benchmarked are reported in the following table:

Device	No. of qubits	2Q gate fidelity (%)	Quantum volume
IBMQ Bogotá	5	99.05	32
IBMQ Lagos	7	99.24	32
IBMQ Lima	5	98.98	8
IBMQ Santiago	5	N/A	32
Quantinuum H1-1	20	99.8	512

Table 5.2: Sources: [PBE22], [AST21], [Qua23b], [Qua23a]. Gate fidelity data for IBMQ Santiago was not available. The quantum volume for H1-1 is the figure for the time that the experiments were run (May 2021).

Therefore we see that in principle spectral methods could be valuable in benchmarking quantum devices by providing multiple figures of merit, and separating different physical noise effects. However, we stress that the experiments presented here are extremely limited in both system size, circuit depth and number of measurement operators sampled. The results cannot be considered as anything beyond a rough assessment of the performance of the different machines in this particular task, and are meant to serve as a proof-of-principle rather than a meaningful comparison.

5.3 Error mitigation based on spectral information

In Appendix B.4 we prove that below a noise threshold, a signal can be perfectly recovered by rounding its Fourier coefficients for a class of quantum circuits. These contain parameterised gates with independent parameters and unparameterised gates that are Clifford but otherwise unknown. However, this theorem does not extend to

general circuits, where typically one has both unparameterised non-Clifford gates and dependent parameters. While the result applies to circuits of practical interest such as the unitary coupled clustered ansatz [RBM$^+$19, CSD20], the classical complexity of the fast Fourier transform scales exponentially in the number of parameters thus limiting the theorem's use to circuits containing few non-Clifford gates. On the other hand, circuits with a number of non-Clifford gates scaling logarithmically in the number of qubits can be efficiently simulated [BG16].

However, given knowledge of the frequency spectrum for a circuit family, we design application-specific error mitigation strategies that allow to approximately reconstruct the signal. In Appendix B.4 we outline two such strategies inspired by well-known classical spectral filtering methods, namely thresholding and filtering. Both aim to recover the clean signal by isolating the most significant components from a noisy spectrum. Thresholding assumes that the noise modes are small in magnitude, and therefore sets a lower bound to the coefficient's absolute value, while filtering uses knowledge of the noiseless spectrum to remove all frequencies that must be induced by noise.

In the next sections we combine these strategies with a decoherence-mitigating method and demonstrate their usefulness in a realistic scenario.

5.3.1 Spectral denoising with near-Clifford circuits

A general feature of incoherent quantum noise is the scaling of signal coefficients, as explained in Sec. 5.2.1. Denoising techniques based on frequency filtering or thresholding do not change the value of the significant coefficients, and therefore would not reduce the effect of decoherent processes on their own. This can be viewed as an indication that Fourier methods should be complemented with other noise mitigation techniques designed to reduce decoherence effects. Here, we explore a combination of Fourier methods with Clifford Data Regression (CDR) [CACC21], which has shown promise as a practical mitigation method for expectation values. CDR has features that make it natural to integrate with Fourier methods, as it relies on learning noise based on the entanglement structure of the circuit, independently from the angles of rotation. The

combination thus leads to a more efficient use of samples.

Denote by $\tilde{f}(\boldsymbol{\theta})$ the noisy expectation value of a given observable O with respect to the parameterised circuit $U(\boldsymbol{\theta})$ acting on input state ρ_0 and by $f^{\#}(\boldsymbol{\theta})$ its restriction to the frequency spectrum $\boldsymbol{\omega} \in \mathcal{S}$ for the input circuit as determined using the techniques illustrated in Appendix B.4. In practice, when implementing these protcols on a quantum device, one deals with sample mean estimators of the noisy expectation value $\tilde{f}$ which, along with implementing a discrete Fourier transforms gives an estimator for $f^{\#}$ that includes shot noise as discussed in Sec. 5.2.3. After filtering noise-induced modes, the second step would be to reduce the effect of incoherent noise by determining the noise scaling $A_{\boldsymbol{\omega}}$ that give the error-mitigated estimators $\hat{c}_{\boldsymbol{\omega}} := A_{\boldsymbol{\omega}}\tilde{c}_{\omega}$ of the exact Fourier coefficients $c_{\boldsymbol{\omega}}$. To this aim, there are several strategies one can employ. In our experiments we have explored the following strategies

The simplest approach is to first apply CDR, as it was introduced in [CACC21], in order to learn scaling factors A and A' that minimise :

$$A, A' := \arg\min_{a,a'} \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |f_C - a\tilde{f}_C - a'|, \tag{5.48}$$

where $\tilde{f}_C$ and f_C denote respectively the noisy and the classically simulated expectation value of the observable O with respect to state $C\rho_0 C^{\dagger}$. C is a near-Clifford circuit derived from the target parameterised circuit $U(\boldsymbol{\theta})$ where all but a fixed number of D non-Clifford gates are replaced with Clifford operations. Once A and A' have been found, we filter out noise-induced modes using the methods in Sec. 5.2.1 and B.4.3 to give an error-mitigated estimator:

$$\hat{f}(\boldsymbol{\theta}) = Af^{\#}(\boldsymbol{\theta}) + A' \tag{5.49}$$

for the target expectation value $f(\boldsymbol{\theta})$. To further reduce quantum compute resources, we include (part of) the circuit evaluations required for the Fourier transform into the training set $\mathcal{C}$.

Alternatively, one may implement CDR by training the model on spectrally filtered data. There is the limitation that the training data must now be in a form conducive

to Fourier mitigation, and thus a minimum of non-Clifford gates must be present. This issue is exacerbated whenever, for a given parameter θ, multiple θ-dependent gates occur in the synthesised target circuit, as they cannot be replaced by independent Clifford gates in the training phase. Lastly, a more involved strategy rests on the observation that not all coefficients corresponding to $\boldsymbol{\omega} \neq 0$ are scaled by the same amount under the action of real hardware noise. This behaviour may be seen in the raw Fourier spectrum obtained on IBMQ devices (Figs. 5.3 and 5.4) but also in the analysis of different noise models in Sec. 5.2. Using a similar training set as in the second approach above, one then would seek to learn a vector $\mathbf{A} = (A_{\boldsymbol{\omega}})$ of parameters that minimise:

$$\mathbf{A} := \arg\min_{a_{\boldsymbol{\omega}}} \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |f_C - \sum_{\boldsymbol{\omega} \in \mathcal{S}} a_{\boldsymbol{\omega}} \tilde{c}_{\boldsymbol{\omega}} e^{i\boldsymbol{\omega} \cdot \boldsymbol{\theta}_C}|, \tag{5.50}$$

where $\boldsymbol{\theta}_C$ corresponds to the near-Clifford circuit C, with each $(\theta_i)_C$ parameter chosen so that the corresponding θ_i parameterised gates are Clifford. This suffers from analogous problems in constructing the training set.

In our experiments all three methods were attempted, however we found that more involved methods did not amount to a significant improvement in the mitigation results, while significantly increasing the computational burden. Therefore, we limit our analysis to the simplest method.

5.3.2 Experiments

To demonstrate the applicability of noise filtering methods, we give proof-of-principle implementations on current quantum hardware for circuit structures used in near-term algorithms.

The magnitude threshold for determining the most significant Fourier coefficients is calculated as:

$$T = B \operatorname*{median}_{\boldsymbol{\omega} \in \tilde{\mathcal{S}}} (|\tilde{c}_{\boldsymbol{\omega}}|), \tag{5.51}$$

for some tunable constant B. Assuming the majority of coefficients in the sampled frequency spectrum are due to noise (which in principle requires the resolved spectrum to be double the size of the noiseless one), and further that these are small and of

similar magnitude, this allows us to select only the outliers and hence recover the signal coefficients. The CDR is performed with a simple linear model as in Eq. (5.49). The training set $\mathcal{C}$ is generated by substituting for each target circuit the closest Clifford to every non-Clifford gate, except for D randomly chosen gates.

As noted before, spectral noise filtering is expected to yield a landscape that has more similar features to the noiseless one, while a scaling technique like CDR is necessary to reduce the error in the magnitudes. Therefore, we focus on two quantities of interest that capture these two different scenarios.

For each noise mitigation experiment we report the similarity measure between the mitigated vector $C^{\#}$ and the noiseless vector C obtained from a unitary simulator (Qiskit Aer's statevector simulator [Qis21]). The simplest measure of similarity between vectors of data points is the Euclidean distance $||C^{\#} - C||$, which has a practical interpretation as the root mean squared error between the data points. Furthermore, it is invariant (up to a constant) under the Fourier transform, due to Parseval's theorem, and therefore may be easily calculated from the spectrum. We then choose to consider the additional measure of cosine similarity:

$$S := \frac{C^{\#\dagger} C}{||C^{\#}||\,||C||}. \tag{5.52}$$

This quantity is robust to global scaling of any one of the two vectors. The quantity can therefore be visualised as a scale-invariant measure of error: indeed, if both vectors are unit-normalised, cosine similarity is monotonically related to Euclidean distance: $||\frac{C^{\#}}{||C^{\#}||} - \frac{C}{||C||}|| = \sqrt{2 - 2S}$.

The experiments are performed on IBMQ superconducting devices.

5.3.2.1 UCCSD

We begin by analysing a UCCSD circuit for the H2 molecule under the STO-3g basis set, which requires four qubits under Jordan-Wigner representation. The circuit has two parameters corresponding to 'singles' excitations, each appearing on two Pauli gadgets, and one parameter governing 'doubles' excitations, composed of eight further

	Cosine similarity				Euclidean distance			
Methods	A	B	C	D	A	B	C	D
Raw	0.824	0.664	0.868	0.881	0.448	0.539	0.448	0.542
CDR	0.823	0.665	0.864	0.873	0.319	0.462	0.331	0.396
Filtering + CDR	0.835	**0.675**	**0.886**	**0.912**	0.316	**0.461**	**0.327**	**0.391**
Thresholding + CDR	**0.841**	0.672	0.871	0.909	**0.314**	0.462	0.329	**0.391**

Table 5.3: Effectiveness of landscape noise mitigation methods on 4-qubit UCC circuit for H_2. Singles excitations parameters are varied, doubles excitations are fixed to random values. A: `ibmq_manila`, $O = ZIZI$; B: `ibmq_manila`, $O = XXYY$; C: `ibmq_lima`, $O = ZIZI$; D: `ibmq_lima`, $O = XXYY$. Highest values highlighed. Note that higher cosine similarity and lower Euclidean distance are better. 8192 shots were taken per grid point. Reporting results from single experimental runs so error bars are not available.

gadgets. The circuit has a total CNOT depth of 52 and no depth-reduction compilation passes were applied. We choose to focus on the two singles parameters for spectral noise mitigation. Since they each control two single-qubit rotation gates, the noiseless support has a maximum frequency of 2, therefore we sample on a 7x7 grid giving a maximum frequency resolution of 3.

In the target circuit, the Pauli gadgets implementing the doubles excitations are chosen independently and set to fixed angles normally distributed around $\pi/4$ with standard deviation 0.1. By randomising the fixed parameters instead of setting them to the optimum we aim to illustrate an improved profile of the noisy energy landscape for general parameter values, allowing to evaluate the mitigation method in a realistic setting where the optimal parameters are not known. For CDR we set $D = 2$ and for thresholding constant we find that $B = 3$ is sufficient.

The results are shown in Table 5.3. Both cosine similarity and Euclidean distance are consistently improved by the combination of spectral noise mitigation and CDR compared to no mitigation and CDR alone. The improvement due to spectral techniques is more pronounced for cosine similarity (though still only about 2%), while the decrease in Euclidean distance is more modest. This reflects the fact that Euclidean distance is highly dependent on relative scaling between the vectors, which is what CDR targets by construction.

5.3.2.2 QAOA

	Cosine similarity					Euclidean distance				
Methods	A	B	C	D	E	A	B	C	D	E
Raw	0.838	0.764	0.930	**0.764**	0.789	1.791	**2.526**	1.803	**2.416**	3.854
CDR	0.863	0.765	0.934	0.751	0.827	1.376	2.581	1.468	2.811	3.229
Filtering + CDR	0.868	**0.766**	0.940	0.752	0.833	1.371	2.581	1.465	2.811	3.225
Thresholding + CDR	**0.903**	0.762	**0.979**	0.753	**0.878**	**1.345**	2.585	**1.451**	2.811	**3.199**

Table 5.4: Effectiveness of landscape noise mitigation methods on MaxCut QAOA circuit for random 3-reg graph. The parameters varied are the last two β, γ. Where $p = 2$, the parameters n the first layer are set to $\pi/4$. The measurement operator is $O = H$. A: `ibmq_guadalupe`, $p = 1$, 8 qubits; B: `ibmq_guadalupe`, $p = 2$, 8 qubits; C: `ibmq_sydney`, $p = 1$, 8 qubits; D: `ibmq_sydney`, $p = 2$, 8 qubits; E: `ibmq_guadalupe`, $p = 1$, 16 qubits. Highest values highlighed. Note that higher cosine similarity and lower Euclidean distance are better. 8192 shots were taken per grid point. Reporting results from single experimental runs so error bars are not available.

We then consider a quantum approximate optimisation algorithm (QAOA) circuit [FGG14] for the MaxCut problem on random 3-regular triangle-free graphs of various sizes, with V and E denoting the vertices and edges. The circuits involve repeated layers labelled by p, whereas each layer alternates between a unitary generated by the cost Hamiltonian $H_C = \sum_{(i,j)\in E} Z_i Z_j$ and a mixing Hamiltonian $H_D = \sum_{i\in V} X_i$, with the initial state an equal superposition over the computational basis. The $p = 1$ layer circuit on n qubits produces the parameterised state $e^{-i\beta H_D} e^{-i\gamma H_C} |+\rangle^{\otimes n}$. The benefits of this class of QAOA problems for benchmarking noise mitigation algorithms are twofold: first, the scaling in depth with system size is favourable compared to UCC, as long as the number of layers is kept small. Second, since the Hamiltonian is diagonal in the Z basis the cost function value can be calculated from the single shots without adding additional rotation gates. We compare the same noise mitigation techniques as UCC, on an 8-qubit QAOA circuit with $p = 1, 2$ on both the 16-qubit `ibmq_guadalupe` and the 27-qubit `ibmq_sydney` quantum computers. We also run a 16-qubit $p = 1$ experiment on `ibmq_guadalupe` as an example of a larger system. As the target function we measure the expectation value of the Hamiltonian.

All experiments are run on two parameters. In the $p = 2$ experiments we choose to focus on the parameters in the last layer, in which case we set both parameters of the first layer to $\pi/4$. By using a lightcone argument one can then show that in all experiments that the frequencies will be limited to 2 for the mixing parameter β, and 5 for the problem parameter γ, which follows from the problem graphs being all 3-regular and triangle-free. Therefore, we sample on a 13x13 grid to ensure a resolution of 6 in frequency space.

Training CDR on this system presents with the additional complication that setting all parameters to give Clifford gates ($\beta, \gamma \in \{0, \pi\}$) leads to an expectation value of zero, as can be verified by explicit calculation. Therefore, in order to obtain training data of sufficient magnitudes the CDR parameter D is increased from 2 to 10, and the data is post-selected to only train on the largest magnitude datapoints. Due to the higher noise level we find that increasing the thresholding constant to $B = 10$ is more appropriate.

The results are shown in Table 5.4. The performance of noise mitigation techniques is seen as highly dependent on the depth of the circuit. Circuits with $p = 1$ show consistent improvements with noise mitigation and especially following spectral filtering. Interestingly, thresholding appears to be the most effective filtering method, which can be justified by the noiseless signal being sparse and much higher than the noise level, as shown in Fig. 5.3. Increasing the system size to 16 qubits does not seem to affect the noise mitigation capability.

On the other hand, circuits with $p = 2$ generally see little improvement in either measures of quality. This is due to the fitting subroutine in CDR failing to find an appropriate scaling factor due to Clifford data being extremely noisy. As a related effect, hardware noise causes the experimental signal to be considerably different from the noiseless one as can be seen in Fig. 5.4. Overall this is a reminder that excessive noise can be an insurmountable obstacle for any kind of error mitigation, whether classic or frequency-based.

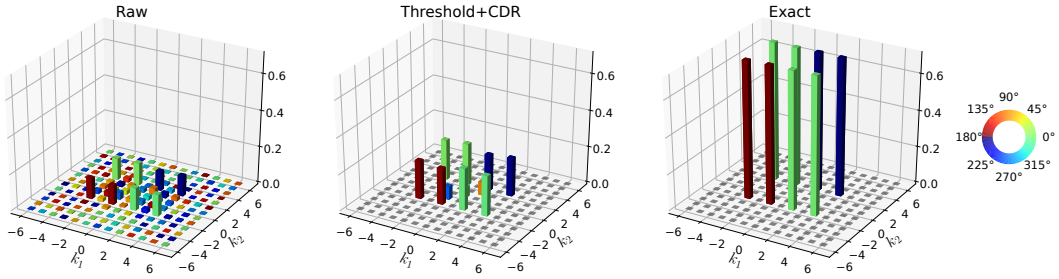


Figure 5.3: QAOA experiment for MaxCut of an 8-node random 3-regular graph, $p = 1$, $O = H$. Magnitudes for Fourier coefficients are shown before and after filtering and CDR, vs exact. Colour represents the complex argument as shown by the colourwheel. Filtering method is thresholding with $B = 10$. Run on `ibmq_sydney`. Spectral filtering is seen to eliminate spurious frequencies, while CDR improves the magnitude of the coefficients, although only partially.

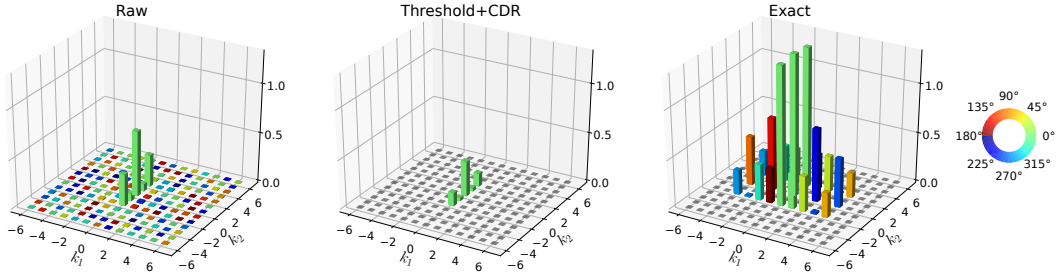


Figure 5.4: QAOA experiment for 8-node random 3-regular MaxCut, $p = 2$, last layer params varied, first layer: $[\beta, \gamma] = [1/4, 1/4]$, $O = H$. Magnitudes for Fourier coefficients are shown before and after filtering and CDR, vs exact. Colour represents the complex argument as shown by the colourwheel. Filtering method is thresholding with $B = 10$. Run on `ibmq_sydney`. Note that both filtering and CDR fail to mitigate for noise since the raw signal is significantly different from the exact one.

5.4 Discussion

In this Chapter, we reframed the study of quantum landscapes in the familiar terms of spectral analysis. This allows us to investigate the quantum noise-induced effects on the Fourier spectrum of parameterised quantum circuits, both in real hardware and through theoretical analysis of common noise models. In turn this led us to new methods for digital error mitigation based on Fourier-space filtering which we demonstrate on experimental samples from a superconducting quantum computer. While the circuits we examined are not large enough to exhibit quantum advantage, they illustrate that denoising in the frequency domain may yield less noisy estimates of expected values from

hardware data. In particular we show how to combine denoising in Fourier space with existing error mitigation based on learning from Clifford circuits. The latter targets incoherent noise, while the former reduces the effect of gate-dependent, correlated errors, including coherent contributions. The compound effect of the Fourier-based error mitigation strategies reduces not only the magnitude of errors but also the noise-induced distortion in the energy landscape.

The theory exposed here may be a fruitful approach to quantum landscape theory [LJGM$^+$23]. Spectral analysis of the landscape can be employed to make statements about its global characteristics; for example, the presence or absence of barren plateaus and narrow gorges [CSV$^+$21]. Furthermore, within this framework noise channels can be included naturally, making it amenable to study phenomena such as noise-induced barren plateaus [WFC$^+$21] and alterations of the quantum training landscape [WCA$^+$21]. Our work also opens up new directions to apply the vast range of signal processing techniques for the purpose of quantifying and reducing quantum noise. We tackled these topics from the perspective of digitised quantum computations, however similar analysis may be applied at the level of pulse sequences. In fact, the latter connects with experimental techniques used in the tuning and calibration of individual gates such as noise spectroscopy [ÁS11, YSH11, YW12].

While here we only explored combining spectral methods with CDR, one may adapt our approach for other techniques such as zero noise extrapolation [TBG17, LB17] and probabilistic error cancellation [TBG17, EBL18], which similarly address decoherence effects. More recent work [CACC21] that improves the efficiency of CDR could also be employed to reduce the total number of training circuits required. Furthermore, investigating the effect of noise on the Fourier coefficients might also give new insight into the extent to which hardware noise satisfies some of the in-built assumptions of error mitigation protocols. For example, digital versions of zero noise extrapolation involve increasing the noise level by introducing additional gates [GTHL$^+$20]. Fourier methods could then be useful to determine appropriate type of such gate folding to validate the different fitting functions used in the extrapolation.

The most glaring disadvantage of Fourier methods is the large sampling overhead

required. We discuss the resource requirements, as well as possible ways to reduce it, in Appendix B.1. There we discuss also how to approach the problem of incomplete or nonuniform sampling, which might also help ameliorate the resource requirement and make the methods more flexible. Overall the work presented here introduces a flexible framework that might extend to numerous other methods of noise characterisation and mitigation and provides a new perspective to study noisy variational quantum algorithms.

Chapter 6

Classical simulations of noisy Variational Quantum Algorithms

---

In this Chapter we describe a classical simulation algorithm, LOWESA *(low weight efficient simulation algorithm), for estimating expectation values of noisy parameterised quantum circuits with a fixed observable. It combines the previously presented results on spectral analysis of parameterised circuits with work on methods for simulations of noisy random circuits using Pauli back-propagation. We show, under some conditions on the circuits and mild assumptions on the noise, that* LOWESA *gives an efficient, polynomial algorithm in the number of qubits (and depth), with approximation error that vanishes exponentially in the physical error rate and a controllable cut-off parameter. We discuss the practical limitations of the method for circuit classes with correlated parameters and its scaling with decreasing error rates.*

The work in this Chapter is based on Ref. [FRDC22a].

6.1 Simulating noisy quantum circuits

A very rough summary of the research presented so far would go like this: far from showing resilience, variational quantum algorithms in the presence of noise present dra-

matic changes in the cost function. In particular we have seen that the landscape tends to become flatter, both experimentally in Chapter 3 and theoretically, as noise-induced barren plateaus in Chapter 4 and as decay of Fourier coefficients in Chapter 5. Barren plateaus have been shown to hinder and potentially prohibit optimisation [AK22] and so they present a significant practical obstacle. However one question about these noisy algorithms arises naturally that overshadows concerns on their trainability: does noise reduce their quantum nature so much that they become tractable by classical computers? This is a natural question, since one of the key aspects of quantum computation is entanglement, and by definition decoherent noise reduces entanglement. Therefore it makes intuitive sense that noise accumulation may at best reduce and at worst eliminate the advantage that quantum algorithms have on classical counterparts.

The theory of general quantum algorithms provides a more nuanced perspective. One of the cornerstone results of quantum computation is the threshold theorem, which states that via error correction, logical error rates may be reduced to arbitrarily small levels provided that the physical error rate is smaller than a certain threshold (depending on the code and noise model) [ABO97, KLZ98, Kit97]. This suggests that noisy quantum computation (with mid-circuit measurements) can be essentially as powerful as error-free computation below the threshold. However it is unclear what can be said regarding the classical simulatability of quantum systems with noise levels above the threshold where no error correction is possible, even though evidence exists that below a (higher) noise threshold, noisy quantum computation with postselection may still be postBQP-complete [Fuj16].

It should be noted that the perspective historically taken for quantum algorithms may not apply to variational quantum algorithms. In the theoretical analysis of simulatability of general quantum algorithms the most frequently used notions are those of *weak classical simulation*, that is sampling from a distribution that is close to the output of the quantum circuit, and *strong simulation* where given a bitstring one approximates its output probability, both to a multiplicative error in probability [TD02]. This makes sense for algorithms like quantum phase estimation or Grover's algorithm where the output is a bitstring, however variational quantum algorithms return an ex-

pectation value of an Hermitian observable and the concept does not directly translate to this setting. This leaves open the possibility that different regimes of simulatability may exist for variational algorithms.

Indeed, frameworks [SFGP21a,DPMRF23] comparing classical algorithms with noisy VQAs by use of entropic quantities concluded that to have any quantum advantage the circuit depth has a lower bound that scales inversely with the physical gate error rate. Beyond this regime, classical methods exist that certifiably outperform the noisy quantum computation [FGP22]. Other classical simulations that target noisy VQAs include decision diagrams [HHM$^+$21] and tensor networks [ALZ$^+$23, ZSW20] and both show effectiveness at very low equivalent noise rates. However, these approaches tend to be heuristic and do not necessarily provide rigorous trade-offs between complexity, approximation error and physical noise.

6.1.1 Simulating noisy RCS

In this Chapter we will discuss an efficient classical algorithm for simulating expectation values of parameterised quantum circuits affected by Pauli noise, for fixed generic observables. We will combine ideas from the spectral decompositions of parameterised noisy quantum circuits that we have seen in the previous Chapters, with a technique for simulating noisy random circuit sampling first introduced in 2022 by Aharanov *et al.* [AGL$^+$22]. Let us briefly review this last work as it will provide useful insights to the main algorithm.

Random circuit sampling (RCS) is a promising task for demonstrating quantum advantage that is amenable to current noisy hardware [BFNV19]. In short, RCS consists in sampling a random circuit from some ensemble (say, Haar random unitaries, or random 2-designs), repeatedly applying the circuit to some initial state on hardware and obtaining a number of output bitstrings. Thus one builds an empirical distribution that can be compared with an expected distribution that is either known or extrapolated using a classical computer, using some measure like cross-entropy fidelity. This is the approach used in by the Google team in 2019 for their "quantum supremacy" demonstration [AAB$^+$19]. It is believed (but has not been proven) that sampling from a distri-

bution with high cross-entropy score is exponentially hard in the number of qubits for any classical device, while a noisy quantum computer may be able to produce sequences of large enough score. The current record for a hard-to-simulate circuit is a 60 qubit experiment by USTC with (estimated) cross-entropy score of 0.0366% [ZCC+22], which would take several years on a supercomputer with a tensor network method [LGL+21].

However, the validity of these results as demonstrations of computational quantum supremacy was cast into doubt by Ref. [AGL+22]. The refutation is not of the hardness of these experiments *per se*, but the extent to which a successful noisy RCS experiment can be considered a violation of the extended Church-Turing thesis, which states that there can be no physical phenomena that require exponential resources to simulate [BV93]. This is because the algorithm they propose runs in polynomial time, albeit with a large exponent. Let us now sketch the algorithm. We begin with the probability of an outcome x from a clean RCS experiment

$$p(U,x) = \mathrm{Tr}\left[|x\rangle\langle x|U|0\rangle\langle 0|U^\dagger\right]. \tag{6.1}$$

Assume the unitary is decomposed into gates as $U = U_d...U_2U_1$. Now we switch to the Pauli basis and adopt the vectorised operator notation introduced in Chapter 1, and the Pauli transfer matrix (PTM) notation for channels. First we define the normalised Pauli basis on n qubits as :

$$|\hat{P}_i\rangle\!\rangle = \frac{1}{2^{n/2}}|P_i\rangle\!\rangle, \tag{6.2}$$

such that $\langle\!\langle \hat{P}_i|\hat{P}_j\rangle\!\rangle = \delta_{ij}$. The corresponding set of all n-qubit Paulis is $\hat{\mathbb{P}}_n$. The probability can be written as a Feynman path integral by expanding the unitary channel and adding resolutions of identity in the Pauli basis:

$$p(U,x) = \sum_{\hat{\boldsymbol{P}}\in\hat{\mathbb{P}}^{d+1}} \langle\!\langle x|\hat{P}_d\rangle\!\rangle\langle\!\langle \hat{P}_d|U_d|\hat{P}_{d-1}\rangle\!\rangle\cdots\langle\!\langle \hat{P}_1|U_1|\hat{P}_0\rangle\!\rangle\langle\!\langle \hat{P}_0|0\rangle\!\rangle, \tag{6.3}$$

where each vector of Paulis $\hat{\boldsymbol{P}}$ defines a path, denote it by $f(U,x,\hat{\boldsymbol{P}})$.

Now we shift to a noisy implementation. Writing the noisy channels in PTM form

as $\tilde{\boldsymbol{U}}_i$, the noisy probability is:

$$\tilde{p}_\gamma(U,x) = \sum_{\hat{\boldsymbol{P}}\in\mathbb{P}^{d+1}} \langle\!\langle x|\hat{P}_d\rangle\!\rangle \langle\!\langle \hat{P}_d|\tilde{\boldsymbol{U}}_d|\hat{P}_{d-1}\rangle\!\rangle \cdots \langle\!\langle \hat{P}_1|\tilde{\boldsymbol{U}}_1|\hat{P}_0\rangle\!\rangle \langle\!\langle \hat{P}_0|0\rangle\!\rangle = \sum_{\hat{\boldsymbol{P}}} \tilde{f}_\gamma(U,x,\hat{\boldsymbol{P}}). \tag{6.4}$$

For now not much has been gained - the number of paths is exponential, equal to 4^{d+1}. The crucial step is now to look at how noise modifies the paths. Assume a noise model where each gate is preceded by identical single-qubit depolarising channels acting on all qubits with noise strength γ: $\tilde{\boldsymbol{U}}_i = \boldsymbol{U}_i\boldsymbol{\mathcal{E}}_{SD}$. As seen in Chapter 1, the behaviour on such a noise channel on a single qubit Pauli operator is to scale the operator by a noise-dependent factor:

$$\boldsymbol{\mathcal{E}}_{SD}|\hat{P}\rangle\!\rangle = \begin{cases} (1-\gamma)|\hat{P}\rangle\!\rangle & \text{if } P\in\{\hat{X},\hat{Y},\hat{Z}\}, \\ |\hat{I}\rangle\!\rangle & \text{if } \hat{P}=\hat{I}. \end{cases} \tag{6.5}$$

Thus, it follows that the noisy paths can be related to the clean ones by a simple formula:

$$\tilde{f}_\gamma(U,x,\hat{\boldsymbol{P}}) = (1-\gamma)^{|\hat{\boldsymbol{P}}|} f(U,x,\hat{\boldsymbol{P}}), \tag{6.6}$$

where $|\hat{\boldsymbol{P}}| = \sum_i |\hat{\boldsymbol{P}}_{\boldsymbol{i}}|$ and $|\hat{\boldsymbol{P}}_{\boldsymbol{i}}|$ is the number of non-identity single-qubit Paulis in $\hat{P}_i$, also called the weight of the Pauli string.

Thus, it follows that one can truncate the summation to all paths with weight less than a cutoff $\ell < n(d+1)/2$, of which there are:

$$\sum_{k=0}^{\ell-1} 3^k \binom{n(d+1)}{k} \leq \ell 3^\ell \binom{n(d+1)}{\ell} \leq (nd)^{O(\ell)}, \tag{6.7}$$

which crucially is polynomial in n, and obtain an approximation:

$$\tilde{q}_{\gamma,\ell}(U,x) = \sum_{|\hat{\boldsymbol{P}}|<\ell} \tilde{f}_\gamma(U,x,\hat{\boldsymbol{P}}). \tag{6.8}$$

One way to compute the accuracy of this approximation is to consider the total variation

distance over all bistrings:

$$\Delta := \sum_{x\in\{0,1\}^n} |\tilde{p}_\gamma(U,x) - \tilde{q}_{\gamma,\ell}(U,x)|, \tag{6.9}$$

which is upper bounded by:

$$\Delta^2 \leq 2^n \sum_{x\in\{0,1\}^n} (\tilde{p}_\gamma(U,x) - \tilde{q}_{\gamma,\ell}(U,x))^2 = 2^n \sum_{x\in\{0,1\}^n} \left(\sum_{|\hat{\boldsymbol{P}}|\geq\ell} (1-\gamma)^{|\hat{\boldsymbol{P}}|} f(U,x,\hat{\boldsymbol{P}}) \right)^2 . \tag{6.10}$$

Interestingly, by considering a circuit with 2-qubit gates drawn independently from a Haar-random distribution (or a 2-design) $\mathcal{D}$, as commonly done in RCS, one can show that if $x = x'$:

$$\mathbb{E}_{U_i\sim\mathcal{D}}[f(U,x,\hat{\boldsymbol{P}})f(U,x',\hat{\boldsymbol{P}})] = 0, \tag{6.11}$$

and therefore the cross-terms in the summation cancel. Further imposing the condition that the output of each circuit is anti-concentrated (a phenomenon commonly observed in random quantum circuit), in Ref. [AGL$^+$22] it is shown that:

$$\mathbb{E}_{U_i\sim\mathcal{D}}[\Delta^2] \leq O(1)\, e^{-2\gamma\ell}, \tag{6.12}$$

and therefore on average the approximation error is exponentially decaying in ℓ.

Thus overall the work in Ref. [AGL$^+$22] shows that RCS output distributions can be approximated with low average total variational distance, with cost only polynomial in n, albeit with a potentially large exponential factor. We shall use these ideas as inspiration to design a simulation algorithm for the expectation values of VQAs. Although conceptually a rather different setting, we will see that the final algorithm shares many similarities with the one just outlined.

6.1.2 Heisenberg picture simulation algorithms for expectation values

The PTM formalism can be used to calculate expectation values in the *Heisenberg picture* via Pauli back-propagation, where the quantum channels are seen as acting on the measurement operator instead of the state [Got98]. In PTM form this adjoint

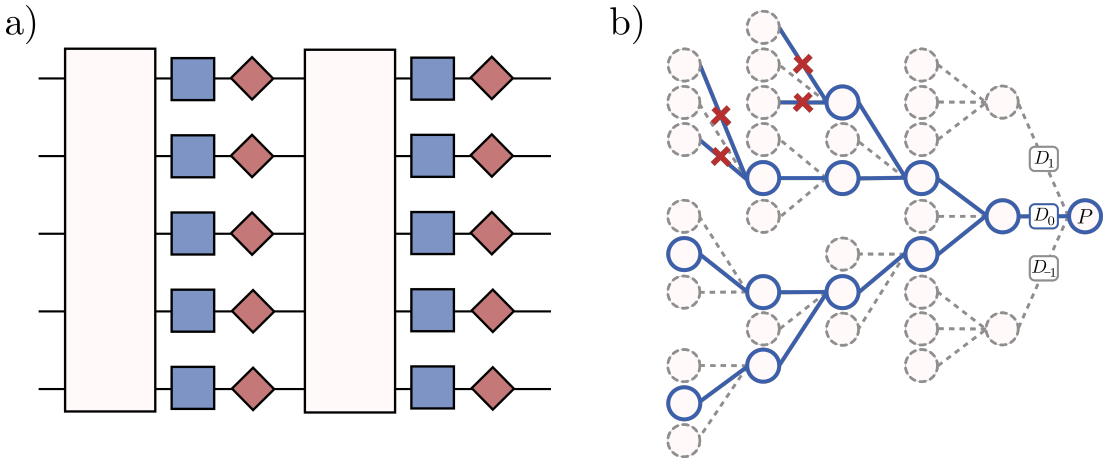


Figure 6.1: a) Schematic of the parameterised quantum circuits that can be simulated by LOWESA. The light boxes are arbitrary (noisy) Clifford gates, the blue boxes are parameterised Pauli rotations and the red kites represent Pauli noise channels.
b) Diagrammatic sketch of LOWESA as described in Algorithm 1 applied to circuits given by Eq. (6.14). The Pauli operator P is propagated backwards through the circuit where every Clifford gate transforms it into another Pauli, and the decomposition of the parameterised Pauli rotations into process modes D_0, D_1, D_{-1} splits the propagation up into paths that may annihilate. A cut-off of $\ell = 2$ is chosen which artificially annihilates paths that branch into D_1, D_{-1} more than 2 times.

operation corresponds to simply taking the transpose of the expectation value:

$$\langle\langle P|\mathbf{E}|\rho\rangle\rangle = \langle\langle \rho|\mathbf{E}^\intercal|P\rangle\rangle, \tag{6.13}$$

which is possible for any $\mathcal{E}$. This perspective provides an efficient approach to classically computing expectation values. Take an n-qubit channel $\mathcal{E}$ and assume it can be decomposed as a sum of N Clifford unitary channels $\mathcal{E}_i$ via $\mathcal{E} = \sum_{i=1}^{N} c_i \mathcal{E}_i$, $\sum_i c_i = 1$. Also consider a stabiliser state [Got98] ρ such that the expectation value with any Pauli operator can be evaluated efficiently. Then, given a Pauli P, the expectation value $\langle\langle P|\mathbf{E}|\rho\rangle\rangle$ can be expanded as a sum of N terms $\langle\langle P|\mathbf{E}_i|\rho\rangle\rangle$. As Clifford unitaries are generalised permutation matrices in the PTM representation we get $\langle\langle \rho|\mathbf{E}_i^\intercal|P\rangle\rangle = \langle\langle \rho|P_i'\rangle\rangle$ (up to a phase), for some other Pauli operator P_i'. When $\mathcal{E}_i$ is an n-qubit Clifford unitary then it can be synthesised into at most $O(n^2/log(n))$ gates [AG04] and the change of Pauli frame from P to P_i' can be efficiently computed in $O(n^2)$ [CIP18, Got98]. Finally, since ρ is assumed a stabiliser state, the expectation value $\langle\langle \rho|\mathbf{E}_i^\intercal|P\rangle\rangle$ can be efficiently

computed in $O(n^2)$. This gives an efficient classical algorithm to compute expectation values when $N \sim \text{poly}(n)$.

This approach is not new. The decomposition of general channels into sums of stabiliser channels (Cliffords and Pauli measurements) for the purpose of quantum circuit simulation was introduced in Ref. [BFH$^+$17]. A similar sum-over-Clifford algorithm for unitary circuits was explored in Ref. [BBC$^+$19]. A PTM-based algorithm for both exact and noisy circuit simulation has been proposed in Ref. [HM22] from a Schrödinger perspective (state propagation). The work in Ref. [RLCK19] is the closest to the method used here, as it covers the PTM representation in conjunction with a Heisenberg picture simulation method. In addition, it discusses the effect on simulatability of adding symmetric depolarising noise on z-rotation gates.

However, something that to our knowledge has not been made explicit before is that the method can be generalised beyond decompositions into Clifford unitaries (or near-Clifford unitaries [BBC$^+$19]) and Pauli measurement channels. Indeed, here we will consider general processes $\mathcal{E}_i$ for which the expectation value $\langle\langle P|\mathbf{E}_i|\rho\rangle\rangle$ can be evaluated efficiently. Notably, the processes $\mathcal{E}_i$ need not even be valid quantum channels (or completely positive trace preserving maps), we only require that its PTM representation is sufficiently sparse. This occurs when the adjoint channel $\mathcal{E}_i^\dagger$ maps every Pauli operator into a combination of small, $O(\text{poly}(n))$, number of Pauli operators. This echoes remarks in Ref. [Nes09], although that work is in the Schrödinger picture. In our case, the $\mathcal{E}_i$ will correspond to compositions of Clifford unitaries and processes that map every Pauli operator to a single (possibly distinct) Pauli operator or to zero.

6.2 Classical simulation of uncorrelated parameter VQAs

6.2.1 Circuit model

A PQC on n qubits is defined as a sequence of m unitary gates, each parameterised by a component of a parameter vector $\boldsymbol{\theta}$. Here, we consider the case where the gates are alternating layers of Clifford operations C_i and Pauli rotations $R_i(\theta_i) = e^{-i\theta_i/2\,P_i}$ where P_i is a multi-qubit Pauli operator. For now we assume that the rotation gates

are independently parameterised. The overall unitary is:

$$U(\boldsymbol{\theta}) = \left(\prod_{i=1}^{m} C_i R_i(\theta_i) \right) C_0 \,. \tag{6.14}$$

This specific form is operationally relevant as it is featured in many common near-term algorithms [KMT+17, FGG14, TCC+22] and proposals for fault-tolerant architectures [Wei13, AMO+24], and since Clifford unitaries and Pauli rotations form a universal gate set, any PQC can be cast in this way (up to fixing a subset of the parameters). As typical for VQAs we assume initialising the quantum computer in $|\mathbf{0}\rangle = |0\rangle^{\otimes n}$, applying the PQC and measuring an observable to obtain a cost function. We write the expectation value for a specific n-qubit Pauli operator $P \in \mathbb{P}^{\otimes n}$ (the *cost function* in VQA terminology) as:

$$f(\boldsymbol{\theta}) := \mathrm{Tr}(P\, \mathcal{U}_{\boldsymbol{\theta}}[|\mathbf{0}\rangle\langle\mathbf{0}|]), \tag{6.15}$$

where the unitary channel is $\mathcal{U}_{\boldsymbol{\theta}}[\cdot] := U(\boldsymbol{\theta})[\cdot]U^{\dagger}(\boldsymbol{\theta})$.

We model the noisy PQC using general Pauli channels, which are probabilistic mixtures of unitary n-qubit Pauli operator evolutions. For a single qubit, a general Pauli channel is given by:

$$\begin{aligned} \mathcal{N}_{Pauli}(p_X, p_Y, p_Z)[\rho] =& (1 - p_X - p_Y - p_Z)\rho \\ &+ p_X X\rho X + p_Y Y\rho Y + p_Z Z\rho Z. \end{aligned} \tag{6.16}$$

These are often used to model local decoherent processes in quantum hardware. The dephasing channel $\mathcal{N}_{Pauli}(0, 0, p)$ is a particular example which models interactions between a qubit and the external environment. The best-fit noise parameters $\{p_X, p_Y, p_Z\}$ for each qubit can be estimated experimentally via procedures like cycle benchmarking [EWP+19].

The noisy circuit model we consider takes the form:

$$\tilde{\mathcal{U}}_{\boldsymbol{\theta}} = \left(\bigcirc_{i=1}^{m} \mathcal{C}_i \circ \tilde{\mathcal{R}}_i(\theta_i) \right) \circ \mathcal{C}_0. \tag{6.17}$$

The resulting noisy cost function is labelled $\tilde{f}(\boldsymbol{\theta})$. Each noisy gate is given by the target unitary followed by a Pauli channel acting on the subset of qubits Q_i where P_i acts nontrivially. Specifically, we have:

$$\tilde{\mathcal{R}}_i(\theta_i) = \bigotimes_{q \in Q_i} \mathcal{N}^{(q)}_{Pauli} \circ \mathcal{R}_i(\theta_i). \tag{6.18}$$

We will later consider more general noise model where the Clifford operations also incur noise: $\tilde{\mathcal{C}}_i = \mathcal{C}_i \circ \mathcal{M}$, where $\mathcal{M}$ are multi-qubit Pauli channels, and where noise is allowed to vary across the circuit. Finally, let us define $p = \min_{\sigma=X,Y,Z} p_\sigma$. Generally real devices will have a symmetric depolarising component to every operation so we can assume $p > 0$ holds [WFC+21].

6.2.2 Strategy

We first show how the noisy variational circuits considered in Eq. (6.14) admit a linear decomposition into processes that are amenable to the classical simulation outlined in Sec. 6.1.2. To that aim, it turns out that a decomposition into Fourier series of the noisy channel $\tilde{\mathcal{U}}_\theta$, and therefore noisy cost function, results in processes that map a Pauli operator into multiple Pauli operators, and thus their composition may lead to an exponential accumulation of terms. However, a different choice of basis involving trigonometric polynomials remedies this to produce a decomposition for which the dominant coefficients in the expansion can be efficiently computed.

We now assume that all rotations are single qubit z-rotations. This is purely to make the exposition easier to follow, the theorems will be valid for any Pauli rotation. Let $\mathbf{R}_z(\theta)$ be the PTM of $\mathcal{R}_z(\theta)$ and let $\mathbf{N}$ be the PTM of the Pauli noise channel $\mathcal{N}_{Pauli}$, $\mathbf{N} = \mathrm{diag}(1, q_X, q_Y, q_Z)$. The eigenvalues of the Pauli channel are related to the error probabilities as $q_X = 1 - 2(p_Z + p_Y)$, $q_Y = 1 - 2(p_Z + p_X)$, $q_Z = 1 - 2(p_X + p_Y)$. Then, the noisy channel $\tilde{\mathcal{R}}_z(\theta) = \mathcal{N}_{Pauli} \circ \mathcal{R}_z(\theta)$ has, with respect to the orthonormal

basis $\{|\hat{I}\rangle\!\rangle, |\hat{X}\rangle\!\rangle, |\hat{Y}\rangle\!\rangle, |\hat{Z}\rangle\!\rangle\}$, the PTM:

$$\mathbf{N}\cdot\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q_X\cos\theta & -q_X\sin\theta & 0 \\ 0 & q_Y\sin\theta & q_Y\cos\theta & 0 \\ 0 & 0 & 0 & q_Z \end{pmatrix}. \tag{6.19}$$

Denote the projectors by $\Pi_0 = |\hat{I}\rangle\!\rangle\langle\!\langle\hat{I}| + |\hat{Z}\rangle\!\rangle\langle\!\langle\hat{Z}|$, $\Pi_X = |\hat{X}\rangle\!\rangle\langle\!\langle\hat{X}|$ and $\Pi_Y = |\hat{Y}\rangle\!\rangle\langle\!\langle\hat{Y}|$. Then we can define new quantum processes $\{\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_{-1}\}$ to be used in the simulation algorithm via their PTM representation $\mathbf{D_0} = \Pi_0\mathbf{NR}\Pi_0$, $\mathbf{D_1} = \Pi_X\mathbf{NR}\Pi_X + \Pi_Y\mathbf{NR}\Pi_Y$ and $\mathbf{D_{-1}} = \Pi_X\mathbf{NR}\Pi_Y + \Pi_Y\mathbf{NR}\Pi_X$ such that:

$$\mathbf{N}\cdot\mathbf{R} = \mathbf{D_0} + \cos\theta\,\mathbf{D_1} + \sin\theta\,\mathbf{D_{-1}}. \tag{6.20}$$

Expanding out these processes, we see that each of them maps any single Pauli operator into at most another single Pauli operator (up to a scaling),

$$\mathbf{D}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_Z \end{pmatrix},\ \mathbf{D}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & q_X & 0 & 0 \\ 0 & 0 & q_Y & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{6.21}$$

$$\mathbf{D}_{-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -q_X & 0 \\ 0 & q_Y & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{6.22}$$

This decomposition allows us to expand the noisy circuits in terms of a multivariate trigonometric basis, which is a more convenient choice for the classical simulation. Consider $\Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}) := \prod_{i=1}^{m}\phi_{\omega_i}(\theta_i)$ where $\phi_0(\theta) = 1,\ \phi_1(\theta) = \cos(\theta),\ \phi_{-1}(\theta) = \sin(\theta)$ are *trigonometric monomials* that encode the $\boldsymbol{\theta}$ dependence. Then, the *noisy* variational

circuits admit the decomposition:

$$\tilde{\mathcal{U}}_\theta = \sum_{\boldsymbol{\omega}\in\{0,\pm1\}^m} \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta})\mathcal{D}_{\boldsymbol{\omega}}, \tag{6.23}$$

where each process $\mathcal{D}_{\boldsymbol{\omega}}$ is labelled by a frequency vector $\boldsymbol{\omega} \in [0, \pm 1]^m$ and given by:

$$\mathcal{D}_{\boldsymbol{\omega}} := \left(\bigcirc_i \tilde{\mathcal{C}}_i \circ \mathcal{D}_{\omega_i}\right) \circ \mathcal{C}_0. \tag{6.24}$$

In keeping with previous work [CJ17] we call these channels *process modes*. Crucially, each process mode maps one Pauli operator onto another Pauli operator, as can be seen from the PTMs above and the defining property of Clifford operations.

Overall, this new decomposition yields the following Fourier series representation for the cost function in Eq. (6.15):

$$\tilde{f}(\boldsymbol{\theta}) = \sum_{\boldsymbol{\omega}} d_{\boldsymbol{\omega}} \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}). \tag{6.25}$$

The *Fourier coefficients* are given by:

$$d_{\boldsymbol{\omega}} := \mathrm{Tr}(P\mathcal{D}_{\boldsymbol{\omega}}[|0\rangle\langle 0|]) = \langle\langle P|\mathbf{D}_{\boldsymbol{\omega}}|0\rangle\rangle. \tag{6.26}$$

Note that in the above, the Clifford unitaries $\mathcal{C}_i$ were noise-free and the parameterised rotation gates carried a time-independent Pauli noise. A similar decomposition arises when we consider the general Pauli noise model for $\tilde{\mathcal{C}}_i = \mathcal{M}_i \circ \mathcal{C}_i$. In this case, we denote the resulting process modes by $\mathcal{D}'_{\boldsymbol{\omega}} := (\bigcirc_i \mathcal{M}_i \circ \mathcal{C}_i \circ \mathcal{D}_{\omega_i}) \circ \mathcal{M}_0 \circ \mathcal{C}_0$ and the corresponding coefficients by $d'_{\boldsymbol{\omega}} = \langle\langle P|\mathbf{D}'\boldsymbol{\omega}|0\rangle\rangle$. We first describe and analyse the proposed classical algorithm for the simpler noise model that only affects the parameterised gates. This is purely to make the exposition easier to follow. The same principle works in the general case (see Sec. 6.2.6). Furthermore, the analysis extends to time-dependent Pauli errors. The noise models considered here also include the local depolarising channels that have been previously used in classical algorithms for noisy random circuit sampling [AGL+22]. Both in our case and in previous work there is an

implicit assumption that the Pauli error probabilities for each gate are known a-priori.

6.2.3 The LOWESA simulation algorithm

We are now in a position to state the simulation algorithm, which shares similar features to the algorithm in Ref. [GD18], but applied to the task of estimating expectation values and to a different family of circuits. We name it LOWESA for LOw Weight Efficient Simulation Algorithm (pronounced "low-EE-sa").

Given a cut-off parameter ℓ, LOWESA returns a *function* $\tilde{g}$ approximating the noisy cost function $\tilde{f}$ constructed from all the low Hamming weight $|\boldsymbol{\omega}| := \|\boldsymbol{\omega}\|_1 \leqslant \ell$ terms. This function is expressed as a trigonometric series and can therefore be used to evaluate the cost estimate for any parameter vector $\boldsymbol{\theta}$ using:

$$\tilde{g}(\boldsymbol{\theta}) = \sum_{|\boldsymbol{\omega}|\leqslant\ell} d_{\boldsymbol{\omega}} \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}), \tag{6.27}$$

with low computational effort. As the algorithm produces all $\{d_{\boldsymbol{\omega}}\}_{|\boldsymbol{\omega}|\leqslant l}$, and as the $\Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta})$ are simple trigonometric functions, after the algorithm has run the complexity of evaluating the function $\tilde{g}(\boldsymbol{\theta})$ at a given $\boldsymbol{\theta}$ does not depend directly on qubit number and depth.

LOWESA involves the following steps:

Algorithm 1 [**LOWESA**] Simulating cost functions of noisy VQAs with uncorrelated angles

Input: Quantum process given by Eq. (6.17) defined by process modes as in Eq. (6.24) with m parameters and Pauli rotation generators $\{P_i\}_{i=1}^m$; measurement Pauli operator P; cut-off parameter ℓ.
Output: $\tilde{g}(\boldsymbol{\theta})$, an approximation of $\tilde{f}(\boldsymbol{\theta})$.

procedure LOWESA
  $\tilde{g}(\boldsymbol{\theta}) \leftarrow 0$
  run BRANCH(P, (), m) recursively to yield $d_{\boldsymbol{\omega}} = \langle\langle 0|\mathbf{D}_{\boldsymbol{\omega}}^{\intercal}|P\rangle\rangle \; \forall \; |\boldsymbol{\omega}| \leq \ell$.
  for all non-zero $d_{\boldsymbol{\omega}}$ **do**
    $\tilde{g}(\boldsymbol{\theta}) \leftarrow \tilde{g}(\boldsymbol{\theta}) + d_{\boldsymbol{\omega}}\Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta})$
  end for
  return $\tilde{g}(\boldsymbol{\theta})$
end procedure

Subroutine: Calculate $d_{\boldsymbol{\omega}} \; \forall \; |\boldsymbol{\omega}| \leq \ell$ via recursion.

procedure BRANCH(Q, $\boldsymbol{\omega}$, i)
  $Q \leftarrow \mathcal{C}_i^{\dagger}(Q)$
  if $i > 0$ **then**
    if $[Q, P_i] = 0$ **then**
      BRANCH($\mathcal{D}_0^{i\dagger}(Q)$, append($\boldsymbol{\omega} \leftarrow 0$), $i-1$)
    else if $|\boldsymbol{\omega}| < \ell$ **then**
      BRANCH($\mathcal{D}_1^{i\dagger}(Q)$, append($\boldsymbol{\omega} \leftarrow 1$), $i-1$)
      BRANCH($\mathcal{D}_{-1}^{i\dagger}(Q)$, append($\boldsymbol{\omega} \leftarrow -1$), $i-1$)
    else
      break
    end if
  end if
  yield $d_{\boldsymbol{\omega}} = \langle\langle 0|Q\rangle\rangle$
end procedure

We shall now explain how Algorithm 1 works step by step and why it is efficient. Note that, while our exposition here deals with expectation values of a single Pauli operator, the results extend immediately to general observables as explained in Sec. 6.2.5.

Start with the target Pauli measurement operator P and propagate in the Heisenberg picture through the circuit. For each Clifford unitary C_i, updating the Pauli operator (by conjugation) takes at most $O(n^2)$. Each process mode $\mathcal{D}_{\omega_i}^i$ within a path $\mathcal{D}_{\boldsymbol{\omega}}$ acts with Pauli generator P_i. Note that the superscript indicates to which gate it corresponds and the subscript $\omega_i \in \{0, \pm 1\}$ labels the type of mode. The gate label is necessary since owing to the different generators P_i the process modes may be dif-

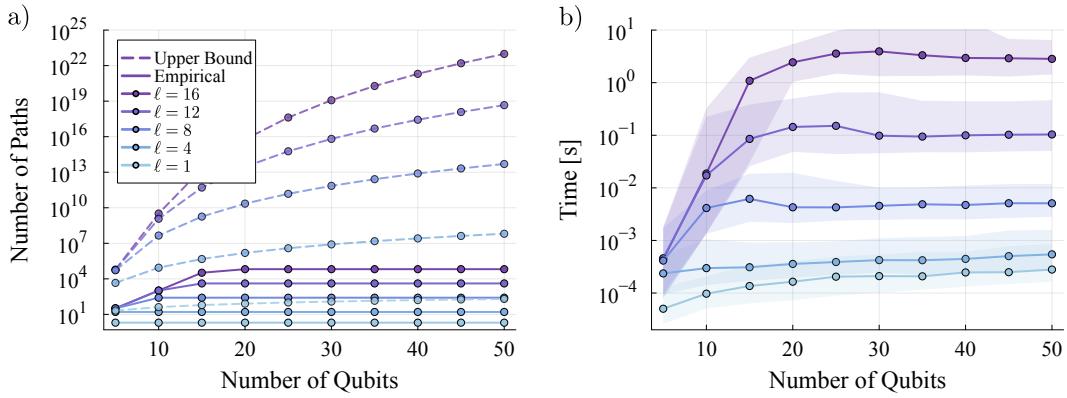


Figure 6.2: Scaling of LOWESA with the number of qubits n and cut-off parameter ℓ. The circuit structure consists of two parameterised layers of $H - Rz(\theta_i) - X - H$ on each qubit, where the Hadamard and X gates are chosen with 0.5 probability, followed by CNOTs placed on a 2D topology. a) Total number of paths for a given ℓ, which equals $\sum_{i=0}^{\ell} \binom{m}{i} 2^i$. Note that the number of paths that LOWESA needs to explore is dramatically lower. b) Wall time to run LOWESA with truncation parameter ℓ on an average laptop without parallelisation. Each data point represents an average over 500 different randomised circuits with Pauli Z measurement operators that act on a random subset of qubits. The shading shows the 90% confidence interval. The simulation of the Clifford gates used a look-up table, meaning that the scaling in n is entirely due the scaling of m with n.

ferent, however their effect on an arbitrary Pauli be easily evaluated making classical simulation possible. If the propagated Pauli operator commutes with the generator P_i then the rotation gate does not act and only $\mathcal{D}_0^i$ leads to a non-zero path, otherwise if it anticommutes either $\mathcal{D}_1^i$ or $\mathcal{D}_{-1}^i$ are valid choices. This can be seen by examining the effect of the PTM of $\tilde{\mathcal{R}}_z$ on the Pauli operators (Eq. (6.21)).

As only $\mathcal{D}_{\pm 1}^i$ contribute to the total weight $|\boldsymbol{w}|$ and since we impose $|\boldsymbol{w}| \leqslant \ell$, it suggests a binary tree-like data structure with ℓ layers to keep track of the change of Pauli frame and the different branching possibilities. This drastically improves the performance, as not all weight vectors produce valid paths that are non-zero: a branch may terminate sooner than if it propagated the Pauli through the entire circuit. By terminating a branch with weight zero we automatically only keep track of the paths leading to non-zero $d_{\boldsymbol{\omega}}$. As the weight of each paths is at most $|\boldsymbol{\omega}| \leqslant \ell$, then there are at most $\ell + 1$ levels in the binary tree. Some of the branches will terminate sooner, but the maximal number of nodes in level i is 2^i for $i \in [0, \ell]$. Note that updating the Pauli frame operator between any two consecutive nodes take $O(n^2\, k)$ where k is the number of $D_0's$ applied in between. However, the number of $D_0's$ applied within any branch satisfies $k \leqslant m$. There are at most 2^i Pauli operators in layer i, therefore, updating the layer $i+1$ given all the Pauli operators in layer i takes at most $O(n^2 2^i m)$. Putting all together it means that propagating P through all valid paths takes at most $\sum_{i=0}^{\ell-1} O(n^2 2^i m) = O(n^2 2^\ell m)$.

We note that the quadratic scaling in n is for general n-qubit Clifford unitaries, and can be improved for k-local (or sparse) unitaries. In that case, if the Clifford layers have maximum depth d, the runtime is $O(k^2 nd)$. Also if one fixes the set of Clifford unitaries that are executed within the circuits (for example the set $\{X, H, CNOT\}$), one can employ time-memory trade-off tools like look-up tables for each k-body Clifford operation and how they act on every k-body Pauli operator. In Fig. 6.2 we illustrate the runtime of LOWESA using this technique on a circuit structure that is typically challenging for classical simulators. Also note that the scaling with m provides a coarse upper bound. If it is attained then only D_0 processes modes are active, meaning that the circuit is trivial and there's no branching in that specific tree, and the complexity

will in that case be independent of the cut-off too. Therefore $k = m$ is never attained for any nontrivial circuit.

Finally, we need to verify that LOWESA yields an accurate simulation of the noisy cost function, i.e. evaluate the accuracy in the approximation $\tilde{g} \approx \tilde{f}$. We define the average L^2-norm error over the space of parameters $\Theta = [0, 2\pi]^m$ or root mean squared error (RMSE):

$$\Delta(\tilde{f}, \tilde{g}) := \left(\frac{1}{|\Theta|} \int_\Theta |\tilde{f}(\boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \right)^{1/2}, \tag{6.28}$$

where the integration measure is $d\boldsymbol{\theta} = d\,\theta_1 d\,\theta_2 ... d\,\theta_m$ and $|\Theta| = (2\pi)^m$ is a normalisation factor so that $\frac{1}{|\Theta|} \int d\,\boldsymbol{\theta} = 1$. We prove the following result:

Theorem 7. *Consider a n-qubit VQA with a PQC as in Eq.* (6.14) *having m independently parameterised rotations affected by a single-qubit Pauli noise channel* $\mathcal{N}_{Pauli}(p_X, p_Y, p_Z)$ *as in Eq.* (6.17)*. Recall that* $p = \min_{\sigma=X,Y,Z} p_\sigma > 0$.

Then, for any weight cut-off $\ell \in \mathbb{N}$*,* LOWESA *(Algorithm 1) returns an approximation* $\tilde{g}$ *for the noisy cost function* $\tilde{f}$ *with RMSE:*

$$\Delta(\tilde{f}, \tilde{g}) \leq (1 - 2p)^{\ell+1} \leq e^{-2p\ell}, \tag{6.29}$$

and runs in time at most $O(n^2 m\, 2^\ell)$.

Proof. Using Eq. (6.25) we can rewrite Eq. (6.28) as:

$$\Delta^2(\tilde{f}, \tilde{g}) = \frac{1}{|\Theta|} \int_\Theta \Big| \sum_{|\boldsymbol{\omega}|>\ell} \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}) d_{\boldsymbol{\omega}} \Big|^2 d\boldsymbol{\theta}. \tag{6.30}$$

Then using the fact that the trigonometric monomials $\Phi_{\boldsymbol{\omega}}$ are orthogonal:

$$\frac{1}{|\Theta|} \int_\Theta \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}) \Phi_{\boldsymbol{\omega}'}(\boldsymbol{\theta}) d\boldsymbol{\theta} = 2^{-|\boldsymbol{\omega}|} \delta_{\boldsymbol{\omega}\boldsymbol{\omega}'}, \tag{6.31}$$

and thus form a basis, we derive the appropriate Parseval's theorem:

$$\Delta^2(\tilde{f}, \tilde{g}) = \sum_{|\boldsymbol{\omega}|>\ell} 2^{-|\boldsymbol{\omega}|} |d_{\boldsymbol{\omega}}|^2. \tag{6.32}$$

Now consider the Fourier coefficients $d_{\boldsymbol{\omega}}$. Recall from Sec. 6.2.2 that these are defined as $\langle\langle P|\mathbf{D}_{\boldsymbol{\omega}}|0\rangle\rangle$, and that the process modes $\mathbf{D}_{\boldsymbol{\omega}}$ (and their adjoints) map one Pauli operator to another Pauli operator, scaled by products of the Pauli channel eigenvalues $q_{X/Y/Z}$. Also note that both Pauli noise and Pauli rotation channels can only act nontrivially on a Pauli operator (i.e. modify the operator) when that operator is not identity on the qubits where the channel is active. Since by our model (Eq. (6.17)) rotation gates are always followed by noise channels acting nontrivially on the same qubits, we can infer that whenever a rotation gate R acts nontrivially on a Pauli operator O, the output operator will also be rescaled by a noise-dependent coefficient. This coefficient will depend on how many components of O are nonidentity on the qubits where R acts nontrivially, but by the argument above there must be at least one such component. Therefore by our assumptions on the noise the operator is scaled by at most $q = 1 - 2p < 1$.

If we define the zero-noise coefficients $d^0_{\boldsymbol{\omega}}$ by setting $p_{X/Y/Z} = 0$, we see that $|d_{\boldsymbol{\omega}}| = Q_{\boldsymbol{\omega}}|d^0_{\boldsymbol{\omega}}|$, where $0 < Q_{\boldsymbol{\omega}} \leq q^{|\boldsymbol{\omega}|}$. Hence we can write:

$$\Delta^2(\tilde{f}, \tilde{g}) = \sum_{|\boldsymbol{\omega}|>\ell} Q^2_{\boldsymbol{\omega}} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2 \tag{6.33}$$

$$\leq q^{2(\ell+1)} \sum_{|\boldsymbol{\omega}|>\ell} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2 \tag{6.34}$$

$$\leq q^{2(\ell+1)} \sum_{\boldsymbol{\omega}} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2. \tag{6.35}$$

Now we recognise that by Parseval's theorem the summation relates to the noise-less cost function $f(\boldsymbol{\theta})$ as:

$$\sum_{\boldsymbol{\omega}} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2 = \frac{1}{|\Theta|} \int_\Theta |f(\boldsymbol{\theta})|^2 d\boldsymbol{\theta} \leq 1. \tag{6.36}$$

where we used the fact that, for Pauli measurements, $|f(\boldsymbol{\theta})| \leq 1$. This implies $\Delta(\tilde{f}, \tilde{g}) \leqslant q^{\ell+1}$, which gives a non-trivial bound whenever $q < 1$, which again is a reasonable assumption. To simplify the expression further, we use the identity $1 - x \leq e^{-x}\, \forall x \in \mathbb{R}$ and find $q^{\ell+1} \leq e^{-2\ell}$. □

It follows from Theorem 7 that LOWESA is both accurate and efficient, as its runtime scales polynomially with m and n and the maximum allowed RMSE; however, the scaling with noise probability is considerably worse. For example, suppose we wish to have an error bounded by ϵ. Then one would choose $\ell \approx \frac{1}{2p} \log \epsilon^{-1}$, giving a runtime $O(\epsilon^{-\frac{\log 2}{2p}} n^2 m)$. While this is asymptotically efficient in the width and depth of the circuit, the dependency on the error rate limits its practicality. Notably the exponent may still be considerably large if the noise is small. When the goal is to simulate the expected outcome of a hardware implementation with a finite number of measurements N_s, the error can be chosen like $\epsilon \in \mathcal{O}(\frac{1}{\sqrt{N_s}})$, thus relaxing the precision requirements.

In Fig. 6.3 we illustrate the mean accuracy of the algorithm for an example circuit of the hardware-efficient family. We observe that the error is typically up to two orders of magnitude lower than the bounds, suggesting these are loose and may be improved for the typical case.

6.2.4 Validity of error measure

The use of RMSE as error measure has limitations, the main one being that the error at any given point is in principle unbounded. However we argue that this limitation is weaker than may appear. Applying Markov's inequality to our Theorem we have the following probabilistic bound:

Corollary 3. *For a fixed circuit, choosing the parameters $\boldsymbol{\theta}$ uniformly at random from $[0, 2\pi]^m$, with probability $\geq 1 - \delta$ the approximation error is bounded by:*

$$|\tilde{f}(\boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta})| \leq e^{-p\ell} \delta^{-1/2}. \tag{6.37}$$

Suppose that we wish to have error bounded by ϵ with probability $1 - \delta$. Then the required cutoff is $\ell \approx p^{-1} \log(\epsilon^{-1} \delta^{-1/2})$, giving again a runtime that scales unfavorably with p. However for fixed p the scaling is logarithmic in both δ and ϵ meaning that the probability of encountering large deviations can be made arbitrarily small by increasing ℓ. This probabilistic formulation has practical relevance as typical VQAs have their parameters initialised uniformly at random [CAB+21] and so it is valid at initialisation;

however, this analysis breaks down when considering the error over the whole path of gradient descent, which may lead into a region of high deviation.

6.2.5 General measurement operators

Up to now we assumed that the measurement operator is Pauli, however in truth most practical algorithms have more complicated measurement operators. Generally, a measurement takes the form:

$$O = \sum_{P_i \in \mathbb{P}^n / \{I^n\}} c_i P_i, \tag{6.38}$$

where we can ignore the identity component as it contributes a constant to the cost function. We get the following result:

Theorem 8. *With a general measurement operator as in Eq.* (6.38)*, LOWESA can simulate the noisy cost function with RMSE* $\Delta \leq \epsilon$ *and with runtime at most:*

$$O((\|\boldsymbol{c}\|_r \epsilon^{-1})^{\frac{\log 2}{2p}} n^2 m). \tag{6.39}$$

Assuming $p \ll \log \sqrt{2}$*,* $r \approx 1$*.*

Proof. By linearity:

$$\tilde{f}(\boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta}) = \sum_i c_i (\tilde{f}_i(\boldsymbol{\theta}) - \tilde{g}_i(\boldsymbol{\theta})), \tag{6.40}$$

where the subscript indicates that the observable P_i is measured. Once again the error is defined as:

$$\Delta^2 = \mathbb{E}_{\boldsymbol{\theta}} (\tilde{f}(\boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta}))^2, \tag{6.41}$$

where again we assume the parameters are sampled from a uniform distribution.

Now we let $e_i(\boldsymbol{\theta}) = \tilde{f}_i(\boldsymbol{\theta}) - \tilde{g}_i(\boldsymbol{\theta})$. Expanding Eq. (6.41) using Eq. (6.40) and using

Cauchy-Schwarz:

$$\Delta^2 = \sum_{i,j} c_i c_j \, \mathbb{E}_{\boldsymbol{\theta}} e_i e_j \tag{6.42}$$

$$\leq \sum_{i,j} |c_i||c_j| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(e_i^2)\mathbb{E}_{\boldsymbol{\theta}}(e_j^2)} \tag{6.43}$$

$$= \left(\sum_i |c_i| \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(e_i^2)} \right)^2 . \tag{6.44}$$

Now using Theorem 7 and letting the cutoff vary with i we get:

$$\Delta \leq \sum_i |c_i| e^{-2p\ell_i}. \tag{6.45}$$

To take into account the simulation cost, we seek to minimise:

$$\sum_i |c_i| e^{-2p\ell_i} + \lambda \sum_i 2^{\ell_i}, \tag{6.46}$$

with λ a Lagrange multiplier. The solution is:

$$\ell_i = \frac{1}{2p + \log 2} \log |c_i| + \log k, \tag{6.47}$$

where k is a positive constant. The error is therefore:

$$\Delta \leq k^{-2p} \sum_i |c_i|^{\frac{\log 2}{\log 2 + 2p}}. \tag{6.48}$$

Alternatively, if we require $\Delta \leq \epsilon$, the total simulation cost is:

$$\sum_i 2^{\ell_i} = k^{\log 2} \sum_i |c_i|^{\frac{\log 2}{\log 2 + 2p}} \tag{6.49}$$

$$\leq \epsilon^{-\frac{\log 2}{2p}} \left(\sum_i |c_i|^{\frac{\log 2}{\log 2 + 2p}} \right)^{1 + \frac{\log 2}{2p}} \tag{6.50}$$

$$= (\epsilon^{-1} \|c\|_r)^{\frac{\log 2}{2p}}, \tag{6.51}$$

with $r = \frac{\log 2}{\log 2+2p}$, which ≈ 1 when $2p \ll \log 2$. Therefore the runtime is in:

$$O((\epsilon^{-1}\|c\|_r)^{\frac{\log 2}{2p}} n^2 m). \tag{6.52}$$

□

The simulation procedure can be parallelised meaning that the actual running time can be reduced considerably from these estimates. Comparing this with the maximum number of shots required by a quantum computer to approximate a composite observable with precision ϵ, $N_s = \|\boldsymbol{c}\|_1^2 \epsilon^{-2}$ [WHT15], we conclude that when $p \ll \log\sqrt{2}$ LOWESA incurs a cost at most polynomially larger than usual sampling cost. Once again, in practical scenarios the factor of p^{-1} will dominate the exponent but this does not invalidate the claim of classical simulatability. Thus we conclude that any expectation value that can be measured efficiently on a quantum computer may be efficiently simulated using our algorithm.

6.2.6 General Pauli noise models

The result can be extended to cover multi-qubit Pauli noise affecting all gates, not just the parameterised ones.

Theorem 9. *Consider an n-qubit VQA under the noise model:*

$$\tilde{\mathcal{U}}_{\boldsymbol{\theta}} = (\bigcirc_{i=1}^{m} \mathcal{M}_i \circ \mathcal{C}_i \circ \mathcal{N}_i \circ \mathcal{R}_i(\theta_i)) \circ \mathcal{M}_0 \circ \mathcal{C}_0, \tag{6.53}$$

where $\{\mathcal{M}_i\}$ are n-qubit Pauli channels with layer-dependent noise parameters and every rotation is followed by a local multi-qubit Pauli noise $\mathcal{N}_i = \bigotimes_{j=1}^{n} \mathcal{N}_{Pauli}^{(j)}(p_X^{ij}, p_Y^{ij}, p_Z^{ij})$ with $p' = \min_{ij\sigma}\{p_\sigma^{ij}\} > 0$, which depends on both layer and qubit. Then, for any weight cut-off $\ell \in \mathbb{N}$, LOWESA (Algorithm 1) with modified process modes $\mathcal{D}'_{\boldsymbol{\omega}} = (\bigcirc_{i=1}^{m} \mathcal{M}_i \circ \mathcal{C}_i \circ \mathcal{D}_{\omega_i}) \circ \mathcal{C}_0 \circ \mathcal{M}_0$ and coefficients $d'_{\boldsymbol{\omega}} = \langle\langle P|\mathbf{D}'_{\boldsymbol{\omega}}|0\rangle\rangle$ returns an approximation $\tilde{g}$ for the cost function $\tilde{f}$ with error:

$$\Delta(\tilde{f}, \tilde{g}) \leq (1-2p')^{\ell+1} \leq e^{-2p'\ell}, \tag{6.54}$$

and runs in time at most $O(n^2 m 2^\ell)$.

Proof. The main difference from the proof of Theorem 7 is that the n-qubit unitaries $\mathcal{C}_i$ are noisy and are replaced by $\mathcal{M}_i \circ \mathcal{C}_i$. However, since $\mathcal{M}_i$ are Pauli channels, then its adjoint acts on any Pauli operator as $\mathbf{M_i}^T |P\rangle\!\rangle \propto |P\rangle\!\rangle$, where the proportionality factor is determined by the eigenvalues of $\mathcal{M}_i$. These are assumed to be accessible, e.g from previous benchmarking experiments. Therefore, the total number of valid, non-zero process modes is also 2^ℓ and there are m Pauli channels $\mathcal{M}_i$ so computing the proportionality factor takes at most $O(m\, 2^\ell)$, which means it does not affect the overall complexity in determining the Fourier coefficients $d'_{\boldsymbol{\omega}} = \langle\!\langle P|\mathbf{D}'_{\boldsymbol{\omega}}|0\rangle\!\rangle$ with $|\boldsymbol{\omega}| \leqslant \ell$ which can be computed, as previously, in $O(n^2 m 2^\ell)$.

It remains to show that the average approximation error $\Delta(\tilde{f}, \tilde{g})$ still decays exponentially with the cut-off parameter l. Like before, the noisy cost function is given by:

$$\tilde{f}(\boldsymbol{\theta}) = \sum_{\boldsymbol{\omega}} d'_{\boldsymbol{\omega}} \Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta}), \tag{6.55}$$

where $d'_{\boldsymbol{\omega}} = Q'(\boldsymbol{\omega}) d^0_{\boldsymbol{\omega}}$, $Q'(\boldsymbol{\omega}) \leq q^{|\boldsymbol{\omega}|}$ if the noise $\mathcal{N}$ on the parameterised gates $\mathcal{R}_i(\theta_i)$ is a fixed, space- and time-independent Pauli channel with eigenvalues (q_X, q_Y, q_Z) and $q := \max_\sigma \{q_\sigma\}$. More generally however, if $\mathcal{N}$ carries a space- and time-dependency with possibly different eigenvalues $(q_X^{ij}, q_Y^{ij}, q_Z^{ij})$ for each of the parameterised gates $\mathcal{R}_i(\theta_i)$ and qubit j, then we have $Q'(\boldsymbol{\omega}) \leq \prod_i (q^i)^{|\omega_i|}$, where $q^i := \max_{j\sigma} \{q_\sigma^{ij}\}$. Note that in this situation, the process modes for each site $\mathcal{D}^i_{\omega_i}$ will have the same form as in the previous analysis but with different parameters that depend on the location.

Finally, orthogonality of the trigonometric functions $\Phi_{\boldsymbol{\omega}}(\boldsymbol{\theta})$ ensures we get:

$$\Delta^2(\tilde{f}, \tilde{g}) = \sum_{|\boldsymbol{\omega}|>\ell} Q'^2(\boldsymbol{\omega}) 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2 \tag{6.56}$$

$$\leq (\max_{|\boldsymbol{\omega}|>\ell} Q'(\boldsymbol{\omega}))^2 \sum_{\boldsymbol{\omega}} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2. \tag{6.57}$$

The term within the brackets is the largest $(\ell+1)$ product of the q^i's. This can be given the trivial upper bound $1 - 2p'$ by defining $p' := \min_{ij\sigma} \{p_\sigma^{ij}\}$. In this case we must have $p' > 0$. Finally, the sum $\sum_{\boldsymbol{\omega}} 2^{-|\boldsymbol{\omega}|} |d^0_{\boldsymbol{\omega}}|^2$ can be bounded by 1 as explained in the

proof of Theorem 7. □

The result relies on the fact that any Pauli channel will map a propagated Pauli operator to itself, up to a proportionality factor that can be at most 1. In other words, this means that each of the modified process modes $\mathcal{D}'_{\boldsymbol{\omega}}$ will act similarly to the previously considered modes $\mathcal{D}_{\boldsymbol{\omega}}$ arising from the simplified error model, so that $\mathbf{D}'_{\boldsymbol{\omega}}|P\rangle\!\rangle \propto \mathbf{D}_{\boldsymbol{\omega}}|P\rangle\!\rangle$. Therefore the proof and the bounds follow in the same way as for Theorem 7. The only modification to the algorithm is that to compute $d'_{\boldsymbol{\omega}}$ one must also keep track of these proportionality factors along with the propagated Pauli.

For generality we have not assumed that the all noise coefficients of $\{\mathcal{M}_i\}$ are bigger than 0, thus it is difficult to improve upon the upper bound on the approximation error Δ since one can be in a situation where along the paths of weight $|\boldsymbol{\omega}| = \ell + 1$ the proportionality factors might all be 1 when propagating the Pauli operator through each Pauli channel $\mathcal{M}_i$. In practical situations the Clifford gates will come with a depolarising component and the bound can be improved. For instance, let's assume that in the decomposition of the n qubit Clifford operator $\mathcal{C}_i$ into primitive (single and two-qubit) gates each incurs a local single-qubit depolarising channel $\mathcal{N}_{dep}$ with error probability η. Then it follows we can find a tighter bound:

$$\Delta(\tilde{f}, \tilde{g}) \leqslant (1-2p')^{\ell+1}(1-\eta)^{\ell+1} \leq e^{-(2p'+\eta)\ell}. \tag{6.58}$$

This comes from the fact that $\mathcal{N}^{\dagger}_{dep}(P) = (1-\eta)P$ if $P \in \{X, Y, Z\}$ and $\mathcal{N}^{\dagger}_{dep}(I) = I$ along with the previous observation that for valid paths leading to non-zero coefficients, $\mathcal{D}_{\pm 1}$ are applied to qubit q_i whenever the propagated Pauli on qubit q_i is not I or Z, since these commute with the generator Z and therefore as we have seen only the mode $\mathcal{D}_0$ is active. Therefore the noise from the Clifford part will contribute and at the very least contract by a factor of $(1-\eta)$ whenever we have a branching possibility to apply either $\mathcal{D}_{+1}$ or $\mathcal{D}_{-1}$, which are the only contributors to the total weight $|\boldsymbol{\omega}|$. Note that this type of noise model has previously been considered in the context of noisy random circuit sampling [AGL$^+$22]. The corresponding result for non-Pauli observables can be obtained similarly as before, giving the same additional factor in the runtime.

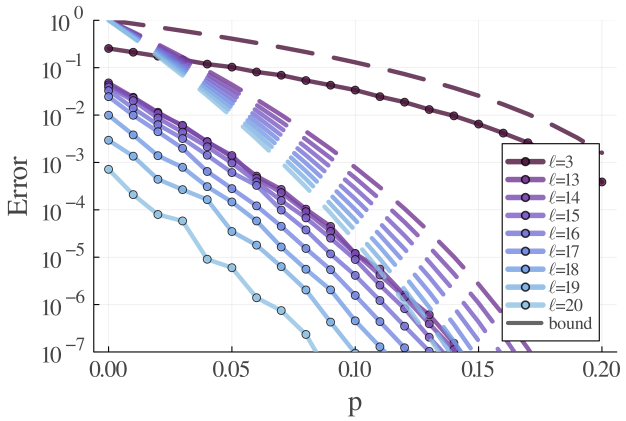


Figure 6.3: Accuracy benchmark of LOWESA compared to the error bounds as predicted in Theorem 7. We show the L^2 error of a single-qubit Pauli Y operator expectation with $\ell < m = 60$ for two layers of a $n = 10$ qubit circuit. The circuit consists of parameterised single-qubit gates $R_z(\boldsymbol{\theta}_i)\, R_x(\boldsymbol{\theta}_{i+1})\, R_z(\boldsymbol{\theta}_{i+2})$ on each qubit followed by CNOT gates in a 2D topology. For this particular circuit, each entangler in the 2D topology was placed with a 0.5 probability. The noise model is symmetric depolarising noise, where the parameters are set $p_X = p_Y = p_Z = p$. Each point is averaged over 1000 random parameterisations of the same circuit to compare to the integral definition of our error bounds. All paths below $\ell = 3$ and above $\ell = 21$ annihilate. Consequently, the simulation with $\ell = 21$ is exact.

6.2.7 Fixed (unparameterised) non-Clifford gates

The extension of LOWESA to the case where non-Clifford unparameterised gates are present is straightforward. As was done in Ref. [FRDC22a], one may treat non-Clifford rotation gates as parameterised rotation gates that have their parameters fixed on at a later stage. A circuit with t fixed z-rotation gates and m parameterised z-rotation gates may be transformed into a circuit with $m+t$ z-rotations for simulation purposes, obtaining a cost function $F(\boldsymbol{\theta}, \boldsymbol{\phi})$. Then the intended cost function is obtained by fixing $\boldsymbol{\phi}$. It follows that any statement on the simulation runtime still applies with the substitution $m \to m + t$. However, getting an error bound with non-Clifford gates is more complicated, since we can no longer average over the expanded parameter space owing to the fixed gates. We can still make a weaker probabilistic statement.

Theorem 10. *Consider a variational circuit consisting of m uncorrelated noisy parameterised rotation gates, and t noisy rotation gates with fixed random angles inde-*

pendently and uniformly distributed in $[0, 2\pi]^t$*. The noise model is that of Theorem 9. Then for weight cut-off* $\ell \in \mathbb{N}$*, with probability* $\geq 1 - \delta$ *the simulation error of* LOWESA *(Algorithm 1) with modified process modes obeys:*

$$\Delta(\tilde{f} - \tilde{g}) \leq e^{-2p'\ell}\delta^{-1/2}, \tag{6.59}$$

and the Algorithm runs in time $O(n^2(m+t)2^\ell)$*.*

Proof. We can use the trick of considering the random t angles to be variables on a space $\Phi = [0, 2\pi]^t$, and replicate the proofs of Theorems 7 and 9 on the expanded parameter space $\Theta \otimes \Phi = [0, 2\pi]^{m+t}$. This gives the bound:

$$\mathbb{E}_{\boldsymbol{\phi}}\Delta^2(\tilde{f}, \tilde{g}) = \frac{1}{|\Theta||\Phi|}\int_\Phi \int_\Theta |\tilde{f}(\boldsymbol{\theta}, \boldsymbol{\phi}) - \tilde{g}(\boldsymbol{\theta}, \boldsymbol{\phi})|^2 d\theta d\phi \tag{6.60}$$

$$\leq e^{-4p'\ell}. \tag{6.61}$$

Now we can use Markov's inequality on the random variable $\Delta(\tilde{f}, \tilde{g})$ (as it now depends on the random $\boldsymbol{\phi}$) and obtain the intended result. The running time is bounded by $O(n^2(m+t)2^l)$ since the number of rotation gates is now $m + t$. □

Theorem 10 implies that for a typical choice of the $\boldsymbol{\phi}$ angles the error is still exponentially suppressed in ℓ. For fixed δ, one would choose $\ell \approx \frac{1}{2p'}(\log \epsilon^{-1} + \frac{1}{2}\log \delta^{-1})$, giving a runtime which is only slightly worse than the one from the previous Theorems, for reasonable choices of δ.

6.3 The case of correlated parameters

The main result has been derived assuming that the parameters controlling the rotation gates in the circuits are uncorrelated. One may therefore wonder whether it extends to correlated parameter circuits, which are ubiquitous in quantum machine learning [BWP+17] as well as forming the basis of algorithms like the Quantum Approximate Optimisation Algorithm (QAOA) [FGG14, HWO+19] or the Hamiltonian Variational Ansatz (HVA) for chemistry problems [WHT15, KMT+17].

However, the argument used in the proof of Theorem 7 does not hold since with correlated angles the basis functions are no longer orthogonal over the correlated parameter space. For example, consider the following case where:

$$\Phi_2(\theta) = \cos^2(\theta), \quad \Phi_{-2}(\theta) = \sin^2(\theta) \tag{6.62}$$

$$\Rightarrow \frac{1}{2\pi}\int \Phi_2(\theta)\Phi_{-2}(\theta)d\theta = \frac{1}{8} \neq 0. \tag{6.63}$$

Interestingly, we find that for correlated angles systems the simulation algorithm frequently returns a trivial result. Consider the following 1-qubit correlated parameter circuit:

$$U_d(\theta) = (R_x(\theta))^d, \tag{6.64}$$

with d some integer > 1. It is simple to show that when $U_d(\theta)$ is applied to the initial state $|0\rangle$, then measuring the Pauli Z produces the cost function:

$$f_d(\theta) = \cos(d\theta) = \sum_{i=0}^{\lfloor d/2 \rfloor} (-1)^i \binom{d}{2i} \sin^{2i}(\theta)\cos^{d-2i}(\theta), \tag{6.65}$$

whose terms are all of weight d. Therefore, any reconstruction with weight $\ell < d$ would trivially return $\tilde{g} = 0$. This behaviour can be generalised to any circuit composed of d repeated, identical, and independently parameterised layers:

$$U(\boldsymbol{\theta}) = \prod_{i=1}^{d} V(\boldsymbol{\theta}_i) = \prod_{i=1}^{d} \left(\prod_{j=1}^{h} e^{-iH_j\theta_{ij}} \right), \tag{6.66}$$

where each layer is generated by the *same* h Hamiltonians. It can be observed that both QAOA and HVA ansätze fit in the prescription. In this situation, for LOWESA to produce a non-zero approximation function $\tilde{g}$, we can show that the cut-off value ℓ has to be greater than the number of repeated layers. We begin with the following Lemma:

Lemma 9. *Consider a circuit in the form of Eq.* (6.66) *and its expansion in process modes. Then given a Pauli operator P, either i) it commutes with all generators $\{H_j\}$, or ii) the process modes that do not annihilate P all have weight $|\boldsymbol{\omega}| \geq d$.*

Proof. First note that, given a Pauli operator P_i and a Hermitian H with $[P_i, H] \neq 0$, then the corresponding unitary acts nontrivially on P, more precisely:

$$e^{-iH\theta} P_i e^{iH\theta} = \sum_j c_j P_j, \tag{6.67}$$

with $c_j \neq 0$ for at least one $j \neq i$. In that case it is easy to show that $[P_j, H] \neq 0$ for all such a P_j.

Therefore we see that if at layer i the unitary $e^{-iH_j\theta_{ij}}$ acts nontrivially on a Pauli P, even if all its products commute with all subsequent unitaries, then they cannot also commute with the next unitary generated by H_j. This means that either P commutes with all generators, or at least one unitary must act nontrivially per layer. In the latter case, it follows that the weight of any process mode that does not annihilate P must necessarily be greater than or equal to the total number of layers d. □

Then we have as a corollary:

Theorem 11. *Given $U(\boldsymbol{\theta})$ as in Eq.* (6.66) *and a Pauli operator P that does not commute with at least one of the generators $\{H_j\}$. If the cut-off $\ell < d$ then* LOWESA *produces a trivial approximation $\tilde{g} = 0$ of the noisy expectation value $\mathrm{Tr}(U(\boldsymbol{\theta})(\rho_0)U^\dagger(\boldsymbol{\theta})P)$ such that $\tilde{g} = 0$ for any noise model and at any noise level.*

The theorem follows simply by observing that if $\ell < d$, Algorithm 1 must return $\tilde{g}(\boldsymbol{\theta}) = 0$ since all process modes annihilate P. The result holds equally if noise is present as noise simply adds up to a constant for each mode in the expansion.

This result implies that the complexity requirements to get a nontrivial result with LOWESA will scale exponentially $\Omega(2^d)$ with the number of layers. Correlating the angles further, for example by setting $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_p$ does not affect the validity of the result. Improvements to the runtime may be possible if the number of valid paths can be reduced, for instance by leveraging symmetries in the circuit.

While the simulation algorithm may appear to fail for correlated angles since the output is constant for $\ell < d$, in fact we have not considered that the simulation RMSE may still be small if the *noisy* cost function $\tilde{f}$ has a small variance, since (assuming

$\ell < d$ and $\mathbb{E}_{\boldsymbol{\theta}} \tilde{f} = 0$):

$$\Delta^2(\tilde{f}, \tilde{g}) = \Delta^2(\tilde{f}, 0) = \mathbb{E}_{\boldsymbol{\theta}} \tilde{f}^2 = \mathrm{Var}_{\boldsymbol{\theta}}(\tilde{f}). \tag{6.68}$$

Indeed this is the case due to the phenomenon of noise-induced barren plateaus (Chapter 4): for our model the cost function variance would decay with depth as $O(e^{-pd/\ln 2})$ [WFC+21, Lemma 1]. Therefore it is still possible that the result in Theorem 7 may hold for correlated parameter VQAs too. For now, however, we are unable to conclusively demonstrate it, so this leaves room for a quantum advantage in QAOA and HVA, as well as in simulating time evolution on noisy quantum devices, as such tasks commonly involve repeated gate patterns.

6.4 Discussion

In this Chapter we introduce LOWESA, an algorithm to approximately classically simulate the cost function of variational quantum algorithms, given any observable that can be efficiently measured on a quantum computer. Crucially, the algorithm is constructive, in that it outputs a *function* of circuit parameters that approximates the entire noisy landscape rather the observable's noisy expectation value at some fixed parameters. We show that for circuits with *independently* parameterised non-Clifford gates and efficiently measurable observables, our procedure gives a polynomial-time algorithm in both the number of qubits and depth, with an upper bound on the average error that decays exponentially with the physical error rate and a controllable cut-off parameter. The implication is that generic variational quantum algorithms with independent parameters, fixed efficiently measurable observable and under constant physical gate error rate can be efficiently simulated classically.

We emphasise that the approximation error measure we employ is an average over the entire parameter space. The claim of efficient classical simulatability for estimating expectation values in the presence of noise should be understood for a *typical circuit* within a family of circuits with fixed structure (i.e. fixed Clifford unitaries on an arbitrary topology interleaved with arbitrary non-Clifford z-rotations) and measurement

operator. For the case of a PQC with uncorrelated parameters, this corresponds to a typical parameter constellation. On the other hand, when the circuit family contains fixed, non-Clifford gates, then our results hold only probabilistically (Theorem 10). Thus, we do not claim the ability to efficiently simulate all noisy Clifford+T circuits. At the same time, the aforementioned result also indicates that, for given circuit parameters, the probability to get an approximation error larger than the target accuracy of our algorithm also decreases exponentially with the cutoff parameter ℓ. While the cutoff is tunable, the algorithm's computational cost scales exponentially in ℓ in the worst case. This behaviour is similar to what was observed in Ref. [GD18], which shows classical simulatability of sampling from generic (random) noisy circuits except a zero-measure subset of (fixed, structured) circuits.

It can be notice that all our results suffer from the same limitation: fixing the simulation error, the scaling of the cutoff with p implies a runtime of $2^{\Theta(1/p)}$, meaning that we require $1/p \in o(m)$ otherwise we recover the noiseless scaling of the algorithm. Another perspective is that there is a *minimum noise probability threshold* of $1/m$ below which LOWESA loses any claim of advantage. This is expected since, if one interprets p as the probability of error per rotation gate, then mp is the total expected number of errors, which we require to be $\gg 1$ otherwise noise will not have a significant effect on the cost function. Thus the quantity mp may represent a crude measure for the capability of LOWESA to simulate a noisy VQA.

The result can be placed within a broader range of research [GD18, AGL$^+$22, SFGP21a, DPMRF23, ZSW20, ABOIN96] that aims to establish the extent to which noise in quantum computations hinders any potential quantum advantage. The works in Refs. [GD18, AGL$^+$22], which inspired our algorithm, are specific to the task of simulating random circuit sampling and thus rely on different assumptions on circuit structure and output state. Recent frameworks [CCHL22] show, up to oracular access, that specific circuit structures can exhibit a noise-robust quantum advantage. Our results are consistent with this because of the intrinsically probabilistic nature of our claims. However, it has also been shown that finite noise can introduce an exponential separation between an algorithm for learning quantum states running on a fault-tolerant

quantum computer vs a NISQ device [HKP20]. Similarly, our results imply that, in the presence of sufficiently large levels of noise, a generic, wide range of VQAs become classically simulatable. This type of conclusion has been reached in Ref. [SFGP21a], where comparisons with classical algorithms lead to trade-offs between physical error rates and depth limitations on variational Hamiltonian optimisation algorithms. For tensor network approaches [ZSW20], truncation error accuracy is impacted by connectivity and has only been empirically related to noise. In contrast, our approach gives a constructive classical algorithm to recover the entire cost function, with provable bounds on accuracy (for the circuit families considered) and does not assume a particular problem or architecture topology. We note that our results in their present form do not apply to variational algorithms that sample from the output state, such as QAOA or quantum generative modeling [POBRGB18,BGPP$^+$19,AAR$^+$18,DDK18]. These may be avenues for future exploration.

Besides the implication for the complexity of noisy VQAs, LOWESA may have a place as a useful simulation algorithm for the NISQ era. While for fixed physical error rate per gate our algorithm scales polynomially in the number of qubits and depth, the complexity grows exponentially with decreasing error rates, in the worst case. However, in practice, it may possible to have better scaling for realistic circuits, for instance if the cost function is dominated by low-weight terms. Our experiments (Fig. 6.3) provide some empirical evidence that this is the case, supplementing similar findings in Ref. [FRDC22a]. LOWESA may be used to generate classical surrogates of VQAs [SEM23, LTD$^+$22], allowing model training without the hybrid optimisation loop. Unlike to other algorithms for this task, it has the advantage of requiring no samples from the quantum computer. As such it may be more directly compared with tensor network methods [RMM$^+$23, TFSS23], with some important differences. Crucially, LOWESA generates the entire cost function, and its efficiency does not directly depend on the entanglement of the state, attributes that may make it preferable to tensor networks in some situations, for instance when the entangling gates are not geometrically local. On the other hand, tensor networks can be used to approximate the circuit's output state on which any expectation value or bitstring probability may

be evaluated, which contrasts with LOWESA's dependency on the measurement operator and specificity to expectation values. In any case, by using LOWESA for the noiseless setting one necessarily abandons the rigorous guarantees of accuracy established here. Nonetheless, there are promising signs that it may still be useful in this regime [RFHC23]. Without the requirement of accuracy guarantees one is also free to implement different heuristics for branch cutting, since our solution was tailored to the noise model. Similar algorithms with different branch cutting heuristics have shown promise [NKF23, BHC23, BC23].

To conclude, classical simulation algorithms such as that presented here can not only serve as benchmarking tools for NISQ devices at larger scales but most importantly, they help establish a threshold where quantum computers, given sufficiently low physical error rates, produce results that are no longer reproducible with classical computing resources. From this perspective, they are essential tools to determine the full picture of resource requirements for practical quantum applications.

Chapter 7

Conclusion

In this Thesis I have aimed to give an overview of my research in the roughly five years of my PhD, striving to be as cohesive as possible. As a conclusion of the Thesis, it feels therefore appropriate to try and distill some overarching message from my findings, and zooming out, to tie such a message to the evolution of the field as a whole during the same period. Indeed I have not worked in a vacuum: as any researcher will attest, producing scientific work is to a great extent an interplay with a wider community.

7.1 Looking back

When I started my PhD in 2019, quantum computing was something in the future, yet there was a not insignificant amount of hope that variational quantum algorithms could turn out to be useful even in the NISQ regime. For the previous five years, there had been growing interest in these algorithms that seemed simultaneously easier to implement, useful (at least for quantum chemistry), not easy to reproduce classically, and resilient to noise. The logical question was therefore to investigate whether these assumptions really held, and my research aimed to partially tackle this question.

More precisely, the goal of my PhD was to assess the effect of noise on variational quantum algorithms and in particular to evaluate the claim that there could be intrinsic resilience. With the natural drift of research, the effective scope of my papers ended up encompassing mainly the effect of noise on the cost function landscape, particularly on

the ability to optimise or classically simulate it. This also required the development of new theoretical techniques, for instance the application of Fourier analysis to the cost function.

Overall the research produced intriguing results that provided a partial answer to the question of noise resilience. As presented in this Thesis, noise introduces significant barriers, which lead to poor quality solutions (see Chapters 3-4), prevent the algorithms from running (Chapter 4), or make them classically simulatable (Chapter 6). For instance, the presence of even a small amount of symmetric depolarising noise leads to a rapid decay of coherence, which manifests itself as a flattening of the cost function that hinders optimisation. Furthermore, even in the absence of a depolarising component the noise is likely to introduce significant distortions of the landscape. The loss of coherence also improves the ability to simulate expectation values, approximately and under some conditions. Remarkably there appears to be a step-like behaviour in the sense that infinitesimal amounts of noise yield polynomial-time algorithms for the cost function, albeit with potentially unwieldy exponents. The overarching message is that noise has a largely negative effect on variational algorithms.

These conclusions certainly come with many caveats, chief among them the fact that noise models are naturally crude approximations of the errors experienced by true quantum devices, which are vastly more complicated, for instance they may be time-varying and non-Markovian, and to some extent still poorly characterised. However there is evidence showing that these simplistic models are still able to capture real behaviours [AAB+19, BSP24]. There are also several tangential directions of research that I could have pursued, but did not for lack of time or vision. For instance, I regret not looking more deeply at how error mitigation techniques may affect these results. Finally, any conclusion that can be drawn from the work presented here is necessarily limited by the fact that I focused almost exclusively on noise resilience treating variational algorithms as black boxed, and avoided the question of the actual usefulness of these algorithms for real-world problems.

7.2 Taking stock of other research

Fortunately, all shortcomings of my work have been compensated many times over by the community. Indeed a large body of research was produced painting a much more complete picture, filling in many of these gaps and providing a vastly broader perspective. Let us now examine some of the most significant results from the past 5 years that are relevant to the topic of research. Following the publication of Ref. [FFR+21a] discussed in Chapter 3, several works were released that largely confirmed the conclusions on the impact of noise of VQAs and in particular VQE [OMML22, SWA21, ZWC+21, LAW23, BDP21] but also QML algorithms [EAGB23, TCD+24]. Largely these worked observed the same suppression of the optimal cost function value and modification to the landscape; some also noted a strong dependence on architecture [OMML22,SWA21], and resilience for moderate noise [OMML22, ZWC+21]. As had been conjectured, resilience is present especially for coherent noise but decoherent or stochastic noise seems to be particularly disruptive, leading to a more complex optimisation landscape [LTM+20]. Of the decoherent noise models the most damaging seem to be symmetric depolarising [ZWC+21], consistently with what found in our own experiments. The phenomena of noise-induced transitions has been replicated in Ref. [SWSW24] for a different setting, showing analytically the existence of an even more abrupt collapse of quantum advantage at a certain noise threshold. Finally, Ref. [TCD+24] brings to light the ability of noise of breaking symmetries in the quantum circuit, reminiscent of the conclusions of [FCA+22] as discussed in Chapter 4. Even though in this case the symmetries pertain to the unitary and not the parameters, the result highlights that the presence of noise may destroy regularities that otherwise hold true in noiseless models.

Overall these numerical and analytical works suggests that noise has a large variety of effects that vary with the noise level, the noise model, and the system size and architecture, significantly complicating any theoretical analysis. However as the noise-induced barren plateau (NIBP) result in Chapter 4 shows, sometimes it is possible to make general statements about the effects of noise. Indeed significant strides

have been made towards a theoretical understanding applicable to a broad range of situations. The NIBP result, which initially assumed a simplified noise model, has been extended to other noise models [SL24, MAG+24, SWC24]. Remarkably, it has been found that non-unital noise such as amplitude damping does not always create barren plateaus [SL24, MAG+24]. Unfortunately non-unital noise comes with other caveats, as we have seen in Chapters 3 and 4. Even more dramatically, Ref. [MAG+24] found that non-unital noise may make the algorithm classically simulatable, since it means that only the last layers of the algorithm up to logarithmic depth contribute to the output. In general, decoherent noise seems to pose limitations to the depth and total number of gates that can be present in the algorithm before the output becomes overly corrupted. Universally the limit seems to be logarithmic depth in the number of qubits [YDCM23], as found in Chapter 4, or $O(1/p)$ noisy gates, where p is the probability of error [GGTC22], which in Chapter 6 was shown to be the threshold for classical simulatability via LOWESA. However, Ref. [SFGP21b] shows that even for $O(1)$ depth the VQE may lose quantum advantage by a different methodology. Specifically noise brings the state to a fixed point which is close to a high-temperature Gibbs state, and depending on the noise level this may be sampled efficiently classically, leading to solutions of approximately the same energy. Finally, even though some works found improvements in the VQA convergence and solution quality with error mitigation [JS22], there exists strong theoretical evidence that NIBPs cannot be eliminated with error mitigation without incurring superpolynomial overheads [TEMG22, DNS+22, TTG23, TSY23, WCA+24, QSFK+24].

Altogether these conclusions suggest that noise is a terminal issue for NISQ as a whole. However, the picture is not as clear-cut. Ref. [CCHL23] took a complexity-theoretic perspective to NISQ and concluded that there exist separations both between NISQ and BPP (classical computation), and between NISQ and BQP (noiseless quantum computation), such that NISQ sits roughly in between the two extremes. Interestingly this is in the regime of constant local depolarising noise that we have observed to be critically bad for VQAs, suggesting that NISQ beyond variational algorithms may still hold some advantage. Indeed there is an extensive body of work

suggesting that random circuit sampling (RCS) retains advantage under local noise sources [BFLL22, ZVBL23, DHJB24]. Furthermore, other non-variational algorithms have been proposed that, while being NISQ-friendly, show some inherent resilience to noise [AUC+22, VJW+23], although it is to be seen if this holds at large problem sizes. Therefore it is still not excluded that quantum computers may be useful before error correction. Still, specifically for VQAs the outlook remains quite bleak.

7.3 Looking ahead

Thus I am led to a crucial question: in light of what is now known about noise and other issues like barren plateaus, is there any possibility for variational algorithms in the near to medium term? Are the hopes that distinguished the early NISQ era doomed to dissipate into nothingness?

Certainly, there is a sense in which the initial promises have not been vindicated. Indeed at this stage we can confidently say that noise resilience in the original sense has been theoretically and experimentally disproven, and even error mitigation techniques cannot be useful at large scale. This is not to mention other technical issues beyond noise that prevent optimisation, and which I have not delved into. Among these I should only mention the presence of barren plateaus, which by now seems to be inexorably tied to hardness of simulatability, in that it appears that the latter almost invariably leads to the former [CLGM+23]. Finally, there remains the overarching problem that even without noise no VQA has been conclusively demonstrated to offer quantum advantage over classical methods (except perhaps QAOA as seen in Ref. [SLC+24], but even then it was coupled with a non-variational, fault-tolerant subroutine). I still see a possibility that NISQ computers themselves may be useful for running RCS and non-variational algorithms, especially if the circuits are shallow and the noise is low enough, and perhaps with error mitigation. After all, many no-go results around error accumulation and mitigation are asymptotic, so there may be leeway for some useful small-scale computation on current devices [ZKH+25]. Nonetheless, due to the combined forces of noise sensitivity, issues with optimisation, and lack of demonstrated advantage, personally I

am not confident that VQAs will become useful anytime soon.

This all happens in a time where error correction is becoming increasingly realistic, seemingly by the day [AABA+24, EMS+24, WSG+24, DSRABR+24, PBP+24]. It is therefore natural to wonder if VQAs will have any utility at all, as they were always meant to be a stopgap measure before the advent of fault tolerance. However this last assumption may prove to be ultimately faulty. Indeed there are reasons to believe that some of the techniques developed for NISQ devices, including VQAs, may ultimately turn out to be useful both in the medium and possibly in the long term too.

In the medium term, despite the recent wave of hope about error correction, it is unlikely that computers will reach the requirements in connectivity, speed and error rate that are necessary to implement large codes efficiently. Therefore we may see a time where computers are only partially fault tolerant [TAF+24], which has been termed intermediate scale quantum (ISQ) [Arr23]. In this time, algorithms will have to be highly optimised since the presence of errors will likely preclude the implementation of large-scale algorithms like Shor's or Grover's [YDCM23]. The flexibility and frugality of variational algorithms may therefore make them relevant again. After all, compact VQAs may address the same problems as much more resource-intensive algorithms (for instance compare VQE with phase estimation) by sacrificing rigorous guarantees for a compact circuit. It is possible therefore that, if errors can be driven low enough to avoid NIBPs, and if the problems around optimisation are addressed by smart initialisation strategies [EMW21, PDTH24], that VQAs may become useful heuristics. In particular QAOA and VQE with problem-aware ansätze like HVA may be useful at intermediate scales, and QAOA combined with quantum minimum finding has been shown to possess scaling advantage [SLC+24]. Regarding error mitigation, these techniques may be employed to some degree to further drive down the noise, in combination with and in assistance to partial error correction [PSB+21, SEFT22].

In the longer term, assuming the availability of large computers capable of full error correction, it is hard to forecast the relevance of variational algorithms in the broader quantum algorithms landscape. This will be highly dependent on the ratio between the discoveries of fault-tolerant algorithms, on the one hand, and of improved

variational algorithms on the other. Even if any new fault-tolerant algorithms are just heuristics and even if the optimisation issue of VQAs are resolved, the former would hold an advantage over the latter due to the fact that arbitrary single-qubit rotations are inherently difficult to implement on logical qubits [TAF+24]. However, it is possible that such future VQAs may not require high precision in the rotations. This would be true particularly if the VQAs employed are small, with few parameters, and if their output is only valuable as a coarse approximation. Indeed one may envision VQAs becoming state-preparation subroutines for larger algorithms, for instance a VQA may be used to prepare a so-called guiding state [GLG22] from which the ground state of the problem Hamiltonian is extracted via fault-tolerant routines, as was done in Ref. [SLC+24]. Finally one can still not exclude the possibility that some quantum advantage is found for quantum machine learning (QML) using variational algorithms, for instance if a mapping is found between a learning algorithm and thermal evolution under a physical Hamiltonian then the results of Ref. [CHPZ24] showing a separation in optimisation capabilities may become relevant. However this prospect, like the ones listed before, remains highly speculative and currently there are no indications that QML would even in principle offer advantages in the fault-tolerant regime.

7.4 One last personal reflection

In conclusion, the duration of my PhD more or less coincided with a full arc, rise followed by fall, of the very same field I was studying: variational quantum algorithms. It would be natural to assume that any pessimism created in the later stages would have also impacted my view of quantum computing as a whole. In fact that is not the case, as I am possibly more hopeful now than I was five years ago. Progress on the hardware side is accelerating and we are beginning to see the first convincing experimental evidence that error correction is possible with the current architectures. The theoretical side is moving accordingly towards fault-tolerant algorithms that actually could bring value to society. Some of the discoveries in the field of VQAs and NISQ as a whole may well contribute to these future algorithms. I yet do not believe that it was a mistake to put faith in a new and untested paradigm of computation. Even though to some

it may have been clear from the beginning that there would probably not be any advantage in variational algorithms, the risk that nothing may be found should never be a barrier to research. Personally, even in the pessimistic scenario that nothing may come of it, studying these algorithms has been an exciting endeavour and an occasion for growth as a researcher, the ultimate goal of any PhD. Indeed, thanks to this experience I now possess a precious piece of first-hand knowledge that by far supersedes any disappointment and that has provided me with motivation to go forward into new challenges. Even in science, when novelty is introduced, at first emotions may intrude and opinions will collide. Given enough time however the scientific community will be able to come to rational conclusions, removing the veil of ignorance, and finally revealing the truth behind the noise.

Appendix A

Numerically evaluating noise resilience

A.1 Linear model of noise propagation

In the interleaved noise model introduced in Sec. 3.1.1 we consider a quantum circuit that we split in d layers. In the absence of noise, the unitary operation applied by the circuit can be written as $U = U_d U_{d-1} ... U_2 U_1$, where U_i is the i^{th} layer unitary operator. In the noisy case we move to the density matrix picture for mixed states, and use maps between density operators (or superoperators) to represent operations on these mixed states, which are defined by $\mathcal{U}(\rho) = U\rho U^\dagger$. This represents a noiseless quantum channel. As outlined in the main text, this corresponding noisy quantum channel is given by :

$$\tilde{\mathcal{U}} = \Lambda_d \mathcal{U}_d ... \Lambda_1 \mathcal{U}_1. \tag{A.1}$$

A.1.1 Example: global depolarising noise

We first consider the case where all the noise channels are given by global symmetric depolarising noise on N qubits, since in this case we can write down an exact model for the noise propagation. The global symmetric depolarising channel is defined as acting

on any state ρ_{in} as follows:

$$\Lambda\left(\rho_{\text{in}}\right) = (1-\gamma)\rho_{\text{in}} + 2^{-N}\gamma\,\mathbb{1}, \tag{A.2}$$

where $\mathbb{1}$ is the identity operator on N qubits (in matrix form it is of dimension $2^N \times 2^N$). The effect of the channel is therefore to replace any state with the completely mixed state with probability γ.

If the circuit under consideration consists of d such channels interleaved with unitary operations, the probability of the state not being affected by any channel is $(1-\gamma)^d$. Any other outcome will lead to the final state being maximally mixed, as the maximally mixed state is invariant under the action of any quantum channel. The final state of the circuit can therefore be written as:

$$\rho = (1-\gamma)^d \mathcal{U}\left(\rho_{\text{in}}\right) + 2^{-N}(1-(1-\gamma)^d)\,\mathbb{1}, \tag{A.3}$$

where $\mathcal{U}$ is the noiseless circuit unitary map.

We can quantify the difference between the state produced by the noisy circuit, ρ, and the state produced without noise, ρ_{id}, using the infidelity, defined in Eq. (3.10). From now on, the subscript *id* will universally refer to a quantity evaluated on an ideal, noiseless circuit. For the global symmetric depolarising noise operator in Eq. (A.3) we obtain:

$$\begin{aligned} R(\rho_{\text{id}}, \rho) &= 1 - \text{Tr}[\rho_{\text{id}}((1-\gamma)^d \rho_{\text{id}} + 2^{-N}(1-(1-\gamma)^d)\mathbb{1})] \\ &= 1 - (1-\gamma)^d \text{Tr}[\rho_{\text{id}}^2] - 2^{-N}(1-(1-\gamma)^d)\text{Tr}[\rho_{\text{id}}] \\ &= \left(1-(1-\gamma)^d\right)\left(1-2^{-N}\right), \end{aligned} \tag{A.4}$$

where we use the fact that ρ_{id} is a pure state, and so $\text{Tr}[\rho_{\text{id}}^2] = 1$. Expanding at first order in γ we get:

$$R(\rho_{\text{id}}, \rho) = (1-2^{-N})\,d\,\gamma + O(\gamma^2). \tag{A.5}$$

This means that, in the low noise regime, the infidelity increases approximately linearly

in both γ and the number of layers d, with a prefactor that is effectively 1 for large systems ($N \to \infty$).

Note that a global symmetric depolarising channel is not equivalent a product of single-qubit depolarising channels, and therefore the relation derived in Eq. (A.5) does not apply for the product channel noise model that we employ in this work. However, the simple result of Eq. (A.5) provides the motivation to seek an analogous equation for more general noise models.

A.1.2 Infidelity propagation for general noise channels

We start by expanding the action of a general N-qubit noise channel on a state ρ to first order as:

$$\Lambda(\rho) = (1-\gamma)\rho + \gamma\lambda(\rho) + O(\gamma^2). \tag{A.6}$$

Here λ is a map that represents the first-order action of the channel, which for global symmetric depolarising channels is simply $\lambda(\rho) = 2^{-N}\mathbb{1}\ \forall\rho$. Note that the matrix $\lambda(\rho)$ is, in general, not a valid quantum state. We term λ the *linear action* of the noise channel. In the expansion we take into account the fact that for a general noise channel there could be higher order effects in the local noise parameter γ. Since the gate noise level in current quantum computers has $\gamma \ll 1$ [DH21], we expect the linear action to always dominate, and hence that one can neglect higher order terms.

We introduce the partially noisy circuit $\tilde{\mathcal{U}}_{:i:} := \mathcal{U}_d \cdots \mathcal{U}_{i+1}\lambda_i\mathcal{U}_i \cdots \mathcal{U}_1$, obtained by inserting a single noise channel in the noiseless circuit and keeping only the terms at order γ. Applying the noisy circuit onto an initial state ρ_0 and expanding in $O(\gamma)$, it reduces to:

$$\rho \approx [(1-\gamma d)\mathcal{U} + \gamma\sum_{i=1}^{d}\tilde{\mathcal{U}}_{:i:}](\rho_0) \tag{A.7}$$

$$= \rho - \gamma\sum_{i=1}^{d}(\mathcal{U} - \tilde{\mathcal{U}}_{:i:})(\rho_0). \tag{A.8}$$

The resulting difference to the noiseless result is an additive contribution proportional to γ.

Appendix A. Numerically evaluating noise resilience

Now we combine the first-order expansion with the layered noise model, and apply it to the variational problem of maximising the fidelity to a pure target state ρ_{T}. In this case, the fidelity between the noisy state output by the circuit after optimisation ρ_{opt} and the target state simplifies to $F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt}}) = \mathrm{Tr}(\rho_{\mathrm{T}}\rho_{\mathrm{opt}})$. In general, the circuit may not reproduce the ideal target state ρ_{T} exactly. We therefore have:

$$F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt}}) = \mathrm{Tr}(\rho_{\mathrm{T}}\rho_{\mathrm{opt}}) \tag{A.9}$$

$$\approx \mathrm{Tr}[\rho_{\mathrm{T}}\rho_{\mathrm{opt,id}} - \rho_{\mathrm{T}}\gamma \sum_{i=1}^{d} (\mathcal{U} - \tilde{\mathcal{U}}_{:i:})(\rho_0)] \tag{A.10}$$

$$= F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt,id}}) - \gamma \sum_{i=1}^{d} (F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt,id}}) - F(\rho_{\mathrm{T}}, \tilde{\mathcal{U}}_{:i:}(\rho_0))). \tag{A.11}$$

We can now define the *relative infidelity* as $R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt,id}}, \rho_{\mathrm{opt}}) \coloneqq F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt,id}}) - F(\rho_{\mathrm{T}}, \rho_{\mathrm{opt}})$. We now define the state at the ith layer $\rho_{\mathrm{opt},i}$, as well as the target state "back-propagated" to ith layer as $\rho_{\mathrm{T},i} \coloneqq \mathcal{U}_{:i}^{-1}\rho_{\mathrm{T}}$, where we use the fact that $\mathcal{U}_{:i}$ is a unitary channel. With this definition and the invariance of fidelity under unitary transformations, we can write:

$$R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt,id}}, \rho_{\mathrm{opt}}) \approx \gamma \sum_{i=1}^{d} R_{\mathrm{rel}}(\rho_{\mathrm{T},i}; \rho_{\mathrm{opt},i}, \lambda(\rho_{\mathrm{opt},i})). \tag{A.12}$$

This result shows that the relative infidelity is approximately linear in the number of noise channels d and the local noise parameter γ. Each noise channel contributes a factor that depends on the target state ρ_{T}, the state through the channel $\rho_{\mathrm{opt},i}$ and its image under the linear action $\lambda(\rho_{\mathrm{opt},i})$. Note that this is consistent with the result for the symmetric depolarising channel in the previous section, by taking $R_{\mathrm{rel}}(\rho_i; \rho_i, \lambda(\rho_i)) = R(\rho_i, \lambda(\rho_i)) = 1 - 2^{-n}\ \forall \rho_i$. Importantly, our total infidelity is state dependent. Hence, the path through the space of states that the circuit takes is very important for the noise resilience of the algorithm. If two different paths lead to the same state in the noiseless case, they may still have very different infidelities in the noisy case, as dictated by these intermediate states.

To simplify the result and compare it to literature, one can estimate the average

infidelity of a noise channel over all states, R_i, and with it obtain a very rough estimate for the total infidelity. First let us assume that the circuit can indeed reach the target state, such that we have $R_{\text{rel}}(\rho_{,i}; \rho_i, \lambda(\rho_i)) = R(\rho_i, \lambda(\rho_i)) = 1 - \text{Tr}[\rho_i\, \lambda(\rho_i)]$ and $R_{\text{rel}}(\rho_{\text{id}}; \rho_{\text{id}}, \rho) = R(\rho_{\text{id}}, \rho) = 1 - \text{Tr}[\rho_{\text{id}}\, \rho]$. Now use the definition of the linear action in Eq. (A.6) to write, to first order in γ, $\gamma R_{\text{rel}}(\rho_i; \rho_i, \lambda(\rho_i)) \approx \gamma - \text{Tr}[\rho_i\, \Lambda(\rho_i)] + (1 - \gamma)\text{Tr}[\rho_i^2] = 1 - \text{Tr}[\rho_i\, \Lambda(\rho_i)]$. In accordance with literature, we define the gate infidelity as $R_i = 1 - \text{Tr}[\rho_i\, \Lambda(\rho_i)]$. In this way with Eq. (A.12) one obtains for the total infidelity of the circuit:

$$R \approx \sum_{i=1}^{d} R_i. \tag{A.13}$$

From this equation we obtain the total state-averaged circuit fidelity as function of the averaged fidelities of each noise channel, $F_i = 1 - R_i$, as:

$$F = 1 - \sum_{i=1}^{d} R_i = \prod_{i=1}^{d} (1 - R_i) + O\left(R_i^2\right). \tag{A.14}$$

Since the infidelities of a noise channels are small, we can neglect higher order terms, and obtain:

$$F \approx \prod_{i=1}^{d} F_i, \tag{A.15}$$

where we have used $F_i = 1 - R_i$. This last expression is widely used in literature. For example, in Ref. [AAB+19] the fidelity is assumed to behave in this way, and subsequent experimental results are shown to be consistent with this assumption. In Ref. [ZWC+20] it is shown that the approximation error of a matrix product state representation of a quantum circuit is approximately multiplicative in the fidelity of a single layer. In Ref. [MWHH20], a similar reasoning shows that this multiplicative decay of fidelity applies for interleaved depolarising channels in QAOA circuits. The decay of the cost function with the number of noisy channel that follows from these equations was rigorously proven in Ref. [WFC+21] for local symmetric depolarising channels. Ref. [CDWE19] shows that infidelity grows at most quadratically with the

number of layers of a circuit, and for decoherent channels it is expected to grow linearly at first order, which is equivalent to what we have shown here. However, the measure used there is the gate infidelity over Haar-distributed states, which might not reflect the outcome of specific experiments [SWS15].

A.1.3 Expected fidelity over an ensemble

Up to now we considered the infidelity for one target state. We extend this to the case where one is provided with an ensemble of n_{T} pure target states sampled from a distribution. As figure of merit we use the average optimal infidelity over the ensemble, which we defined in Eq. (3.12) as:

$$\bar{R} = \langle R(\rho_{\mathrm{T}}, \rho_{\mathrm{opt}}) \rangle_{\mathrm{T}}, \tag{A.16}$$

where ρ_{opt} is optimised with noise for each target state. An $\bar{R}$ of zero would imply that the quantum circuit can represent any N-qubit state exactly. In practice $\bar{R}$ is usually larger than zero even in the noiseless case due the inherent limitations of a given circuit, and the addition of noise further increases $\bar{R}$. We define the average optimal relative infidelity as:

$$\bar{R}_{\mathrm{rel}} \coloneqq \bar{R} - \bar{R}_{\mathrm{id}} = \langle R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt,id}}, \rho_{\mathrm{opt}}) \rangle_{\mathrm{T}}, \tag{A.17}$$

where as before the subscript id indicates the infidelity and density matrix evaluated without noise: $\bar{R}_{\mathrm{id}} = \langle R(\rho_{\mathrm{T}}, \rho_{\mathrm{opt,id}}) \rangle_{\mathrm{T}}$.

Next we derive an approximate relation for $\bar{R}_{\mathrm{rel}}$ allowing us to qualitatively understand its behaviour and relate it to existing literature. Using Eq. (A.12) and the fact that in our noise model all the noise channels are identical, we obtain:

$$\bar{R}_{\mathrm{rel}} \approx \gamma \left\langle \sum_{i=1}^{d} R_{\mathrm{rel}}(\rho_{\mathrm{T},i}; \rho_{\mathrm{opt},i}, \lambda(\rho_{\mathrm{opt},i})) \right\rangle_{\mathrm{T}}. \tag{A.18}$$

We now assume that at every level the noiseless $\rho_{\mathrm{T},i}$ and $\rho_{\mathrm{opt},i}$ are distributed identically

to the final states ρ_{T} and ρ_{opt}. Thus we can write:

$$\begin{aligned}\bar{R}_{\mathrm{rel}} &\approx \langle R_{\mathrm{rel}}(\rho_{\mathrm{T}};\rho_{\mathrm{opt}},\lambda(\rho_{\mathrm{opt}}))\rangle_{\mathrm{T}}\,\gamma\,d \\ &\approx \alpha\,\gamma\,d,\end{aligned} \tag{A.19}$$

where we defined the constant of proportionality:

$$\alpha \coloneqq \langle R_{\mathrm{rel}}\left(\rho_{\mathrm{T}};\rho_{\mathrm{opt}},\lambda(\rho_{\mathrm{opt}})\right)\rangle_{\mathrm{T}}. \tag{A.20}$$

Note that since $\lambda(\rho_{\mathrm{opt}})$ is generally not a valid quantum state by itself, since usually $\mathrm{Tr}[\lambda(\rho)] \neq 1$, α may also be greater than 1.

With the additional assumption about the distributions of the intermediate states, the variance of the relative infidelity can be estimated. Defining the variance of the optimal relative infidelity over the target states:

$$\Delta^2_{\mathrm{rel}} \coloneqq \langle R^2_{\mathrm{rel}}(\rho_{\mathrm{T}};\rho_{\mathrm{opt,id}},\rho_{\mathrm{opt}})\rangle_{\mathrm{T}} - \bar{R}^2_{\mathrm{rel}} = \operatorname*{Var}_{\mathrm{T}}\left(R_{\mathrm{rel}}(\rho_{\mathrm{T}};\rho_{\mathrm{opt,id}},\rho_{\mathrm{opt}})\right). \tag{A.21}$$

Using Eq. (A.12) and a standard property of the variance, within these approximations we obtain:

$$\begin{aligned}\Delta^2_{\mathrm{rel}} &\approx \operatorname*{Var}_{\mathrm{T}}\left(\gamma\sum_{i=1}^{d} R_{\mathrm{rel}}(\rho_{\mathrm{T},i};\rho_{\mathrm{opt},i},\lambda(\rho_{\mathrm{opt},i}))\right) \\ &= \gamma^2 \operatorname*{Var}_{\mathrm{T}}\left(\sum_{i=1}^{d} R_{\mathrm{rel}}(\rho_{\mathrm{T},i};\rho_{\mathrm{opt},i},\lambda(\rho_{\mathrm{opt},i}))\right).\end{aligned} \tag{A.22}$$

Now we must introduce some information about the correlation between states at every noise channel. Since states going through successive channels have a certain degree of similarity, in the sense that they are related by a short-depth sequence of unitaries, it should be expected that they retain a high level of correlation. Therefore, we assume that the states at every noise channel $\{\rho_{\mathrm{opt},1},\rho_{\mathrm{opt},2},\cdots,\rho_{\mathrm{opt,d}}\}$ are perfectly correlated. It follows that:

$$\Delta^2_{\mathrm{rel}} \approx \beta\,\gamma^2 d^2, \tag{A.23}$$

with:

$$\beta \coloneqq \operatorname*{Var}_{\mathrm{T}}(R_{\mathrm{rel}}(\rho_{\mathrm{T}}; \rho_{\mathrm{opt}}, \lambda(\rho_{\mathrm{opt}}))). \tag{A.24}$$

Note that, if instead we assumed that the states are uncorrelated with each other, we would get $\Delta^2_{\mathrm{rel}} \approx \beta\, \gamma^2 d$.

Eqs. (A.19) and (A.23) define the stochastic model for noise propagation.

A.1.4 Estimation of α and β

In this section we show that α and β can be estimated with knowledge of only the distribution of target states and the noise channel properties. First of all, we note that both constants depend on the output state ρ_{opt}, which depends on the capability of the circuit to approximate the target state. Since we wish to remove the dependence on the circuit entirely, we use as an approximation $\lambda(\rho_{\mathrm{opt}}) \approx \lambda(\rho_{\mathrm{T}})$. This is justified by the fact that, given a sufficiently expressive circuit, ρ_{opt} will not be much different from ρ_{T}. Therefore, using the definitions in Eqs. (A.20) and (A.24) the parameters can be estimated from the target states as:

$$\alpha \approx \langle R_{\mathrm{rel}}\left(\rho_{\mathrm{T}}; \rho_{\mathrm{T}}, \lambda(\rho_{\mathrm{T}})\right)\rangle_{\mathrm{T}} = \langle R\left(\rho_{\mathrm{T}}, \lambda(\rho_{\mathrm{T}})\right)\rangle_{\mathrm{T}}, \tag{A.25}$$

$$\beta \approx \operatorname*{Var}_{\mathrm{T}}(R(\rho_{\mathrm{T}}, \lambda(\rho_{\mathrm{T}}))), \tag{A.26}$$

where we can switch the relative infidelity for regular infidelity as we have removed all dependence on the circuit.

Using Eq. (A.6) and taking the derivative of $\Lambda(\rho)$ about $\gamma = 0$ we obtain:

$$\frac{d\Lambda(\rho)}{d\gamma}\Big|_{\gamma=0} = \lambda(\rho) - \rho \quad \rightarrow \quad \lambda(\rho) = \frac{d\Lambda(\rho)}{d\gamma}\Big|_{\gamma=0} + \rho. \tag{A.27}$$

Therefore we can write:

$$\begin{aligned} R(\rho_{\mathrm{T}}, \lambda(\rho_{\mathrm{T}})) &= 1 - \mathrm{Tr}\left(\rho_{\mathrm{T}}\lambda(\rho_{\mathrm{T}})\right) \\ &= 1 - \mathrm{Tr}\left(\rho_{\mathrm{T}}\left(\frac{d\Lambda(\rho_{\mathrm{T}})}{d\gamma}\Big|_{\gamma=0} + \rho_{\mathrm{T}})\right)\right) \\ &= 1 - \mathrm{Tr}\left(\rho_{\mathrm{T}}\frac{d\Lambda(\rho_{\mathrm{T}})}{d\gamma}\Big|_{\gamma=0}\right) - \mathrm{Tr}\left(\rho_{\mathrm{T}}^2\right) \\ &= -\mathrm{Tr}\left(\rho_{\mathrm{T}}\frac{d\Lambda(\rho_{\mathrm{T}})}{d\gamma}\Big|_{\gamma=0}\right). \end{aligned} \tag{A.28}$$

Substituting back into Eqs. (A.25) and (A.26) we obtain:

$$\alpha \approx -\left\langle \frac{d}{d\gamma}\mathrm{Tr}(\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}}))\Big|_{\gamma=0}\right\rangle_{\mathrm{T}}, \tag{A.29}$$

$$\beta \approx \underset{\mathrm{T}}{\mathrm{Var}}\left(\frac{d}{d\gamma}\mathrm{Tr}(\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}}))\Big|_{\gamma=0}\right). \tag{A.30}$$

If states can be efficiently sampled from the distribution, the constants α and β can be estimated given an exact density matrix simulator. Given a sampled target state, the derivative can be evaluated in practice using finite differences as:

$$\frac{d}{d\gamma}\mathrm{Tr}(\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}}))\Big|_{\gamma=0} = \lim_{\epsilon\to 0}\frac{\mathrm{Tr}(\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}}))|_{\gamma=\epsilon} - 1}{\epsilon}. \tag{A.31}$$

The results obtained with Eqs. (A.29, (A.31)) for the distribution of 4-qubits real states used in our simulations are shown in Table A.1.

Noise channel	α	β
Phase	0.888	0.00585
Amplitude	1.88	0.119
depolarising	2.78	0.0132

Table A.1: Estimated α and β for local noise channels on four qubits, averaged over the real Haar distribution states. 10000 randomly chosen states were used for the numerical estimation.

A.1.5 Scaling of α with the number of qubits

For an approximate scaling estimate of α with the number of qubits N, we consider simplified case of a target state being a tensor product of N independent identically distributed single-qubit states. With this assumption we can write:

$$\begin{aligned}
\alpha &\approx -\frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}[\rho_{\mathrm{T}}\Lambda(\rho_{\mathrm{T}})] \\
&= -\frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}\left[\bigotimes_i \rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})\right] \\
&= -\frac{d}{d\gamma}\Big|_{\gamma=0} \prod_i \mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})] \\
&= -\sum_i \frac{d}{d\gamma}\Big|_{\gamma=0} \left(\mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})]\right) \prod_{j\neq i} \mathrm{Tr}[\rho_{\mathrm{T}}^{(j)}\Lambda^{(j)}|_{\gamma=0}(\rho_{\mathrm{T}}^{(j)})] \\
&= -\sum_i \frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})] \\
&= \sum_{i=1}^{N} \alpha^{(\mathrm{i})},
\end{aligned} \tag{A.32}$$

where we defined the single-qubit $\alpha^{(\mathrm{i})} := -\frac{d}{d\gamma}\Big|_{\gamma=0} \mathrm{Tr}[\rho_{\mathrm{T}}^{(i)}\Lambda^{(i)}(\rho_{\mathrm{T}}^{(i)})]$. Since we assumed that the single-qubit product states are identically distributed, these terms are equal for all i and we obtain a scaling of $\alpha \sim O(N)$. Thus, we expect the effect of noise to grow linearly with the number of qubits.

A.2 Additional results

A.2.1 Noise-induced state transitions for further random states

In Fig. 3.9 we show how fidelity and concurrence as function of γ abruptly change slope at at threshold γ value for one randomly chosen target state. To show that this behaviour is general, in Fig. A.1 we present analogous results for four more randomly chosen target states. We also consider different circuit depths. The transitions are overall very different to one another, at times discontinuous in their fidelity and at other times smooth and barely detectable. In these latter cases, the concurrence acts as a

clearer indicator for such transitions. The varied phenotype of transitions suggests that the phenomenon is complex and depends heavily on the circuit and the chosen target state. Indeed, in some situations (see the last row of Fig. A.1) no sharp transitions are observed, and both measures of state quality vary smoothly. The results illustrate that generally one always observes critical noise level thresholds, and that these can be either abrupt and discontinuous or else smoothed out in a continuous way.

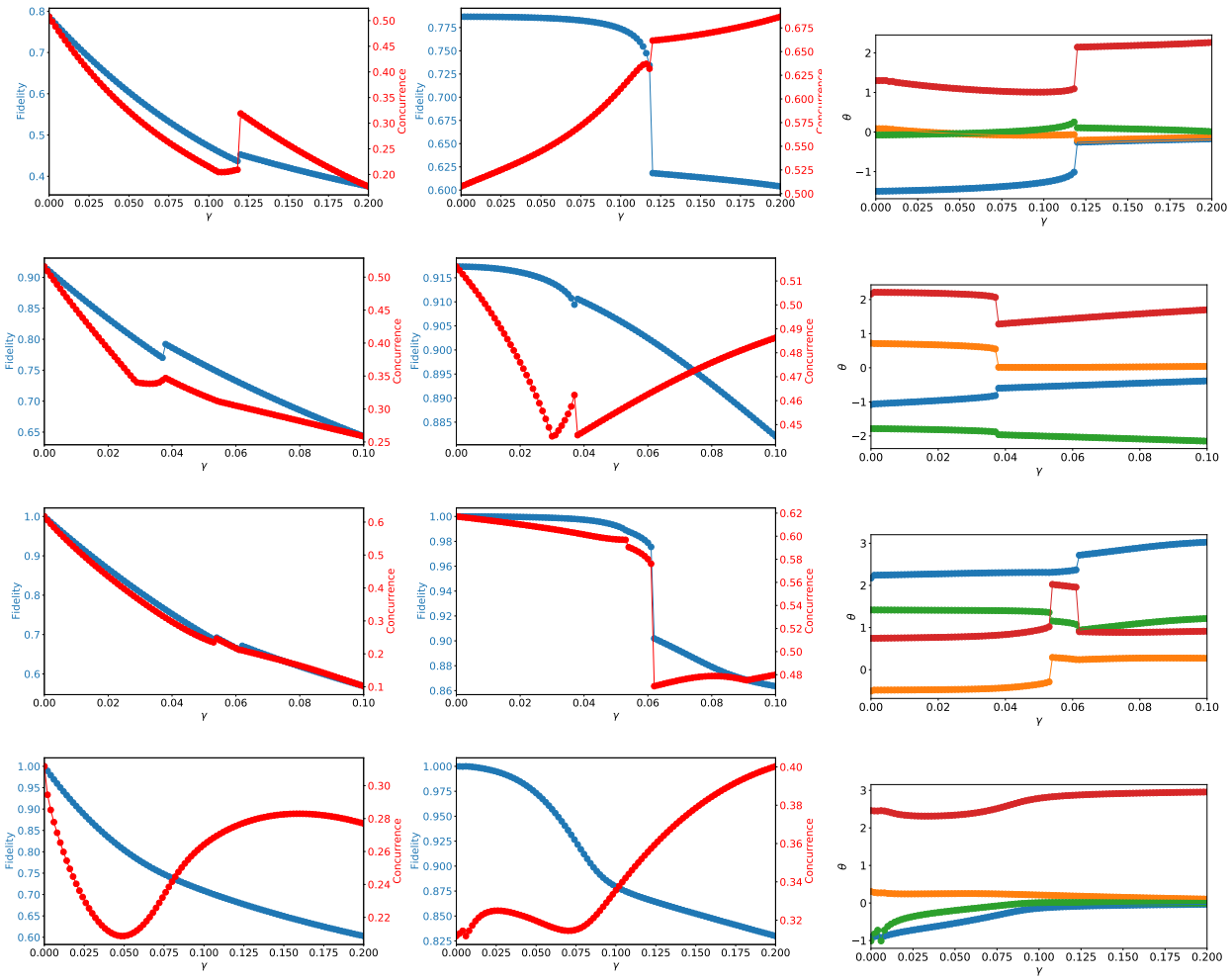


Figure A.1: Evidence of noise-induced state transitions for different random states. Various measures are plotted at convergence for a range of values of the phase noise parameter γ. The plotted data are described in more detail in Fig. 3.9. Left panels: fidelity (blue/dark line) and concurrence (red/gray line) at convergence, evaluated with noise; centre panels: fidelity and concurrence evaluated without noise, but for the converged optimised angles at every noise level; right panels: value of four of the circuit rotation parameters at convergence. The first two rows have a circuit of depth $L = 3$, the last two have $L = 4$.

Appendix B

Fourier spectrum of noisy Variational Quantum Algorithms

B.1 Resource requirements and nonuniform sampling

As mentioned in the text, sampling overhead is the largest obstacle to a pratical implementation of spectral analysis methods. One-dimensional DFT up to a frequency of N would require sampling $d = 2N + 1$ points, itself a linear scaling. However, the scaling in the dimension is unfavourable as it is exponential, practically limiting the applicability of the method to small sets of parameters. In a practical setting the runtime is also affected by the requirement of high shot counts to reduce shot noise. In itself, however, the method could still be useful even in these conditions. A variational circuit with several parameters could be broken down into sets of parameters which then could be mitigated for noise individually. Furthermore, one might expect that different parameterised gates would be affected differently by noise based on factors such as their order and the noise characteristics of the qubits they act on. For example the effect of the first gates in the circuit would be diluted by successive noise channels. Therefore, one might choose to focus on the most noisy gates in order to maximise the benefit of mitigation.

In addition, we remark that many variational algorithms of interest do not feature many individual parameters, but rather a few parameters that govern larger sets of

correlated gates. In that case, the sampling points would be much reduced, even when accounting for the larger set of frequencies that need to be sampled. For example, while the spectrum of M uncorrelated single-qubit rotations would have size 3^M, correlating the angles into a single parameter would require sampling a number of points linear in M.

Finally, the use of advanced signal extraction techniques might enable these methods to work with much reduced sets of data. There exists an array of techniques that allow for the retrieval of the most significant Fourier coefficients from a signal, using a subexponential amount of samples [CRT06b, TG07, Don06, CRT06a]. In the case of sparse signals, techniques such as the sparse fast Fourier transform (sFFT) actually allow for complete reconstruction with sampling much below the Nyquist rate [HIKP12]. Therefore, if the quantum signal consists of a few large Fourier coefficients or, even better, if it is indeed sparse, these techniques could allow to circumvent the exponential sampling requirements in the case of many parameters. Interestingly, they would also apply an implicit thresholding by selecting only the most significant coefficients, and therefore is expected to provide some denoising effects.

B.1.1 Non-integer and non-equidistant samples or frequency spectra

In some cases, it might be desirable to get an estimate of the signal at non-integer frequencies, or spectra that are not equidistant. This might be relevant in case, for example, the error channel is suspected to depend in a nontrivial way on a parameter, leading to frequencies that are non-integer multiples of the parameter's frequency. This problem is known as non-uniform DFT or NUDFT and various approaches have been developed that implement an efficient version (NUFFT), based on interpolation [Boy92, DR93, PS03] or low-rank approximations [RAT18].

The discrete Fourier transform can also be implemented on incomplete data or nonuniform sampled data, while returning frequency data on a uniform support, known as type II NUDFT [RAT18]. Fast algorithms exist also in this scenario, which similarly use approximations and therefore come at the cost of reduced accuracy [RAT18, PS03], however might be sufficient as a rough error estimation technique. There exist scenar-

ios where a large amount of nonuniform data points are generated as a byproduct, for example in the optimisation of a variational algorithm. Using nonuniform estimation, it might be possible to implement spectral noise evaluation at no additional sampling cost.

Finally, we note the connection between the spectral analysis of quantum algorithms and signal processing in NMR. The latter similarly features a costly sampling procedure that generates multidimensional data, which then needs to be processed into its spectral components. Techniques such as nonuniform sampling [MMS+11] and maximum entropy reconstruction [HMM+14] have been devised to minimise the sampling and therefore the running time, while reconstructing the image with accuracy. One can therefore envisage directly utilising similar techniques for quantum algorithm.

B.2 Effect of different channels on Fourier modes

Lemma 10 (Depolarising channel)**.** *The noisy expectation value is:*

$$\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})} = \sum_{\mathbf{k}\in\boldsymbol{\Lambda}} (1-p)^m c_{\mathbf{k}} e^{i\omega_{\mathbf{k}}\cdot\boldsymbol{\theta}} + \frac{(1-(1-p)^m)}{2^n}\mathrm{Tr}(O). \tag{B.1}$$

Therefore, the noisy Fourier coefficients $\tilde{c}_{\mathbf{k}}$ are homogeneously contracted.

Proof. $\tilde{\mathcal{U}} = \mathcal{E}\circ\mathcal{U}$ where $\mathcal{E}(\rho) = (1-p)^{m+1}\rho + (1-(1-p)^{m+1})\frac{\mathbb{1}}{2^n}$. Then, using Eq. (5.35) the result follows. □

Consider a single Pauli noise channel $\mathcal{N}_p := (1-p)\mathcal{I} + p\mathcal{P}$, which appears within a quantum process:

$$\tilde{\mathcal{U}} = \mathcal{V}_0 \circ \mathcal{U}_1(\theta_1) \circ \cdots \circ \mathcal{N}_p \circ \cdots \circ \mathcal{U}_m(\theta_m) \circ \mathcal{V}_m. \tag{B.2}$$

For this we have the following

Lemma 11 (Pauli channel)**.** *Under a Pauli channel, the frequency spectrum does not change. For the specific case of O a Pauli operator, the signal's coefficients are contracted: $|\tilde{c}_{\mathbf{k}}(O)| \leq |c_{\mathbf{k}}(O)|$, and so there are no extra coefficients. For a general*

O, some new modes that were previously zero due to cancellation of paths might find themselves with a small non-zero value under noise.

Proof. First we consider the case of $p = 1$, that is, a Pauli error P that occurs within the quantum process:

$$\tilde{\mathcal{U}} = \mathcal{V}_0 \circ \mathcal{U}_1(\theta_1) \circ \cdots \circ \mathcal{P} \circ \cdots \circ \mathcal{U}_m(\theta_m) \circ \mathcal{V}_m. \tag{B.3}$$

Assume that each $\mathcal{U}_i(\theta_i) = e^{-iP_i\theta_i/2}$ where P_i are Pauli operators themselves, and that the unparameterised unitaries are Clifford as well. Then since Pauli operators either commute or anti-commute then propagating a Pauli error P through the circuit will change a subset of $\theta_i \to -\theta_i$. In terms of the individual process paths $\mathcal{V}_0 \circ \mathcal{C}^1_{k_1} \circ ... \circ \mathcal{C}^m_{k_m} \circ \mathcal{V}_m$ labelled by vector $\mathbf{k} = (k_1, ..., k_m) \in \mathbb{Z}_3^{\times m}$, the effect of a Pauli error within the circuit will result in a "noisy" path where $\mathcal{P}' \circ \mathcal{V}_0 \circ \mathcal{C}^1_{k'_1} \circ ... \circ \mathcal{C}^m_{k'_m} \circ \mathcal{V}_m$ labelled by vector $\mathbf{k}'$ for which $k'_i = -k_i$ on the fixed (path-independent) subset of $i \in \{1, ..., m\}$. What happens is that since the lattice $\mathcal{S}$ contains both ω and $-\omega$ then this means that the spectrum of frequencies for noisy $\tilde{\mathcal{U}}$ will, just like those for $\mathcal{U}$, also be included in $\mathcal{S}$. However, some new modes that were previously zero due to cancellation of paths might find themselves with a small non-zero value under noise for specific target observables. For the corresponding Fourier coefficient $c_{\mathbf{k}}(O)$ for the expected value of operator O over some input state, its noisy version will be $\tilde{c}_{\mathbf{k}}(O) = c_{\mathbf{k}'}(¶'(O))$. Similarly, in the case of a Pauli channel $\mathcal{N}_p = (1-p)\mathcal{I} + p\mathcal{P}$ the noisy Fourier coefficients become $\tilde{c}_{\mathbf{k}}(O) = (1-p)c_{\mathbf{k}}(O) + pc_{\mathbf{k}'}(\mathcal{P}'(O))$.

Going back to the case where O is a Pauli operator, we have $\mathcal{P}'(O) = \pm O$. Now, $c_{\mathbf{k}'}$ is equal to $c_{\mathbf{k}}$ or its conjugate. To see this, consider any one of the components, say k_j. For the case of $k_j = \pm 1$, one has:

$$\begin{aligned} c_{\mathbf{k}}(O) &= \mathrm{Tr}[O\mathcal{V}_0 \circ \cdots \circ \mathcal{C}^j_{\pm 1} \circ \cdots \circ \mathcal{V}_m(\rho)] && \text{(B.4)} \\ &= \mathrm{Tr}[O\mathcal{V}_0 \circ \cdots \circ \mathcal{C}^j_{r} \circ \cdots \circ \mathcal{V}_m(\rho)] && \text{(B.5)} \\ &\quad \pm i\,\mathrm{Tr}[\mathcal{O}\mathcal{V}_0 \circ \cdots \circ \mathcal{C}^j_{i} \circ \cdots \circ \mathcal{V}_m(\rho)], \end{aligned}$$

where $\mathcal{C}_r^j = \frac{1}{4}\left(\mathcal{U}(0) - \mathcal{U}(\pi)\right)$ and $\mathcal{C}_i^j = \frac{1}{8}\left(\mathcal{U}(0) - 2\mathcal{U}(\pi/2) + \mathcal{U}(\pi)\right)$ are two real channels. Therefore, $k_i \to -k_i$ is equivalent to conjugating the coefficient, and by induction, it is so for any $\mathbf{k}'$ as long as the number of affected nontrivial indices is odd. It follows that $|c_{\mathbf{k}'}(\mathcal{P}'(O))| = |c_{\mathbf{k}}(O)|$. Therefore by the triangle inequality the noisy coefficients will obey $|\tilde{c}_{\mathbf{k}}(O)| \leq |c_{\mathbf{k}}(O)|$. □

Thus we see that a contraction of the Fourier coefficients is a feature of Pauli channels, and this result extends the previous one on global depolarising channels.

The case of general $\mathcal{V}$s, as well as noise channels that are linear combinations of different Pauli channels, and multiple Pauli channels, follow analogously from these considerations. Furthermore, state preparation and measurement (SPAM) errors can also be modelled by Pauli noise channels.

A special case of the above for which more can be said is the aligned Pauli error, which is when a channel is followed by a Pauli channel which is identical to its generator.

Lemma 12 (Aligned Pauli error)**.** *For any operator O, an aligned Pauli error leads to a contraction of some coefficients without extra modes. Contrary to the previous result, this holds in the presence of non-Clifford non-parameterised gates.*

Proof. Take the operator $U(\theta_j) = e^{-iP_j\theta_j/2} = \mathcal{C}_0^{(j)} + e^{i\theta_j}\mathcal{C}_1^{(j)} + e^{-i\theta_i}\mathcal{C}_{-1}^{(j)}$. If this channel is followed by a Pauli channel which is identical to its generator, the overall channel is $P_jU(\theta_j) = U(\theta_j + \pi) = \mathcal{C}_0^{(j)} - e^{i\theta_j}\mathcal{C}_1^{(j)} - e^{-i\theta_i}\mathcal{C}_{-1}^{(j)}$, and therefore for a Pauli noise channel we get:

$$\mathcal{P}_j\mathcal{U}_i(\theta_i) = \mathcal{C}_0^{(i)} + (1-2p)e^{i\theta_i}\mathcal{C}_1^{(i)} + (1-2p)e^{-i\theta_i}\mathcal{C}_{-1}^{(i)}. \tag{B.6}$$

Seen from the spectrum's perspective, there will be a contraction of some coefficients without extra modes. □

Lemma 13 (Non-unital channel)**.** *For non-unital channels, the Fourier coefficients will be modified and new modes might be present that only include a subset of the parameters following the channel.*

Proof. As explained in Sec. 5.2.1, Clifford channels are not sufficient to represent arbitrary non-unital channels such as amplitude damping. One must also include Pauli

reset channels, which trace out a qubit and initialise it with a specific eigenstate of a Pauli operator. Therefore, from the process path point of view such channels will include additional paths that are independent of the initial state, and only depend on gates following the channel:

$$\begin{aligned}\mathrm{Tr}[O\tilde{\mathcal{U}}(\rho_0)] =& \mathrm{Tr}[O\mathcal{V}_0 \circ \cdots \circ \mathcal{N}_U \circ \cdots \circ \mathcal{V}_m(\rho_0)] \\ &+ \mathrm{Tr}[O\mathcal{V}_0 \circ \cdots \circ \mathcal{V}_i(\sigma)],\end{aligned} \tag{B.7}$$

where $\mathcal{N}_{NU}$ is the non-unital channel, while $\mathcal{N}_U$ is the unital part and σ is the state which the channel resets to. The latter two need not be physical, however they still allow a description in terms of process paths. □

Finally, we explore what happens if the noise channels now depend on some of the parameters.

Lemma 14 (Parameter-dependent and correlated channels)**.** *For channels which depend on some parameters, we can have* $\tilde{\mathcal{S}} \not\subset \mathcal{S}$*.*

This is in contrast with all other types of noise explored previously.

Proof. For simplicity, consider a single noise channel that depends on a single parameter elsewhere in the circuit. Due to causality, the noise channel must follow the unitary with that parameter:

$$\tilde{\mathcal{U}} = \mathcal{V}_0 \circ \cdots \circ \tilde{\mathcal{N}}(\theta_k) \circ \mathcal{U}_l(\theta_l) \circ \cdots \circ \mathcal{U}_k(\theta_k) \circ \cdots \circ \mathcal{V}_m. \tag{B.8}$$

For noise channels and unitaries acting on a subset of the qubits, it is not guaranteed that the measurement operator falls into both light cones. Let us analyse the two distinct cases. In case it does, there may be process paths which depend on both channels, and therefore the range of frequencies will be increased to include the additional ones from the noise channel. Denoting by ω the frequencies in the clean spectrum $\mathcal{S}$, and by $\tilde{\omega}_k$ the frequencies from $\tilde{\mathcal{N}}(\theta_k)$, the new spectrum will be $\tilde{\mathcal{S}} = \{\omega + \tilde{\omega}_k\}$. In case it does not, the two channels will not interact, and therefore the frequencies are not

added together, and the new spectrum will be $\tilde{\mathcal{S}} = \mathcal{S} \cup \{\tilde{\omega}_k\}$. In either case we have $\tilde{\mathcal{S}} \not\subset \mathcal{S}$. □

B.2.1 Shot noise

Lemma 15 (Shot noise)**.** *Given a number of shots n_s, each Fourier coefficient is normally distributed around the noiseless mean with standard deviation:*

$$\sigma(Re\mathcal{F}[\tilde{x}]_{\boldsymbol{k}}) = \frac{\sqrt{d^m - ||x||_2^2}}{d^m\sqrt{2n_s}}, \tag{B.9}$$

and similarly for the imaginary part. Note that the distributions at different $\boldsymbol{k}$'s are not independent.

Proof. First let us review the DFT convention we use in this paper. Given an m-dimensional signal $\{x_{\boldsymbol{i}}\}$, its DFT is defined as:

$$\mathcal{F}[x]_{\boldsymbol{k}} = \frac{1}{d^m}\sum_{\boldsymbol{j}=\boldsymbol{0}} x_{\boldsymbol{j}} e^{-i\boldsymbol{\theta_j}\cdot\boldsymbol{\omega_k}}. \tag{B.10}$$

With this convention Parseval's theorem is $||x||^2 = d^m||\mathcal{F}[x]||^2$. For simplicity we assume that both the samples and the frequencies are considered on regular lattices, and we set $\boldsymbol{\theta_i} = 2\pi\boldsymbol{i}/d$ and $\boldsymbol{\omega_k} = \boldsymbol{k}$. In the quantum case, the signal is corrupted by sampling noise, which follows a binomial distribution. For a sufficient number of samples n_s, one can show that the signal is approximately Gaussian distributed around the expectation value $x_{\boldsymbol{i}}$, with a variance $\sigma_s(x_{\boldsymbol{i}}) = \frac{\sqrt{1-x_{\boldsymbol{i}}^2}}{\sqrt{n_s}}$. Therefore, we can write the signal as a noiseless signal plus a Gaussian noise:

$$\tilde{x}_{\boldsymbol{i}} = x_{\boldsymbol{i}} + G_{\sigma_s(x_{\boldsymbol{i}})}. \tag{B.11}$$

Since the DFT is a linear operation,

$$\mathcal{F}[\tilde{x}]_{\boldsymbol{k}} = \mathcal{F}[x]_{\boldsymbol{k}} + \mathcal{F}[G_{\sigma_s(x)}]_{\boldsymbol{k}}. \tag{B.12}$$

Let us focus on the noise signal. This is equivalent to a Gaussian signal with standard

deviation $1/\sqrt{n_s}$ scaled by a function of x: $G_{\sigma_s(x)} = G_{1/\sqrt{n_s}}\sqrt{1-x^2}$. The Fourier transform of the noise can therefore be written as a convolution:

$$\mathcal{F}[G_{\sigma_s(x)}]_{\boldsymbol{k}} = \left[\mathcal{F}[G_{1/\sqrt{n_s}}] * \mathcal{F}[\sqrt{1-x^2}]\right]_{\boldsymbol{k}}. \tag{B.13}$$

Now the following lemma can be proven:

Lemma 16. *The Fourier transform of a 1D normally distributed signal of size d with standard deviation σ is normally distributed with standard deviation $\sigma/\sqrt{2d}$:*

$$\mathcal{F}[G_\sigma]_k = \left[G_{\sigma/\sqrt{2d}}\right]_k + i\left[G_{\sigma/\sqrt{2d}}\right]_k. \tag{B.14}$$

For an m-dim normal distributed signal of over a grid of volume d^m, the same holds with a standard deviation of $\sigma/\sqrt{2d^m}$.

Proof. Focus on the real part of $\mathcal{F}[G_\sigma]_k$ (the result for the imaginary part follows by symmetry):

$$\operatorname{Re}\mathcal{F}[G_\sigma]_k = \frac{1}{d}\sum_{j=0}^{d} x_j \cos(2\pi\frac{kj}{d}). \tag{B.15}$$

First first notice that $\langle \operatorname{Re}\mathcal{F}[G_\sigma]_{\boldsymbol{k}}\rangle = 0$. Now consider the variance:

$$\sigma^2(\operatorname{Re}\mathcal{F}[G_\sigma]_k) = \frac{1}{d^2}\sum_{jj'}\langle x_j x_{j'}\rangle \cos(2\pi\frac{kj}{d})\cos(2\pi\frac{kj'}{d}). \tag{B.16}$$

Since each x_m is independently distributed, $\langle x_j x_{j'}\rangle$ is only nonzero when $j = j'$:

$$\sigma^2(\operatorname{Re}\mathcal{F}[G_\sigma]_k) = \frac{1}{d^2}\sum_{j}\sigma^2\cos^2(2\pi\frac{kj}{d}) \tag{B.17}$$

$$= \frac{\sigma^2}{2d}. \tag{B.18}$$

All other moments are zero as the higher moments of a normal distribution are zero. Therefore, the distribution is normal with mean 0 and standard deviation $\frac{\sigma}{\sqrt{2d}}$. A similar argument leads to the result for higher dimensions. □

Finally, it remains to be shown that the convolution of a normally distributed signal

and an arbitrary signal is still normally distributed. We can show

Lemma 17. *Take an arbitrary deterministic signal $\{y_k\}$. Then $[G_\sigma * y]_k$ is a normally distributed signal with mean 0 and standard deviation $\sigma||y||_2$. Note that this time, the distributions at different k's are no longer necessarily independent.*

Proof. By definition of the convolution,

$$[G_\sigma * y]_k = \sum_l [G_\sigma]_l y_{k-l} \tag{B.19}$$

Now we have $\langle [G_\sigma * Y]_k \rangle = 0$ as the normal distribution has zero mean. For the variance we can use a similar argument as before:

$$\sigma^2([G_\sigma * y]_k) = \sum_{ll'} \langle [G_\sigma]_l [G_\sigma]_{l'} \rangle y_{k-l} y_{k-l'} \tag{B.20}$$

$$= \sum_l \sigma^2 y_{k-l}^2 = \sigma^2 ||y||_2^2. \tag{B.21}$$

As before, all the higher moments vanish, implying a normal distribution.
To see that the distributions are no longer independent in general, consider the two-point correlation:

$$\langle [G_\sigma * y]_k [G_\sigma * y]'_k \rangle = \sum_{ll'} \langle [G_\sigma]_l [G_\sigma]_{l'} \rangle y_{k-l} y_{k'-l'} \tag{B.22}$$

$$= \sigma^2 \sum_l y_{k-l} y_{k'-l}. \tag{B.23}$$

Therefore the correlation between the distributions at k, k' depends on the autocorrelation characteristics of the coefficients y_k. □

Combining the results of these two lemmas, one deduces that each Fourier coefficient $\mathcal{F}[\tilde{x}]_{\boldsymbol{k}}$ is normally distributed around the noiseless mean $\mathcal{F}[x]_{\boldsymbol{k}}$ with standard deviation:

$$\sigma(\mathrm{Re}\,\mathcal{F}[\tilde{x}]_{\boldsymbol{k}}) = \frac{||\mathcal{F}[\sqrt{1-x^2}]||_2}{\sqrt{2n_s d^m}} \tag{B.24}$$

$$= \frac{\sqrt{d^m - ||x||_2^2}}{d^m \sqrt{2n_s}}, \tag{B.25}$$

where in the second equality we used Parseval's identity. The imaginary part is identically and independently distributed because of symmetry. □

B.3 Figures of merit

Lemma 18 (Average fidelity)**.** *The average fidelity of the output state over all parameters is proportional to the inner product of the vectors of Fourier coefficients:*

$$\langle F \rangle_{\boldsymbol{\theta}} = \frac{1}{2^n} \boldsymbol{c}^{\dagger} \tilde{\boldsymbol{c}}. \tag{B.26}$$

Proof. As a reminder, we use the vectorised notation $\boldsymbol{c}$ for the Fourier coefficients over all Paulis: $[\boldsymbol{c}]_{\boldsymbol{k},i} = c_{\boldsymbol{k}}(P_i)$. Consider the following expression for the fidelity between a pure state and an arbitrary state [FL11]:

$$F(\sigma, \rho) = \sum_i \mathrm{Tr}[\chi_i \sigma] \mathrm{Tr}[\chi_i \rho], \tag{B.27}$$

where the summation is over some informationally complete set of observables, which we can take to be the normalised Pauli observables: $\chi_i = P_i/\sqrt{2^n}$. Now consider the fidelity between a state produced by a noisy parameterised quantum circuit $\tilde{\rho}(\boldsymbol{\theta})$ and its exact counterpart $\rho(\boldsymbol{\theta})$.

In the Fourier picture, the coefficients of the fidelity as a function of $\boldsymbol{\theta}$ are given by a convolution between the Fourier coefficients of the expectation values of the Pauli observables:

$$\mathcal{F}[F]_{\boldsymbol{k}} = \frac{1}{2^n} \sum_i \left[c(P_i) * \tilde{c}(P_i) \right]_{\boldsymbol{k}}, \tag{B.28}$$

where the discrete convolution is defined as:

$$[c * \tilde{c}]_{\boldsymbol{k}} = \sum_{\boldsymbol{x} \in \boldsymbol{\Lambda}} c_{\boldsymbol{x}} \tilde{c}_{\boldsymbol{k}-\boldsymbol{x}}. \tag{B.29}$$

Therefore, the set of Fourier coefficients over all Pauli observables determines the fidelity as a function of the parameters.

If we define the average fidelity over all sampled parameters:

$$\langle F\rangle_{\boldsymbol{\theta}} := \frac{1}{d^m}\sum_{\boldsymbol{j}} F(\rho(\boldsymbol{\theta_j}),\tilde{\rho}(\boldsymbol{\theta_j})) \tag{B.30}$$

we see that it is equivalent to $\mathcal{F}[F]_{\mathbf{0}}$. Hence:

$$\langle F\rangle_{\boldsymbol{\theta}} = \frac{1}{2^n}\sum_i\sum_{\boldsymbol{x}\in\boldsymbol{\Lambda}} c_{\boldsymbol{x}}(P_i)\tilde{c}_{-\boldsymbol{x}}(P_i) = \frac{1}{2^n}\sum_i\sum_{\boldsymbol{x}\in\boldsymbol{\Lambda}} c_{\boldsymbol{x}}(P_i)\tilde{c}^*_{\boldsymbol{x}}(P_i) = \frac{1}{2^n}\boldsymbol{c}^{\,\dagger}\tilde{\boldsymbol{c}}, \tag{B.31}$$

which concludes the proof. □

As a remark, for the average fidelity defined as above to correspond to the average over all possible parameters $\int F(\rho(\boldsymbol{\theta}),\tilde{\rho}(\boldsymbol{\theta}))d^m\boldsymbol{\theta}$ a sufficient condition is for the sampling rate to be twice the maximum frequency of the signal, via Nyquist-Shannon sampling theorem.

We also have a corresponding lemma for the purity:

Lemma 19 (Average purity)**.** *The average purity of the output state over all parameters is proportional to the squared norm of the vector of Fourier coefficients:*

$$\langle P\rangle_{\boldsymbol{\theta}} = \frac{1}{2^n}|\boldsymbol{c}|^2. \tag{B.32}$$

Proof. Notice that in the definition of fidelity above, $P(\rho) = F(\rho,\rho)$. The result follows therefore by taking $\tilde{c} = c$ in the previous Lemma. □

B.4 Filtering-based error mitigation

B.4.1 Perfect filtering for classically simulatable circuits

Theorem 12. *Take a noisy parameterised quantum state $\tilde{\rho}(\boldsymbol{\theta}) = \tilde{\mathcal{U}}(\boldsymbol{\theta})(\rho_0)$ such that all the unparameterised unitaries in $\mathcal{U}(\boldsymbol{\theta})$ are Clifford and ρ_0 is stabiliser. Then given a Pauli measurement operator O, there exists a noise threshold, under which the noiseless landscape $\langle O\rangle_{\rho(\boldsymbol{\theta})}$ can be perfectly reconstructed by sampling from the noisy landscape $\langle O\rangle_{\tilde{\rho}(\boldsymbol{\theta})}$. Above the threshold, such reconstruction is not possible.*

Proof. In Sec. 5.1.2 we saw that the noiseless cost function, when expressed as a trigonometric polynomial, has coefficients $c'_{\boldsymbol{k}} \in \{0, \pm 1\}$ with support $\boldsymbol{\Lambda}$ as defined in Theorem 5. We also saw in Sec. 5.2.1 that noise channels introduce a perturbation in these coefficients, such that under noise they become $\tilde{c}'_{\boldsymbol{k}}$ with potentially different support $\tilde{\boldsymbol{\Lambda}}$. Therefore, given perfect access to the noisy cost function, as long as $|\tilde{c}'_{\boldsymbol{k}} - c'_{\boldsymbol{k}}| \leq \frac{1}{2}\ \forall \boldsymbol{k} \in \boldsymbol{\Lambda}$, a thresholding algorithm can perfectly reconstruct the noiseless trigonometric coefficients by:

$$S(\tilde{c}'_{\boldsymbol{k}}) = \begin{cases} \text{round}(\tilde{c}'_{\boldsymbol{k}}) & \text{if } \boldsymbol{k} \in \boldsymbol{\Lambda}; \\ 0 & \text{otherwise.} \end{cases} \tag{B.33}$$

Conversely, if noise is larger than the threshold, then it is not guaranteed that the outcome will be the correct noiseless landscape. This is a consequence of the fact that parameterised quantum circuits are universal approximators, as proven in [SSM21]. Therefore a landscape produced by a highly noisy state could be corrected to one produced by another, incorrect, circuit. □

In Fig. B.1 we show a practical example of reconstructing such a landscape. The landscape is sampled at different numbers of shots and therefore different levels of noise, showing the existence of a well-defined noise threshold for perfect reconstruction.

B.4.2 Filters for noise-induced frequency modes

In line with classical signal processing, some of the simplest filters that can be applied are band-pass filters, that allow particular frequencies while eliminating the others. These can be designed based on knowledge of the theoretical frequency spectrum, which in the most general case is given by Theorem 5. In general, given an observed function $\tilde{f}$ of quantum expectation values on a support spanned by a basis $\{\psi_{\boldsymbol{k}}\}$ (which may be the Fourier basis or any other basis of choice) indexed by a lattice $\tilde{\boldsymbol{\Lambda}}$, as well as a known exact support indexed by a lattice $\boldsymbol{\Lambda} \subset \tilde{\boldsymbol{\Lambda}}$, the denoised function is given by:

$$f^{\#} = \sum_{\boldsymbol{k} \in \boldsymbol{\Lambda}} \langle f, \psi_{\boldsymbol{k}} \rangle \psi_{\boldsymbol{k}}. \tag{B.34}$$

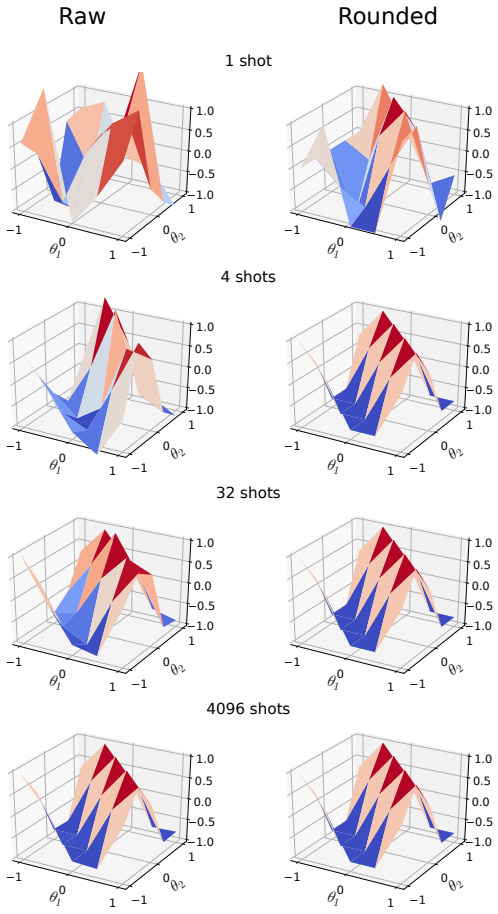


Figure B.1: Reconstructing the landscape of a classically simulatable circuit as run on Quantinuum H1-1 system, at different numbers of shots. Left is raw landscape, right is after rounding of the trigonometric coefficients. Colour indicates the height. Perfect reconstruction is obtained for all experiments except the 1-shot one, the exact (noiseless) landscape corresponds to the last three diagrams in the right column.

In other words, such a filter cuts off all frequencies beyond those expected to be output by the circuit. Such a method would be especially helpful to mitigate correlated noise that can give rise to high-frequency modes. However, it would most likely not be helpful for other non-correlated noise channels, as well as being more resource-intensive by requiring the resolution of a larger set of frequencies.

For specific circuits, it is possible to deduce additional characteristics that the noiseless spectrum should have, allowing for an improved noise mitigation. For example, take the case where the measurement operator is local, i.e. acts on only a subset of qubits. Then it is clear that its expectation value can only depend on the parameters in its backwards light-cone. For a more involved example, assume the input state ρ_0 is known, and the parameters of interest are those of the first r unitaries, which are Pauli gadgets $\mathcal{U}_i(\theta_i) = e^{-iP_i\theta_i/2}$, $i = 1, 2, \cdots, r$. Further, suppose that all P_i's commute, such that the ordering of the Pauli gadgets does not matter. All the subsequent gates, as well as the measurement operator, are arbitrary. Then, if $[\rho_0, P_i] \neq 0$ for all $i \leq r$, it follows that f must contain terms where either all parameters $\theta_1, \cdots, \theta_r$ appear, or none of them do. This is because a parameterised Pauli gadget that does not commute with its input state gives an output state that is a non-constant function of that parameter. A similar reasoning can be applied on the measurement operator and by extension to the expectation value. These conditions apply in many situations of practical interests, thereby affording a more refined error mitigation than general circuits.

B.4.3 Thresholding

The previous filtering methods relied on specific knowledge of the circuits. It is desirable to devise mitigation methods that would apply in general, and therefore would need to be agnostic to the circuit structure. One such method is based on isolating the most significant coefficients. In classical signal processing, this is known as thresholding, and was first studied in Ref. [DJ94]. As a basic review of the method, thresholding applies a non-linear transform to the coefficients of some basis:

$$f^{\#} = \sum_{\boldsymbol{k} \in \tilde{\boldsymbol{\Lambda}}} S(\langle f, \psi_{\boldsymbol{k}} \rangle) \psi_{\boldsymbol{k}}. \tag{B.35}$$

There exist two common choices for the transform S, known as 'hard' and 'soft' thresholding. Hard thresholding has $S_T^H(x) = \begin{cases} x & \text{if } |x| > T \\ 0 & \text{if } |x| \leq T \end{cases}$, while soft thresholding has $S_T^S(x) = \max(1 - T/|x|, 0)x$ [Mal09].

In [DJ94] it was proven that, given an N-sparse signal perturbed by additive Gaussian noise with mean 0 and standard deviation σ, the optimal threshold is $T = \sigma\sqrt{2\log N}$, however in practice it is observed that $T = O(\sigma)$ is sufficient for both hard and soft thresholding [Pey21].

Thresholding methods rely on the basic assumption that the noiseless signal is sparse, or approximately sparse, in the basis of interest. As this in general is not the case for quantum algorithms, we only claim that thresholding be useful on a heuristic level. Another assumption that might not hold is that of additive Gaussian white noise. In the Fourier basis, this is only true for shot noise, however quantum noise will not generally obey these requirements. At the same time, from an experimental point of view, it does make at least practical sense to assume Gaussian noise to reflect the lack of knowledge of these noise processes. To this end, estimates of σ could be derived from a combination of the predicted sources of error, including known decoherence and measurement error rates.

Bibliography

[AA20] Frank Arute and al Arya, et. Hartree-fock on a superconducting qubit quantum computer. *arXiv preprint arXiv:2004.04174*, 2020. URL: `https://arxiv.org/abs/2004.04174`.

[AAB+19] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[AAB+20a] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Andreas Bengtsson, Sergio Boixo, Michael Broughton, Bob B Buckley, et al. Observation of separated dynamics of charge and spin in the fermi-hubbard model. *arXiv preprint arXiv:2010.07965*, 2020.

[AAB+20b] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, David A Buell, et al. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *arXiv preprint arXiv:2004.04197*, 2020. URL: `https://arxiv.org/abs/2004.04197`.

[AABA+24] Rajeev Acharya, Laleh Aghababaie-Beni, Igor Aleiner, Trond I Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw,

Nikita Astrakhantsev, Juan Atalaya, et al. Quantum error correction below the surface code threshold. *arXiv preprint arXiv:2408.13687*, 2024.

[AAR+18] Mohammad H Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum boltzmann machine. *Physical Review X*, 8(2):021050, 2018.

[ABO97] Dorit Aharonov and Michael Ben-Or. Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 176–188, 1997.

[ABOIN96] Dorit Aharonov, Michael Ben-Or, Russell Impagliazzo, and Noam Nisan. Limitations of noisy reversible computation. *arXiv preprint quant-ph/9611028*, 1996.

[AG04] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

[AGDLHG05] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.

[AGL+22] Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. *arXiv preprint arXiv:2211.03999*, 2022.

[AHCC22] Andrew Arrasmith, Zoë Holmes, Marco Cerezo, and Patrick J Coles. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Science and Technology*, 7(4):045015, 2022.

[AK22] Eric R Anschuetz and Bobak T Kiani. Quantum variational algorithms are swamped with traps. *Nature Communications*, 13(1):7760, 2022.

[ALZ+23] Thomas Ayral, Thibaud Louvet, Yiqing Zhou, Cyprien Lambert, E Miles Stoudenmire, and Xavier Waintal. Density-matrix renormal-

ization group algorithm for simulating quantum circuits with a finite fidelity. *PRX Quantum*, 4(2):020304, 2023.

[AMO+24] Yutaro Akahoshi, Kazunori Maruyama, Hirotaka Oshima, Shintaro Sato, and Keisuke Fujii. Partially fault-tolerant quantum computing architecture with error-corrected clifford gates and space-time efficient analog rotations. *PRX Quantum*, 5(1):010337, 2024.

[AOAGC19] Eric Anschuetz, Jonathan Olson, Alán Aspuru-Guzik, and Yudong Cao. Variational quantum factoring. In *Quantum Technology and Optimization Problems: First International Workshop, QTOP 2019, Munich, Germany, March 18, 2019, Proceedings 1*, pages 74–85. Springer, 2019.

[Arr23] Juan Miguel Arrazola. From nisq to isq, June 2023. Accessed: 2025-01-19. URL: `https://pennylane.ai/blog/2023/06/from-nisq-to-isq`.

[ÁS11] Gonzalo A Álvarez and Dieter Suter. Measuring the spectrum of colored noise by dynamical decoupling. *Physical review letters*, 107(23):230501, 2011.

[AST21] Giovanni Acampora, Roberto Schiattarella, and Alfredo Troiano. A dataset for quantum circuit mapping. *Data in Brief*, 39:107526, 2021.

[ASZ+21] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.

[AUC+22] Ismail Yunus Akhalwaya, Shashanka Ubaru, Kenneth L Clarkson, Mark S Squillante, Vishnu Jejjala, Yang-Hui He, Kugendran Naidoo, Vasileios Kalantzis, and Lior Horesh. Towards quantum advantage on noisy quantum computers. *arXiv preprint arXiv:2209.09371*, 2022.

[AW05] George B. Arfken and Hans J. Weber. *Mathematical Methods For Physicists*. Academic Press, 6th edition, 2005.

[BBC+19] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019.

[BBC+24] Kostas Blekos, Dean Brand, Andrea Ceschini, Chiao-Hui Chou, Rui-Hao Li, Komal Pandya, and Alessandro Summer. A review on quantum approximate optimization algorithm and its variants. *Physics Reports*, 1068:1–66, 2024.

[BBF+20] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. *Nature Communications*, 11(1):1–6, 2020. URL: `https://www.nature.com/articles/s41467-020-14454-2`, `doi:10.1038/s41467-020-14454-2`.

[BC23] Tomislav Begušić and Garnet Kin Chan. Fast classical simulation of evidence for the utility of quantum computing before fault tolerance. *arXiv preprint arXiv:2306.16372*, 2023. URL: `https://doi.org/10.48550/arXiv.2306.16372`.

[BCLK+22] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022.

[BD08] Angelo Bassi and Dirk André Deckert. Noise gates for decoherent quantum circuits. *Phys. Rev. A*, 77(3), 2008.

[BDG+24] Michael Bogobowicz, Kamalika Dutta, Martina Gschwendtner, Anna Heid, Mena Issler, Niko Mohr, Henning Soller, Rodney Zemmel, and Alex Zhang. Steady progress in approaching the quantum advantage, April 2024. URL: `https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/steady-progress-in-approaching-the-quantum-advantage`.

[BDP21] Sebastian Brandhofer, Simon Devitt, and Ilia Polian. Error analysis of the variational quantum eigensolver algorithm. In *2021 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pages 1–6. IEEE, 2021.

[Bel57] Richard E. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, 1957.

[Ben80] Paul Benioff. The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *Journal of statistical physics*, 22:563–591, 1980.

[BFH+17] Ryan S Bennink, Erik M Ferragut, Travis S Humble, Jason A Laska, James J Nutaro, Mark G Pleszkoch, and Raphael C Pooser. Unbiased simulation of near-clifford quantum circuits. *Physical Review A*, 95(6):062337, 2017.

[BFLL22] Adam Bouland, Bill Fefferman, Zeph Landau, and Yunchao Liu. Noise and the frontier of quantum supremacy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1308–1317. IEEE, 2022.

[BFNV19] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, 2019.

[BG16] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Physical review letters*, 116(25):250501, 2016.

[BGPP+19] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Information*, 5(1):45, 2019.

[BHC23] Tomislav Begušić, Kasra Hejazi, and Garnet Kin Chan. Simulating quantum circuit expectation values by clifford perturbation theory. *arXiv preprint arXiv:2306.04797*, 2023. URL: `https://doi.org/10.48550/arXiv.2306.04797`.

[BK21] Lennart Bittel and Martin Kliesch. Training variational quantum algorithms is np-hard. *Physical review letters*, 127(12):120502, 2021.

[BKGN+13] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. *arXiv preprint arXiv:1310.4492*, 2013.

[BLSF19] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

[BM07] Rodney J Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291, 2007. URL: `https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.79.291`, `doi:10.1103/RevModPhys.79.291`.

[BMSSO18] Xavi Bonet-Monroig, Ramiro Sagastizabal, M Singh, and TE O'Brien. Low-cost error mitigation by symmetry verification. *Physical Review A*, 98(6):062339, 2018.

[Boy92] John P Boyd. A fast algorithm for chebyshev, fourier, and sinc interpolation onto an irregular grid. *Journal of Computational Physics*, 103(2):243–257, 1992.

[BPLC+23] Carlos Bravo-Prieto, Ryan LaRose, Marco Cerezo, Yigit Subasi, Lukasz Cincio, and Patrick J Coles. Variational quantum linear solver. *Quantum*, 7:1188, 2023.

[Bro70] C G Broyden. The Convergence of a Class of Double-rank Minimization Algorithms: 2. The New Algorithm. *IMA J. Appl. Math.*, 6(3):222–231, 1970. URL: https://doi.org/10.1093/imamat/6.3.222.

[BSF94] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[BSP24] Dean Brand, Ilya Sinayskiy, and Francesco Petruccione. Markovian noise modelling and parameter extraction framework for quantum devices. *Scientific Reports*, 14(1):4769, 2024.

[BV93] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 11–20, 1993.

[BWP+17] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[BWP23] Joseph Bowles, David Wierichs, and Chae-Yeun Park. Backpropagation scaling in parameterised quantum circuits. *arXiv preprint arXiv:2306.14962*, 2023.

[CAB+21] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(1):625–644, 2021. URL: https://www.nature.com/articles/s42254-021-00348-9, doi:10.1038/s42254-021-00348-9.

[CACC21] Piotr Czarnik, Andrew Arrasmith, Patrick J Coles, and Lukasz Cincio. Error mitigation with clifford quantum-circuit data. *Quantum*, 5:592, 2021.

[CCHL22] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of nisq. *arXiv preprint arXiv:2210.07234*, 2022.

[CCHL23] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of nisq. *Nature Communications*, 14(1):6001, 2023.

[CD11] Bob Coecke and Ross Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016, 2011. URL: `https://iopscience.iop.org/article/10.1088/1367-2630/13/4/043016`, `doi:10.1088/1367-2630/13/4/043016`.

[CDWE19] Arnaud Carignan-Dugas, Joel J. Wallman, and Joseph Emerson. Bounding the average gate fidelity of composite channels using the unitarity. *New J. Phys.*, 21(5), 2019.

[CHPZ24] Chi-Fang Chen, Hsin-Yuan Huang, John Preskill, and Leo Zhou. Local minima in quantum systems. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1323–1330, 2024.

[CIP18] Christopher Chamberland, Pavithran Iyer, and David Poulin. Fault-tolerant quantum computing in the pauli or clifford frame with slow error diagnostics. *Quantum*, 2:43, 2018.

[CJ17] Cristina Cirstoiu and David Jennings. Global and local gauge symmetries beyond lagrangian formulations. *arXiv preprint arXiv:1707.09826*, 2017.

[CKW00] Valerie Coffman, Joydip Kundu, and William K Wootters. Distributed entanglement. *Phys. Rev. A*, 61(5):052306, 2000.

[CLGM+23] M Cerezo, Martin Larocca, Diego García-Martín, NL Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, et al. Does provable absence of barren plateaus

imply classical simulability? or, why we need to rethink variational quantum computing. *arXiv preprint arXiv:2312.09121*, 2023.

[CN97] Isaac L Chuang and Michael A Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997.

[Cro18] Gavin E Crooks. Performance of the quantum approximate optimization algorithm on the maximum cut problem. *arXiv preprint arXiv:1811.08419*, 2018.

[CRO+19] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. *Chemical reviews*, 119(19):10856–10915, 2019. URL: `https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00803`, `doi:10.1021/acs.chemrev.8b00803`.

[CRT06a] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[CRT06b] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[CSD20] Alexander Cowtan, Will Simmons, and Ross Duncan. A generic compilation strategy for the unitary coupled cluster ansatz. *arXiv preprint arXiv:2007.10515*, 2020.

[CSV+21] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow

parametrized quantum circuits. *Nature communications*, 12(1):1–12, 2021.

[DCEL09] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.80.012304`.

[DDK18] Pierre-Luc Dallaire-Demers and Nathan Killoran. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.

[Deu85] David Deutsch. Quantum theory, the church–turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117, 1985.

[DH21] Megan L Dahlhauser and Travis S Humble. Modeling noisy quantum circuits using experimental characterization. *Physical Review A*, 103(4):042603, 2021.

[DHJB24] Alexander M Dalzell, Nicholas Hunter-Jones, and Fernando GSL Brandão. Random quantum circuits transform local noise into global white noise. *Communications in Mathematical Physics*, 405(3):78, 2024.

[DJ92] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

[DJ94] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.

[DMN13] Simon J Devitt, William J Munro, and Kae Nemoto. Quantum error correction for beginners. *Reports on Progress in Physics*, 76(7):076001, 2013.

[DNS+22] Abhinav Deshpande, Pradeep Niroula, Oles Shtanko, Alexey V Gorshkov, Bill Fefferman, and Michael J Gullans. Tight bounds on the convergence of noisy random circuits to the uniform distribution. *PRX Quantum*, 3(4):040329, 2022.

[Don06] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[DPMRF23] Giacomo De Palma, Milad Marvian, Cambyse Rouzé, and Daniel Stilck França. Limitations of variational quantum algorithms: a quantum optimal transport approach. *PRX Quantum*, 4(1):010309, 2023.

[DR93] Alok Dutt and Vladimir Rokhlin. Fast fourier transforms for nonequispaced data. *SIAM Journal on Scientific computing*, 14(6):1368–1393, 1993.

[DSRABR+24] MP Da Silva, C Ryan-Anderson, JM Bello-Rivas, A Chernoguzov, JM Dreiling, C Foltz, F Frachon, JP Gaebler, TM Gatterman, L Grans-Samuelsson, et al. Demonstration of logical qubits and repeated error correction with better-than-physical error rates. *arXiv preprint arXiv:2404.02280*, 2024.

[EAGB23] Erik Terres Escudero, Danel Arias Alamo, Oier Mentxaka Gómez, and Pablo García Bringas. Assessing the impact of noise on quantum neural networks: An experimental analysis. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 314–325. Springer, 2023.

[EAŻ05] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347, 2005.

[EBL18] Suguru Endo, Simon C Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8(3):031027, 2018. URL: `https://journals.aps.org/prx/abstract/10.1103/PhysRevX.8.031027`, `doi:10.1103/PhysRevX.8.031027`.

[EMS+24] Alec Eickbusch, Matt McEwen, Volodymyr Sivak, Alexandre Bourassa, Juan Atalaya, Jahan Claes, Dvir Kafri, Craig Gidney, Christopher W Warren, Jonathan Gross, et al. Demonstrating dynamic surface codes. *arXiv preprint arXiv:2412.14360*, 2024.

[EMW21] Daniel J Egger, Jakub Mareček, and Stefan Woerner. Warm-starting quantum optimization. *Quantum*, 5:479, 2021.

[EWP+19] Alexander Erhard, Joel J Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *Nature communications*, 10(1):5347, 2019.

[FCA+22] Enrico Fontana, Marco Cerezo, Andrew Arrasmith, Ivan Rungger, and Patrick J Coles. Non-trivial symmetries in quantum landscapes and their resilience to quantum noise. *Quantum*, 6:804, 2022.

[Fey82] Richard P Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6/7), 1982.

[FFR+21a] Enrico Fontana, Nathan Fitzpatrick, David Muñoz Ramo, Ross Duncan, and Ivan Rungger. Evaluating the noise resilience of variational quantum algorithms. *Phys. Rev. A*, 104:022403, Aug 2021. URL: `https://link.aps.org/doi/10.1103/PhysRevA.104.022403`, `doi:10.1103/PhysRevA.104.022403`.

[FFR+21b] Enrico Fontana, Nathan Fitzpatrick, David Muñoz Ramo, Ross Duncan, and Ivan Rungger. Evaluating the noise resilience of variational quantum algorithms. *Physical Review A*, 104(2):022403, 2021. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.104.022403`, `doi:10.1103/PhysRevA.104.022403`.

[FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*,

2014. URL: https://arxiv.org/abs/1411.4028, doi:10.48550/arXiv.1411.4028.

[FGP22] Daniel Stilck França and Raul Garcia-Patron. A game of quantum advantage: linking verification and simulation. *Quantum*, 6:753, 2022. URL: https://arxiv.org/abs/2011.12173.

[FH16] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv preprint arXiv:1602.07674*, 2016.

[FHC+23] Enrico Fontana, Dylan Herman, Shouvanik Chakrabarti, Niraj Kumar, Romina Yalovetzky, Jamie Heredge, Shree Hari Sureshbabu, and Marco Pistoia. The adjoint is all you need: Characterizing barren plateaus in quantum ans\" atze. *arXiv preprint arXiv:2309.07902*, 2023.

[FL11] Steven T Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Physical review letters*, 106(23):230501, 2011.

[Fle70] R Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13(3):317–322, 1970. URL: https://doi.org/10.1093/comjnl/13.3.317.

[FN18] Edward Farhi and Hartmut Neven. Classification with quantum neural networks on near term processors. *arXiv preprint arXiv:1802.06002*, 2018.

[FRDC22a] Enrico Fontana, Ivan Rungger, Ross Duncan, and Cristina Cîrstoiu. Efficient recovery of variational quantum algorithms landscapes using classical signal processing. *arXiv preprint arXiv:2208.05958*, 2022.

[FRDC22b] Enrico Fontana, Ivan Rungger, Ross Duncan, and Cristina Cîrstoiu. Spectral analysis for noise diagnostics and filter-based digital error mitigation. *arXiv preprint arXiv:2206.08811*, 2022.

[Fuj16] Keisuke Fujii. Noise threshold of quantum supremacy. *arXiv preprint arXiv:1610.03632*, 2016.

[GD18] Xun Gao and Luming Duan. Efficient classical simulation of noisy quantum computation. *arXiv preprint arXiv:1810.03176*, 2018.

[GGTC22] Guillermo González-García, Rahul Trivedi, and J Ignacio Cirac. Error propagation in nisq devices for solving classical optimization problems. *PRX Quantum*, 3(4):040326, 2022.

[GLG22] Sevag Gharibian and François Le Gall. Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum pcp conjecture. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 19–32, 2022.

[GMLC24] Diego García-Martín, Martín Larocca, and Marco Cerezo. Effects of noise on the overparametrization of quantum neural networks. *Physical Review Research*, 6(1):013295, 2024.

[Gol70] Donald Goldfarb. A Family of Variable-Metric Methods Derived by Variational Means. *Math. Comput.*, 24(109):23–26, 1970. URL: `http://www.jstor.org/stable/2004873`.

[Got98] Daniel Gottesman. The heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

[Got19] Daniel Gottesman. Maximally sensitive sets of states. *arXiv preprint arXiv:1907.05950*, 2019.

[Gre15] Daniel Greenbaum. Introduction to quantum gate set tomography. *arXiv:1509.02921*, 2015.

[Gro96] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.

[GTHL+20] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J Zeng. Digital zero noise extrapolation for quantum error mitigation. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 306–316. IEEE, 2020.

[GVT20] Francisco Javier Gil Vidal and Dirk Oliver Theis. Input redundancy for parameterized quantum circuits. *Frontiers in Physics*, 8:297, 2020.

[GWOB19] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019. URL: `https://quantum-journal.org/papers/q-2019-12-09-214/`.

[GZB+20] Bryan T. Gard, Linghua Zhu, George S. Barron, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *NPJ Quantum Inf.*, 6(1), 2020. URL: `http://dx.doi.org/10.1038/s41534-019-0240-1`.

[GZD17] Xun Gao, Zhengyu Zhang, and Luming Duan. An efficient quantum algorithm for generative machine learning. *arXiv preprint arXiv:1711.02038*, 2017.

[HBK21] Tobias Haug, Kishor Bharti, and MS Kim. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum*, 2(4):040309, 2021.

[HdSS02] Xuedong Hu, Rogerio de Sousa, and S Das 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*, pages 3–11. World Scientific, 2002.

[HHM+21] Yipeng Huang, Steven Holtzen, Todd Millstein, Guy Van den Broeck, and Margaret Martonosi. Logical abstractions for noisy variational

quantum algorithm simulation. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 456–472, 2021.

[HIKP12] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Simple and practical algorithm for sparse fourier transform. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 1183–1194. SIAM, 2012.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[HL21] Yifei Huang and Peter Love. Feynman-path-type simulation using stabilizer projector decomposition of unitaries. *Physical Review A*, 103(2):022428, 2021.

[HM22] Qunsheng Huang and Christian B Mendl. Classical simulation of quantum circuits using a multiqubit bloch vector representation of density matrices. *Physical Review A*, 105(2):022409, 2022.

[HMM+14] Jeffrey C Hoch, Mark W Maciejewski, Mehdi Mobli, Adam D Schuyler, and Alan S Stern. Nonuniform sampling and maximum entropy reconstruction in multidimensional nmr. *Accounts of chemical research*, 47(2):708–717, 2014.

[HMO+21] William J Huggins, Sam McArdle, Thomas E O'Brien, Joonho Lee, Nicholas C Rubin, Sergio Boixo, K Birgitta Whaley, Ryan Babbush, and Jarrod R McClean. Virtual distillation for quantum error mitigation. *Physical Review X*, 11(4):041036, 2021.

[HNdJB20] Andre He, Benjamin Nachman, Wibe A de Jong, and Christian W Bauer. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Physical Review A*, 102(1):012426, 2020.

[HNM+21] Akel Hashim, Ravi K. Naik, Alexis Morvan, Jean-Loup Ville, Bradley Mitchell, John Mark Kreikebaum, Marc Davis, Ethan Smith, Costin Iancu, Kevin P. O'Brien, Ian Hincks, Joel J. Wallman, Joseph Emerson, and Irfan Siddiqi. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *Phys. Rev. X*, 11:041039, Nov 2021. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.041039, doi:10.1103/PhysRevX.11.041039.

[Hol24] W. Coffeen Holton. Quantum computer. *Encyclopedia Britannica*, October 10 2024. URL: https://www.britannica.com/technology/quantum-computer.

[Hon21] Honeywell. Honeywell sets another record for quantum computing performance, 2021. URL: https://www.honeywell.com/us/en/news/2021/07/honeywell-sets-another-record-for-quantum-computing-performance.

[HPK13] Markus Heyl, Anatoli Polkovnikov, and Stefan Kehrein. Dynamical quantum phase transitions in the transverse-field Ising model. *Phys. Rev. Lett.*, 110(13):135704, 2013.

[HSCC22] Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022.

[HW97] Scott Hill and William K Wootters. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.*, 78(26):5022, 1997.

[HWO+19] Stuart Hadfield, Zhihui Wang, Bryan O'Gorman, Eleanor G Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2):34, 2019. URL: https://www.mdpi.com/1999-4893/12/2/34.

[JB22] Tyson Jones and Simon C Benjamin. Robust quantum compilation and circuit optimisation via energy minimisation. *Quantum*, 6:628, 2022. URL: https://quantum-journal.org/papers/q-2022-01-24-628/, doi:10.22331/q-2022-01-24-628.

[JIBE23] Aditya Jain, Pavithran Iyer, Stephen D Bartlett, and Joseph Emerson. Improved quantum error correction with randomized compiling. *Physical Review Research*, 5(3):033049, 2023.

[Joz94] Richard Jozsa. Fidelity for Mixed Quantum States. *J. Mod. Opt.*, 41(12):2315–2323, dec 1994. URL: https://doi.org/10.1080/09500349414552171.

[JS22] Sharu Theresa Jose and Osvaldo Simeone. Error-mitigation-aided optimization of parameterized quantum circuits: Convergence analysis. *IEEE Transactions on Quantum Engineering*, 3:1–19, 2022.

[KEA+23] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505, 2023.

[Kit97] A Yu Kitaev. Quantum error correction with imperfect gates. In *Quantum communication, computing, and measurement*, pages 181–188. Springer, 1997.

[KKR05] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. In *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science: 24th International Conference, Chennai, India, December 16-18, 2004. Proceedings 24*, pages 372–383. Springer, 2005.

[KLGS22] Ashish Kakkar, Jeffrey Larson, Alexey Galda, and Ruslan Shaydulin. Characterizing error mitigation by symmetry verification in qaoa. *arXiv preprint arXiv:2204.05852*, 2022.

[KLP+19] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles. Quantum-assisted quantum compiling. *Quantum*, 3:140, May 2019. doi:10.22331/q-2019-05-13-140.

[KLR+08] Emanuel Knill, Dietrich Leibfried, Rolf Reichle, Joe Britton, R Brad Blakestad, John D Jost, Chris Langer, Roee Ozeri, Signe Seidelin, and David J Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 2008.

[KLZ98] Emanuel Knill, Raymond Laflamme, and Wojciech H Zurek. Resilient quantum computation: error models and thresholds. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):365–384, 1998.

[KMT+17] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

[Koc21] Bálint Koczor. Exponential error suppression for near-term quantum devices. *Physical Review X*, 11(3):031057, 2021.

[LAW23] Thibaud Louvet, Thomas Ayral, and Xavier Waintal. Go-no go criteria for performing quantum chemistry calculations on quantum computers. *arXiv preprint arXiv:2306.02620*, 2023.

[LB17] Ying Li and Simon C Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, 7(2):021050, 2017. URL: https://journals.aps.org/prx/abstract/10.1103/PhysRevX.7.021050.

[LCS+22] Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J Coles, and Marco Cerezo. Diagnosing barren

plateaus with tools from quantum optimal control. *Quantum*, 6:824, 2022.

[LGL+21] Xin Liu, Chu Guo, Yong Liu, Yuling Yang, Jiawei Song, Jie Gao, Zhen Wang, Wenzhao Wu, Dajia Peng, Pengpeng Zhao, et al. Redefining the quantum supremacy baseline with a new generation sunway supercomputer. *arXiv preprint arXiv:2111.01066*, 2021.

[LHHGW18] Joonho Lee, William J Huggins, Martin Head-Gordon, and K Birgitta Whaley. Generalized unitary coupled cluster wave functions for quantum computation. *Journal of chemical theory and computation*, 15(1):311–324, 2018. URL: `https://pubs.acs.org/doi/10.1021/acs.jctc.8b01004`, `doi:10.1021/acs.jctc.8b01004`.

[LJGM+23] Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J Coles, and Marco Cerezo. Theory of overparametrization in quantum neural networks. *Nature Computational Science*, 3(6):542–551, 2023.

[LKL+17] Henri Lyyra, Göktuğ Karpat, Chuan Feng Li, Guang Can Guo, Jyrki Piilo, and Sabrina Maniscalco. Symmetry in the open-system dynamics of quantum correlations. *Sci. Rep.*, 7(1), 2017.

[Llo96] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

[LTD+22] Jonas Landman, Slimane Thabet, Constantin Dalyac, Hela Mhiri, and Elham Kashefi. Classically approximating variational quantum machine learning with random fourier features. *arXiv preprint arXiv:2210.13200*, 2022. URL: `https://doi.org/10.48550/arXiv.2210.13200`.

[LTM+20] Wim Lavrijsen, Ana Tudor, Juliane Müller, Costin Iancu, and Wibe De Jong. Classical optimizers for noisy intermediate-scale quantum devices. In *2020 IEEE international conference on quantum computing and engineering (QCE)*, pages 267–277. IEEE, 2020.

[LTW+24] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J Coles, Lukasz Cincio, Jarrod R McClean, Zoë Holmes, and M Cerezo. A review of barren plateaus in variational quantum computing. *arXiv preprint arXiv:2405.00781*, 2024.

[LVS+07] Peter J. Love, Alec Maassen Van Den Brink, A. Yu Smirnov, M. H.S. Amin, M. Grajcar, E. Il'ichev, A. Izmalkov, and A. M. Zagoskin. A characterization of global entanglement. *Quantum Inf. Process.*, 6(3):187–195, 2007.

[LW18] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Physical review letters*, 121(4):040502, 2018.

[LYLW20] Jing Liu, Haidong Yuan, Xiao-Ming Lu, and Xiaoguang Wang. Quantum fisher information matrix and multiparameter estimation. *Journal of Physics A: Mathematical and Theoretical*, 53(2):023001, 2020.

[LYPS17] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun. Hybrid quantum-classical approach to quantum optimal control. *Phys. Rev. Lett.*, 118(15):150503, 2017.

[M+18] Nikolaj Moll et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Sci. Technol.*, 3(3), 2018.

[MAG+24] Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. Noise-induced shallow circuits and absence of barren plateaus. *arXiv preprint arXiv:2403.13927*, 2024.

[Mal09] Stephane Mallat. *A wavelet tour of signal processing: the sparse way.* Academic Press, 2009.

[Man80] Yuri Manin. Computable and uncomputable. *Sovetskoye Radio, Moscow*, 128:15, 1980.

[MB24] Qiang Miao and Thomas Barthel. Equivalence of cost concentration and gradient vanishing for quantum circuits: an elementary proof in the riemannian formulation. *arXiv preprint arXiv:2402.07883*, 2024.

[MBS+18] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018. URL: `https://www.nature.com/articles/s41467-018-07090-4`, `doi:10.1038/s41467-018-07090-4`.

[MEAG+20] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C Benjamin, and Xiao Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1):015003, 2020.

[MGE12] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85(4):042311, 2012.

[MGS+13] Seth T Merkel, Jay M Gambetta, John A Smolin, Stefano Poletto, Antonio D Córcoles, Blake R Johnson, Colm A Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Physical Review A*, 87(6):062119, 2013.

[MGV+23] MS Moreira, Gian Giacomo Guerreschi, Wouter Vlothuizen, Jorge F Marques, Jeroen van Straten, Shavindra P Premaratne, Xiang Zou, Hany Ali, Nandini Muthusubramanian, Christos Zachariadis, et al. Realization of a quantum neural network using repeat-until-success circuits in a superconducting quantum processor. *npj Quantum Information*, 9(1):118, 2023.

[MKSCD17] Jarrod R. McClean, Mollie E. Kimchi-Schwartz, Jonathan Carter, and Wibe A. De Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A*, 95(4):1–10, 2017.

[MKW21] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement-induced barren plateaus. *PRX Quantum*, 2(4):040316, 2021. doi:10.1103/PRXQuantum.2.040316.

[MM19] Alexander Meill and David A. Meyer. Pairwise concurrence in cyclically symmetric quantum states. *Phys. Rev. A*, 100(4), 2019.

[MMS+11] Mark W Maciejewski, Mehdi Mobli, Adam D Schuyler, Alan S Stern, and Jeffrey C Hoch. Data sampling in multidimensional nmr: fundamentals and strategies. *Novel sampling approaches in higher dimensional NMR*, pages 49–77, 2011.

[MNKF18] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.

[Mor16] Valter Moretti. Mathematical foundations of quantum mechanics: An advanced short course. *International Journal of Geometric Methods in Modern Physics*, 13(Supp. 1):1630011, 2016.

[MRBAG16] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, 2016. URL: https://iopscience.iop.org/article/10.1088/1367-2630/18/2/023023/meta, doi:10.1088/1367-2630/18/2/023023.

[MS21] Ashley Montanaro and Stasja Stanisic. Error mitigation by training with fermionic linear optics. *arXiv preprint arXiv:2102.02120*, 2021.

[MWHH20] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg. Characterizing local noise in qaoa circuits. *IOP SciNotes*, 1(2):025208, 2020.

[NC10] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, 2010. URL: https://www.cambridge.org/

core/books/quantum-computation-and-quantum-information/01E10196D0A682A6AEFFEA52D53BE9AE.

[NC11] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[Nes09] M Nest. Simulating quantum computers with probabilistic methods. *arXiv preprint arXiv:0911.1624*, 2009.

[NFT20] Ken M Nakanishi, Keisuke Fujii, and Synge Todo. Sequential minimal optimization for quantum-classical hybrid algorithms. *Physical Review Research*, 2(4):043158, 2020. URL: https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.2.043158.

[NGR+21] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography. *Quantum*, 5:557, 2021.

[NKF23] Nikita A Nemkov, Evgeniy O Kiktenko, and Aleksey K Fedorov. Fourier expansion in variational quantum algorithms. *arXiv preprint arXiv:2304.03787*, 2023.

[OBK+16] Peter JJ O'Malley, Ryan Babbush, Ian D Kivlichan, Jonathan Romero, Jarrod R McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, et al. Scalable quantum simulation of molecular energies. *Physical Review X*, 6(3):031007, 2016.

[OGB21] Mateusz Ostaszewski, Edward Grant, and Marcello Benedetti. Structure optimization for parameterized quantum circuits. *Quantum*, 5:391, 2021.

[OMA+17] Johannes S Otterbach, Riccardo Manenti, Nasser Alidoust, A Bestwick, M Block, B Bloom, S Caldwell, N Didier, E Schuyler Fried, S Hong,

et al. Unsupervised machine learning on a hybrid quantum computer. *arXiv preprint arXiv:1712.05771*, 2017.

[OMML22] Marita Oliv, Andrea Matic, Thomas Messerer, and Jeanette Miriam Lorenz. Evaluating the impact of noise on the performance of the variational quantum eigensolver. *arXiv preprint arXiv:2209.12803*, 2022.

[OPG+04] Jeremy L O'Brien, Geoff J Pryde, Alexei Gilchrist, Daniel FV James, Nathan K Langford, Timothy C Ralph, and Andrew G White. Quantum process tomography of a controlled-not gate. *Physical review letters*, 93(8):080502, 2004.

[PBE22] Elijah Pelofske, Andreas Bärtschi, and Stephan Eidenbenz. Quantum volume in practice: What users can expect from nisq devices. *IEEE Transactions on Quantum Engineering*, 3:1–19, 2022.

[PBP+24] Lukas Postler, Friederike Butt, Ivan Pogorelov, Christian D Marciniak, Sascha Heußen, Rainer Blatt, Philipp Schindler, Manuel Rispler, Markus Müller, and Thomas Monz. Demonstration of fault-tolerant steane quantum error correction. *PRX Quantum*, 5(3):030326, 2024.

[PCDR+19] Timothy J Proctor, Arnaud Carignan-Dugas, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Direct randomized benchmarking for multiqubit devices. *Physical review letters*, 123(3):030503, 2019.

[PDTH24] Ricard Puig, Marc Drudis, Supanut Thanasilp, and Zoë Holmes. Variational quantum simulation: a case study for understanding warm starts. *arXiv preprint arXiv:2404.10044*, 2024.

[Pey21] Gabriel Peyré. *Mathematical foundations of data sciences*. 2021. [Online; accessed 10-August-2021].

[PF23] Lirandë Pira and Chris Ferrie. Explicability and inexplicability in the interpretation of quantum neural networks. *arXiv preprint arXiv:2308.11098*, 2023.

[PIOM19] Robert M Parrish, Joseph T Iosue, Asier Ozaeta, and Peter L McMahon. A jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization. *arXiv preprint arXiv:1904.03206*, 2019.

[PMS+14] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man Hong Yung, Xiao Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5:4213, 2014. URL: `https://www.nature.com/articles/ncomms5213`, `doi:10.1038/ncomms5213`.

[PNGY21] Taylor L Patti, Khadijeh Najafi, Xun Gao, and Susanne F Yelin. Entanglement devised barren plateau mitigation. *Physical Review Research*, 3(3):033090, 2021.

[POBRGB18] Alejandro Perdomo-Ortiz, Marcello Benedetti, John Realpe-Gómez, and Rupak Biswas. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Science and Technology*, 3(3):030502, 2018.

[Pre18] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.

[PS03] Daniel Potts and Gabriele Steidl. Fast summation at nonequispaced knots by nfft. *SIAM Journal on Scientific Computing*, 24(6):2013–2037, 2003.

[PSB+21] Christophe Piveteau, David Sutter, Sergey Bravyi, Jay M Gambetta, and Kristan Temme. Error mitigation for universal gates on encoded qubits. *Physical review letters*, 127(20):200505, 2021.

[PSR+22] Timothy Proctor, Stefan Seritan, Kenneth Rudinger, Erik Nielsen, Robin Blume-Kohout, and Kevin Young. Scalable randomized benchmarking of quantum computers using mirror circuits. *Physical Review Letters*, 129(15):150502, 2022.

[Qis21] Qiskit. Statevectorsimulator, 2021. URL: `https://qiskit.org/documentation/stubs/qiskit.providers.aer.StatevectorSimulator.html`.

[QSFK+24] Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. Exponentially tighter bounds on limitations of quantum error mitigation. *Nature Physics*, 20(10):1648–1658, 2024.

[Qua23a] Quantinuum system model h1 product data sheet, version 5.40. Technical report, Quantinuum, June 2023. Version 5.40. URL: `https://assets.website-files.com/62b9d45fb3f64842a96c9686/648c742dd3e744dfeeb7cd06_Quantinuum%20H1%20Product%20Data%20Sheet%20v5.4%2015Jun23.pdf`.

[Qua23b] Quantinuum. Quantinuum h-series quantum computer accelerates through 3 more performance records for quantum volume, June 2023. Accessed via Quantinuum blog; publication date: June 30, 2023 :contentReference[oaicite:0]index=0. URL: `https://www.quantinuum.com/blog/quantinuum-h-series-quantum-computer-accelerates-through-3-more-perfor`

[RAT18] Diego Ruiz-Antolin and Alex Townsend. A nonuniform fast fourier transform based on low rank approximation. *SIAM Journal on Scientific Computing*, 40(1):A529–A547, 2018.

[RBKSC04] Joseph M Renes, Robin Blume-Kohout, Andrew J Scott, and Carlton M Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

[RBM+19] Jonathan Romero, Ryan Babbush, Jarrod R McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, 2019. URL: https://iopscience.iop.org/article/10.1088/2058-9565/aad3e4/meta?casa_token=pvqYhPQeNtwAAAAA:D0BG_T_-eh6wGRAi0sD9JmEvOzgohwa_-fMdP2KRz0HEmXyJ6Xy2N4JMxK5iUk47a9hkrPTleg.

[RBS+23] Michael Ragone, Bojko N Bakalov, Frédéric Sauvage, Alexander F Kemper, Carlos Ortiz Marrero, Martin Larocca, and M Cerezo. A unified theory of barren plateaus for deep parametrized quantum circuits. *arXiv preprint arXiv:2309.09342*, 2023.

[Res] Wolfram Research. CircularRealMatrixDistribution - Wolfram Language Documentation. https://reference.wolfram.com/language/ref/CircularRealMatrixDistribution.html.

[RFC+19] Ivan Rungger, Nathan Fitzpatrick, Honxiang Chen, CH Alderete, Harriett Apel, Alexander Cowtan, Andrew Patterson, D Munoz Ramo, Yingyue Zhu, Nhung Hong Nguyen, et al. Dynamical mean field theory algorithm and experiment on quantum computers. *arXiv preprint arXiv:1910.04735*, 2019. URL: https://arxiv.org/abs/1910.04735, doi:10.48550/arXiv.1910.04735.

[RFHC23] Manuel S Rudolph, Enrico Fontana, Zoë Holmes, and Lukasz Cincio. Classical surrogate simulation of quantum systems with lowesa. *arXiv preprint arXiv:2308.09109*, 2023. URL: https://doi.org/10.48550/arXiv.2308.09109.

[RGM22] Eliott Rosenberg, Paul Ginsparg, and Peter L McMahon. Experimental error mitigation using linear rescaling for variational quantum

eigensolving with up to 20 qubits. *Quantum Science and Technology*, 7(1):015024, 2022.

[RJV+20] Brian Rost, Barbara Jones, Mariya Vyushkova, Aaila Ali, Charlotte Cullip, Alexander Vyushkov, and Jarek Nabrzyski. Simulation of thermal relaxation in spin chemistry systems on a quantum computer using inherent qubit decoherence. *arXiv preprint arXiv:2001.00794*, 2020.

[RLCK19] Patrick Rall, Daniel Liang, Jeremy Cook, and William Kretschmer. Simulation of qubit quantum circuits via pauli propagation. *Physical Review A*, 99(6):062337, 2019.

[RMM+23] Manuel S Rudolph, Jacob Miller, Danial Motlagh, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. Synergistic pretraining of parametrized quantum circuits via tensor networks. *Nature Communications*, 14(1):8367, 2023.

[RSA78] Ronald L Rivest, Adi Shamir, and Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[SBG+19] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.99.032331`, `doi:10.1103/PhysRevA.99.032331`.

[SBMS+19] Ramiro Sagastizabal, Xavier Bonet-Monroig, Malay Singh, M Adriaan Rol, CC Bultink, Xiang Fu, CH Price, VP Ostroukh, N Muthusubramanian, A Bruno, et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Physical Review A*, 100(1):010302, 2019.

[SCCC22] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J Coles. Trainability of dissipative perceptron-based quantum neural networks. *Phys-*

ical Review Letters, 128(18):180505, 2022. doi:10.1103/PhysRevLett.128.180505.

[SEFT22] Yasunari Suzuki, Suguru Endo, Keisuke Fujii, and Yuuki Tokunaga. Quantum error mitigation as a universal error reduction technique: Applications from the nisq to the fault-tolerant quantum computing eras. *PRX Quantum*, 3(1):010345, 2022.

[SEM23] Franz J Schreiber, Jens Eisert, and Johannes Jakob Meyer. Classical surrogates for quantum learning models. *Physical Review Letters*, 131(10):100803, 2023.

[SFGP21a] Daniel Stilck França and Raul Garcia-Patron. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, 2021. URL: https://www.nature.com/articles/s41567-021-01356-3, doi:10.1038/s41567-021-01356-3.

[SFGP21b] Daniel Stilck França and Raul Garcia-Patron. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, 2021.

[Sha70] D F Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Math. Comput.*, 24(111):647–656, 1970. URL: http://www.jstor.org/stable/2004840.

[Sha21] Andrew Shaw. Classical-quantum noise mitigation for nisq hardware. *arXiv preprint arXiv:2105.08701*, 2021.

[Sho95] Peter W Shor. Scheme for reducing decoherence in quantum computer memory. *Physical review A*, 52(4):R2493, 1995.

[Sho99] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[Sim97] Daniel R Simon. On the power of quantum computation. *SIAM journal on computing*, 26(5):1474–1483, 1997.

[SJAG19] Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019.

[SKCC20] Kunal Sharma, Sumeet Khatri, Marco Cerezo, and Patrick J Coles. Noise resilience of variational quantum compiling. *New J. Phys.*, 22(4):043006, 2020.

[SL24] Phattharaporn Singkanipa and Daniel A Lidar. Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and fixed points. *arXiv preprint arXiv:2402.08721*, 2024.

[SLC+24] Ruslan Shaydulin, Changhao Li, Shouvanik Chakrabarti, Matthew DeCross, Dylan Herman, Niraj Kumar, Jeffrey Larson, Danylo Lykov, Pierre Minssen, Yue Sun, et al. Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem. *Science Advances*, 10(22):eadm6761, 2024.

[SMM+20] Andrea Skolik, Jarrod R McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *arXiv preprint arXiv:2006.14904*, 2020. URL: `https://arxiv.org/abs/2006.14904`.

[SN20] Jun John Sakurai and Jim Napolitano. *Modern quantum mechanics*. Cambridge University Press, 2020.

[SSM21] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021.

[SSMAG16] Nicolas P.D. Sawaya, Mikhail Smelyanskiy, Jarrod R. McClean, and Alán Aspuru-Guzik. Error sensitivity to environmental noise in quantum circuits for chemical state preparation. *J. Chem. Theory Comput.*, 12(7):3097–3108, 2016.

[SSP14] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. *Quantum Information Processing*, 13(11):2567–2586, 2014. URL: `https://link.springer.com/article/10.1007/s11128-014-0809-8`, `doi:10.1007/s11128-014-0809-8`.

[SSP15] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.

[SWA21] Waheeda Saib, Petros Wallden, and Ismail Akhalwaya. The effect of noise on the performance of variational algorithms for quantum chemistry. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 42–53. IEEE, 2021.

[SWC24] Marco Schumann, Frank K Wilhelm, and Alessandro Ciani. Emergence of noise-induced barren plateaus in arbitrary layered noise models. *Quantum Science and Technology*, 9(4):045019, 2024.

[SWS15] Yuval R Sanders, Joel J Wallman, and Barry C Sanders. Bounding quantum gate error rate based on reported average fidelity. *New J. Phys.*, 18(1):12002, 2015.

[SWSW24] Weixiao Sun, Fuchuan Wei, Yuguo Shao, and Zhaohui Wei. Sudden death of quantum advantage in correlation generations. *Science Advances*, 10(47):eadr5002, 2024.

[TAF+24] Riki Toshio, Yutaro Akahoshi, Jun Fujisaki, Hirotaka Oshima, Shintaro Sato, and Keisuke Fujii. Practical quantum advantage on partially

fault-tolerant quantum computer. *arXiv preprint arXiv:2408.14848*, 2024.

[Tal96] Michel Talagrand. A new look at independence. *The Annals of Probability*, 24(1):1 – 34, 1996. doi:10.1214/aop/1042644705.

[TBG17] Kristan Temme, Sergey Bravyi, and Jay M Gambetta. Error mitigation for short-depth quantum circuits. *Physical review letters*, 119(18):180509, 2017. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.180509, doi:10.1103/PhysRevLett.119.180509.

[TCC+22] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H Booth, et al. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, 2022.

[TCD+24] Cenk Tüysüz, Su Yeon Chang, Maria Demidik, Karl Jansen, Sofia Vallecorsa, and Michele Grossi. Symmetry breaking in geometric quantum machine learning in the presence of noise. *PRX Quantum*, 5(3):030314, 2024.

[TD02] Barbara M Terhal and David P DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *arXiv preprint quant-ph/0205133*, 2002.

[TEMG22] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1):114, 2022.

[TFSS23] Joseph Tindall, Matt Fishman, Miles Stoudenmire, and Dries Sels. Efficient tensor network simulation of IBM's kicked ising experiment. *arXiv preprint arXiv:2306.14887*, 2023. URL: https://doi.org/10.48550/arXiv.2306.14887.

[TG07] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.

[TMN+21] Yuto Takaki, Kosuke Mitarai, Makoto Negoro, Keisuke Fujii, and Masahiro Kitagawa. Learning temporal data with a variational quantum recurrent neural network. *Physical Review A*, 103(5):052414, 2021.

[TQ18] Swamit S Tannu and Moinuddin K Qureshi. A case for variability-aware policies for nisq-era quantum computers. *arXiv preprint arXiv:1805.10224*, 2018.

[TSY23] Kento Tsubouchi, Takahiro Sagawa, and Nobuyuki Yoshioka. Universal cost bound of quantum error mitigation based on quantum estimation theory. *Physical Review Letters*, 131(21):210601, 2023.

[TTG23] Ryuji Takagi, Hiroyasu Tajima, and Mile Gu. Universal sampling lower bounds for quantum error mitigation. *Physical Review Letters*, 131(21):210602, 2023.

[TWA+23] Giacomo Torlai, Christopher J Wood, Atithi Acharya, Giuseppe Carleo, Juan Carrasquilla, and Leandro Aolita. Quantum process tomography with unsupervised learning and tensor networks. *Nature Communications*, 14(1):2858, 2023.

[UB21] AV Uvarov and Jacob D Biamonte. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*, 54(24):245301, 2021.

[VBM+19] Guillaume Verdon, Michael Broughton, Jarrod R McClean, Kevin J Sung, Ryan Babbush, Zhang Jiang, Hartmut Neven, and Masoud Mohseni. Learning to learn with quantum neural networks via classical neural networks. *arXiv preprint arXiv:1907.05415*, 2019. URL: `https://arxiv.org/abs/1907.05415`.

[VC21] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, 2021.

[VDBMKT23] Ewout Van Den Berg, Zlatko K Minev, Abhinav Kandala, and Kristan Temme. Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors. *Nature Physics*, 19(8):1116–1121, 2023.

[VJW+23] Harish J Vallury, Michael A Jones, Gregory AL White, Floyd M Creevey, Charles D Hill, and Lloyd CL Hollenberg. Noise-robust ground state energy estimates from deep quantum circuits. *Quantum*, 7:1109, 2023.

[VKG+21] Joseph Vovrosh, Kiran E Khosla, Sean Greenaway, Christopher Self, Myungshik S Kim, and Johannes Knolle. Simple mitigation of global depolarizing errors in quantum simulations. *Physical Review E*, 104(3):035309, 2021.

[VSB+01] Lieven MK Vandersypen, Matthias Steffen, Gregory Breyta, Costantino S Yannoni, Mark H Sherwood, and Isaac L Chuang. Experimental realization of shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature*, 414(6866):883–887, 2001.

[VT18] Javier Gil Vidal and Dirk Oliver Theis. Calculus on parameterized quantum circuits. *arXiv preprint arXiv:1812.06323*, 2018.

[Wal18] Joel J Wallman. Randomized benchmarking with gate-dependent noise. *Quantum*, 2:47, 2018.

[WCA+21] Samson Wang, Piotr Czarnik, Andrew Arrasmith, M Cerezo, Lukasz Cincio, and Patrick J Coles. Can error mitigation improve trainability of noisy variational quantum algorithms? *arXiv preprint arXiv:2109.01051*, 2021.

[WCA+24] Samson Wang, Piotr Czarnik, Andrew Arrasmith, Marco Cerezo, Lukasz Cincio, and Patrick J Coles. Can error mitigation improve trainability of noisy variational quantum algorithms? *Quantum*, 8:1287, 2024.

[WE16] Joel J Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5):052325, 2016. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.94.052325`.

[Wei13] Yaakov S Weinstein. Pseudorandom circuits from clifford-plus-t gates. *Physical Review A—Atomic, Molecular, and Optical Physics*, 88(6):062303, 2013.

[WFC+21] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1):6961, 2021.

[WGK20] David Wierichs, Christian Gogolin, and Michael Kastoryano. Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer. *Physical Review Research*, 2(4):043246, 2020.

[WGMAG14] Jonathan Welch, Daniel Greenbaum, Sarah Mostame, and Alan Aspuru-Guzik. Efficient quantum circuits for diagonal unitaries without ancillas. *New Journal of Physics*, 16(3):033040, 2014.

[WHE+04] Yaakov S Weinstein, Timothy F Havel, Joseph Emerson, Nicolas Boulant, Marcos Saraceno, Seth Lloyd, and David G Cory. Quantum process tomography of the quantum fourier transform. *The Journal of chemical physics*, 121(13):6117–6133, 2004.

[WHJR18] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel. Quantum approximate optimization algorithm for MaxCut: A fermionic view. *Phys. Rev. A*, 97(2):022304, 2018. `doi:10.1103/PhysRevA.97.022304`.

[WHT15] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92(4):042303, 2015.

[WIWL22] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, 2022.

[Woo98] William K Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80(10):2245, 1998.

[Woo01] William K Wootters. Entanglement of formation and concurrence. *Quantum Inf. Comput.*, 1(1):27–44, 2001.

[WSG+24] Yang Wang, Selwyn Simsek, Thomas M Gatterman, Justin A Gerber, Kevin Gilmore, Dan Gresh, Nathan Hewitt, Chandler V Horst, Mitchell Matheny, Tanner Mengle, et al. Fault-tolerant one-bit addition with the smallest interesting color code. *Science Advances*, 10(29):eado9024, 2024.

[WZdS+20] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring entanglement and optimization within the hamiltonian variational ansatz. *PRX Quantum*, 1(2):020319, 2020. URL: `https://journals.aps.org/prxquantum/pdf/10.1103/PRXQuantum.1.020319`, `doi:10.1103/PRXQuantum.1.020319`.

[XSE+21] Xiaosi Xu, Jinzhao Sun, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan. Variational algorithms for linear algebra. *Science Bulletin*, 66(21):2181–2188, 2021.

[YDCM23] Yuxuan Yan, Zhenyu Du, Junjie Chen, and Xiongfeng Ma. Limitations of noisy quantum devices in computational and entangling power. *arXiv preprint arXiv:2306.02836*, 2023.

[You96] Peter Young. Jackknife and bootstrap resampling methods in statistical analysis to correct for bias. *Statistical Science*, 11:189–228, 1996.

[YSH11] Tatsuro Yuge, Susumu Sasaki, and Yoshiro Hirayama. Measurement of the noise spectrum using a multiple-pulse sequence. *Physical review letters*, 107(17):170504, 2011.

[YW12] Kevin C Young and K Birgitta Whaley. Qubits as spectrometers of dephasing noise. *Physical Review A*, 86(1):012314, 2012.

[YYE07] M Yönaç, Ting Yu, and J H Eberly. Pairwise concurrence dynamics: a four-qubit model. *J. Phys. B*, 40(9):S45—-S59, apr 2007. URL: `https://doi.org/10.1088{%}2F0953-4075{%}2F40{%}2F9{%}2Fs02`.

[ZCC+22] Qingling Zhu, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, et al. Quantum computational advantage via 60-qubit 24-cycle random circuit sampling. *Science bulletin*, 67(3):240–245, 2022.

[ZKH+25] Zoltán Zimborás, Bálint Koczor, Zoë Holmes, Elsi-Mari Borrelli, András Gilyén, Hsin-Yuan Huang, Zhenyu Cai, Antonio Acín, Leandro Aolita, Leonardo Banchi, et al. Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't. *arXiv preprint arXiv:2501.05694*, 2025.

[ZSW20] Yiqing Zhou, E Miles Stoudenmire, and Xavier Waintal. What limits the simulation of quantum computers? *Physical Review X*, 10(4):041038, 2020.

[ZVBL23] Alexander Zlokapa, Benjamin Villalonga, Sergio Boixo, and Daniel A Lidar. Boundaries of quantum supremacy via random circuit sampling. *npj Quantum Information*, 9(1):36, 2023.

[ZWC+20] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D Lukin. Quantum approximate optimization algorithm: Per-

formance, mechanism, and implementation on near-term devices. *Physical Review X*, 10(2):021067, 2020.

[ZWC+21] Jinfeng Zeng, Zipeng Wu, Chenfeng Cao, Chao Zhang, Shi-Yao Hou, Pengxiang Xu, and Bei Zeng. Simulating noisy variational quantum eigensolver with local noise models. *Quantum Engineering*, 3(4):e77, 2021.

[ZYZP19] Linxi Zhang, Yan Yu, Changhua Zhu, and Changxing Pei. Noise tailoring for quantum circuits via unitary 2t-design. *Sci. Rep.*, 9(1):1–11, 2019. URL: `http://dx.doi.org/10.1038/s41598-018-38158-2`.