INAUGURAL - DISSERTATION
zur Erlangung der Doktorwürde der
Gesamtfakultät für Matkematik,
Ingeniur-und Naturwissenschaften der
Ruprecht–Karls–Universität Heidelberg

QUANTUM CHEMISTRY
ON QUANTUM COMPUTERS

Vorlegt von: Daniel Bultrini

Tag der mündlichen Prüfung: 21.03.2025

Gutachter: Prof. Dr. Oriol Vendrell

Gutachter: Prof. Dr. Andreas Dreuw

Contents

APPENDIX

List of Figures

List of Tables

Dedicated to all those whose paths have constructively interfered with mine.

12

Acknowledgements

I would like to take a moment to thank the many people who have always been there the whole time for me, and that's my parents Orit and Raimondo, who are always there to lend an ear and support through the many difficult parts of the PhD journey and beyond. To my brother Ben who popped by here and there, but it was definitely nice to know he was there somewhere. And of course, to Elena who made the difficult times easier and much more fun! I wouldn't have it any other way.

This whole doctorate wouldn't have been possible at all without my Doktorvater Oriol Vendrell, who was always available to discuss and managed to put up with my many doubts on quantum computers amongst other things. Since I have written this thesis now I would say that the job was completed successfully thanks in large part to him.

I am also grateful that I had a lot of collaborators and teachers along this journey, German Sierra and Esperanza Lopez of the institute of theoretical physics in Madrid and of course, to my colleagues at Los Alamos National Laboratory, Lukasz Cincio and Patrick Coles, who taught me the importance and the process of extreme scientific rigor, Piotr Czarink for showing an amazing technical aptitude and unparalleled work ethic, and also to Max Hunter-Gordon with whom it was a pleasure to work with and explore Madrid for the time we were coworkers.

I would like to give a special thank you to all the students who have joined my 'Heidelberg Quantum' group so that I could talk to and teach some other enthusiastic students about the field, in particular Sophia Gurtler, Dominik Tonne and Justus Lau. Jing Sun started out as my office mate at the start of my Heidelberg time and actually ended up doing a postdoc in quantum computing as well as a multi-year joint project in board games with me.

For my fantastic stay IBM Zürich, a big thank you to Laurin Fischer for being an amazing co-worker and friend, Francesco Tacchino, Ivano Tavernelli and Stefan Woerner for their supervision. In addition, the fantastic opportunity to run my algorithms on hardware and for in general being generous lovely people, Malay Singh and Federico Roy. I would write a lot more about the many people who have been important to this PhD, but I will at least mention my other co-authors and collaborators, Samson Wang, Marco Cerezo, Andrew Arransmith, Nick Koukoulekidis and Tom O'Leary.

I have learned a lot through the process of working with you all, so thank you for the time shared!

Abstract

DESIGNING AN ALGORITHM for quantum chemistry for a quantum computer requires a lot more than the pure algorithmic design. Quantum computers, especially the ones that exist at the time of writing, are not perfect and are subject to many limitations in terms of the types and amount of operations that can be done. Due to this it is important to design resource-efficient algorithms and strategies to mitigate the effect of noise. To do the latter it is also important to study what actually is the effect of noise on a computation. Armed with this knowledge it is then possible to tackle the actual implementation of a quantum algorithm on hardware.

Quantum chemistry, and specifically chemical dynamics is a particularly exciting canvas upon which to explore the potential of quantum computing. This thesis aims to present the full journey from algorithmic design for chemical problems to their implementation on quantum computers and the extraction of useful values through error mitigation.

To achieve this Part I of the work introduces all the context and concepts required to follow the research presented. The Chapter 1 discusses the reason why one would want to apply quantum computers to problems in general and in chemistry. The Chapter 2 presents the foundational mathematics, quantum theory and high level quantum computing prospects. This allows the third chapter on quantum algorithms, with a focus on the variational quantum eigensolver to be followed. Because quantum algorithms are subject to noise, Chapter 4 introduces the theory and effects of noise and a selection of fundamental techniques used to mitigate noise or correct for errors. Chapter 5 then presents how quantum chemistry is dealt within quantum computing.

Part II presents the research and results, beginning with Chapter 6 which presents a quantum algorithm to perform non-adiabatic mixed quantum-classical dynamics with a fixed depth circuit. The Time-Dependent Variational Quantum Propagation (TDVQP) algorithm is then tested on the Shin-Metiu Model, where it achieves promising results as a proof of concept study. However, that study was noiseless, so in anticipation of the inevitable hardware noise, Chapter 7 presents an error mitigation technique called UNIfied Technique for Error mitigation with Data (UNITED) which was tested on both variational and random circuits showing that the combination is better than the parts. To further understand the effect of noise Chapter 8 presents a different approach to the problem; starting with a perfect quantum computer, what happens when noisy qubits are added one at a time? Here it was shown that a single noisy qubit is enough to make the whole machine behave as though it was noisy, although with an exponentially suppressed error rate. Finally, Chapter 9 closes off this section by showing the process of implementing TDVQP on a 4 qubit quantum computer at the Walter Meisner Institute, at the time of writing experimental results were not yet available, but here all the considerations, caveats and compromises required to run on real hardware are shown, as well as the performance of UNITED on this problem.

Closing off the work Part III holds the concluding remarks, an outlook to the future of quantum computing and a reflection on the work that was done. Actually computing useful quantum chemistry on quantum computers is a tough business, and perhaps it will require fault-tolerant machines and clever algorithms together, but there are promising attempts at the problem with noisy machines all the same.

14

Abstrakt

Die Entwicklung eines Algorithmus für die Quantenchemie für einen Quantencomputer erfordert viel mehr als nur die reine algorithmische Entwicklung. Quantencomputer, insbesondere die, die zum Zeitpunkt der Erstellung dieses Artikels existieren, sind nicht perfekt und unterliegen vielen Beschränkungen in Bezug auf die Art und Menge der durchführbaren Operationen. Aus diesem Grund ist es wichtig, ressourceneffiziente Algorithmen und Strategien zu entwickeln, um die Auswirkungen des Rauschens abzuschwächen. Zu diesem Zweck muss auch untersucht werden, wie sich Rauschen tatsächlich auf eine Berechnung auswirkt. Mit diesem Wissen ist es dann möglich, die tatsächliche Implementierung eines Quantenalgorithmus auf Hardware in Angriff zu nehmen.

Die Quantenchemie, insbesondere die chemische Dynamik, ist ein besonders spannendes Feld, auf dem das Potenzial des Quantencomputers erforscht werden kann. Ziel dieser Arbeit ist es, den gesamten Weg vom Entwurf der Algorithmen für chemische Probleme bis zu ihrer Implementierung auf Quantencomputern und der Gewinnung nützlicher Werte durch Fehlerbegrenzung darzustellen.

Um dies zu erreichen, führt Teil I der Arbeit in alle Zusammenhänge und Konzepte ein, die erforderlich sind, um die vorgestellte Forschung zu verfolgen. In Kapitel 1 werden die Gründe erörtert, warum man Quantencomputer auf Probleme im Allgemeinen und in der Chemie anwenden möchte. In Kapitel 2 werden die mathematischen Grundlagen, die Quantentheorie und die Aussichten für Quantencomputer auf hohem Niveau vorgestellt. Daran schließt sich das dritte Kapitel über Quantenalgorithmen an, wobei der Schwerpunkt auf dem Variationsquanten-Eigensolver liegt. Da Quantenalgorithmen mit Rauschen behaftet sind, werden in Kapitel 4 die Theorie und die Auswirkungen von Rauschen sowie eine Auswahl grundlegender Techniken zur Rauschminderung oder Fehlerkorrektur vorgestellt. In Kapitel 5 wird dann dargestellt, wie die Quantenchemie im Rahmen der Quanteninformatik behandelt wird.

In Teil II werden die Forschungsarbeiten und Ergebnisse vorgestellt, beginnend mit Kapitel 6, in dem ein Quantenalgorithmus zur Durchführung einer nicht-adiabatischen gemischten quantenklassischen Dynamik mit einem Schaltkreis fester Tiefe vorgestellt wird. Der Algorithmus TDVQP (Time-Dependent Variational Quantum Propagation) wird dann am Shin-Metiu-Modell getestet, wo er in einer Machbarkeitsstudie vielversprechende Ergebnisse erzielt. Diese Studie war jedoch geräuschlos, so dass in Erwartung des unvermeidlichen Hardware-Rauschens in Kapitel 7 eine Technik zur Fehlerminderung vorgestellt wird, die UNIfied Technique for Error mitigation with Data (UNITED) genannt wird und sowohl an variablen als auch an zufälligen Schaltungen getestet wurde. Um die Auswirkungen des Rauschens besser zu verstehen, wird in Kapitel 8 ein anderer Ansatz für das Problem vorgestellt: Was passiert, wenn man ausgehend von einem perfekten Quantencomputer ein verrauschtes Qubit nach dem anderen hinzufügt? Hier wurde gezeigt, dass ein einziges verrauschtes Qubit ausreicht, damit sich die gesamte Maschine so verhält, als sei sie verrauscht, allerdings mit einer exponentiell unterdrückten Fehlerrate. Schließlich schließt Kapitel 9 diesen Abschnitt ab, indem es den Prozess der Implementierung von TDVQP auf einem 4-Qubit-Quantencomputer am Walter-Meisner-Institut zeigt. Zum Zeitpunkt der Abfassung dieses Artikels lagen noch keine experimentellen Ergebnisse vor, aber hier werden alle Überlegungen, Vorbehalte und Kompromisse, die erforderlich sind, um auf echter Hardware zu laufen, sowie die Leistung von UNITED bei diesem Problem gezeigt.

Zum Abschluss der Arbeit enthält Teil III die Schlussbemerkungen, einen Ausblick auf die Zukunft des Quantencomputers und eine Reflexion über die geleistete Arbeit. Nützliche Quantenchemie auf Quantencomputern zu berechnen, ist ein schwieriges Unterfangen und erfordert vielleicht fehlertolerante Maschinen und clevere Algorithmen, aber es gibt dennoch vielversprechende Versuche, das Problem mit lauten Maschinen zu lösen.

Declaration of Originality

I, Daniel Bultrini, declare that this thesis and the work presented in it are my own original writing. There is a large section introducing fundamental concepts in quantum computing, which are textbook knowledge in quantum information and linear algebra. Although these have been compiled from various sources and cited as necessary, the foundational knowledge is mostly based on my reading of Nielsen and Chuang's Quantum Information and Quantum Computation [1], as well as Sheldon Axler's Linear Algebra Done Right [2], but many notes I have taken over the years can come from sources I could not find. The various research presented in this work is either my own or the result of a collaboration in which I have been a first author unless otherwise stated, and has not been submitted by myself for any other degree or professional qualification. Where I have consulted the work of others, this is always cited, and for my original research, the contributions of others are always acknowledged at the start of the chapter.

Figures are either created by myself or are used with permission under the Creative Commons license, which will be denoted as CCX, where X is the version number, in general this will be CC4. Whenever a figure that is not original is used, it will have a citation to the original work, which in most cases will be one of my papers written throughout the course of the doctoral program.

Notes on Notation

Throughout the text, the following notation is commonly used. Terms and uncommonly used symbols are defined as they are introduced throughout the text. The convention that all constants are set to unity holds, in chemistry this is often termed atomic units. Equations are referred to numbers in brackets, for example (equation number).

Mathematical Notation

Table 1: Common mathematical notation.
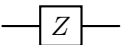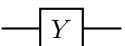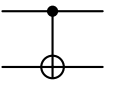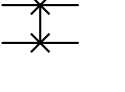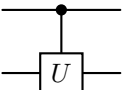
Notation	Definition
$(\cdot)^*$	Conjugate of $(\cdot)$.
$\lvert\psi\rangle$, $\langle\psi\rvert$	A vector and its dual, known as 'ket' and 'bra'.
A	A matrix or operator A.
$\langle\psi\vert\phi\rangle$	Inner product between $\lvert\psi\rangle$ and $\lvert\phi\rangle$
$\lvert\psi\rangle\otimes\lvert\phi\rangle$	Tensor product of two vectors.
$\lvert\psi\rangle\lvert\phi\rangle$, $\lvert\psi\phi\rangle$	Abbreviations for the above tensor product.
$(\cdot)^{\mathrm{T}}$	Transpose of $(\cdot)$.
$(\cdot)^\dagger$	Hermitian conjugate or Adjoint of $(\cdot)$, where $(\cdot)^\dagger = ((\cdot)^{\mathrm{T}})^*$.
$\lVert\psi\rVert$	Norm of a vector $\lvert\psi\rangle$.

Quantum Gates

There will also be several quantum circuits shown throughout the work. Here are several commonly used gates, with a description on deciphering quantum gates given in section 2.4.1. It should be noted that the matrix to diagram mapping is dependent on the ordering of the qubits, which can be little endian or big endian ordering. This work tends to use big endian ordering, but software such as Qiskit use little endian ordering. In practice as long as one is consistent it does not matter which is used, and the one must only

reverse the order of the output bits to transform the output of one circuit to the other ordering. Qubits are also normally initialized in the $|0\rangle^{\otimes n}$ state, where $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Table 2 is a reference to circuit diagrams.

Table 2: Common qauntum gates.

Notation	Matrix Representation	Notes
$|\psi\rangle \,\diagup^{n}$	-	n qubits initialized at $|\psi\rangle$, single lines represent single qubits.
$-\boxed{U}-$	$\begin{bmatrix} \cos(\frac{\alpha}{2}) & -e^{i\gamma}\sin(\frac{\alpha}{2}) \\ e^{i\beta}\sin(\frac{\alpha}{2}) & e^{i(\beta+\gamma)}\cos(\frac{\alpha}{2}) \end{bmatrix}$	Arbitrary single qubit gate U parameterized by $\alpha, \beta, \gamma \in \mathcal{R}$.
$-\boxed{H}-$	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$	**Hadamard gate**, acts as $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.
$-\boxed{X}-$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	**Pauli-X** gate or a **bit flip** gate since $X|0\rangle = |1\rangle$.
$-\boxed{Z}-$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	**Pauli-Z** gate or **phase flip** gate since $Z|1\rangle = -|1\rangle$
$-\boxed{Y}-$	$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	**Pauli-Y** gate, referred to as a bit and phase flip gate since $Y|0\rangle = i|1\rangle$.
CNOT	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	**CNOT** or **CX** gate, conditional flip from the control qubit at $\bullet$ to the target qubit at $\oplus$
SWAP	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	**Swap gate** which swaps two qubit states such that SWAP $|ab\rangle = |ba\rangle$, equivalent to 'crossing the wires'.
Controlled-U	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & & \\ 0 & 0 & & U \end{bmatrix}$	**Controlled-U gate**, like a CNOT but with any unitary matrix.
$\diagup^{n}\!-\boxed{U}$	-	Arbitrary n qubit unitary gate U.
$-\boxed{\measuredangle}$	-	Measurement of a qubit in the computational basis unless otherwise specified.

Part I

Introduction and Theory

1
Why Quantum Computing?

As hangs the flexible line, so but inverted will stand the rigid arch.
Description of Helioscopes, Robert Hooke, 1675

THE FIRST KNOWN COMPUTATIONAL DEVICE is arguably the
ancient Greek Antikythera mechanism, which through a complicated
gearing system, could predict astronomical features far in the future.
Many tools to perform and aid arithmetic and geometric compu-
tations have been invented since, but there is another approach to
solving complicated systems of equations that is perhaps somewhat
more curious. This is using another physical system with a partic-
ular physics governing it to simulate another less accessible system.
Although the differential gearing systems of these old machines are
in a sense an analog computer, just a slide rule might be considered
one, but a more interesting example for us is the hanging chain. The
shape of a hanging chain, which was 'programmable' via hanging
weights on it is the same as the optimal shape of an arch that would
have to support those weights. This was formalized by Hooke, and
is among the first examples of a physical system that can be used to
simulate another, and it has been used to design buildings as grand
as Gaudi's Sagrada Familia in Barcelona.

Before the advent of the powerful computers that we have now,
exact arithmetic computations were the singular domain of complex
mechanical calculators, but they could only compute with a limited
number of digits and with their slow manual operation, meant that
they would mostly be used for accounting. As such, more complex
system had to be modeled again by differently complex mechanisms.
Continuing in the tradition of the Antikythera mechanism, differ-
ential analyzers were dynamical mechanisms that could integrate
certain classes differential equations, with their parameters scaled
such that they could fit within the allowed ranges of motion. With
the advent of electronics, it was also possible to use an expanded
vocabulary of physics to simulate ever more complex systems, such
as alternating current power network analyzers to simulate energy
delivery networks and problems in nuclear physics. Of course, a vast
number of analogue computers were built and in use until the 1970s,
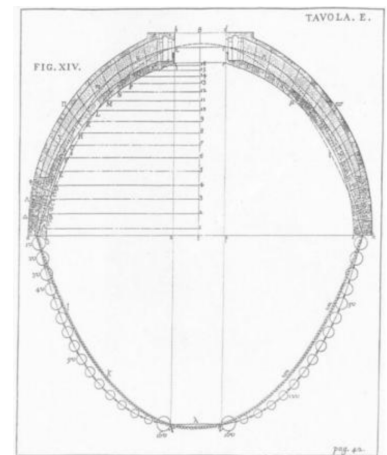
Figure 1.1: **Dome of St. Pe-
ter's in rome** over a 'simula-
tion' of its shape by a hanging
chain. (Poleni, 1748).

when the digital computer became the dominant computational tool.

QUANTUM THEORY is thought to underlay all physical phenomena, although at the moment there isn't a quantum theory of gravity. Both Manin, in 1980 [3] and Feynman in 1981 [4] independently came to the conclusion that to simulate quantum mechanical phenomena one must use quantum mechanical systems. This has lead to a new field of research with one ultimate goal: creating the universal quantum simulator. Such a machine, also called a quantum computer (QC), is able to operate on a state space that scales exponentially with respect to the number of informational units, of which two level quantum systems called *qubits* are the most commonly used. A perfect machine would be able to explore a state space larger than our most powerful supercomputers with only 100 qubits. In reality there are many issues plaguing the technology from being a viable at this moment, but there are good reasons to research alternative computational approaches.

The necessity of computation for scientific progress is undeniable, and it would be a disservice to the reader to describe how essential simulations, finding optimal solutions and processing data is. It follows naturally that being able to compute ever larger simulations and to make sense of ever-growing datasets is desirable. Our everyday binary digital computers, which in the quantum computing literature are referred to as *classical computers*, have been extremely successful in all of these areas. But they are reaching fundamental limitations in growth, Moore's law [5] , which has guided the development of processors and planning of what computers can be expected to do in future is no longer as valid [6]. This has lead to a revolution of alternative computing strategies, including massive parallelization [7], the development and implementation of advanced coprocessors such as the graphical processing unit (GPU) to scientific problems [8] and more recently, so-called tensor cores have been added to GPUs, alongside other specialized hardware that are very often specifically tailored to running machine learning algorithms [9]. These approaches are all transistor chip based, so do not bypass the energy cost of ever larger computers and the physical limits of fitting more transistors on a chip are a likely unsurmountable barrier to this continued growth in computational power [6].

There has been an explosion of alternative computing strategies beyond the classical paradigm which aim to augment various computational tasks. Although not an exhaustive list, a very promising candidate is neuromorphic computing, which aims to generate analog circuits that represent neural networks, which trade machine precision (which is not essential for all machine learning tasks [10]) for large improvements in power efficiency [11]. There are many more physics based computational strategies, such as thermodynamic computing [12], which has been shown to be able to find good approximations of matrix inverses [13], but the scaling of these approaches is yet to be proven. These techniques all try to remove some of the

Note: Moore's law is not a physical law.

more expensive computational burdens from the main processor, much like a GPU, and aim to achieve at least one of two things: reduce power consumption and increase performance on the specific computations they tackle. But the trend in GPUs is to use lower precision arithmetic, which may be detrimental to non-ML scientific computing [14].

If we remain in the world of classical computation, we are limited to binary operations. By definition, the type of problems that are solvable in polynomial time via binary operations are within the complexity class *polynomial-time* (P). Problems of higher complexity can be tackled, but either the problem size must remain within the computational constraints or heuristic techniques and various approximations have to be made. It is definitely not out of the question that these approximations can be good enough for all practical purposes due to the availability and ease of use of classical hardware.

QCs differ from the classical methods in that the complexity of the problems solvable with quantum computations is expected to be higher than those of classical computers. The argument stems from the fact that there are different classes of complexity for a problem to be in. QCs can efficiently solve (BQP) [1]. This means that a quantum computer can simulate a classical computer efficiently, while the reverse is not believed to be true.

The way quantum computers are envisioned to be used is to tackle a set of problems that quickly become intractable to classical computers. In that sense, the quantum computer will likely be used as a coprocessor rather than a normal computer. Indeed, it is impossible to construct a fully programmable quantum computer [15]. This is because encoding two different unitaries in quantum memory requires two different orthogonal program states, since there are infinite possible unitaries. Even a single qubit "quantum computer" would require an infinite number of addresses for the infinite number of rotations that are possible. This may sound like a death-knell, but one can reach arbitrarily precise approximations to any desired operation with a non-complete set of operations. There is also probabilistic computing that utilizes probabilistic bits for advantages in certain algorithms [16], but this is more niche - and can be thought of as the stage between a classical bit and a quantum bit.

TO ANSWER why quantum computers are promising then is an easy question. They are promising because they should be able to deal with problems with favorable asymptotic scaling since the addition of a single qubit doubles the computational space. This is not the case with classical computing, and although the computational space is one part, algorithmically there are proven primitives that ensure that this is indeed the case given the current understanding of complexity.

On the other hand, there is no reason to think that insights about the structure of problems and better approximate techniques cannot keep up with advances in quantum algorithms. Furthermore, at
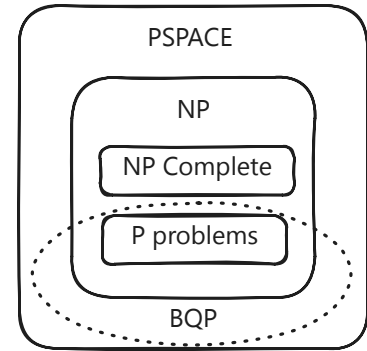
Figure 1.2: **A cartoon of the complexity classes** for relevant problems in computing. PSPACE is the set of decision problems whose input and computation can be solved by a Turing machine using polynomial space. NP are those problems that are not solvable in polynomial time but whose answer can be verified in polynomial time, while P are those that can be solved within polynomial time. NP complete problems are a subclass of NP problems whose solutions can solve all problems in NP. BQP is bounded-error quantum polynomial time and is where quantum Turing machines are thought to exist in - where decision problems are solved in polynomial time with an error probability of at most $\frac{1}{3}$.

the moment and in the foreseeable future quantum computers also suffer from errors that cannot easily be dealt with, and as such the idea that an error-prone machine can give useful results is on shaky grounds.

Whether the immense effort put into quantum computing would better be spent elsewhere is another question entirely, and this thesis hopes to shed light on the difficulties and considerations in developing algorithms for near term machines, including the immense work required to mitigate errors and the compromises that have to be made when running algorithms on actual quantum hardware.

1.1 Why Chemistry?

THEORETICAL CHEMISTRY and the study of quantum systems in general is one of the most promising uses of the technology. The original raison d'être of quantum computing were based on the idea that to truly simulate a quantum system one would need an analogous quantum system, but the relationship goes deeper than that. Experimental quantum computing actually has a very close connection to chemistry, specifically to chemical physics and nuclear magnetic resonance (NMR) spectroscopy. The first ever quantum computation was performed on a NMR machine by Somaroo et al. [17] for quantum harmonic oscillators, with many other early experiments being performed on these machines [18, 19].

Theoretical chemistry is a mature field of its own with numerous techniques, both rigorous and approximate, which have notable success in the areas of studying electronic configurations, structural optimization and dynamics of both the electronic and nuclei of various molecules. *Quantum chemistry* is the term that is more associated with electronic structure while *theoretical chemistry* encompasses all aspects of the theoretical description of molecules, which may be classical models as well. The degree of accuracy that can be achieved is in general inversely proportional to the size, so a smaller system can be studied very accurately, while larger systems must make use of various approximations.

The Hilbert space for quantum systems scales exponentially with the number of particles being considered, but it is clear that the Hilbert space is unlikely to be densely occupied. Any molecule is only able to act within the confines of physical laws and this already cuts down on both the amount of information that is theoretically required to be stored. If one were to have access to the manifold of possible configurations a-priori and an efficient encoding it might be that being able to compute on the full Hilbert space is unnecessary. This is not the case for arbitrary molecules at the moment and hence being able to do so is thought to be of great benefit.

Perhaps one extremely successful theory which somewhat bypasses the complexity of accounting for the whole of the Hilbert space for electronic problems it density functional theory (DFT) [20, 21] which

integrates out all degrees of freedom apart from the electron density. This is certainly correct if not for the fact that the quality of the final solution depends on the exchange-correlation functional whose exact form is unknown. Of course, entire books have been written about quantum chemistry and its many techniques for electronic structure.

Within theoretical chemistry one must also be able to describe and compute the time evolution of chemical systems, which is the domain of dynamics. There is the possibility of studying dynamics solely within the electronic subsystem, but what is of particular interest here is the nuclear dynamics. The latter problem is much more complicated as one must keep a description of the full nuclear and electronic systems and their combination. Approximations must be made - some are trajectory based, offloading some of the complexity of storing a full nuclear wavefunction on top of electronic ones by statistically sampling many classical nuclear paths, or the electronic potential energy surfaces can be precomputed, a step limited by the quality of electronic structure computations, upon which the nuclear subsystem is propagated as in (Multi-Layer) Multi-configurational time-dependent Hartree [22, 23]. This is extremely computationally taxing, so a simplification of the full nuclear dynamics can also be made, which is that of using classical trajectories for the nuclei. This is the domain of mixed quantum-classical dynamics, which is the focus of this thesis.

It can be broadly stated that all classical techniques try and avoid two things: maintaining the full wavefunction in memory and computing and performing time evolution by matrix vector multiplication. These things are natural to do on a quantum computer, since the first representation, although inefficient for classical computers, is natural on quantum computers, and the operations required to simulate time-(in)dependent Hamiltonians exists. This is why it is believed that there is a great chance to use these machines for quantum chemistry.

At the same time, exact techniques for quantum computing are not a panacea. A common benchmark is the molecule FeMoCo which can require at least around $5 \cdot 10^{12}$ expensive fault-tolerant operations (T-gates) to prepare its wavefunction [24]. The number of fault-tolerant gates is not a good proxy, since gate times vary between quantum computers and the specific error correction code. Assuming an active space of 76 orbitals, a high quality superconducting qubit quantum computer would require on the order of 10^6 qubits and 3 years to compute this including the overheads of error correction [25]. As such the current state of the art generally attempts to run algorithms on either small or model systems, as is done here with the Shin-Metiu model [26], which is a minimal model that captures many interesting phenomena in dynamics.

1.2   Why this Thesis?

It is no secret that quantum computers are not ready to be picked up and used as they currently are. For quantum chemistry especially the algorithms which give guaranteed results, which is to say, either probabilistic or deterministic algorithms or variational algorithms that are guaranteed to be able to express correct solutions, require huge numbers of near error free qubits and operations to yield good results. At the moment there are no qubits that are of high enough quality or the operations near the quality required for useful quantum computation.

But is that truly the case? Perhaps there is some way that current machines, despite their limitations can be used for simulating interesting models and problems in chemistry. Or at least, algorithms that could eventually do this can be promised. Although quantum computing is arguably at a stage where the aim is to see if classical techniques can be at least matched with existing machines. If so how can this be done, and is it efficient?

This thesis aims to showcase the entire process of designing an algorithm for chemical dynamics to running it on a quantum computer. This includes all the considerations and technique development required to cope with the errors and noise present on today's quantum computers. The research presented then consists of the design of an algorithm to perform dynamics in first quantization for mixed quantum-classical systems, combining disparate error mitigation techniques together to make one better than the sum of its parts, and analyzing the performance of quantum computers as they become more fault-tolerant. Finally, the process of moving a theoretical algorithm to run on an actual quantum computer is presented.

Hopefully this thesis will serve as a marker for how research in this field is done during the in between time before fault-tolerant machines take over and before a clear pathway to quantum advantage with noisy machines has been found, if it is indeed possible.

2
Quantum Information Processing

> Now what is the meaning of the terms "uncertain," "statistics," "probability"?
> *Die gegenwartige Situation in der Quantenmechanik, Schrödinger, 1935*

This chapter begins with a brief non-pedagogical overview of the mathematical tools and notation that will be used throughout the work, it then introduces the specific constructions and conventions that are used in quantum computing that will be used to build the algorithms and error mitigation strategies in the part II of the thesis.

QUANTUM ADVANTAGE, unlike spherical cows, cannot be defined in a vacuum, and thus a baseline must be drawn. The well known computer which from now on will be delegated to the name *classical computer* will be the natural comparison. A classical computer works by processing binary data that can have two states, generally denoted 0 and 1. Operations on classical bits are performed by logic gates that are dictated by truth tables, they take in two bits and output a single bit. It turns out that this single operation if done nearly countless times in very specific patterns is enough to perform any boolean operation of any inputs. Sets of operations with this property are known as *universal gate sets*, in this case, the minimal universal gateset for classical computing contains a single operation.

Now, a quantum computer works under slightly different rules, a qubit does have states 0 and 1, but it can have any superposition of these two states. If it could only be a real superposition it would be a probabilistic bit (pbit), and it would be representable as a point on a line, but since it can have complex values a qubit it can be represented as points on a sphere (known as the Bloch sphere). Operations on quantum computers are different to the classical computing we are used to, as qubit operations take in as many qubits as they output. There is a classical computing paradigm that also has this feature, which is called reversible computing, which takes in 3 bits and outputs 3 bits in its minimal universal form via a so-called Toffoli gate.

Quantum computing performs operations through rotations of the state on this sphere and interactions between two spheres, which,

The NAND gate, or the not AND gate is defined by the following truth table with inputs A and B:

A	B	OUT
0	0	1
0	1	1
1	0	1
1	1	0

as will be shown, is universal. Now, reversibility does not imply a larger computational power, as a reversible classical computer is as powerful as a non-reversible one, but this matter of operating on spheres holds a bit more promise. It turns out that alone, this is also not quite enough for a computational advantage, and to truly explain the differences there are some formalisms to dig through.

2.1 Mathematical Preliminaries

The study of quantum mechanics is the study of finite dimensional complex vector spaces with a defined inner product. This is called either a Hilbert space or an inner product space, which are equivalent in finite dimensions. In the following definitions and preliminaries, we restrict ourselves to a complex $(\mathcal{C})$ vector spaces.

Definition 1 (Finite dimensional complex vector space). *A finite dimensional complex vector space V is a vector space over $\mathcal{C}$ with a basis of size $dim(V)$ along with addition and scalar multiplication on V.*

Addition is commutative,

$$a + b = b + a \ and \ ab = ba \ \forall \ a, b \ \in V.$$

Addition and multiplication are associative,

$$u + (v + w) = (u + v) + w \ and \ (ab)v = a(bv) \ \forall \ a, b \in \mathcal{C} \ and \ \forall \ u, v, w \in V.$$

There exist additive and multiplicative identities,

$$\exists \ 0, 1 \in V \ s.t. \ v + 0 = v \ \forall \ v \in V \ and \ 1v = v \ \forall \ v \in V.$$

There exists an additive inverse such that

$$\forall w \in V \ \exists! \ v \in V \ s.t. \ w + v = 0.$$

Finally, there are the following multiplicative and distributive properties,

$$a(u + v) = au + av \ and \ (a + b)v = av + bv \ \forall \ a, b \in \mathcal{C} \ and \ \forall \ u, v \in V.$$

Definition 2 (Inner product space). *A vector space V with a defined function $\langle \cdot | \cdot \rangle : V \times V \to \mathcal{C}$ called the inner product that satisfies the following properties:*

Positive definite,

$$\langle v | v \rangle \geqslant 0 \ and \ \langle v | v \rangle = 0 \iff v = 0$$

.

Additive in the first entry,

$$\langle u + v | w \rangle = \langle u | w \rangle + \langle v | w \rangle \ \forall \ u, v, w \in V.$$

Homogeneous in the first entry,

$$\langle av | w \rangle = a \langle v | w \rangle \ \forall \ v, w \in V \ and \ \forall \ a \in \mathcal{C}.$$

Conjugate symmetric,

$$\langle v|w \rangle = (\langle w|v \rangle)^*,$$

is called an inner product space or a finite dimensional Hilbert space. The distinction between the two is only relevant in infinite dimensions.

Definition 3 (Basis). *A basis for a vector space V is a set of linearly independent vectors that span V. A basis is denoted by $\{|v_i\rangle\}_{i=1}^{\dim(V)}$ and any vector $w \in V$ can be described by a list of tuples of coefficients $a_i \in \mathcal{C}$ and $|v_i\rangle$ as*

$$w = \sum_{i=1}^{\dim(V)} a_i |v_i\rangle.$$

In quantum computing the *computational basis* is very often used and as it is the natural measurement basis for qubits. It is defined by the eigenvalues of the Z Pauli matrix.

Theorem 1 (Gram-Schmidt). *If $\{\,|v_1\rangle, \ldots, |v_{\dim V}\rangle\,\}$ is a linearly independent set of vectors in a vector space V, then we can always compute an orthonormal basis $\{\,|e_1\rangle, \ldots, |e_{dimV}\rangle\,\}$ inductively by*

$$|e_1\rangle = \frac{|v_1\rangle}{\||v_1\rangle\|},$$

$$|e_i\rangle = \frac{|v_i\rangle - \sum_{j=1}^{i-1} \langle e_j|v_i\rangle |e_j\rangle}{\left\| |v_i\rangle - \sum_{j=1}^{i-1} \langle e_j|v_i\rangle |e_j\rangle \right\|}.$$

The Gram-Schmidt theorem ensures that we are always able to form an orthonormal basis given any linearly independent basis, and as such whenever a basis is mentioned, it will always mean an orthonormal basis as it is guaranteed that it can be generated.

Definition 4 (Outer Product). *Given two vectors $|v\rangle$ and $|w\rangle$ in a vector space V, the outer product is defined as the operator $|v\rangle \langle w|$ which acts on a vector $|u\rangle$ as $|v\rangle \langle w| |u\rangle = \langle w|u\rangle |v\rangle$. For complex vectors it can be thought as matrix multiplication with the dual as*

$$|v\rangle \langle w| = |v\rangle (|w\rangle^T)^*.$$

The outer product is particularly important as it is often used to describe the measurement in a basis, acting as a projector. It is of course also possible to describe many linear operators as a sum of outer products, known as the outer product representation.

Theorem 2 (Completeness Relation). *For any orthonormal basis $\{\,|e_i\rangle\,\}$ of a vector space V, the completeness relation states that*

$$\sum_{i=1}^{dimV} |e_i\rangle \langle e_i| = \mathbb{I}, \tag{2.1}$$

where $\mathbb{I}$ is the identity operator. An operator that satisfies this is said to be complete.

This is easily seen by remembering that if there is a vector $|v\rangle \in V$ which can be written as $|v\rangle = \sum_{i=1}^{\dim V} |e_i\rangle \langle e_i|v\rangle = \sum_{i=1}^{\dim V} v_i |e_i\rangle$, this allows any operator to be written as a sum of outer products for some map $M : V \to W$, $M = \mathbb{I}_W M \mathbb{I}_V$, where $\mathbb{I}_W$ is the identity operation of $\mathbb{I}$ on the vector space W. With these facts then the projector can be defined, which is the type of measurement most used in this work.

Definition 5 (Projector). *A projector P onto some subspace W of dimension k of a vector space V of dimension d such that $k \leqslant d$ is*

$$P = \sum_{i=1}^{k} |e_i\rangle \langle e_i|,$$

where $\{\,|e_i\rangle\,\}_i^k$ is an orthonormal basis of W. By the Gram-Schmidt procedure it is always possible to construct a basis for V $\{\,|e_i\rangle\,\}_i^d$ that contains the basis for W.

The projector is *Hermitian*, which is to say $P = P^\dagger$, and *idempotent*, $P^2 = P$. This means that the projector is a linear operator that when applied twice gives the same result as applying it once, it makes sense that if a projector maps something to a subspace, then mapping the subspace to itself should not alter the state.

The next set of mathematical preliminaries are less formally presented, but are referred to often. These will be presented in the context of quantum operations so that they can be more easily applied.

Definition 6 (Trace). *The trace of an operator A can be defined as the sum of the diagonal elements of the matrix representation of A in some basis. The trace is denoted*

$$Tr(A) = \sum_i A_{ii}.$$

If the trace is invariant under the action of an operator $\hat{O}$, that is to say $Tr(\hat{O}A) = Tr(A)$, that operator is said to be trace-preserving.

The trace map is in essence the trace multiplied by some outer product of a basis vector in another space, and so identical to the trace function. This makes the trace a quantum operation, but allows for the definition of the partial trace, which is effectively what happens when a subset of qubits of a larger system is measured,

Definition 7 (Partial Trace). *The partial trace of a density operator ρ over a subsystem B with basis $\{\,|b_i\rangle\,\}_i^{dim(B)}$ of a composite system AB is defined as a function takes some $H_{AB} \to H_A$ as*

$$Tr_B(\rho) = \sum_i^{dim(B)} (\mathbb{I}_a \otimes \langle b_i|)\rho(\mathbb{I}_a \otimes |b_i\rangle),$$

for some orthonormal basis of H_B $|b_j\rangle$ for $j = 1, \dots, dim(B)$.

The partial trace is extremely important in the study of noise and colloquially is called *tracing out* a subsystem. There are many definitions and related definitions of both the trace and partial trace, such as the trace map, but these are not used in this work.

2.2   Fundamental Postulates

Historically quantum mechanics was not discovered linearly, unlike
the theory, and although its development is fascinating, with a very
human account of those times being given by today we have four
postulates that encompass the enormity of the field elegantly, there
is no meaning to the order these are given in. These postulates are
presented in both a pure state and density operator formalism, as
both are used frequently in this work.

Postulate 1 (State Space)**.** Isolated systems (also known as closed
systems) have an associated complex vector space with an inner
product, termed a Hilbert Space, which is the system's state space.
A particular configuration of a system is completely described by a
state vector $|\psi\rangle$, which is a unit vector in this space, which we will
refer to as the "state". Equivalently there is a positive operator ρ
that describes the system, with the additional information that if the
system is in state ρ_i with probability p_i, then the density operator
describing this is $\sum_i p_i \rho_i$.

Quantum mechanics doesn't tell us what this state space or asso-
ciated vector is for a given system. This is then a problem tackled
individually for any given system, be it interacting qubits as in this
work, or quantum electrodynamics in the study of light-atom inter-
actions.

Postulate 2 (Unitary Evolution)**.** A system like those described
in postulate 1 can only change through a unitary transformation.
Although time is not essential to describe this postulate, it is helpful
to say that at time t_1 we have a state $|\psi\rangle$, and it is related to $|\psi'\rangle$ at
time t_2 by some operator $U(t_1, t_2)$ by

$$|\psi'\rangle = U(t_1, t_2) |\psi\rangle \,,$$

and in the density operator formalism,

$$\rho' = U(t_1, t_2)\rho U(t_1, t_2)^\dagger$$

Quantum mechanics imposes no restrictions beyond this in general,
and for qubits we allow ourselves to use any unitary operator, but
this is not the case for all systems. It should be added that the time
evolution of a state is always governed by the Schrödinger equation,
the equivalent expression for density operators is the Liouville-Von
Neumann equation, which famously are:

$$i\hbar\frac{\mathrm{d}\,|\psi\rangle}{\mathrm{d}t} = \hat{H}\,|\psi\rangle \,, \tag{2.2}$$

$$i\hbar\frac{\mathrm{d}\rho}{\mathrm{d}t} = [\hat{H}, \rho], \tag{2.3}$$

where $\hat{H}$ is a special fixed Hermitian operator known as the Hamilto-
nian of the system and $\hbar$ is the reduced Planck's constant, which in
practice is set to 1 and never thought about again.

Postulate 3 (Measurement). Measurement of a quantum state is described by some set of measurement operators $\{\, M_m \,\}$ that act on the state space of the system and have potential outcomes indexed by m. The measurement operator is complete $\left(\sum_m M_m^\dagger M_m = I\right)$ If the state was $|\psi\rangle$ before measurement, then the probability $p(\cdot)$ that m occurred is

$$p(m) = \langle\psi|\, M_m^\dagger M_m \,|\psi\rangle\,,$$

and the state post measurement becomes

$$\frac{M_m \,|\psi\rangle}{\sqrt{\langle\psi|\, M_m^\dagger M_m \,|\psi\rangle}},$$

as a consequence of completeness, $\sum_m p(m) = 1$.

For density operators this is instead

$$p(m) = \mathrm{Tr}(M_m^\dagger M_m \rho), \tag{2.4}$$

with the post measurement system becoming

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}(M_m^\dagger M_m \rho)}. \tag{2.5}$$

Postulate 4 (Composite Systems). Distinct quantum systems form a state space that is the tensor product of the individual state spaces, this is also the case for the density operator formalism. With respect to the number of qubits N the number of elements of the statevector scales as 2^N and those of the density matrix scale as 2^{2N}.

2.3 The Qubit

In classical computing the bit is the minimal unit of information, which takes one of two states, denoted 0 and 1. The quantum information equivalent is called the qubit, and is represented by a complex two-dimensional vector $|\psi\rangle \in \mathbb{C}^2$. We choose a canonical basis with which to represent this vector, and it makes sense that this will be the same as the one in which the quantum computer measures the qubits. This happens to be the Z-basis, often called the computational basis, and borrowing from the classical representation, we denote its two eigenstates $|0\rangle$ and $|1\rangle$. The qubit state can be expressed in the computational basis by applying the projector $|0\rangle\langle 0| + |1\rangle\langle 1|$

$$|\psi\rangle = (|0\rangle\langle 0| + |1\rangle\langle 1|)\,|\psi\rangle\,,$$
$$|\psi\rangle = \underbrace{\langle 0|\psi\rangle}_{c_0}\,|0\rangle + \underbrace{\langle 1|\psi\rangle}_{c_1}\,|1\rangle\,,$$
$$|\psi\rangle = c_0\,|0\rangle + c_1\,|1\rangle\,,$$
$$|\psi\rangle = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}.$$

Where c_0 and c_1 are complex coefficients with the property that $|c_0|^2 + |c_1|^2 = 1$, such that $|\psi\rangle$ is normalized. This allows the qubit

state to also be written (to a global phase factor) as

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)|1\rangle. \qquad (2.6)$$

When measuring the qubit we do not have access to the coefficients, but only that we measure 0 with probability $|c_0|^2$ or that we measure 1 $|c_1|^2$ of the time. Measurement is described in more detail in section 2.2.

This "unobservability" of the quantum state makes it quite different from the usual models of computation, where there is a very direct link between theory and implementation. Of course, it is possible to transform the state of the qubit such that a sequence of transformations and measurements leads to a complete characterization of the qubit state. To do this we must define some allowable operations that can be performed on the qubit, but in theory we must simply measure in different bases to extract the complex coefficients.

All the density operators of qubit states can be represented by density matrices built from the orthogonal basis that is a combination of the Identity and the three Pauli matrices:

$$\sigma_0 := \sigma_I := \mathbb{I} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 := \sigma_X := X := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\sigma_2 := \sigma_Y := Y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \quad \sigma_3 := \sigma_Z := Z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
$$(2.7)$$

The Pauli matrices decompose the qubit density matrix as

$$\rho = \frac{1}{2}(\sigma_I + a\sigma_X + b\sigma_Y + c\sigma_Z), \qquad (2.8)$$

and the state can be written as a vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ with $a^2 + b^2 + c^2 \leqslant$ 1. This is the Bloch vector and is constrained to lie within a unit sphere. When $a^2 + b^2 + c^2 = 1$, the qubit is a pure state and lies on the sphere's surface. This representation is called the Bloch sphere and is shown in fig. 2.1.

When the vector does not lie on the surface of the Bloch sphere, it can be thought of as a weighted sum of pure states. This is the most general representation of a quantum state, and are known as mixed states, which are described by the density matrix ρ defined in (2.8). Each pure state is distinguishable and labelled by i as $|\psi\rangle_i$ and has an associated probability P_i. If the states were not distinguishable, they could add coherently yielding a pure state again.

A single qubit is more useful than one might expect at first glance, and with clever algorithms, it can be made to act as a universal classifier [27, 28] or to approximate bounded complex functions [29]. As interesting as these results are, it is desirable to have more than a single qubit for more general computations.
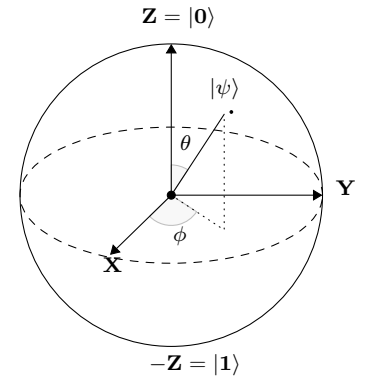
Figure 2.1: **The Bloch sphere**, a representation of a qubit showing and arbitrary state $|\psi\rangle$ and the axes defined by the eigenvectors of the Pauli matrices in eq. 2.7.

2.3.1 Multiple Qubits

The real power of quantum computation comes from being able to use multiple qubits. For a student of theoretical chemistry or physics, it comes as no surprise that as the number, N, of qubits is increased, the dimension, dim of the Hilbert space spanned by the basis generated by the constituent subsystems grows as 2^N. The composition of multiple systems is simply the Kronecker tensor product of the individual systems, as described in Postulate 4 in section 2.2. In the standard initial state of two qubits we write

$$|0\rangle \otimes |0\rangle \equiv |00\rangle \qquad (2.9)$$

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad (2.10)$$

and will use the shorthands $|0\rangle^{\otimes n}$ or $|0_n\rangle$. Of course, it's not unusual to see other common qubit initializations such as $|1\rangle^{\otimes n}$, or the Hadamard basis $|+\rangle$, $|-\rangle$. Usually alternative formulations are defined in the literature when used.

Multiple qubits then can have any superposition of their basis states. This is to say that for N qubits, which can be written as a sum over 2^N of the basis states, each with its own complex coefficient a_i as

$$|\psi\rangle = \sum_i = 0^{2^N-1} a_i |i\rangle, \qquad (2.11)$$

where $|i\rangle$ is the binary representation of the integer i. The fact that a quantum computer can act on this vector of all combinations of inputs is sometimes called quantum parallelism, as one can think of it as acting on all possible inputs at once. The reality is perhaps a little more complex than this, but it can be said with certainty that in some way N qubits can store exponentially more information than N bits, since N bits can only store a single bit string of length N rather than some arbitrary combination of all of them.

This is also when all the interesting aspects of quantum mechanics come into play, as although a single qubit is a curious object on its own, it is only when at least two systems come together, as the full breadth of what makes quantum information interesting begins to reveal itself. This is the world of entanglement, which is excellently defined for two systems - and although it is an essential quantum resource, it will not be discussed in detail. When moving to many qubits, entanglement is less well-defined, and there is a measure for the "complexity" of the entanglement termed *magic* which is much more complicated to measure [30].

2.4 Operations on qubits

On a QC, the qubits are usually *initialized* in the $|0\rangle$ state and then manipulated by a sequence of quantum gates. These gates are unitary operators that act on the qubit state, some important examples

are the Pauli gates, the Hadamard gate, the single qubit rotation gate and the CNOT gate. The Pauli matrices are the most basic quantum gates and are defined in (2.7). The Hadamard gate is a single qubit gate that creates superpositions of the computational basis states, and is defined as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{2.12}$$

$$UU^{\dagger} = \mathbb{I} \tag{2.13}$$

Any operation can be composed to act on a larger qubit space by taking the tensor product of itself with other operators until the dimensionality is equal to that defined by the number of qubits. Formally this means that to apply H to the second qubit in a 3 qubit system one must perform the operation

$$(\mathbb{I} \otimes H \otimes \mathbb{I}) |000\rangle,$$

this can be shortened to $H_1 |000\rangle$.

Now, the above string of operations is also shortened to a string like $\mathbb{I}H\mathbb{I}$. Normally the identity operator isn't written out, but this string becomes very useful. If one consults (2.7), there will be entries for Pauli matrices. There is a very important type of operator string called the Pauli string, which for N qubits is usually denoted

$$P = \sigma_{i_1} \otimes \sigma_{i_2} \otimes \cdots \otimes \sigma_{i_N}, \tag{2.14}$$

where i_j for $j = 1, \ldots, N$ is an integer from 0 to 3, which are analogous to the more commonly used $\mathbb{I}, X, Y, Z$. These Pauli strings are often parts of weighted sums which represent operators. All Hermitian operators can be decomposed into weighted sums of Pauli strings [1].

There are operations that act on more than a single qubit, and although in hardware there might be physical limitations as to which two qubits can directly interact, in theory any two systems could interact. As such they are almost always written as a labelled operation between two of the systems (in this case hypothetical qubits 0 and 1) as

$$\mathrm{CNOT}_{01} |000\rangle.$$

Unitaries can act on an arbitrarily large number of qubits, but this can quickly get unwieldy, so it is much more common to see the quantum circuit diagram to represent these operations.

2.4.1   Quantum circuit diagrams

Quantum circuit diagrams are a visual representation of the many operations that take place over the course of a quantum program. They help to visualize the complexity of a quantum program at a glance. They are not the only representation nor the most useful, indeed there is also the directed acyclic graph representation that is much more useful in practice when programming. Nonetheless,
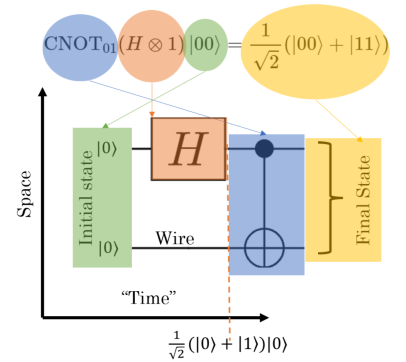
Figure 2.2: **Reading a quantum circuit** The circuit is read from left to right, with the qubits being the horizontal lines and the operations being the boxes. The order of operations is the opposite to linear algebra, and the mapping is shown in this cartoon.

circuit diagrams are the most human-friendly and readable representations of quantum programs, and are essential to understanding the literature. A short example circuit which produces a Bell state from an initial $|00\rangle$ state is shown in figure 2.2. Different qubits are represented by wires that start at the leftmost side at some initial state and the state 'travels' rightward where it is acted on by labelled boxes called gates. Gates are then just the (usually) unitary operations acting on the qubits over which the box is drawn.

Many commonly used gates have their own letters or symbols (such as the CNOT as shown), and may also represent entire subprocesses and subalgorithms. Because of this some multi-qubit gates do not act on all the qubits the box is drawn over.

To compare the action of the gates to the potentially more familiar language of linear algebra, a representative mapping is also shown in figure 2.2, everything but the equality is in opposite order to the linear algebra representation. Another nice feature of the circuit representation is that the adjoint of a circuit is the circuit run backwards with all operators being their adjoint as well.

It may seem that any unitary of any size can just be written and implemented on a quantum computer, but this is not the case as physical implementations of quantum computer only use some single and two qubit gates. On top of this, limitations imposed by the connectivity of the particular QC also come into play.

2.5   Towards operations on hardware

Hardware has to contend with the limits of the real world, so although one would like to be able to write any unitary and have it be implemented on some quantum state in the quantum registers of a quantum computer, it would be impossible to encode every single operation for an arbitrarily large number of qubits.

It is not obvious at all, but two quantum gates are all that is required to have a universal gateset for quantum computing. These are the single qubit rotation, denoted U_3, as it takes 3 parameters, and the CNOT gate. This means that any unitary, regardless of its size, can be factored into a sequence of these two gates [1]. Now, this is not the only universal gateset, which is very good news for hardware implementations of quantum computers, as some operations are more natural than others.

There are two more things that are going to go in the favor of quantum computers and one that goes against. Beginning with the good news, single qubit rotations can be decomposed into rotations on two axes. Specifically, one can always write

$$U_3(a, b, c) = D(a)E(b)D(c), \qquad (2.15)$$

for some arbitrary parameterized unitary matrices D and E, which represent rotations about two orthogonal axes. In hardware this may be further decomposed into more operations, some of which are fixed and only one of which is tunable, so that as little error as possible is

introduced.

The second benefit is that all two qubit gates can be transformed into any other two qubit gates. There is an optimal decomposition of any 2 qubit gate into at most 3 CNOTS and 8 U_3 gates [31]. Similarly, CNOTS can be represented by other two qubit gates with some other specific transformation. This means that it is relatively straightforward to go from a CNOT representation to any other 2 qubit gate - since the rotations are much more trivial to translate. This is nothing new since in essence we are stating that there are more than one universal gateset.

What is new is that these things can be done approximately with discrete single qubit operations. Sequences of these discrete rotations can be made to yield an operation that is arbitrarily close to any unitary operation one would wish to implement. This is the Solovay-Kitaev theorem [32], which although not so useful for noisy machines, is essential for error corrected machines.

The one problem comes from decomposing larger unitaries into the basic CNOT and single qubit rotation. It is guaranteed that it is possible due to the universality of the gateset, but it is not efficient [1]. Indeed, it is an NP-hard problem to do so efficiently. As such it is always possible, but one may end up with extremely long chains of fully connected circuits.

This is a huge area of study, which is known as compilation, where a circuit is taken from its abstract unitary representation and cut down and optimized as much as possible to run as quickly as possible on a given machine. This requires taking into consideration things like the topology of the machine in the first place, but also error rates between the different qubits can be taken into account [33].

2.6   Fundamental Theorems

Quantum physics has brought with it elegant theorems, such as Bell's theorem [34] proving that no hidden variable theorem can describe quantum mechanics, and many others that are essential to our understanding of quantum physics. Here, only several theorems that are particularly relevant to quantum computing and algorithms are presented. Many of these theorems are no-go theorems which set fundamental limits on what can and cannot be done within quantum information processing.

Theorem 3 (No cloning). *There is no unitary operation C that can perform the transformation $C \left| \psi \right\rangle \left| 0 \right\rangle = \left| \psi \right\rangle \left| \psi \right\rangle$ for an unknown state $\left| \psi \right\rangle$. As such, it is impossible to clone or copy an unknown quantum state.*

THE NO CLONING, may be thought of as straightforward since it emerges from the linearity of quantum mechanics. Although this was first shown in Wootters' and Zurek's "A single quantum state cannot be cloned" in 1982 [35], there were multiple hints that these

theorems were fundamental to quantum mechanics in earlier work, notably in Wiesner's "Conjugate Coding" [36], written in the 70's, but only published in '83. Without the knowledge of such a theorem, it is possible to incorrectly theorize superluminal communications as in Herbert's 1982's "FLASH - A superluminal communicator based upon a new kind of measurement" [37].

Proof: take an operation C which copies two orthogonal states $|\psi\rangle$ and $|\phi\rangle$ such that $C|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$ and $C|\phi\rangle|0\rangle = |\phi\rangle|\phi\rangle$. Here $|0\rangle$ represents a state that is primed for cloning. From linearity:

$$C(\alpha|\psi\rangle + \beta|\phi\rangle)|0\rangle = \alpha C|\psi\rangle|0\rangle + \beta C|\phi\rangle|0\rangle$$
$$= \alpha|\psi\rangle|\psi\rangle + \beta|\phi\rangle|\phi\rangle$$

but it is also true that C acting on the composite state $(\alpha|\psi\rangle + \beta|\phi\rangle)|0\rangle$ gives

$$C(\alpha|\psi\rangle + \beta|\phi\rangle)|0\rangle = (\alpha|\psi\rangle + \beta|\phi\rangle)(\alpha|\psi\rangle + \beta|\phi\rangle)$$
$$= \alpha^2|\psi\rangle|\psi\rangle + \beta^2|\phi\rangle|\phi\rangle + \alpha\beta|\psi\rangle|\phi\rangle + \alpha\beta|\phi\rangle|\psi\rangle.$$

Since one wants the resulting state to be $\alpha|\psi\rangle|\psi\rangle + \beta|\phi\rangle|\phi\rangle$, then taking the two expressions before gives

$$\alpha|\psi\rangle|\psi\rangle + \beta|\phi\rangle|\phi\rangle = \alpha^2|\psi\rangle|\psi\rangle + \beta^2|\phi\rangle|\phi\rangle + \alpha\beta|\psi\rangle|\phi\rangle + \alpha\beta|\phi\rangle|\psi\rangle \iff \alpha = \beta = 0. \tag{2.16}$$

Thus by contradiction, it is impossible to clone an unknown quantum state. This leads to important corollaries which together can define some limits to quantum computations. The inability to clone alone is equivalent to not being able to copy a state to be further processed and compared to the original without evolving two states separately, for example.

Corollary 1 (No broadcasting). *With an unknown state ρ_i drawn from the set $\{\rho_i\}_{i\in\{1,2\}}$ where $[\rho_1, \rho_2] \neq 0$, there is no method that can create state $\rho_{\{AB\}}$ in $H_A \otimes H_B$ that has partial traces $\mathrm{Tr}_A \rho_{AB}$ and $\mathrm{Tr}_B \rho_{AB} = \rho_i$.*

THE NO BROADCASTING THEOREM is an extension from a pure quantum state to mixed quantum states proven in [38]. The result only holds for two mixed states, but given more mixed states it is possible to violate the theorem via a process known as super broadcasting [39]. The proof to this theorem goes beyond the scope of this thesis.

Theorem 4 (No deleting). *We define the transformation $|\psi\rangle \to |0\rangle$ as deletion. There is no unitary operation D that can perform the transformation $D|\psi\rangle|\psi\rangle = |\psi\rangle|0\rangle$ for an unknown state $|\psi\rangle$.*

THE NO DELETING THEOREM is similar to the no cloning theorem in the sense that it can be thought of as the time reversed version of it. The theorem was discovered by Dieks [40] as a direct response to the FLASH paper [37] independently of the no cloning theorem. The proof is similar to the no cloning theorem.

Theorem 5 (No programming). *There is no universal finite gate array that can implement an arbitrary quantum program. This is because the number of possible unitary operations on even a single qubit is infinite, and the number of possible programs is also infinite.*

THE NO PROGRAMMING THEOREM is more subtle than the others. It is indeed impossible to construct an arbitrarily programmable quantum gate array that can implement any operation exactly [15]. A quantum gate array is defined as a gate that is applied given the state of some (quantum) program register. But it is possible to have a programmable quantum computer with some finite subset of possible operations. It has been shown that any operation can be approximated to arbitrary precision with a finite number of gates. This is also the basis as to why a fault-tolerant quantum computer can be built. As we will see in section 4.3, it is possible to extend the non-universal fault-tolerant gateset to an arbitrarily good approximation to a universal gateset by using so called 'magic states' to implement a $Z^{\frac{1}{4}}$ gate, called the T gate [41].

Proof: Given programs $|\mathcal{P}\rangle$ and $|\mathcal{Q}\rangle$ which implement the distinct unitary operators U_p and U_q and the unitary programmable gate array G, which encodes the total dynamics of the system, then when operating on a state $|\psi\rangle$:

$$G|\mathcal{P}\rangle|\psi\rangle = |\mathcal{P}'\rangle|U_p\psi\rangle, \qquad (2.17)$$

$$G|\mathcal{Q}\rangle|\psi\rangle = |\mathcal{Q}'\rangle|U_q\psi\rangle. \qquad (2.18)$$

The prime states are the outgoing state of the program registers, which might have been changed by G. The inner product of the above equations gives

$$\langle\mathcal{Q}|\mathcal{P}\rangle = \langle\mathcal{Q}'|\mathcal{P}'\rangle\langle\psi|U_q^\dagger U_p|\psi\rangle. \qquad (2.19)$$

We can divide by the primed registers if the inner product is not 0, giving

$$\frac{\langle\mathcal{Q}|\mathcal{P}\rangle}{\langle\mathcal{Q}'|\mathcal{P}'\rangle} = \langle\psi|U_q^\dagger U_p|\psi\rangle. \qquad (2.20)$$

Since the left-hand side is not dependent on $|\psi\rangle$, we can determine that $U_q^\dagger U_p = \gamma\mathbb{I}$, $\gamma \in \mathcal{C}$. Then $\langle\mathcal{Q}'|\mathcal{P}'\rangle \neq 0$ is only true if U_p and U_q are equal to a global phase, otherwise $\langle\mathcal{Q}'|\mathcal{P}'\rangle = 0$. This then implies that in (2.19) $\langle\mathcal{Q}|\mathcal{P}\rangle = 0$ and the programs are orthogonal. As such, no universal gate array that is deterministic can be constructed.

3
Quantum Algorithms

> It can do whatever we know how to order it to perform.
>
> *"Note G" on her translation of L.F. Menabrea's article of "Scientific Memoirs", Ada Lovelace, 1843*

This chapter will briefly present the concept of quantum algorithms, a brief history, some influential algorithms and recent developments in their implementations. Following this, the algorithms that are used in this thesis are presented in their own sections, so that they may be referred to while reading the rest of the thesis.

QUANTUM ALGORITHMS are defined much like their mathematical/classical counterparts; perform a series of pre-defined operations on a given input. A quantum algorithm is then an algorithm that is designed to process quantum states mainly through quantum operations, and their implementations are meant to be run on a quantum computer. This makes quantum algorithms a subset of all possible algorithms, with a particular distinction between two approaches to quantum algorithms, one that uses the quantum computer almost exclusively and another that uses both classical and quantum resources.

These two high levels classifications are often called *fully quantum algorithms* and *quantum-classical algorithms* respectively. Fully quantum algorithms are those which run almost entirely on the QC with minimal post-processing, these are the ones that usually have clearer provable advantages over all known classical algorithms. Fully quantum algorithms could also be used as subroutines in an otherwise classical paradigm where the advantages of the QC are used in conjunction with classical computers. The second approach is with *hardware efficient algorithms* that are usually hybrid quantum-classical by design. These are designed to be best suited to current noisy intermediate-scale Quantum (NISQ) devices, but have less certain advantages over fully quantum algorithms [42, 43]. They are characterized by a strong partition of the work between a quantum and classical processor, generally by classically optimizing parameters on a quantum circuit measured on a QC. The quantum-classical approach is a more recent development borne of limitations in current hardware, but the history of quantum algorithms, and likely far

Speedup	Algorithms
Exponential	3
Superpolynomial	31
Polynomial	30
Unknown or varies	5

Table 3.1: **Speedups of known quantum algorithms** as of November 2024 tabulated with information from the Quantum Algorithm Zoo `https://quantumalgorithmzoo.org/`. The polynomial category includes constant speedups. The table is not exhaustive, merely illustrative.

future, will lean more on fully quantum algorithms.

The history of quantum algorithms does not span as far back as many other fields. Indeed, the first true quantum algorithm is Deutsch's algorithm in 1985 [44], which is not much later than the first ideas on quantum computation by Feynman and Manin. This algorithm could discover global properties of a $n \rightarrow 1$ function in a single query rather than two classically. From this early discovery several other black box algorithms had been found, until a turning point in 1994 when Peter Shor discovered his eponymous algorithm [45]. Shor's algorithm could factorize numbers in polynomial time and was the first end-to-end quantum algorithm. It is this algorithm that likely kicked off the immense interest in quantum computing since then due to its potential to crack cryptographic protocols like RSA. It is not the point of this thesis to give an exhaustive list of all algorithms, but suffice it to say that the field has grown immensely since then, with some excellent reviews being [46, 47, 48, 49]. According to the currently up to date Quantum Algorithm Zoo there are around 69 fully quantum algorithms (although in some, parameter optimization is generally done classically), not including subvariants, with different speedups which are tabulated in table 3.1.

These algorithms span a large set of applications in mathematics, chemistry and beyond, but there are some that should be kept in mind, at least in name and application. Grover's algorithm [51] for unstructured search, the quantum Fourier transform (QFT) [52] which does not have a speedup over the classical discrete Fourier transform, but is an important subroutine in what are probably the most useful or influential algorithms in the literature. It allows for the quantum phase estimation (QPE) algorithm [53] which enables the measurement of any value encoded in the phase of a quantum state. This is particularly useful to measure important properties, such as the energy, of some prepared quantum state. These subroutines together are key components of Shor's algorithm.

The key issue with fully quantum algorithms is that they disregard any practical limitations of current and realistic fault-tolerant machines. The issue that arises is due to their requirement for arbitrary controlled unitaries acting on many qubits, which could require many swaps of qubit positions for computers with limited connectivity, and in both cases, deep decompositions into native gate sets for the machines in question. Figure 3.1 shows the decomposition of a fairly basic Toffoli gate with full and linear connectivity to illustrate the increase in depth brought about implementing a 3 qubit gate with two qubit operations with (a) full connectivity and (b) linear connectivity. The larger the unitary, the higher the overhead,
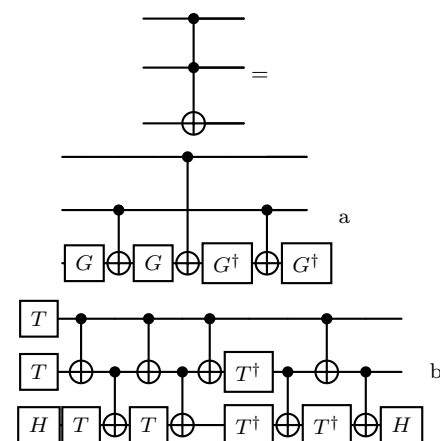
Figure 3.1: **Illustrative decompositions of a Toffoli gate** with all to all connectivity (a) and linear connectivity (b) according to [1]. It should be noted that more optimal decompositions could exist for (b), but not (a) [50]. G and T are specific single qubit gates.

especially in architectures with limited connectivity. In general the optimal factorization of a generic unitary into two qubit gates is an NP-hard problem [1].

The high depth requirements for these algorithms can be somewhat eased for general circuits by clever compilation and transpilation [33], but due to the computational intractability of the problem this approach cannot make significant improvements for general circuits. As such many of the most important algorithms have been studied and re-expressed via methods that greatly reduce the gate counts on common limited connectivity architectures. An example of the depth reduction that is possible is the QFT algorithm, whose depth has been reduced from an already optimized $\mathcal{O}(n^2)$ depth circuit to the new state-of-the-art depth of order $\mathcal{O}(n)$ [54]. Such reductions in depth will be useful in both the current NISQ era and the future fault-tolerant eras (hopefully).

As has been mentioned in passing, current machines cannot run fully quantum algorithms well for any system size of interest, and while the exact reasons why it will be discussed in chapter 4, this limits the kind of algorithms that can be run on current machines. Thus, hardware efficient algorithms are used, they are often characterized by the use of shorter circuits, classical optimization and post-processing of the QC results.

Hardware efficient algorithms are much more recent in the history of quantum algorithms, with the first example arguably being the variational quantum eigensolver (VQE) [55] in 2013. This algorithm will be described in depth later on in this chapter, but it has spawned a plethora of derivative algorithms that try to exploit various features of the problems that flavor of VQE is tackling. Around the same time, the quantum approximate optimization algorithm (QAOA) [56] was also developed, it is similar to VQE in its use of a parameterized hardware efficient ansatz that is optimized classically, but it is specifically designed to solve combinatorial optimization problems.

The development of these techniques is vast for the short amount of time they have existed, but the general family that these algorithms exist in is called variational quantum algorithms (VQA), as all members of this algorithmic class vary parameters in some predefined circuit. An excellent review of many of the relevant methods is found in [57]. To make VQAs work well, it is important to find a statement of the problem that requires as few measurements as possible for the expectation values, as well as the design of the variational circuits, which can be thought of as a subfield of its own [58]. It is unknown whether variational techniques will be advantageous over classical techniques as many questions remain on the trainability of these circuits [59] and the influence of noise on the simulability of the circuits [60, 61]. Indeed, it may be that a whole class of ansatze are simulable even in the noiseless case [62], or that if there are no barren plateaus, that the problem is not classically hard [42]. Despite these fundamental challenges, research in this

area is still active, and more thorough descriptions of the VQE and QAOA are presented in the following sections.

3.1   Variational Quantum Eigensolver

Finding the ground state of a system is an important problem in many areas of physics and chemistry. It can also be the initial step in the study of more complex dynamics. It is in general a difficult problem to solve due to many factors. For one, with limited memory with which to store a wavefunction, one cannot accurately describe the exact ground state for a general system. If one had enough memory to do so, the computational time required to invert a large matrix would be infeasible. One can also attempt to start with some structure to the state vector initialized in some non-ground state and evolve it to the ground state within the limits of the expressivity of the structure imposed on the state vector due to computational or other constraints. On a quantum computer, this would be the amount of time that the state may be evolved for before noise takes over, or in the absence of noise, how patient (and rich) the user is. Fortunately, most problems that chemistry and physics concern themselves with have the helpful property that a given system's ground state is the eigenvector with the lowest eigenvalue of the system's Hamiltonian.

This allows for the iterative lowering of the energy on successive trial wavefunctions via a variational method, which are based on the well known variational principle [63].

$$\frac{\langle \Psi_{\text{trial}} | H | \Psi_{\text{trial}} \rangle}{\langle \Psi_{\text{trial}} | \Psi_{\text{trial}} \rangle} \geqslant \frac{\langle \Psi_{\text{true}} | H | \Psi_{\text{true}} \rangle}{\langle \Psi_{\text{true}} | \Psi_{\text{true}} \rangle} \qquad (3.1)$$

Thus, one may find an upper bound for the energy of a physical system given an appropriate Hamiltonian. The VQE in its simplest form uses some Hamiltonian H and generates a trial wavefunction on N qubits using some number of parameters, usually denoted as the parameter vector $\boldsymbol{\theta}$, which are tied to gates in a quantum circuit which acts as a trial wavefunction denoted $\Psi(\boldsymbol{\theta})$. $\boldsymbol{\theta}$ generally refers to the angle with which a particular rotation gate is applied, but could refer to any tunable parameter. In essence the particular rotations applied are not important and many schemes exist depending on the capabilities of the quantum computer [55, 64, 65].

In the most general case, an initial circuit C with *bound* parameters is run once per term in H, and measured in a basis related to the term. Generally, the terms require some rotations on the wavefunction to actually measure the Hamiltonian terms. Since qubits do not necessarily share the correct statistics as the problem (although one can have fermionic [66] or bosonic [67] qubits), some mapping $\mathcal{M}$ can be applied to the Hamiltonian such that $\mathcal{M}(H) \to H_q$, the standard definition $H_q | \Psi(\boldsymbol{\theta}) \rangle = E | \Psi(\boldsymbol{\theta}) \rangle$ holds, note that the subscript will be dropped since it is a one-to-one mapping. Each term of

a total of T terms of the qubit Hamiltonian

$$H = \sum_i^T c_i H_i, \tag{3.2}$$

may be measured separately given that the same state is prepared, this means that a relatively short circuit can be used rather the same circuit tied to a direct energy measurement like with QPE [68], which would require a lot more quantum resources than NISQ can currently provide. The energy is the sum of each constituent measurement related to a Hamiltonian term H_i multiplied by the appropriate prefactor. The number of ideal measurements to resolve the standard second quantization electronic structure Hamiltonian grows as $\mathcal{O}(N^4)$. One key aim is to make approximations or develop methods to reduce the number of measurements required. This requirement is further compounded in the near and midterm, as one must take numerous measurements of each term to increase the signal-to-noise ratio and to potentially apply quantum error mitigation (QEM)[1].

[1] Quantum error mitigation will be discussed in depth in section 4.2.
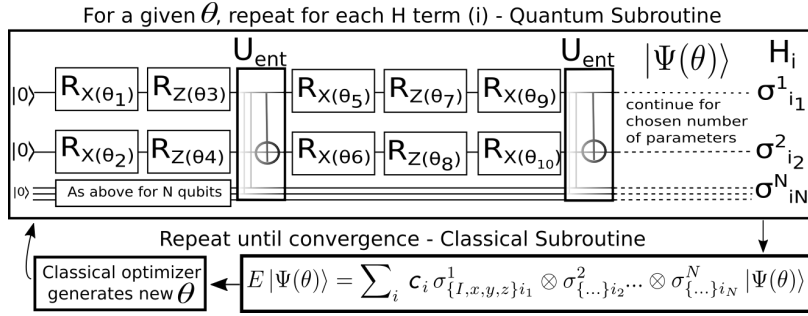
Figure 3.2: **The variational quantum eigensolver method** illustrated for a simple hardware efficient ansatz for a qubit Hamiltonian already converted into a tensor product of Pauli rotations before measurement. $|0\rangle$ *represents the initial computation state, but could be any classically computed initial guess [69].*

Algorithm 1: Variational quantum eigensolver (VQE)

1. Begin with a random choice of $\boldsymbol{\theta}_0$

2. Prepare the quantum state on the quantum computer by applying a chosen sequence of parameterized gates interspersed with entangling operations.

3. Measure a given term of the Hamiltonian H_i. Step 2 has to be repeated with the same parameters a chosen number of shots for each term. The measurements for the term give an expectation value which is multiplied with a prefactor as in (3.2), the sum of all these measurements is the energy $E(\boldsymbol{\theta}_j)$.

4. Input the energy into a noise resistant classical optimizer which finds a new set of $\boldsymbol{\theta}$s, this will likely require more measurements to find ΔE. Repeat from 2. until convergence.

This approach is actually very general, and any problem that can be mapped to some cost-function that can be translated into a sum

of observables can in theory be solved via something like a VQE. This is indeed what is done in research and is the family of VQAs which are well reviewed here [57]. As such, now the algorithm will be described in more detail to better define what is meant by a VQE on a device with properties like a superconducting processor with a stronger emphasis on the potential pitfalls.

3.1.1 The cost function

The cost function of the VQE is a minimization problem derived from the variational principle (3.1), which is

$$E = \min_{\boldsymbol{\theta}} \langle 0|^{\otimes n} U^{\dagger}(\boldsymbol{\theta}) H U(\boldsymbol{\theta}) |0\rangle^{\otimes n} \qquad (3.3)$$

for some Hamiltonian H and parameterized unitary U on n qubits. The Hamiltonian ought to be measurable on a quantum device, and as such it should be decomposable into a sum of Pauli strings $H = \sum_{i}^{T} = c_{i P_i}$, where P_i is a tensor product of single qubit Pauli operators (2.7) with length n. Many problems must be mapped to this form via a transformation - the oldest of which is the Jordan-Wigner (JW) transform [70], but more efficient transformations have been found, such as the Bravyi-Kitaev (BK) [66] and some not yet named transformations like [71, 72]. It should be noted that the JW is completely local but produces many Pauli terms, while the BK is non-local. The newer methods can be thought of as compromises between the two which might add additional ancillary qubits. The state of the art [72] requires 1.016 qubits per encoded fermion. Whatever mapping is chosen, to actually compute the cost function one must measure all terms T as

$$E = \sum_{i}^{T} c_i \min_{\boldsymbol{\theta}} \langle 0|^{\otimes n} U^{\dagger}(\boldsymbol{\theta}) P_i U(\boldsymbol{\theta}) |0\rangle^{\otimes n}. \qquad (3.4)$$

The cost of this grows with the number of terms, which for chemical problems such as electronic structure is N^4 [55].

The first thing that should be apparent is that one does not have the Hamiltonian for free. Thus, it is standard that the Hamiltonian be precomputed via *ab-initio* methods that are standard for the problem at hand. This limits the accuracy of the resulting eigenvector to the accuracy of the method used to produce the Hamiltonian in the first place. As such, if an accurate physically motivated Hamiltonian is too expensive to compute or is unavailable the resulting eigenvector from a VQE calculation will be limited to being the best eigenvector for a 'bad' Hamiltonian. For electronic structure methods, this means that the choice of basis set and method will have a strong impact on the results [73].

The problems with the large number of terms can also somewhat be mitigated by grouping commuting terms into single measurements. This can require a rotation into an appropriate basis that is diagonal with respect to all the strings in the grouping [74], which comes at the cost of some circuit depth. Other optimizations in the

measurements can be made, such as ignoring low weight terms and exploiting symmetries in the problem [75]. All the mentioned processes, except for the expectation value measurements, all require significant classical processing. This means that the VQE is only going to yield an advantage if the computation of the expectation values is the most expensive part of the calculation and that the time saved here outweighs computing the integrals on classical hardware. Of course, if the calculation is much more accurate than the classical computation then perhaps the computational time is less relevant.

3.1.2   The ansatz

The quantum circuit ansatz is at the heart of the VQE. Although any parameterized unitary could in theory be used, in practice one aims to limit the number of tunable parameters to be polynomial in qubit number. This is because if the variational space is too large it becomes difficult to solve the minimization problem classically [76]. Another common limitation is that the circuit should be efficient on the hardware, although this limits the physicality of the ansatz in representing the problem. The state-of-the-art lies in balancing the two such that the circuit is both physically motivated and efficient.

PHYSICALLY MOTIVATED ANSATZE can be explored form the lens of electronic structure problems, where ansatze that have been derived from electronic structure theory have been developed. The most widely studied is the family unitary coupled cluster (UCC) ansatze [77] which construct their circuits based on excitation operators acting on a reference state. These are the most accurate in theory, but generate deep circuits. As such, strategies inspired by this such as adaptive VQE (ADAPT-VQE) [78, 79] which select a subset of UCC or other problem-dependent operators to include[2]. This yields a more efficient circuit that is still physically motivated but much friendlier to hardware. Other techniques that allow for physically motivated constructions might take information from known correlations in the problem, such that qubits that represent non-interacting parts are far away and weak interactions are not represented in the circuit [80].

HARDWARE EFFICIENT ANSATZE are heuristic ansatze that use the existing connectivity and gate set of a given computer. Generally they are constructed with single qubit rotations on one or more axes followed by the native two qubit gate of the device in an alternating pattern. More intricate constructions can be made, such as choosing gates so that some properties (like particle number) are conserved [82] with a slight cost to depth, but without additional SWAP operations. For this class of ansatz, the most important characteristic is termed the *expressivity* (sometimes called expressibility) [83], which is a measure of how much the ansatz can explore the Hilbert space.

[2] The exact technique used by ADAPT-VQE for the selection criteria is a topic of research on its own.
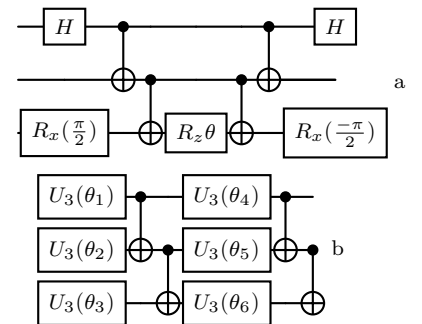
Figure 3.3: **3 qubit VQE ansatze** showing one of many potential terms, $e^{i\frac{\theta}{2}XZY}$, in a UCC ansatz (a), and the linear connectivity efficient $SU2$ ansatz [81] (b).

Figure 3.3 shows a single term that may appear in a UCC ansatz in (a), while (b) shows a standard hardware efficient ansatz with linear connectivity. It should be noted that although the UCC ansatz seems efficient as well, the depth and width can be very high even for hydrogen. For example, for the standard STO-3G basis set on hydrogen [84] one already requires a circuit depth of 97 on 4 qubits. With higher levels of theory, these numbers easily balloon [85][3]. So it is important to find balance between the two approaches. This is what something like ADAPT-VQE attempts to do, for example.

[3] Aug-cc-pVTZ requires 97 qubits and circuit depths of around 1.5 million.

3.1.3 Optimization

Optimization is the classical subroutine central to the algorithm. There are different types of optimizers - continuous variable and integer optimization, but this discussion restricts itself to continuous variable optimization in the context of gradient descent as required for understanding the VQE. As might be clear from the huge reduction in scope, the study of optimization is vast and underpins many modern technologies, notably machine learning. The main idea is that given a cost function C , an initial set of parameters $\boldsymbol{\theta}_i$ and a function of $\boldsymbol{\theta}$, $f(\boldsymbol{\theta})$, find a new set of parameters $\boldsymbol{\theta}_{i+1}$ such that $C(f(\boldsymbol{\theta}_{i+1})) < C(f(\boldsymbol{\theta}_i))$ [86].

This can be done by taking or estimating the gradient of the cost function with respect to the parameters and moving in the direction of the negative gradient. This is the basis of the gradient descent algorithm. The size of the step taken in the direction of the gradient and how it is efficiently estimated or computed is the main difference between optimization algorithms. The optimization problem in the VQE is to minimize the energy of the system (3.3) with respect to the parameters of the ansatz. While, to give a well known comparison, in machine learning it may be to tune the so-called weights and biases of a neural network or other model - in this field Adam [87] and other Adam-derived optimizers are often used [86].

There are two general methods of computing the gradients - for classical computing and certain efficient circuits this may be done analytically, but it is by far more common to use numerical methods. The simplest case is a finite differences' method, but these are often prone to numerical instability. Fortunately in quantum computing there is a parameter shift rule which gives numerically exact derivatives for many parameterized gates [88]. This is a very powerful tool, but requires many measurements of the circuit to measure the gradient. As such, stochastic techniques such as SPSA [89] and COBYLA [90] have proven to be successful in the current NISQ era [91].

The main issue with optimization is the existence of barren plateaus in either the cost function or the ansatz [92]. The barren plateau is defined by a region of the cost function where the gradient becomes exponentially small. It is worsened by an increase in the size of the system, but can also occur due to noise, in so-called
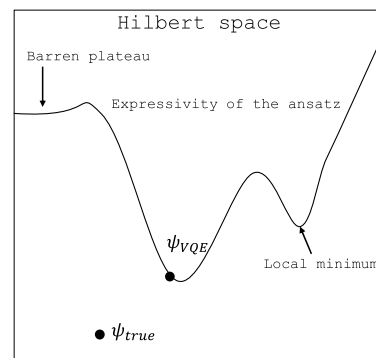
Figure 3.4: **Some features of the VQE** landscape portrayed in an illustrative manner, with the expressivity of the ansatz represented by a line in some Hilbert space, where points are the vectors. The ansatz cannot express the true state exactly, and features a barren plateau and local minima which could trap an unsuspecting optimizer.

noise induced barren plateaus [93]. Furthermore, the problem of local minima, which trap optimizers in suboptimal regions of the parameter space also exist in quantum variational problems [94]. These phenomena are illustrated in figure 3.4.

3.2   Quantum Alternating Operator Ansatz

The QAOA is not much different than the VQE or other VQAs. It is at its essence a highly specialized ansatz that is designed to solve combinatorial optimization problems. Its original formulation was an improvement on the Trotterized quantum adiabatic algorithm, where, like with the adiabatic theorem, a ground state of a time dependent Hamiltonian will stay the ground state through the evolution of the Hamiltonian if it is slow enough. It should be noted that this is the approach taken by adiabatic quantum computing [95]. But the straightforward Trotterization of time evolution did not seem to function well [96].

The solution that was found [56] initially for a MaxCut problem, but since extended to an immense amount of problems and variants [97], was to alternate between two types of non-commuting operators, a mixing Hamiltonian H_M and a cost Hamiltonian H_C. In a circuit these are applied by constructing corresponding parameterized unitary evolutions of the form

$$U_C(\gamma) = e^{-i\gamma H_C}, \tag{3.5}$$

$$U_M(\beta) = e^{-i\beta H_M}. \tag{3.6}$$

The complexity of the circuits is completely dependent on these unitaries, the parameters could take any real value, but due to their periodicity are usually constrained to be between 0 and 2π.

Now the ansatz state can be written as a number of layers L of these alternating operators as

$$|\psi(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = U_C(\gamma_L)U_M(\beta_L)\cdots U_C(\gamma_1)U_M(\beta_1)|+\rangle^{\otimes n}, \tag{3.7}$$

which when written as a circuit looks like[4] including the measure-

ment at the end. The measurement is actually in the computational basis in the standard QAOA approach, and so it is unnecessary to measure in multiple basis for each term of the Hamiltonian as in the general case for a VQE. This is because the function being measured is generally a function on a bitstring that represents the solution to the problem. This could be, for example, the edges that are cut in a graph for MaxCut.
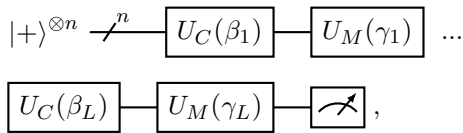
The actual process of the algorithm is as before, where the circuit is run, then an optimizer iteratively changes the parameters to

[4] It should be noted that there is a variant of QAOA called multi-angle QAOA [98] which uses more parameters per layer.

minimize (or maximize) some cost function. Most of the problems
that exist for VQE and VQAs in general also exist here, but due to
the ansatz being very related to the problem by construction makes
this approach more algorithmically sound - and indeed, when not
considering the impact of noise, this algorithm does have a proven
advantage for some classes of problems for variants of QAOA [97].

4
Quantum Errors, their Mitigation and Correction

> It can't do anything we try to order it to perform.
>
> *After trying to use noisy quantum computers, Ada Lovelace 1842i.*

This chapter introduces one of the main challenges of quantum computing, which is the presence of errors. It begins with an introduction followed by a presentation of the basic theory of errors from a algorithmic perspective. It then continues to discuss methods of error mitigation that have been developed, ending with a brief section on error correction.

CURRENT QUANTUM COMPUTERS are far from the Platonic ideal required to implement the quantum algorithms described in the previous section. Although the tools used in modern manufacture are far from the tools of the 19th century, the sentiment remains the same. Both the tools and materials used to construct quantum computers and our control of the interactions are imperfect. This leads to errors that occur throughout a program's runtime, these are generally brought about either through unwanted interactions with the environment or imperfections with the control hardware.

To make a quantum computer functional, there are three approached to dealing with noise with varying likelihoods of success. If probability of an error occurring per operation are miniscule (on the order of 10^{-11} or below) [1], then most algorithms would have a high enough probability of success. How is one supposed to get to such rates? The first is to somehow minimize the amount of error in a device via hardware design - this is limited by engineering, and is unlikely to reach such low levels of inherent error, currently the lowest error rates are much higher than this (10^{-2}). As such, one can try to measure the same properties multiple times with the hope of strengthening the signal-to-noise ratio, known as error mitigation, or to correct the errors as they happen on the processor, known as error correction, which is effective at much higher hardware error rates.

Overall the high error rates of current devices do not allow for any interesting programs to be run directly on the machine. Algorithms for fault-tolerant devices are vanishingly unlikely to return any useful results for any problem size of interest. For these to run successfully,

we must run error correction. But there are ways to use current QCs, and return useful results for much larger problem sizes. This requires algorithms designed to fit within existing limits. These are termed NISQ(-friendly) algorithms.

This chapter begins with a section that introduces quantum errors and their associated formalisms. Then, a section on how errors are mitigated is presented which focuses on three exemplary techniques of three different schools of error mitigation. This is a distinct practice from actually correcting errors, which is known as error correction, which will conclude the chapter. This will be the foundation for a large part of the research work presented later, and will be referred to quite frequently throughout the work.

4.1   Errors and Noise

Although the sources of error are a very interesting field of research on their own, the physical source of what eventually becomes noise is very device specific. The physics that governs the different systems also governs the specific phenomena that become errors in the computation. Noise is the overall effect of accumulated errors throughout the computations, which tend to wash out the 'signal' that would be the expected result from a given quantum program. From the computer scientist perspective, the different physical phenomena that lead to errors can be abstracted away into forms that are uniform throughout the machines.

Of course, in the actual practical use of various techniques it is indeed a benefit to design device specific techniques. This can be achieved on an algorithmic level and in the post-processing, and certainly to correct some very specific errors, such as physical qubit loss in trapped ion and neutral atom systems, something that does not happen in superconducting devices.

There are two classes of errors that can occur in a quantum computer, *coherent errors* and *incoherent errors*. Coherent errors are those that can be described by a unitary operator acting on the state of the qubit, and incoherent errors act stochastically on the qubit. Coherent errors are usually caused by imperfect control hardware, and incoherent errors are usually caused by interactions with the environment.

4.1.1   Coherent errors

As mentioned previously, coherent errors are in essence the application of a unitary that is not the unitary that is intended. For a single qubit the only allowed operation is an arbitrary rotation U which is parameterized by three real variables (α, β, γ), as in table 2 this operation, if implemented perfectly, is denoted $U(\alpha, \beta, \gamma)$. In the case of a coherent error we introduce ϵ_i where i is the relevant parameter, and the most general effect is that the net effect on the qubit is $U'(\alpha, \beta, \gamma) = U(\alpha + \epsilon_\alpha, \beta + \epsilon_\beta, \gamma + \epsilon_\gamma)$. A visualization of
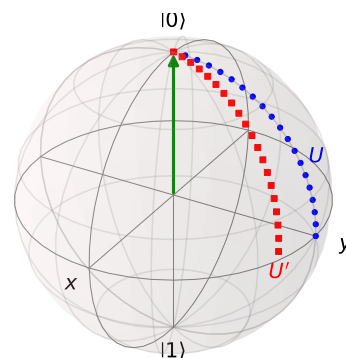
Figure 4.1: **Coherent errors** occur when the intended operation on a qubit is U (blue path) but another unitary U' (red) is physically applied.

this is given in figure 4.1, the net effect of a coherent error in one qubit is a different rotation of the vector than intended.

In the most general case the magnitude of ϵ_i isn't limited. In practice the error is relatively well controlled, especially in the single qubit case. As such, the simplifying assumption that ϵ_i is small can be made. Of course, this means that small changes in angle would be more affected by these errors than larger changes. Another issue is that ϵ_i is not a constant, it may have a dependence on the desired angle, effectively making it some function $\epsilon_i(i)$. It is also known that these are not constant from run to run, and could also change within the runtime of a single program should the program run for long enough. This effect is known as *drift*.
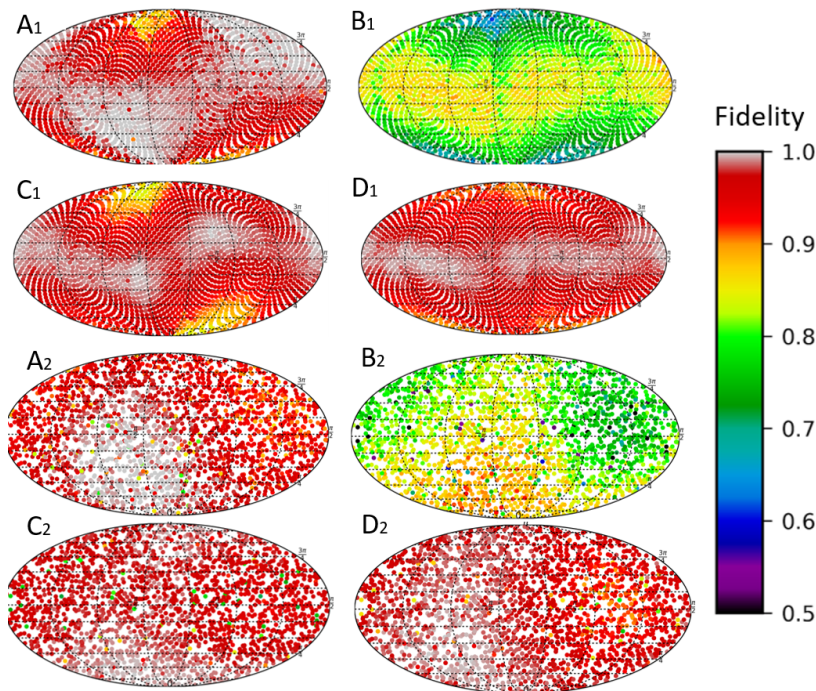
Figure 4.2: **The fidelity of the preparation of a qubit in uniformly distributed states**. Color represents fidelity, and the states are projected via the Mercator projection onto the plane, with 0 at the bottom and 1 at the top. Each point is the result of reconstructed single-qubit tomography at 10,000 shots per basis on IBM Poughkeepsie qubit 1,2,3,4 (denoted A, B, C, D), and the fidelity is with respect to the ideal vector generated by the operation $U(\alpha, \beta, \gamma)|0\rangle = |\psi\rangle$. The subscript denotes the experiment, with 1 being done sequentially using a Fibonacci covering of the sphere, and 2 in a random order via uniform sampling one month after.

It is very difficult in a real machine to distinguish purely coherent errors from purely incoherent errors. This is because the quantum computer is in essence a black box with an output that is affected by the gate errors, which can be coherent and incoherent, environmental interactions that are generally incoherent, and readout errors which are incoherent. As such, the best gauge of coherent errors is to assume that for very short circuits, the probability of incoherent errors is negligible, and if readout errors are corrected for (as will be described in section 4.2.1), then it is possible to gauge the magnitude of coherent errors. Sadly, it is difficult to truly correct for this, as they are likely to be angle dependent and drift over time. This is why in general, digital quantum computers will actually only tune very specific pulses, which vary from machine to machine, and then use tunable virtual Z gates, that alter the relative phase of the tuned pulses, to actually implement the desired gate parameters [99].

To demonstrate this, an experiment on real quantum hardware
has been run that prepares a qubit in states that are uniformly dis-
tributed around the Bloch sphere prepared by a single $U(\alpha, \beta, \gamma)$
gate acting on a qubit initialized at $|0\rangle$. The fidelity of the prepared
state with the ideal state is shown in figure 4.2. The fidelity is calcu-
lated as the overlap of the prepared state with the ideal state, using
the Schumacher fidelity [100]

$$\text{Tr}(\rho_a^{1/2}\rho_b\rho_a^{1/2})^{1/2}, \tag{4.1}$$

although any fidelity measure is appropiate. Assuming the error is
completely coherent, the fidelity measures $|\langle 0| U^\dagger(\alpha, \beta, \gamma) U'(\alpha, \beta, \gamma) |0\rangle|^2$
for values of α, β, γ that are uniformly distributed around the sphere
for a total of 1000 points. One was done in order following a fi-
bonacci covering of the sphere, and the other was done in a random
order via uniform sampling of the angles at a different time.

It is clear from figure 4.2 that the state preparation is not uni-
form, with some bands appearing of high fidelity and others of poor
fidelity - this strongly implies that the error rates are highly angle
dependent, as a constant angle error would make the low angle ro-
tations have a lower fidelity to the high angle rotations, which is not
the case. It is also clear that the two timepoints have very different
charachteristics with regards to the error, notwithstanding the differ-
ence between the qubits. This is a clear demonstration of the drift of
the error rates and charactersitics on top of the error. It should be
noted that it is nearly impossible to characterize the $\epsilon_i(i)$ properly,
as an additional rotation is required to perform tomography which is
usually performed as a modification to the desired U, but this anal-
ysis shows the magnitude of the coherent errors as well as possible.
To mitigate all possible other sources of error, the experiment run
in sets within the IBM job limit a structure like that of figure 4.3.
This means that the readout error should be accounted for as well
as possible, as well as a pulse calibration sequence to minimize pulse
errors based on a previous work [101]. Irrespective of this, it is clear
that the errors persist and that different qubits have very different
inherent errors. Qubit B in figure 4.2 was likely a bad qubit since it
has a much lower overall fidelity, but the bands are likely the result
of coherent errors.

Coherent errors are very tricky to work with, and there are pa-
pers dealing with them specifically [102], and are not truly possible
to mitigate directly. Even in quantum error correction codes, if a
coherent error were to somehow enter the system, it would not be
possible to correct for it as it would still be within the code-space
of the error correction. This was hinted at when another previous
work [103] implied that combining a physical qubit with coherent
and non-coherent errors with a perfect logical qubit would eventually
lead to the complete deterioration of the whole computer's state, this
will be discussed in depth in chapter 8. A follow up work with the
Steane code also showed this to be the case [104]. As such, strategies
have been developed to turn coherent errors into incoherent errors

Figure 4.3: **Job structure for
state preparation**. The job is
divided within the limit of the
runtime as shown. R is the first
readout and pulse calibration,
S is the state preparation to-
mography, followed by a pulse
calibration round C.

via techniques like Pauli-twirling [105, 106]. There is another layer of coherent errors that somewhat overlaps with incoherent errors if one assume they have a true qubit, which is that of leakage into higher energy states, such as $|2\rangle$ and above. This is beyond the scope of what is considered in this thesis, but it is worth keeping in mind.

4.1.2   Incoherent Errors

Incoherent errors are errors that are not unitary, but rather stochastic. The physical sources of incoherent errors are unintended interactions with the environment or control hardware. These may be spurious excitations or relaxations of the qubit, or in atomic based hardware, even the physical loss of a qubit. Errors are generally modelled as *quantum channels* which perform operations on a system as it is 'sent' through them, and the channel can mix in other systems. Note that the mathematical formalism used to describe incoherent processes is powerful enough to encompass both incoherent and coherent errors.

The word channel in this context comes from communication theory. Here Alice attempts to send ρ through some means to Bob, who recieves $\varepsilon(\rho)$, where ε is a superoperator, also called the quantum channel. It is a completely positive, trace preserving map completely positive, trace preserving (CPTP) that acts on the density matrix of the qubit. The description of used in this work for a quantum channel is the *Kraus representation* [107], which is a *sum of operators* acting on the qubit. Some other commonly used equivalent representations are the Choi or *relative state representation*, or the (Stinespring dilated) *unitary representation* [108].

The simplest way to study incoherent errors when one only has access to the system of interest is within the density matrix formalism, since incoherent errors lead to a classical mixture of states with the form $\rho = p\rho_1 + (1-p)\rho_2$. ρ_1 is our desired (likely pure) state and ρ_2 is the error state, which may itself be an ensemble. $p \in [0,1]$ is the probability that the state is in ρ_1. This arises from a process that takes place in the *Church of the Higher Hilbert Space*, which is a tongue in cheek name for the fact that many 'noisy' processes can be imagined as a unitary process to an n dimensional environment, where n is the number of error types that are captured by this transition. Both approaches will be described, beginning with the treatment of density matrices.

THE KRAUS FORMALISM introduces Kraus operators $\{K_i\}$, $i = 1, 2..., m$ with the trace preserving property $\sum_{i=1}^{m} K_i^\dagger K_i = \mathbb{I}$ and complete positivity. They map some state ρ as

$$\rho \to \sum_i K_i \rho K_i^\dagger. \qquad (4.2)$$

The channel can be represented by a non-unique sum of operators that are all related by a unitary basis transformation V_{ab}, where $|a\rangle$ is the basis of the system of interest that lives in A and $|b\rangle$ is the

basis of the auxilliary system B acting on the system $|\psi\rangle_a |0\rangle_b$ such that

$$|b\rangle = \sum_a |a\rangle V_{ba}. \tag{4.3}$$

Any Kraus operators acting on the joint state AB are related by the above via

$$\sum_{a,b} K_a |\psi\rangle_A |b\rangle_B V_{ba} = \sum_b \tilde{K}_i |\psi\rangle_A |b\rangle_B, \tag{4.4}$$

where $\tilde{K}_a = \sum_a K_a V_{ba}$ are the 'new' Kraus operators, note that the Ks are not necessarily unitary. This Kraus map is general in the sense that any physical map on finite dimensional Hilbert spaces can be represented by it. As the representation is not unique, the art of creating noise models is to find the simplest physically motivated set of Kraus operators that can represent the expected error channels.

This is because a channel that maps a system on A of dimension d to A' of dimension d' (note, in general the dimensionality considered here in quantum computing is the same) requires $d^2(d'^2 - 1)$ parameters. This is because $(dd')^2$ parameters are required to define a density operator, but there are d^2 constraints due to the trace preserving property for the d^2 inputs. For a qubit this happens to be 12 parameters, and this grows quickly with more qubits. Compared to a unitary map for a qubit, which has 3 parameters, it is clear that completely characterizing noise channels is computationally impossible for larger systems, and this is why the incoherent errors must be approximated by a small number of expected phenomena. A formal proof for the number of parameters required to fully characterize a quantum channel is beyond the scope of this thesis, but it can be found in [109]. Intuitively this is similar to the minimal number of pure states required to represent an ensemble of states for a density operator. Hence even though this number is minimal, a good model that uses even less parameters must be found to be practical.

STINESPRING DILATION is paradoxically easier to explain with Kraus operators and explains Kraus operators. Starting with a operator-sum representation of a channel $\varepsilon_{A \to A'}$ an equivalent unitary map may be written that acts on an external system with some orthonormal basis $|i\rangle$ for $i = 1, ..., n$, where n is the number of Kraus operators. Thus the additional environment system E has the same dimension as the number of Kraus operators, and an unitary isometry $\mathcal{U}$ can be written as

$$\mathcal{U}_{A \to A'E} : |\psi\rangle \to \sum_i^n M_i |\psi\rangle |i\rangle. \tag{4.5}$$

Here the Kraus operators M_i are complete and due to the existence of the environment, the fact that they could be non-unitary for a map of $\varepsilon_{A \to A'}$, a trivial extension to unitary operators is possible with the extended Hilbert space with the environment. Now taking the environment is traced out returns the original map as

$$\mathrm{Tr}_E[\mathcal{U}_{A \to A'E}(|\psi\rangle)] = \varepsilon_{A \to A'}(|\psi\rangle). \tag{4.6}$$

This is the reason why when access to the environment is impossible the unitary evolution on the product space of the system and environment can only be described by a generally uninvertable quantum channel, unless it is unitary (and thus coherent).

This is quite an abstract take, but by studying some important error channels that are referenced in this work the effect of this unitary perspective should be more clear. There are many error channels that are important to characterising noise in quantum computers, and some important (although not exhaustive) errors and their Kraus operators are tabulated in table 4.1, but only a more thorough analysis of the depolarizing channel will be presented, where Kraus operators are constructed from these more basic error types. The identity is always included in the set of Kraus operators and represents the absence of an error.

Table 4.1: Common Single Qubit Errors

Name	Kraus Operators
Bit flip	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, also denoted as X and σ_1.
Phase flip	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, also denoted as Z and σ_3.
Phase and bit flip	$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, also denoted as Y and σ_2.
Dephasing	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.
Amplitude damping	$\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}$ and $\begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$.
Phase damping	$\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix}$.

4.1.3   Local Depolarizing Noise

The local depolarizing error channel is the most commonly used worst case scenario in theoretical work [110, 111, 104] as if a model is able to mitigate local depolarizing noise, it is able to mitigate any Pauli channel [1]. Indeed, the depolarizing channel consists of three types of errors that are characterized by the first three entries table 4.1, the bit, phase and bit-phase flips which are all represented by the three corresponding Pauli matrices.

If a state is initialized perfectly into $|\psi\rangle$ and a depolarizing error occurs, then it evolves into an ensemble of states $\sigma_1 |\psi\rangle$, $\sigma_2 |\psi\rangle$ and $\sigma_3 |\psi\rangle$ with equal probability, with the remainder staying as $\mathbb{I} |\psi\rangle$. As mentioned, it can be beneficial to consider an environment with the same number of states as the number of error types plus the identity. As such, the minimal environment E has 4 states labelled $|0\rangle_E , |1\rangle_E , |2\rangle_E , |3\rangle_E$. The unitary channel is then represented as an

isometry (or a unitary operator acting on both the system and the environment) as

$$U_{\text{depol}} : |\psi\rangle \to \sqrt{1-p}\,|\psi\rangle_A\,|0\rangle_E + \sqrt{\frac{p}{3}}\left(\sigma_1\,|\psi\rangle_A\,|1\rangle_E + \sigma_2\,|\psi\rangle_A\,|2\rangle_E + \sigma_3\,|\psi\rangle_A\,|3\rangle_E\right).$$
(4.7)

If one had access to the environment it would be possible to reverse the effect of this, as each of the basis states of the environment could be measured and would yield whether the error had occurred or not, which could be reversed. When only system A is accessible though, this is the same as taking the partial trace over E, which will derive the operator-sum or Kraus representation. Defining

$$M_i =_E \langle i|\,U_{\text{depol}}, \ i = 0, 1, 2, 3,$$
(4.8)

yields

$$M_0 = \sqrt{1-p}\,\mathbb{I}, \ M_1 = \sqrt{\frac{p}{3}}\sigma_1, \ M_2 = \sqrt{\frac{p}{3}}\sigma_2, \ M_3 = \sqrt{\frac{p}{3}}\sigma_3.$$
(4.9)

Since all Pauli matrices are self adjoint then the completeness relation and normalization are preserved and

$$\sum_{i=0}^{3} M_i^\dagger M_i = \mathbb{I}.$$
(4.10)

As the Kraus operators evolve a state as $\rho \to \sum_i M_i \rho M_i^\dagger$, the general density matrix for a single qubit evolves as

$$\rho \to (1-p)\rho + \frac{p}{3}\sum_{i=1}^{3}\sigma_i \rho \sigma_i.$$
(4.11)

Usually this noise channel is applied to all qubits individually with no correlation. More advanced noise models and simulations are essential to a complete understanding of the hardware and are the realm of open system dynamics [1], which is another fascinating field of study.

4.2   Quantum Error Mitigation

The presence of noise does not allow us to run arbitrary quantum algorithms and expect a useful result. As such, both algorithms and sampling must be tailored to work within the NISQ hardware limitations. Although designing noise resilient or noise minimizing algorithms is essential for this class of algorithms, these are accompanied by one or more error mitigation techniques. These allow us to strengthen the signal-to-noise ratio and compute better expectation values from the experiments. This section will describe the error mitigation techniques used in this work in detail which are: zero-noise extrapolation (ZNE) [112], Clifford data regression (CDR) [113] and virtual distillation (VD) [114, 115].

The question quantum error mitigation tries to answer is if we can hope to achieve quantum advantage before the arrival of fault-tolerant quantum computers. Quantum hardware has seen major improvements in both the quality and control of many qubit systems. A full list would be too long to include, as it spans different quantum platforms with multiple fronts and directions. The idea is to translate this continuous progress into immediate advantages for quantum computations, while accepting limits in the complexity of the algorithms used.

The main tools of error mitigation require algorithms to be of short depth, as to not allow errors to accumulate and bring the computer into the fully mixed state, which when measured will return a random bit string. This is why current hardware is mostly used for variational type algorithms such as the family of VQAs like VQE [55] and QAOA [56], which use short, hardware efficient circuits. The second point is that the output of the algorithm should be a small set of observables of the prepared state compared to full tomography. The expectation values of these observables must then be estimated to a good enough precision to be useful, such as within chemical accuracy for chemistry problems. Thus, the goal of QEM for NISQ machines is to reduce the error-induced bias in expectation values.

A moment should be taken to define the overall benefit and cost of using QEM. This description is foundational knowledge and the following paragraphs will follow a similar language and presentation of the problem as [116]. Ideally a QC running some program would produce a perfect state ρ_0, but with noise the output will instead be the noisy state ρ. In most cases, the actual value of interest is that of some observable of interest O of ρ_0. Even in the ideal case, we would need to run a number of experiments to approximate the expectation value of O. Each experiment is called a *shot*, and the number of shots is S will always be a finite number in practice. This finite S implies a finite accuracy on the expectation value, and is called *shot noise*. If there were no noise, there would be no bias and the precision would scale as $\frac{1}{\sqrt{S}}$. Noise induces a bias in the measurement, and this is the bias that QEM aims to reduce. Formally this means that given an estimator $\tilde{O}$ of the true value $\text{Tr}[O\rho_0]$,

QEM wishes to reduce the mean square error, angle brackets denote the expectation value,

$$\mathrm{MSE}[\tilde{O}] = \langle \tilde{O} - \mathrm{Tr}[O\rho_0]\rangle^2 = \mathrm{Bias}[\tilde{O}] + \mathrm{Variance}[\tilde{O}] \qquad (4.12)$$

Bias and variance are defined in the usual way as $\mathrm{Bias}[\tilde{O}] = \langle \tilde{O}\rangle - \mathrm{Tr}[O\rho_0]$ and $\mathrm{Variance}[\tilde{O}] = \langle \tilde{O}^2\rangle - \langle \tilde{O}\rangle^2$. Measuring O on ρ_0 will yield the *unmitigated estimator*. This will yield a biased estimator of $\mathrm{Tr}[O\rho_0]$, and at the limit of many shots, the error will tend towards being dominated by the bias. QEM techniques all aim to fulfill the condition $\mathrm{Bias}[\tilde{O}_{\mathrm{QEM}}] \leqslant \mathrm{Bias}[\mathrm{Tr}[O\rho]]$. The construction of the error mitigated estimator $\tilde{O}_{\mathrm{QEM}}$ is generally more complex, and will always lead to an increase in variance when S is constant, namingly $\mathrm{Variance}[\tilde{O}_{\mathrm{QEM}}] \geqslant \mathrm{Variance}[\mathrm{Tr}[O\rho]]$.

It should be mentioned that QEM is somewhat ill-defined. This thesis will use the term algorithmic QEM to refer to algorithmic techniques that construct circuits and estimate error-mitigated expectation values by combining the results of various experiments. Single circuits that are run within a QEM technique will not be less noisy, and may indeed be noisier than without algorithmic QEM, as in the case of ZNE. The thesis will explore ZNE and CDR as examples of this. The alternative approach which will be referred to as coherent QEM or error suppression, and also that of quantum error correction (QEC), is to make any single run of a circuit be less effected by noise. A classic technique for this is dynamical decoupling [117], alongside others which come from optimal quantum control theory. One can also run circuits that exploit subspace expansions or purify the desired state, with VD being an example of this - although the proof of this is subtle [118].

In the above paragraphs a lot of classifications and types of QEM were shown, but it is important to classify and name representative techniques that are commonly used in the field. A thorough description of each is beyond the scope of this thesis, but the following paragraphs will separate the different families of error correction and give a short description, some representative techniques and references so that the rest of the work can be put in context. In addition to this, a more detailed description of readout error mitigation is given - usually this is done separately from the other techniques, as the correcting the readout process is more like tuning the measurement apparatus than mitigating the errors in the computation, and this is usually abstracted away from the rest of the error mitigation process. Indeed, readout 'correction' is almost always performed before any other processing is done.

POST SELECTION is another important and conceptually simple tool in an error mitigation toolkit. If it is known that the output must obey certain properties, as for example, a certain parity in bits or a known number of 1's in the output, one may opt to throw out results that do not obey these rules [119]. This can be done with or without measurement error mitigation and to filter a dataset before
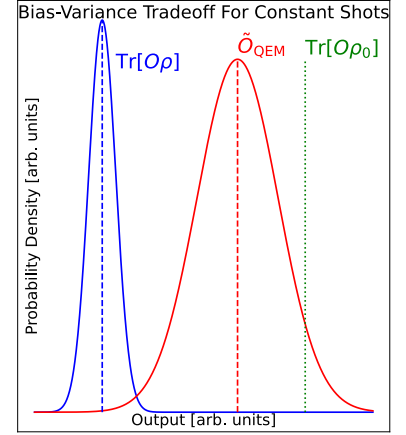
Figure 4.4: **Bias-variance tradeoff** for a constant number of shots. Units and exact values are not shown as this plot is only meant to illustrate the various terms used. The leftmost Gaussian represents the noisy expectation value, the middle the error mitigated one and the leftmost line is the ideal value.

any other techniques are applied [120]. It should be noted that to then accurately gauge the performance of a given algorithm on a given machine, one must consider the ratio of post-selected shots from the total number of shots.

LEARNING BASED MITIGATION is perhaps the currently more studied area of QEM. The idea of all of these techniques is to prepare a model of the noise that is valid for long enough that after the model is prepared, experiments can be run and mitigated with the technique. There are noise agnostic techniques like CDR, which will be covered in depth in a later section, and noise model dependent techniques like probabilistic error cancellation (PEC), which require some theoretical model of the noise a-priori. PEC is the reference method in this category. PEC attempts to alter the circuit with additional gates that probabilistically cancel out noise channels. This involves running the circuit many times with a different random placement of these gates [121]. Tensor-network error mitigation is a more recent development that reverses the effect of the noise in post-processing via a purpose built tensor-network on the output of the QC [122]. The general idea for all of these techniques is that they must learn a parameter for an idealized model of the noise and correct for it, or learn a map from noisy to 'exact' results. The issues come from whether the idealized model of the noise adequately captures the true device noise and whether the tunable parameters can be learned efficiently and accurately.

SUBSPACE EXPANSION techniques cover a vast array of techniques that have been used throughout many fields over many years. In this section, we limit ourselves to their use in quantum error mitigation. It was recently shown that many techniques once thought disparate can be described under a framework called *generalized subspace expansion* [118]. The general idea is that a subspace may be spanned by some set of expansion operators. These are applied to the quantum state of interest and the relevant expansion operator expectation values are measured. This allows for the construction of matrices that allow for the solution to the *generalized expectation value problem*, whose solution can be computed classically due to the much smaller size of these matrices compared to the full Hilbert space. Finding this after the construction of a state can give much better results than measuring the state directly either due to additional symmetries being present in a 'symmetry expansion' [123], or as in VD which will be covered later in this section.

4.2.1 *Measurement error mitigation*

Measurement error is the conceptually simplest error to mitigate, but has its subtleties. Measurement errors arise in the readout phase, where the qubit measurement is classified as having been a 1 or 0. Since the output of these measurements is analogue (although

the specifics vary from platform to platform), classifiers are trained on the output voltages [124], but these aren't perfect. The difficulty in readout error mitigation lies in the fact that separating the errors occurring during the operation of an algorithm and those errors that occur solely during readout while only having access to the readout.

Nonetheless, the simplest method is to build a *confusion matrix*[1] where one prepares every possible bitstring on their quantum computer, which can be done with depth 1 and a negligible error rate, and measure it until a reasonable map to the probability of all possible outputs is gained. The full confusion matrix for N qubits would be $2^N \times 2^N$. In general, one might take the single qubit matrix or clusters of few qubits and prepare tensor products of the smaller confusion matrices to mitigate error on larger systems. This is, however, not state of the art, with more intricate techniques focusing on two major approaches: more efficient constructions of the confusion matrix that use neural networks [125] or known noisy subspaces [126], and tomography of the measurement process called *detector tomography* which involves use of the positive operator-valued measure (POVM) formalism [127, 128] to decouple the readout mechanism from other errors. Nonetheless, this is an essential component of using actual quantum computers, and so a recipe is provided below to describe how this would be done in practice.

1. A single qubit i (or if assuming uncorrelated readout error, many qubits, either simultaneously or in sequence) is (are) sequentially prepared and then measured in the $|0\rangle$ and $|1\rangle$ states.

2. The frequency at which $|0\rangle$ states give a readout of 0 over the total readouts is P_0; P_1 is analogously defined, building the confusion matrix.

3. From this the confusion matrix for a single qubit i is
$\mathcal{M}_i = \begin{bmatrix} P_0^i & 1 - P_1^i \\ 1 - P_0^i & P_1^i \end{bmatrix}$. This is optimal and can be built from measuring probabilities only from the $|0\rangle$ and $|1\rangle$ states.

4. For multiple qubits one may use a tensor product of the form $\mathcal{M}_{0,1,\dots,i} = \mathcal{M}_0 \otimes \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_i$. Getting the measurements to construct the tensor product matrix can be done in $\mathcal{O}(1)$.

5. To correct some result vector $\boldsymbol{v}$, one would multiply it by the correction matrix, which is the inverse of the confusion matrix, $\mathcal{M}^{-1}\boldsymbol{v} = \boldsymbol{v}_{\text{mitigated}}$.

The reason why one would need to construct the full matrix is that true measurement procedure couples the qubits to a resonator, and in the case of multiple qubits one cannot ignore crosstalk (spurious correlations between different qubits induced by operations that ought to be localized to a select number of qubits) [129]. Although we will not concern ourselves with this directly, it is worth noting that the solution to this is to measure the probabilities of all

[1] A *confusion matrix* A_1 for one qubit is $\begin{bmatrix} P(0|0) & P(0|1) \\ P(1|0) & P(1|1) \end{bmatrix}$, where $P(E|M)$ is the probability of measuring M when one would have expected E.

combinations of $|0\rangle$ and $|1\rangle$ states. Assuming no additional state dependence then a matrix of the kind below (shown for two qubits) can be constructed in $\mathcal{O}(2^N)$ time, where N is the number of qubits.

$$\begin{bmatrix} P_{00\to00} & P_{00\to01} & P_{00\to10} & P_{00\to11} \\ P_{01\to00} & P_{01\to01} & P_{01\to10} & P_{01\to11} \\ P_{10\to00} & P_{10\to01} & P_{10\to10} & P_{10\to11} \\ P_{11\to00} & P_{11\to01} & P_{11\to10} & P_{11\to11} \end{bmatrix}.$$

This type of matrix (where $P_{10\to01}$ represents the probability of qubit 0 with value 1 flipping to 0 and qubit 1 with value 0 flipping to 1) can be mixed with itself or the single qubit matrix using the tensor product procedure described in entry 3. of the previous list. There is the possibility that the confusion matrix is not invertible, in which case one must use a pseudo-inverse or some more physically motivated technique, such as bounded least squares optimization:

$$\min_{\boldsymbol{v}_{corrected}} |\mathcal{M}\boldsymbol{v}_{corrected} - \boldsymbol{v}_{experimental}|, \qquad (4.13)$$

here conditions can be put to ensure that the probabilities of $\boldsymbol{v}_{corrected}$ are positive and normalized. Of course, significant errors in state preparation will make $\mathcal{M}$ contain incorrect probabilities. Generally state preparation errors and readout errors are ignored in other error mitigation techniques, and as such it is often seen in QEM and QEC papers that state preparation and measurement (SPAM) errors are ignored.

4.2.2 Zero Noise Extrapolation

ZNE takes the noisy expectation values of desired observables O from a quantum circuit C that is run at various controlled noise levels in the QC which allows one to extrapolate to the noiseless limit. To increase the error rate in a controlled manner there are two main techniques. The most accurate but time-consuming method is to increase the length and decrease the amplitude of the pulses which form the gates of the quantum circuit as with the first large scale experiment that utilized this technique [75], illustrated in figure 4.5. This requires fine-tuning every gate pulse used such that the increase in error follows some known relation, in general this tries to be linear. The second method is to add multiple copies of the same gate preceding or following any desired gate such that the overall action of these added gates forms an identity operation, known as gate repetition [130] and is shown in figure 4.6.

The goal of ZNE is to estimate the noise-free expectation value μ on an N qubit QC

$$\mu = \langle 0|^{\otimes N} C^\dagger O C |0\rangle^{\otimes N}$$

from measurements on a device initialized at $\langle 0|^A \otimes N$ subject to noise. As such μ is inaccessible, and we must estimate it from the noisy expectation values $\tilde{\mu}_j$ obtained from the physical device. ZNE assumes that the noise strength can be characterized by a single
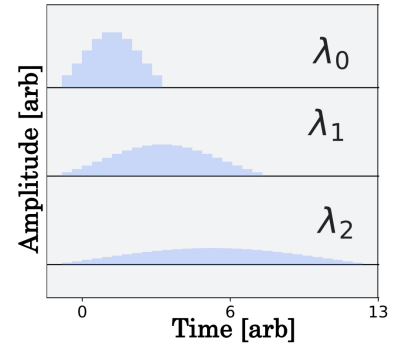
Figure 4.5: **Pulse stretching** to controllably increase the error rate of a quantum gate. The amplitude of the pulse is decreased, and the length is increased to preserve the intended action of the gate, but to increase the error rate. λ_i are the error rates of the gate at each level.

parameter λ. There is no way to decrease the error rate below a machine's characteristic error rate λ_0 easily, but it can be increased somewhat controllably within some range. A number n of noise levels are chosen, and we alter the parameter λ such that the noise strength is increased by monotonically increasing factors c_j, $j = 1, 2, \ldots, n$ from the base level λ_0 to $\lambda_j = c_j \lambda_0$. The $n+1$ noisy expectation values $\tilde{\mu}_j$ are then obtained from the noisy device at each noise level. The noiseless expectation value μ is then estimated by extrapolating the noisy expectation values to the noiseless limit via some extrapolation technique, with Richardson extrapolation being widely used [131]. We create the dataset $\mathcal{D}^{\text{ZNE}} = \{\tilde{\mu}_j^{\text{ZNE}}\}_{j=0}^n$. Richardson extrapolation gives the estimated noise-free value μ via $\mathcal{D}^{\text{ZNE}}$ as

$$\mu_n^{\text{ZNE}} = \sum_{j=0}^n \gamma_j \tilde{\mu}_j^{\text{ZNE}}, \tag{4.14}$$

where

$$\sum_{j=0}^n \gamma_j = 1, \quad \sum_{j=0}^n \gamma_j c_j^k = 0 \text{ for } k = 1, \ldots, n. \tag{4.15}$$

The noiseless value μ is approximated with an error $|\mu^{\text{ZNE}} - \mu| \in \mathcal{O}(\lambda_0^{n+1})$ [112]. The order that the error can be suppressed depends directly on the number of noisy observables computed. By writing the series expansion of the noisy observable dependent on λ, with expansion coefficients α_i we get

$$\tilde{\mu}(\lambda) = \mu + \sum_{i=1}^\infty \alpha_i \lambda^i, \tag{4.16}$$

$$\tilde{\mu}_j^{\text{ZNE}} = \tilde{\mu}(\lambda_j) = \mu + \sum_{i=1}^\infty \alpha_i \lambda_j^i. \tag{4.17}$$

Given two noisy observables $\tilde{\mu}_0^{\text{ZNE}}$ and $\tilde{\mu}_1^{\text{ZNE}}$, at error rates c_0 and c_1, we can write the expansion of the noisy observables, as

$$\tilde{\mu}_0^{\text{ZNE}} = \mu + \alpha_0 \lambda_0 + \mathcal{O}(\lambda_0^2), \tag{4.18}$$

$$\tilde{\mu}_1^{\text{ZNE}} = \mu + \alpha_0 \lambda_1 + \mathcal{O}(\lambda_1^2). \tag{4.19}$$

We can now assign some arbitrary coefficients γ_j to the expectation values and see that we can cancel some terms

$$\begin{aligned} \mu_1^{\text{ZNE}} &= \gamma_0 \tilde{\mu}_0 + \gamma_1 \tilde{\mu}_1 \\ &= (\gamma_0 + \gamma_1)\mu + \alpha_0 \lambda_0 (\gamma_0 + c_1 \gamma_1) + \mathcal{O}(\lambda_0^2), \end{aligned} \tag{4.20}$$

up to an order $\mathcal{O}(\lambda_0^2)$ with

$$\gamma_0 + \gamma_1 = 1, \tag{4.21}$$

$$\gamma_0 + c_1 \gamma_1 = 0. \tag{4.22}$$

For an arbitrarily large number of error rates, one takes the combination of $n+1$ $\tilde{\mu}_j^{\text{ZNE}}$ observables

$$\mu_n^{\text{ZNE}} = \sum_{j=0}^n \gamma_j \tilde{\mu}_j^{\text{ZNE}}. \tag{4.23}$$
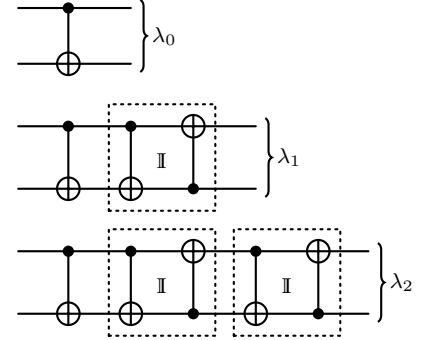
Figure 4.6: **Gate repetition** to controllably increase error rates. This is done by appending copies of the desired gate such that the action of the added gates would ideally be an identity.

This can be expanded into powers of λ_j

$$\mu_n^{\text{ZNE}} = \sum_{j=0}^{n} \gamma_j \mu + \sum_{j=0}^{n} \gamma_j \sum_{i=1}^{\infty} \alpha_i \lambda_j^i$$
$$= \sum_{j=0}^{n} \gamma_j \mu + \sum_{i=1}^{\infty} \alpha_i \sum_{j=0}^{n} \gamma_j \lambda_j^i. \qquad (4.24)$$

In terms of error rates this becomes

$$\mu_n^{\text{ZNE}} = \sum_{j=0}^{n} \gamma_j \mu + \sum_{i=1}^{\infty} \alpha_i \lambda_0^i \sum_{j=0}^{n} \gamma_j c_j^i. \qquad (4.25)$$

This allows for the error to be suppressed to $\mathcal{O}(\lambda_0^{n+1})$. γ_j are chosen to cancel the leading n orders of the error in (4.25), leading to the linear equations for γ_j

$$\sum_{j=0}^{n} \gamma_j = 1, \quad \sum_{j=0}^{n} \gamma_j c_j^i = 0 \ \text{ for } i = 1, \dots, n. \qquad (4.26)$$

ZNE has proven itself to be very valuable in NISQ workloads due to its relative simplicity and its ability to be combined with many QEM techniques with relative ease [110, 132, 133, 134, 135, 136], but it is limited in scaling to larger system sizes and very sensitive to the noise amplification coefficients, which when not estimated correctly can lead to large errors in the final result, but variants have of course been used in large 127 qubit experiments [137].

4.2.3 Clifford Data Regression

Although a traditional regression based approach to error mitigation such as ZNE can be effective on smaller circuits, there is an alternative approach which attempts to learn the noise in the machine indirectly, which is the class of learning-based error mitigation techniques.

CDR [113] is a framework that uses classically simulable *training circuits* to build a mapping between the noisy expectation values of a quantum circuit C and the noiseless expectation values. The set of N_t training circuits $\{T_i\}_{i=1}^{N_t}$ are generally composed of mostly Clifford gates, but in principle, any classically simulable circuit class could be used. The training circuits are run on the quantum computer, and their expectation values alongside the exact classical simulation results, are used to train a regression model. This model is then used to predict the noiseless expectation values of the quantum circuit. The training circuits are chosen to be a projection of the actual circuit of interest onto the Clifford group. This ensures that the topology and number of gates match the desired circuit of interest, with the intuition being that the noise channels experienced by the device is not dependent on whether the gate is Clifford or not, thus allowing for the trained regression model to be transferred to the unsimulable circuit results.
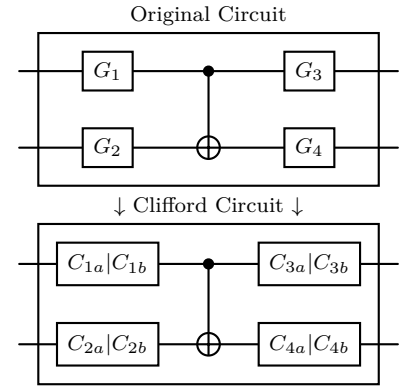
Figure 4.7: **Clifford projection** of a circuit formed by arbitrary gates G_i to one of their closest corresponding Clifford gates C_{ix}, one would sample the gates since most gates might be between two Clifford gates C_{ia} or ($|$) C_{ib}. It is also possible to retain some original gates G and still remain simulable.

The Clifford circuits, known as training circuits are ensured to be classically computable via the Gottesman-Knill theorem [138], which allows us to compute noiseless expectation values at scale for our chosen observable O as $\mu_i = \langle 0|^{\otimes N} T_i^\dagger O T_i |0\rangle^{\otimes N}$ and our noisy expectation values $\tilde{\mu}_i = \langle 0|^{\otimes N} \tilde{T}_i^\dagger \tilde{O} \tilde{T}_i |0\rangle^{\otimes N}$. This creates the dataset $\mathcal{D}^{\mathrm{CDR}} = \{(\tilde{\mu}_i, \mu_i)\}_{i=1}^{N_t}$ which forms the training set for a function f that maps expectation values $\mathcal{E}$ as $f^{\mathrm{CDR}} : \mathcal{E}_{\mathrm{noisy}} \to \mathcal{E}_{\mathrm{noiseless}}$. f can be as simple as linear regression, or as complex as a neural network. It turns our that the simplest model, linear regression, is effective and thus the relation that approximates the noiseless expectation value from the noisy expectation value is

$$f^{\mathrm{CDR}}(x) = a_1 x + a_2, \tag{4.27}$$

where a_1 and a_2 are the regression coefficients which are found via least squares regression on the dataset $\mathcal{D}^{\mathrm{CDR}}$:

$$\arg\min_{a_1, a_2} \sum_{i=1}^{N_t} \left(\mu_i - (a_1 \tilde{\mu}_i + a_2) \right)^2. \tag{4.28}$$

In practice, it is possible to use some non-Clifford single qubit gates and still have a simulable circuit, these *near-Clifford circuits* are used as they greatly enhance the spread of the possible expectation values [113]. The problem with this is that then CDR suffers from a problem inherent to near-Clifford circuits, which is that when they are randomly constructed, the expectation values cluster around 0, as shown in figure 4.8.

The clustering of expectation values is detrimental to scaling of CDR since it means that many shots are required to properly estimate distinct expectation values [139]. As such, it is important to ensure that the variance of the expectation values of O is wide enough to train the regression model with a reasonable number of shots. There is no solution to build such training circuits at the moment, but one can post select a good set of training circuits from a larger set of randomly generated circuits [110], but this is likely not scalable, and so more clever constructions such as Markov Chain Monte Carlo [113]. There is as of now no general solution to this problem [110].

4.2.4 Virtual Distillation

VD [115, 114] is a coherent method of error suppression, unlike the previous ones which rely mostly on post-processing of the data, and it belongs in the class of subspace expansion methods, as it relies on powers of the state. We have a circuit we want to mitigate C which produces a noisy state described by a density matrix ρ on N qubits. VD instructs us to prepare more than one (M) copies of ρ, which will not be prepared identically. These are then *distilled* so that we may obtain expectation values of an observable O on this purer state via

$$\tilde{\mu}^{\mathrm{VD}} = \frac{\mathrm{Tr}[\rho^M O]}{\mathrm{Tr}[\rho^M]}. \tag{4.29}$$
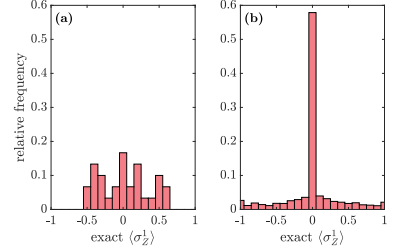
Figure 4.8: **Clustering of near-Clifford circuits.** (a) plots the relative frequencies of exact expectation values of an observable in a sample of 30 4 qubit, 4 layer random quantum circuits. (b) shows 3000 derived near-Clifford circuits. For each random circuit, 100 near-Clifford circuits are generated. Reproduced from [110] under CC4.
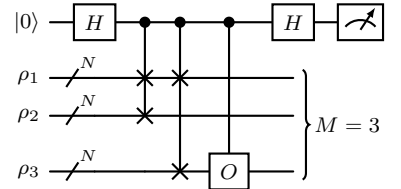
Figure 4.9: VD circuit with $M = 3$ copies of ρ that each live on N qubits. The ancilla qubit is prepared in the $|0\rangle$ state and the copies of ρ are prepared in MN qubits. The ancilla controls a swap gate between the copies of ρ and finally a controlled gate for desired observable, followed by a measurement of the ancilla. Although it may look like only M controlled swaps are necessary, each one is acting on a cluster of N qubits, hence why MN two qubit controlled swaps are required.

We use the ordered eigen-decomposition of the density matrix $\rho = \sum_{i=0}^{2^N-1} p_i |\psi_i\rangle \langle\psi_i|$, where p_i decreases as i increases. Then (4.29) can be rewritten as

$$\tilde{\mu}^{\text{VD}} = \frac{\langle\psi_0|O|\psi_0\rangle + \sum_{i=1}^{2^N-1}(p_i/p_0)^M \langle\psi_i|O|\psi_i\rangle}{1 + \sum_{i=0}^{2^N-1}(p_i/p_0)^M}. \tag{4.30}$$

In (4.30) ψ_0 is called the dominant eigenvector, and assuming that $p_0 > p_{i>0}$, the contributions from $\psi_{i>0}$ are exponentially supressed with respect to the number of copies M.

The dominant eigenvector is not necessarily the desired noiseless state $|\psi_{\text{exact}}\rangle$, but it is assumed that it is very close to it. As such, there is a fundamental noise floor that limits the correction that can be achieved with VD. The error in the expectation value ε is given by

$$\varepsilon = \langle\psi_0|O|\psi_0\rangle - \langle\psi_{\text{exact}}|O|\psi_{\text{exact}}\rangle, \tag{4.31}$$

which is bounded as $|\varepsilon| \leqslant ||O||_\infty 2\sqrt{1 - |\langle\psi_{\text{exact}}|\psi_0\rangle|^2}$. Here $||O||_\infty$ is the operator norm of O, i.e. the norm of largest eigenvalue of O [140]. If ψ_0 happens to be an eigenvector of O then the bound is tighter, but within the constraints of NISQ hardware, ε is greater than 0 [115]. Nonetheless, even current error rates are sufficiently small that the noise floor is not the limiting factor for VD [140].

To prove this, we must perform an eigendecomposition of (4.29), rewritten here for convenience. Yielding

$$\tilde{\mu}_M^{\text{VD}} = \frac{\text{Tr}[\rho^M O]}{\text{Tr}[\rho^M]} = \frac{\sum_{i=0}^{2^N-1} p_i^M O_i}{\sum_{i=0}^{2^N-1} p_i^M}, \tag{4.32}$$

for some observable O with expectation values $O_i = \langle\psi_i|O|\psi_i\rangle$. With some algebraic manipulation,

$$\begin{aligned}
\tilde{\mu}_M^{\text{VD}} &= \frac{p_0^M O_0 + \sum_{i=1}^{2^N-1} p_i^M O_i}{p_0^M + \sum_{i=1}^{2^N-1} p_i^M} \\
&= O_0 + \frac{\sum_{i=1}^{2^N-1} p_i^M (O_i - O_0)}{p_0^M + \sum_{i=1}^{2^N-1} p_i^M}.
\end{aligned} \tag{4.33}$$

Now extract the difference between the dominant noisy eigenvector's expectation value O_0 and our mitigated expectation value as

$$|\tilde{\mu}_M^{\text{VD}} - O_0| = \frac{\sum_{i=1}^{2^N-1} p_i^M |O_i - O_0|}{p_0^M + \sum_{i=1}^{2^N-1} p_i^M} \leq \frac{2||O||_\infty \sum_{i=1}^{2^N-1} p_i^M}{p_0^M + \sum_{i=1}^{2^N-1} p_i^M}, \tag{4.34}$$

now the following can be stated,

$$\frac{\sum_{i=1}^{2^N-1} p_i^M}{p_0^M + \sum_{i=1}^{2^N-1} p_i^M} \leq \frac{\sum_{i=1}^{2^N-1} p_i^M}{p_0^M} \leq \frac{(\sum_{i=1}^{2^N-1} p_i)^M}{p_0^M}, \tag{4.35}$$

which leads to

$$|\tilde{\mu}_M^{\text{VD}} - O_0| \leq \frac{2||O||_\infty \xi^M}{(1-\xi)^M}. \tag{4.36}$$

Where $\xi = 1 - p_0 = \sum_{i=1}^{2^{N-1}}$ is the sum of mitigated eigenvalues except for the dominant one. Assuming $\|O\|_\infty$ is finite, then

$$|\tilde{\mu}_M^{\mathrm{VD}} - O_0| = |\hat{\mu}_M^{\mathrm{VD}} - \mu - \varepsilon| \in \mathcal{O}(\xi^M), \qquad (4.37)$$

which finally proves the exponential error suppression of VD [115, 110].

What makes VD difficult to implement experimentally is the fact that one must prepare M copies of ρ that itself consists of N qubits. All of these copies must be connected to an ancilla that performs a controlled swap of the copies with each other. A direct implementation of VD requires $NM + 1$ qubits which then implement NM controlled swaps to permute all the positions of the copies, as seen in figure 4.9. This brings an additional layer of errors to the system which is not accounted for in the error suppression of VD. To mitigate this, one may need to use algorithmic QEM to supplement VD [115]. This interconnected circuit topology makes such a scheme unwieldy for superconducting processors which are limited to two dimensions, but might be quite natural for trapped ion or neutral atom systems, which can implement all to all connectivity.

4.2.5   The Limits of Error Mitigation

It is well known within the quantum computing literature that quantum error mitigation is limited. It is commonly agreed that due to the bias-variance tradeoff of QEM shown in figure 4.4, that not only does one need to increase the number of shots to reduce the variance for a given system size, all mitigation techniques also have a penalty of an exponentially growing shot cost with respect to the system size, and more efficient techniques tend to merely decrease the prefactor of this scaling [116]. Techniques that perform error suppression like VD do not suffer from this, as they suppress errors. More recently, it has been discovered that there are potentially more fundamental limitations to how well QEM can mitigate errors [141, 142].

4.3   Quantum Error Correction

Error correction has long been known to be necessary for most quantum algorithms that have proven quantum advantages. The overheads brought on by implementing quantum error correction codes are high, and are more expensive the worse the underlying physical qubits are. Although it is thought that practical quantum computers that implement fault tolerance are always 5 years away, impressive developments in both theory and experiment have been made in the field [143, 144, 145, 146]. This paves the way to building what is known as a fault-tolerant quantum computer. The particular experiments and their importance will be presented as the necessary background is introduced.

This section is intended to introduce some core concepts of fault tolerance, including the fundamental ideas of error correction, the all important threshold theorem and the limitations imposed by error corrections, alongside the proposed strategies to overcome these limitations. A thorough pedagogical introduction to the formalisms of error corrections can be found in the following works [147, 148]. A thorough treatment of error correction is beyond the scope of this thesis, but it is an essential component to the future of quantum computing, and it is made use of in the research presented in chapter 8.

It is important to keep in mind that to perform any existing quantum algorithm error probabilities must be kept very low in the absence of QEC, around the order of 10^{-15} [149]. The most promising technology in terms of inherent error rate is the trapped ion architecture which might achieve error rates of 10^{-6} [150]. There is a little bit of terminology that must be introduced - a *logical qubit* and *logical operation* are the qubits and operations that are encoded in via a quantum error correction code. They consist of multiple physical qubits and operations which are mapped in a code-specific way. For any given QEC code, there is a threshold error rate p_{th}, which is the maximum error rate that can be tolerated before the error correction code fails to correct the errors and error correction becomes a hindrance. This threshold is dependent on the particular error correction code and the number of physical qubits used to encode the logical qubit. This is true for both classical [151] and quantum error correction and is known as the threshold theorem [152, 153]. Indeed, it experimentally verified that it is true that below threshold correction is possible [144, 143].

It is assumed that any operation in a circuit will fail with some probability, so every operation within a circuit must be encoded as well as the underlying qubits. This encoding will be represented by a bar over the component, so a fault-tolerant circuit C will be denoted $\overline{C}$. Although the specifics are code dependent, in general this means that there must be rounds of *syndrome measurements* that determine if and what error occurred followed by a *recovery round* if an error did indeed occur. The encoding and QEC come at

some additional overhead cost associated with each operation, which are known as fault-tolerant gadgets. Figure 4.10 shows the high level costs associated with applying QEC to a circuit. Notably, after any operation a cycle of QEC must be applied to each qubit, and each operation must likewise be encoded.
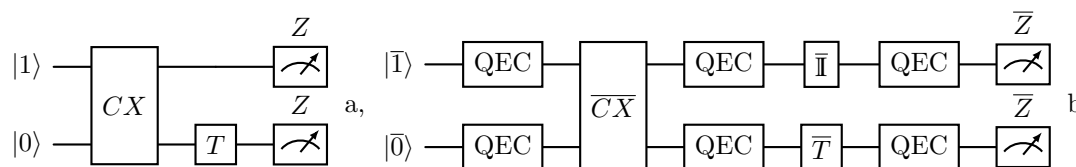
There are some operations that are native to a given error correction code, which are known as *transversal gates*, and these are just multiple copies of the underlying physical operation acting on some or all of the physical qubits of a logical qubit. But, transversal gates alone for any QEC code are *not universal*, which is to say that one cannot run a general quantum algorithm with any potential for advantage using only transversal gates. This result is known as the Eastin-Knill theorem [154], and the reason why only transversal gates cannot give quantum advantage is due to the Gottesman-Knill theorem [155] as they are equivalent to the Clifford group which is classically simulable.

So to implement something like the $\overline{T}$ gate to give an arbitrarily close to universal gateset (guaranteed by the Solovay-Kitaev theorem [32]), one must use alternative techniques. There are a variety of ways to go about this, but one that has been demonstrated experimentally is via the use of magic state distillation [146], which involves the preparation of many of these so-called *magic states* that must be purified and teleported to the logical qubit. This is perhaps one of the most expensive parts of any error correction procedure.

Figure 4.10: **The effect of applying QEC** on circuit C at a, giving the encoded circuit $\overline{C}$ at b. To perform error correction every qubit is encoded and every operation, including the identity, is also encoded. These are called fault-tolerant gadgets and can incur significant overhead if the operation is not native to the error correction code, such as T.

5

Quantum Chemistry on Quantum Computers

'You can't always get what you want, but if you try sometimes you might find you get what you need.'
Rolling Stones, 1969

This chapter will introduce how electronic structure and dynamics are currently being researched in quantum computing with a strong focus on NISQ machines, beginning with a short introduction to the general problems and discussing a brief history and current state of the art quantum algorithms. This is followed up by the particular complexities of studying quantum chemistry on quantum computers. It should be noted that all parts quantum chemistry can be treated in what are called the first and second quantization, but in practice certain phenomena are studied almost exclusively with one or the other, this thesis presents electronic structure in second quantization and then dynamics in first quantization.

QUANTUM CHEMISTRY is one of the areas in which quantum computers are expected to be able to bring novel insight and improved performance in either the accuracy or speed of computations. This is in line with the original idea from Feynman and Manin that one needs a quantum device to simulate the quantum physics which underlies all chemical processes. The field of theoretical and computational chemistry attempt, in different but related ways, to simulate everything from molecules to materials with an optimal tradeoff between the accuracy of the result and its computational cost. Through this it is possible to predict properties of interest or outcomes of various processes, such as in a reaction between two molecules or an interaction such as the absorption of a photon.

Although the equations and numerical techniques exist to solve these problems perfectly, the memory and time requirements for these scale unfavorably with the number of electrons. Even the smallest molecules beyond simple diatomics quickly begin to have such large numbers of electrons that make the dimensionality of the problem beyond the reach of computers without the use of various simplifying assumptions and approximations. Herein lies the art of the fields of theoretical chemistry, which has found multiple tailored approaches that make different compromises to solve very large

problems with lower accuracy or smaller problems with a high level of accuracy.

The knowledge built by the decades of research in this field has shed light on the structure of chemical problems and there is a general understanding of what effects are important for given classes of problems. Quantum computing jumped into this field with quite high expectations that have been tempered over the last few years as it is clear that if the machines do end up working, they will augment the computational capabilities and perhaps also lead to new discoveries thanks to the loosening of computational boundaries and a new approach to encoding problems.

5.1 A Brief History

Although the main driver of NISQ research has been various versions of the VQE introduced in section 3.1, it is important to visit the early work on quantum chemistry which assumed more idealized machines. These were the first algorithms to be designed which hinted at the power of quantum computing, although they are as of now, still unachievable on existing hardware. It is either surprising or unsurprising that the first quantum algorithm to show a theoretical exponential speedup was discovered in 1998 [156] to find eigenvectors and eigenvalues via QPE. This was shortly followed up on with an algorithm for the thermal rate constant [157]. The authors in both works acknowledged the importance of noise and error correction, in 1999 the first ever quantum simulation experiment was run [17] on NMR machines was run, so the extent of this enterprise was as yet unknown.

It was very quickly clear that quantum algorithms would need to be tailored to work with existing machines, and the most promising candidate algorithm was theorized to solve pairing Hamiltonians via the adiabatic theorem [158]. Here, one can also find the first reply from the tensor network community called the claimed advantage of the technique into question. An experiment implementing this algorithm on a NMR machine was published in 2006 [19]. In these early days adiabatic relaxation was expected to be a very promising approach, and algorithms were expected to run on adiabatic quantum computers. The first experiments performed these types of algorithms in what is termed *adiabatic state preparation* on NMR machines for hydrogen was performed in 2010 [18] and on a photonic QC [159]. Of course, digital QCs were still being studied, notably around 2005 early numerical results of the ideas brought forth in [156] and extensions of said algorithm to compute energies [160]. Furthermore, the first algorithms for chemical dynamics using a split-operator approach were also presents [161].

The problem with a quantum annealer or any annealing strategy for finding ground states is that as the energy gap between the ground state and the next lowest lying states becomes small, the adiabatic algorithm must be run proportionally longer such that the

system always stays in the ground state [1]. As such, early reviews and proposals [162, 163] also started considering digital quantum computers more seriously. Some early examples focused on finding efficient gate based algorithms for molecular states that need only polynomially many gates [164] and early ideas for simulating open system dynamics [165].

The algorithm that very directly kicked off a veritable race for quantum chemistry was developed in 2014 and is the VQE [55], which was proposed and experimentally tested for hydrogen. This technique is described in detail in section 3.1, it was hardware efficient, extensible and was thought to be able to converge to the full configuration interaction (FCI) solution. These properties propelled it to be one of the most studied quantum algorithms and generated a whole family of related algorithms known as VQAs [57]. Although quantum annealers were commercially usable in 2011 (from D-Wave), the first easily accessible quantum computers was IBM's Quantum experience in 2016, which was well timed with this development of efficient digital QC algorithms.

Classically the era of VQEs can be compared to variational methods often found in tensor network techniuqes like density matrix renormalization group [166] and variational Monte-Carlo [167]. The advantage was later formalized as coming from the higher potential expressivity of quantum circuits as opposed to low-rank tensor approximaitons [83], hence the early ideas of FCI-like results. However, with a focus on these hardware-friendly algorithms that were meant to be rin came a much stronger focus on the handling of errors and being efficient with the shot noise. It is around this time that it was announced to be the start of the NISQ era [168].

Although research in more fault tolerant algorithms did not stop, the amount of research that has gone into NISQ algorithms alone is immense and can hardly be covered in a single review paper or thesis, but several specific reviews exist from around this time that give an excellent overview when read together [169, 170, 171, 172, 57]. NISQ algorithms also could not be truly useful alone, and here is where the study of error mitigation took a central role with the introduction of ZNE, which was first showcased on a VQE [75]. Section 4.2 is dedicated to error mitigation techniques for more information.

The study of VQE split off into many branches focusing on the different constituent parts of the algorithm. This includes the circuits which aim to be both efficient and physics constrained [173] of which a first example is the ADAPT-VQE [79] introduced in 2019, which itself became a subject of study [174, 175]. A plethora of variants to prepare excited states [176, 177], do real-time dynamics [178, 179, 180] and many other facets.

This explosion of interest is however being dampened by the first hints of fault-tolerant quantum computers could be possible was in 2021 with the first realization of the surface code [181], but beyond break even performance has been achieved only as recently as 2024

[143, 144, 145, 146] and indeed it has been shown that many NISQ algorithms, but specifically VQAs suffer from likely insurmountable problems. These can stem from the noise leading to barren plateaus [111] which stop the optimizer, to noise in the measurements leading to issues with optimization. Not even error mitigation is expected to solve this [182]. As such, the future will likely look more towards making efficient fault tolerant algorithms, which will need to restrain themselves to limitations that the early pioneers did not attempt to do[1].

[1] Although they were aware that this would be the case, and some early papers mention estimating the quantum resources required [183].

5.2 The Electronic Structure Problem in Second Quantization

Many molecular properties arise from electronic structure alone. If one had access to the eigenstates of the Hamiltonian is an important step in the prediction of reaction rates, structural optimization and optical properties [184]. Electronic structure calculations are usually taken under the Born-Oppenheimer approximation (BOA). This is much simpler to solve than the full Hamiltonian, and is where early research in quantum algorithms focused. The question to solve is: given a static nuclear potential, compute the ground state energy E of a Hamiltonian H_{el} acting on a system of interacting electrons. Ψ:

$$H_{el} \left| \Psi \right\rangle = E \left| \Psi \right\rangle$$

The electronic Hamiltonian is usually introduced in its first quantization form, and although first quantization will be revisited in more depth, for now a brief presentation suffices. Taking the ∇ as the partial derivative with respect to the coordinates, $\mathbf{R}_i$ as the position vector of the ith nucleus and $\mathbf{r}_i$ as the position vector for the ith electron in atomic units, the simplest BOA Hamiltonian that is generally considered is, in atomic units, [2]

[2] Many terms such as Zeeman, spin-orbit, spin-spin, orbit-orbit and diamagnetic interactions and relativistic effects are completely ignored [184].

$$H_{el} = -\sum_i \frac{\nabla_{r_i}^2}{2} - \sum_{i,j} \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}_j|} + \sum_{i,j>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (5.1)$$

In this form of the Hamiltonian, of an N particle system, defined as $\Psi(1, 2, \ldots, N) = |1, 2, \ldots, N\rangle$ must be expanded into a complete set of symmetric and antisymmetric wavefunctions to solve the Schrödinger equation. Thus, to prepare the molecular state one must encode the appropriate symmetries in the wavefunction directly. Several algorithms exist that achieve this, but current publications favour encoding fermionic statistics directly into the operators. This is known as the second quantized representation. The form of the Hamiltonian from Eq. (5.1) in second quantization requires a definition based fermionic creation and annihilation operators $a^\dagger, a$ which will be described in more detail in the next section, but for now the following equalities and definitions should be kept in mind

$$H = \sum_{ij} h_{ij} a_i^\dagger a_j + \frac{1}{2} \sum_{ijkl} h_{ijkl} a_i^\dagger a_j^\dagger a_k a_l \quad (5.2)$$

with $h_{ij} = \int \phi_i^*(\mathbf{r}) \left(\frac{\nabla_{r_i}^2}{2} - \sum_i \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} \right) \phi_j(\mathbf{r}) d\mathbf{r}$,

$$h_{ijkl} = \int \frac{\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_k(\mathbf{r}_1)\phi_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2$$

and $\quad \{a_i^\dagger, a_j\} = \delta_{i,j} \quad \{a_i^\dagger, a_j^\dagger\}_+ = \{a_i, a_j\}_+ = 0 \quad$ (5.3)

h_{ij} and h_{ijkl} are the one and two electron integrals over the nth spin orbital $\psi_n(r_i)$ occupied by electron i which is expressed through its spatial degrees of freedom, $\{\cdot, \cdot\}$ is the anticommutator[3] . These N orbitals are generated through a classical mean field theory, such as Hartree-Fock and act as a basis for the molecular system. This Hamiltonian contains N^4 terms, which will become relevant later. Unlike the first quantization Hamiltonian, the correspondence of the second quantization representation with qubits is evident in the form of the wavefunction, which is a number state in the form $|n_{p_1}, n_{p_2}, \ldots, n_{p_N}\rangle$. Here n_{p_i} represents the number n of particles with property pi. In the qubit case, n can be a seen as an occupation of an orbital: $\alpha|0\rangle + \beta|1\rangle$. The creation and annihilation operators are also simply representable in the qubit basis as $\sigma^+ = |1\rangle\langle 0|, \sigma^- = |0\rangle\langle 1|$. When dealing with a fermionic qubit, this would be sufficient to represent $a^\dagger, a$ but in the case of superconducting qubits one must enforce fermionic parity. To this end, the operator must be encoded through mappings such as the Bravyi-Kitaev (BK) or the Jordan-Wigner (JW) transformations [66, 70].

An informative exercise is to go through a simple example for hydrogen with a minimal basis set. If one pictures the molecular diagram as in Figure 5.1 for hydrogen then we know there is a σ and σ^* bonding/anti-bonding orbital. The electrons may occupy one of two orbitals, with spin being denoted by an arrow - a total of four states: $\{|\sigma_\uparrow\rangle, |\sigma_\downarrow\rangle, |\sigma_\uparrow^*\rangle, |\sigma_\downarrow^*\rangle\}$. Together these states can be combined in six ways (due to the Pauli exclusion principle forbidding the same quantum state for more than one fermion). These six combinations have total spin numbers $S_{\uparrow\uparrow} = S_{\downarrow\downarrow} = 1, S_{\uparrow\downarrow} = S_{\downarrow\uparrow} = 0$. To reduce the amount of qubits we choose to only focus on the $S = 0$ (singlet) states, which leave us with only four fermionic states which we may map onto two qubits as follows:

Figure 5.1: The molecular diagram for hydrogen. The σ and σ^* orbitals are shown. The electrons may occupy one of two orbitals, with spin being denoted by an arrow.

$$|\sigma_\uparrow \sigma_\downarrow\rangle \to a_{\sigma\uparrow}^\dagger a_{\sigma\downarrow}^\dagger |\text{vac}\rangle \to |00\rangle \quad |\sigma_\uparrow \sigma_\downarrow^*\rangle \to a_{\sigma\uparrow}^\dagger a_{\sigma^*\downarrow}^\dagger |\text{vac}\rangle \to |01\rangle$$
$$|\sigma_\uparrow^* \sigma_\downarrow\rangle \to a_{\sigma^*\uparrow}^\dagger a_{\sigma\downarrow}^\dagger |\text{vac}\rangle \to |10\rangle \quad |\sigma_\uparrow^* \sigma_\downarrow^*\rangle \to a_{\sigma^*\uparrow}^\dagger a_{\sigma^*\downarrow}^\dagger |\text{vac}\rangle \to |11\rangle$$

If one applies the Hamiltonian in (5.2) and transforms it using any parity mapping it is possible to arrive to a qubit Hamiltonian H_q that consists of the integral prefactors followed by a series of Pauli matrices. By exploiting symmetries in the one and two electron integrals and discarding any 0 values the potential N^4 measurements that appear to be necessary to perform can be reduced to as few as five [185] [186]. In the now also used BK transformation, which is carried through explicitly in Seeley et al. [187], this results in a Hamiltonian

$$H_{qH_2} = c_0 \mathbb{I} \otimes \mathbb{I} + c_1 \sigma_z^1 \otimes \sigma_z^2 + c_2 \sigma_z^1 \otimes \mathbb{I} + c_3 \mathbb{I} \otimes \sigma_z^2 + c_4 \sigma_x^1 \otimes \sigma_x^2 1 \quad (5.4)$$

where c_i are coefficients given by the overlap integrals h_{ij} and h_{ijkl} in (5.2). Each therm represents a measurement that must be taken on the machine and although any commuting observables can be read out on the same run, the quantum state degrades between these operations. Thus, it is important to reduce the total number of measurements.

5.2.1 Fermion to Qubit Mappings

There is some choice on how the fermionic creation and annihilation operators are translated to spin operators as those found in many classes of qubits. As with any mapping there are tradeoffs to be made which impact the practical cost of implementing them. The oldest and most well known has already been introduced in passing and is the Jordan-Wigner mapping [70]. Originally it was proposed to map spin operators onto fermionic ones, but it was inverted to use in quantum computing applications.

Any system with N fermionic modes has a set of annihilation operators of the same number which are commonly denoted a_i for $= 0, ..., N - 1$ which satisfy the anticommutator relations presented in 5.3

This leads to some interesting consequences, such as the fact that $[a_i^\dagger a_i, a_j^\dagger a_j] = 0$ and have eigenvalues 0, 1 - this leads to the occupation number operator which is often used. There is also a so-called vacuum state denoted $|\text{vac}\rangle$ which the same as a 0 occupation number on all modes. On a quantum computer this is commonly going to be the $|0\rangle^{\otimes N}$ state. But perhaps the two most pertinent consequence is that the sign of the wavefunction changes upon annihilation (or creation) of two fermions (and thus also for exchanges, which are composites of creation and annihilation operators) as

$$a_i |\ldots, n_{i-1}, 1, n_{i+1}, \ldots\rangle = (-1)^{\sum_{j=0}^{i-1} n_j} |\ldots, n_{i-1}, 0, n_{i+1}, \ldots\rangle, \quad (5.5)$$

$$a_i |\ldots, n_{i-1}, 0, n_{i+1}, \ldots\rangle = 0. \quad (5.6)$$

Furthermore, any $(a_i^{(\dagger)})^2 = 0$, which means that for a given mode one cannot annihilate or create a fermion twice. All of these are direct consequences of the anti-commutator relations, but are perhaps the most complex to map as these are not the properties of spin chains, which are distinguishable and do not follow these statistics.

Algorithm 2: Fermion to qubit mapping checklist

---

1 If the qubit at $n_i = 0$, return $|0\rangle^{\otimes n}$ due to an invalid operation.

2 Else $n_i = 1$ and this can be flipped to be $n_i = 0$, then

3 potentially update all other bits, n_j, for $j \neq i$, depending on the encoding rules.

4 Finally update the parity by some function of $\sum_{j=0}^{i-1} n_j$

Then there is a simple algorithm any mapping must perform with any computational statevector $|\cdot\rangle$ its mapped fermionic annihilation operator a_i as shown in algorithm 2.

THE JORDAN-WIGNER TRANSFORM [70] maps the annihilation operator as

$$a_i \to \frac{1}{2}(X_i + iY_i)Z_0 \dots Z_{i-1}. \tag{5.7}$$

This transformation leads to comparable basis vectors in the qubit computational basis (Z) in exactly the same way as (5.5) and (5.6) by redefining a_i to what is shown in 5.7. One important thing to note is that the fermionic operators are local for both occupancy (the flipping of a 0 to a 1) and in parity (the whole state doesn't need to be acted on for the sign to switch). In the JW transform is still local in the occupancy, the chain of Z operators in 5.7 is non-local. This is a problem since it necessitates many gates to implement or more measurements to estimate the mapped expectation values.

In terms of the algorithm 2 checklist, steps 1,2, and 3 are represented by $X_i + iY_i$, which is more evident when one realizes that the operator may be written as the projector $(|0\rangle\langle 1|)_i$, where the subscript i is the index of the qubit. The parity of step 4 is maintained by the string of Z operators, and this step is linear in the number of operations with the number of qubits. It is also linear in the number of sites as the number of qubits. It should be noted that there is a parity transform that is instead linear with the number of occupations but constant in parity operators (in essence the 'opposite' of the JW).

THE BRAVYI-KITAEV TRANSFORM [66, 188] is an attempt to find a better than linear scaling. This is the case, but there is an initial overhead that makes this technique only practically useful at more than around 32 orbitals, and specifically at powers of 2 [189].

This was done by working with the Majorana operators $c_i = a_i + a_i^\dagger$ and $d_i = -i(a_i - a_i^\dagger)$ with some algebra the annihilation operator can be written as

$$a_i = \frac{1}{2}(c_i + id_i), \tag{5.8}$$

with the adjoint merely flipping the sign of the second term. The action of both c_i and d_i on a fermionic string are identical to a change in phase brought about by the imaginary part in the parity multiplier. This is then for c_i going to be

$$c_i |\dots, n_{i-1}, n_i, n_{i+1}, \dots\rangle = (-1)^{\sum_{j=0}^{i-1} n_j} |\dots, n_{i-1}, 1 - n_i, n_{i+1}, \dots\rangle. \tag{5.9}$$

This operator combines both a flip of the occupation site i and also multiplies it by the parity. Since the action is more akin to an update of the occupation rather than a flip, steps 1,2 and 3 of algorithm 2 can be combined to be one step that updates a string

$|\cdot\rangle \rightarrow |\cdot'\rangle$ and then updates the parity. Thus, the mapping can be written as

$$c_i \rightarrow X_{F(i)} Z_{P(i-1)} \tag{5.10}$$

where $F(i)$ is a precomputed set of bits that must be flipped given an index and $P(i-1)$ is the same for the parity. For JW and the parity transform these are $F(i) = i and P(i) = 0, \ldots, i-1$, for example. The opposite goes for the parity transform. But for the brilliance of the BK it was discovered that using a Fenwick tree [190], which is a type of indexed binary tree, both could be encoded with $O(log(N))$ qubits. The specifics of the encoder are important for the optimal depth, but the true insight comes from looking at the problem through the lens of Majorana operators and this simpler string update operator to an occupation change.

Now, since then there have been other types of transformations that have been discovered and studied. Although a short table is provided to allow a quick comparison of the techniques discussed, other works should be consulted when choosing a mapping for an actual problem, for example the work of O'Brien et al. [72] which is currently the most efficient mapping, but various approaches exist [189, 71, 191, 192]. It should be noted that some mappings lead to controlled unitaries being required, and potentially less qubits than the one-to-one qubit occupancy tradeoff in the examples given, making them interesting but potentially inefficient for near term devices.

Table 5.1: Basic Fermion to Qubit Mappings

Name	Occupancy	Parity	Gate type
JW [70]	$O(1)$	$O(N)$	Pauli
Parity [187]	$O(N)$	$O(1)$	Pauli
BK [66, 188]	$O(\log N)$	$O(\log N)$	Pauli

5.3 First Quantization

Although traditionally first quantization is briefly introduced before second quantization, most research in both classical and quantum quantum[4] chemistry focuses on electronic structure (and sometimes dynamics) in second quantization. This might not necessarily always be so, and certainly it is not the case for nuclear dynamics , but regardless, a small reminder of first quantization methods is required. In first quantization the parity is not held within the operators' action on a particular wavefunction, but the wavefunction is explicitly anti-symmetrized to ensure the exchange symmetry. This leads to a (discretized) representation, where for simplicity, one can assume that each spatial dimension is discretized into P points and the system contains N particles.

$$|\Psi\rangle = \sum_{x_1, \ldots, x_N}^{max} c_i(x_1, \ldots, x_N) A(|x_1, \ldots, x_N\rangle) \tag{5.11}$$

[4] Sorry, this was necessary.

is the wavefunction with each position vector

$$|x_i\rangle = |x_i, y_i, z_i, s_i\rangle \begin{cases} i \in \{1, ..., N\} \\ x_i, \ y_i, \ z_i \in \{1, ..., P\} \\ s_i \in \{0, 1\} \ \text{for electrons} \end{cases} \tag{5.12}$$

representing the 3D position in space x, y, z on the discretized grid and the particle spin s. Although nuclei do have spin, this is generally ignored. A is then the anti-symmetrization. This leads to a very large set of amplitudes required to describe the wavefunction of $2^N P^{3N}$. Of course, in practice one can confine nuclei to a much smaller (but potentially denser) space in which they can move and techniques like discrete variable representation can be used to greatly reduce the number of grid points required in the classical case [193].

That being said, usually a full nuclear-electron grid representation is not done in practice, and in techniques for nuclear quantum dynamics such as the multi-configurational time-dependent Hartree (MCTDH) family of approaches [23] generally precompute electron potential energy surfaces and focus on the nuclear dynamics. Nonetheless, in this high level picture one must apply the full Hamiltonian or the electronic Hamiltonian or just the nuclear Hamiltonian depending on which system or subsystem one wishes to study. An alternative representation in first quantization also uses a proper basis much like in second quantization, with plane waves being particularly studied in the QC literature [194].

5.3.1   Mapping the first quantization representation

There are two main ways one can think of when mapping the first quantization representation to a quantum computer. For a grid based mapping it is sufficient to assign each grid point to a string of qubits which immediately leads to representing the entire state with $3N \log_2(P) + N$ qubits. There is no doubt that there are clever ways to decide which grid point to map to which qubit string to simplify operations, but in essence one simply needs to perform some mapping $|\boldsymbol{x}_1, \ldots, \boldsymbol{x_N}\rangle \rightarrow |q_1, \ldots, q_{(3N \log_2(P)+N)}\rangle$, remembering that the position vectors themselves have to be expanded, q_j, $j \in \{1, ..., (3N \log_2(P) + N)\}$ is the qubit state which can be 0 or 1.

Next, and important step and a great advantage to using the first quantization approach is that although there is an associated cost to anti-symmetrize (or symmetrize) the wavefunction, once it is prepared, the action of any valid Hamiltonian will preserve the symmetry property, removing the expensive overhead brought by the various mappings required in the second quantization case. This requires on the order of $O(Q \log Q \log \mathcal{P})$, where Q is the number of qubits used and $\mathcal{P}$ is the total number of grid points or basis functions [195]. This is in general a much shorter procedure than any following valid time evolution or operation that might follow the initial setup of the (anti)symmetrized state.

5.4 Some Algorithmic Primitives

5.4.1 State Preparation

Quantum state preparation is the first step and sometimes only step, besides the measurement of observables, in many problems that deal with studying specific states. This is definitely the case with ground states, since excited states sometimes are found as an evolution from the ground state [177]. The most widely used algorithm is the VQE[5]. Almost all other algorithms for near term state preparation are variants or extensions thereof [196].

[5] See section 3.1

However, the feeling amongst many researchers, including one of the creators of the VQE algorithm, is that it is probably best to leave this technology behind. Indeed, for larger scale experiments it appears that newer subspace expansion and sampling based techniques are taking over for larger experiments with a much larger dependence on supercomputing power, in particular sample-based quantum diagonalization [197, 198] is now able to solve for active spaces of 36 orbitals on 77 qubits, and their electronic excitations.

Other alternatives that can, but do not have to, use something like VQE include imaginary time evolution

5.4.2 Dynamics

Chemical dynamics is an exciting field which often touches the edge of computability. This is doubly true when the BOA does not hold, as here so-called non-adiabatic transitions can occur, as well as coupling of the electrons to nuclear motion. The first step in any such system is to prepare an initial state on which the dynamics play out.

Although a lot can and has been said about state preparation in the previous sections and chapters, sometimes it is interesting or important to do something with the initial state. This is then the realm of dynamical simulation - where many techniques are shared with state preparation algorithms and are generally referred to as *Hamiltonian simulation* algorithms [199].

If one begins by the assumption that most Hamiltonians are not dense in their Pauli string representation, which is to say that when decomposing a Hamiltonian H as $H = \sum_i h_i P_i$, where P_i is a tensor product of Pauli operators and h_i is a real number, most will be Identity. This ensures that many approximations hold better, and a very common one is the so called Trotterization, which is a unitarization of a sum of products that can be implemented as a gate.

Trotterization is a fundamental approximation in dynamics as it is impossible to implement a non-unitary operation on a quantum computer. The specific proof as to why the approximations work is available [200], but it is known that the traditional time evolution operator for a Hamiltonian has the form e^{-iHt}, but H is a sum of potentially non-commuting matrices. If this point is brushed aside, the exponentiation of a matrix is then $e^A = U e^D U^\dagger$, where U is

some unitary and D is the diagonalization of A, which is inefficient. But by the definition of the exponential and for large N it can be said that $e^{A/N} \approx I + \frac{A}{N}$. Then some sum of non-commuting matrices (two are shown as an example) can be approximated as

$$e^{A+B} \approx \left(e^{A/N}e^{B/N}\right)^N \approx \left[\left(I+\frac{A}{N}\right)\left(I+\frac{B}{N}\right)\right]^N \approx \left[I+\frac{A+B}{N}+\cancelto{0}{\frac{AB}{N^2}}\right]^N.$$
(5.13)

This fact was exploited for a long time to theoretically understand the action of Hamiltonians but also in the earlier days of quantum computing [201]. The actual form that used in practice is the Suzuki-Trotter formula [202] in various orders, but for the near term the first order approximation is mostly used, which gives a unitary U for a Hamiltonian with K terms,

$$U(t) = \left(\prod_i^K e^{-i\frac{t}{N}H_i}\right)^N.$$
(5.14)

From here there is a standard construction that leads to simple looking but deep circuits of a form as shown in figure 5.2.

From the trotter decomposition there are many tricks that can be played to get lower depths. The first is to optimize the order of the Pauli terms in the sum so that as many gates cancel out as possible [203], this can also include optimizing the order of the approximation [204]. One can also probabilistically choose what terms to use over many simulations forming more of a *time evolution channel* than an approsimate unitary evolution in a technique called qDRIFT [205] and a variant known as qSWIFT [206].

Of course, many other techniques also exist in this area. Many are variational, either exploiting a variational principle such as adaptive Hamiltonian simulation [207], and others taking more complicated approaches such as variational fast forwarding [208]. Then there are approximate diagonalization techniques such as Variational Hamiltonian Diagonalization [209]. This topic will be discussed in greater depth in chapter 6.

Figure 5.2: **One term of a Suzuki-Trotter circuit** on 3 qubits, showing the general ladder construction and parameterized rotation. The Unitaries are a rotation that depends on which Pauli string is applied, and CNOTs may be between different qubits depending on the weight of the Pauli string.

5.5   *Problems and Pathways*

If the aim is a perfect simulation of electronic structure, the closest numerical approximation we have is the FCI. This is a problem which, as mentioned before, grows so quickly that the current maximum number of active electrons and orbitals that has been simulated is 24 for only a single iteration, a simulation requiring close to 10 trillion Slater Determinants [210]. Such growth in size is what limits the level of study to small molecules or single atoms. As such the gold standard of electronic structure calculations is the coupled cluster method (CC) with more excitations being more accurate. Here, a reference wavefunction $|\Phi\rangle$ is acted upon by the exponential operator T based on a number of electron excitations, leading to a
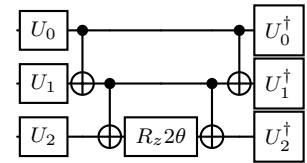
variational ansatz $e^T |\Phi\rangle$. In the classical case, T is not necessarily Hermitian nor unitary and the resulting expression is not guaranteed to be amenable to the variational method. In quantum computers it has been shown that it is possible to efficiently prepare a UCC operator [211], something that is intractable classically [212], as the necessary truncation of the expression $(e^{T-T\dagger} = e^{-iH})$ leads to uncontrollable errors. So, in electronic structure at least, there is a very clear standard that can be improved upon, and for ground state calculations at least, it is possible with certainty to say when one has overcome classical simulation techniques for the time. Of course, classical simulation can always improve.

However, it must always be remembered that for many of these electronic structure techniques the Hamiltonians are still generated via classical methodology and matrix elements are computed from classical basis functions that themselves are approximations [184]. Perhaps it will be the case that everything will be quantum from the computation of the matrix elements to state preparation and subsequent dynamics. For now though, the darling of quantum computing, VQAs is finding a lot of resistance and potentially insurmountable problems for as long as there is noise. This has been spoken about before, but noise induced barren plateaus [60] which get worse as system sizes increase will be problems for larger VQAs. Alternatives have since been developed, such as sample-based quantum diagonalization [197].

For dynamics the question is more difficult, without good state preparation one cannot truly do useful dynamics. For certain classes of Hamiltonians with low weight Pauli strings it is possible to have quite efficient algorithms, and indeed, with good error mitigation strategies, some of the largest so-called utility scale quantum computations have been performed with various flavours of Ising models [213, 214], but chemical simulations are still beyond the scope of having advantage over state-of-the-art classical techniques like multi-layer multi-configurational time-dependent Hartree [215, 22].

Part II

Research Work

6

Mixed Quantum-Classical Dynamics for the NISQ era

Everything changes and nothing stands still.
Heraclitus, as quoted in Plato's Cratylus, 402a.

This chapter presents the construction and study of a newly proposed NISQ algorithm for chemical dynamics in first quantization in a mixed quantum classical setting. It begins with an introduction to the problem it is trying to solve, followed by the algorithm and its ability to solve a model system. Results are then presented and discussed. The work was published in [180].

Quantum computers have found great success in electronic structure theory through the variational quantum eigensolver [55] and subsequent algorithms known as variational quantum algorithms (VQAs) [216]. Adding additional electrons to a system greatly increases its complexity, this is something quantum computers are expected to handle. If one were to also consider the full nuclear dynamics, the problem becomes unmanageable much faster, potentially even for quantum computers [217]. One way to reconcile this is to partition the system into interacting quantum and classical parts. This is the realm of mixed quantum-classical (MQC) approaches, which are a widely used set of tools for understanding chemical systems [218, 219]. In quantum computing, this area is less researched than the electronic structure problem, but it is actively being explored [220, 217, 221]. Here, a NISQ friendly algorithm that can be used to study mixed quantum-classical (MQC) dynamics is presented and explored.

Using quantum computers alongside classical computers is the backbone of VQAs, but splitting a system into sections treated separately by each machine is not new. A DFT embedding scheme with a quantum computer expansion of the active space [222] and has been found to outperform certain types of state-of-the-art approximate techniques such as CASSCF [223] in finding ground state energies. Furthermore, ground state dynamics, geometry relaxation, and force measurements for MD applications have been explored

Contributions: This work was ideated together with and discussed at length with Oriol Vendrell. Although the majority of the writing, development, simulation and algorithmic ideation was carried out by myself, the use of the Shin-Metiu model in particular and many small technicalities would have not been noticed or resolved if it were not for the advice of Oriol.

with success in [221]. In [217], dynamics are explored in both first and second quantization, but using a time-independent Hamiltonian. This is also the case for various other time propagation techniques [224, 225, 226]; drawing inspiration from and use projected variational quantum dynamics (p-VQD) [225] in this work.

Here, a general algorithmic structure to tackle non-adiabatic molecular dynamics (NAMD) by offloading the QM part to a quantum computer and evolving the classical system by the Ehrenfest method is presented. Observables from the quantum mechanical (QM) subsystem are measured and used to update the classical system, which in turn will update the time-dependent Hamiltonian that is used to evolve the QM state in turn. This is all done in first quantization, which saves the algorithm from needing to measure nonadiabatic couplings, as these are treated directly within the wave function and its evolution in this setting. The algorithm is demonstrated in the Shin-Metiu model [26], which is often used to test various non-adiabatic techniques [227, 228, 229, 230]. This is modified to be a NAMD-like problem by partitioning the system into a classical nucleus and quantum electron. The major contribution is the study of how the interaction between observable measurement and system updates play out as well as introducing a scheme that is suited to begin exploring other time-dependent phenomena on NISQ machines.

The theoretical advantage of using quantum computers is that they have access to an exponentially growing computational space for each additional qubit in the system [1]. Current machines have access to hundreds of qubits, which would ideally allow them to already outperform current supercomputers. This is not the case due to noise coming from interactions with the environment and imperfect gate implementations. As such, NISQ algorithms [231] have to contend with limits on the number of imperfect operations that can be made. But even if this were not the case and full quantum dynamics could be simulated, researchers probably will always want to tackle a problem bigger than current machines can handle, so these kinds of approximations will always be used.

It should be noted that existing supercomputers by far outperform existing quantum computers in handling large quantum-chemistry problems, and applications to chemical problems will have to be deferred until there is a provable quantum advantage. Approximations like limiting the simulation to a selected active space [222] can extend the reach of quantum computers, but analogues of these ideas apply to classical computers as well. For now, quantum algorithms in chemistry mostly study systems that are comfortably computable on current classical hardware.

6.1 *The Shin-Metiu Model*

The Shin-Metiu model is a numerically exactly solvable minimal model which captures essential nonadiabatic effects [26]. It is often
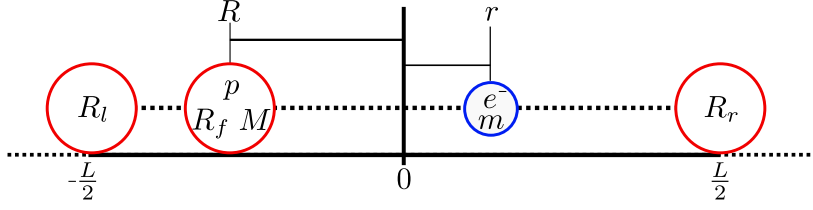
Figure 6.1: **Illustration of the Shin-Metiu model** with fixed ions at $\frac{-L}{2}$, $\frac{L}{2}$ as stationary boundaries, the mobile ion p of mass M at distance R from the origin and the electron e^- at distance r from the origin. R_l, R_r and R_f are constants for the regularized Coulomb potential in (6.1). Reproduced from [180] under CC4.

used as a benchmark system for new techniques and is used to study the effects of different environments as has been done for polaritonic dynamics, coupling to cavities, and the effect of electromagnetic fields [227, 228, 229, 232]. It is simple to change its parameters for it to exhibit adiabatic to strongly non-adiabatic dynamics.

In its simplest and original conception, the model shown in Fig. 6.1 consists of two stationary ions separated by a distance of L, specifically located at $\frac{L}{2}$ and $\frac{-L}{2}$. These enclose a mobile ion p of mass M at distance R from the origin and an electron e^- at distance r. The modified Coulomb potential is parameterized by the constants R_l, R_r and R_f, as shown in Eq. 6.1. This is done to avoid singularities and make the system numerically simpler to simulate.

The full Hamiltonian of the system is

$$H = -\frac{1}{2M}\frac{\partial}{\partial R^2} + H_e(r, R),$$

with the electronic part being

$$H_e = -\frac{1}{2m}\frac{\partial^2}{\partial r^2} + \frac{1}{\left|\frac{L}{2}-R\right|} + \frac{1}{\left|\frac{L}{2}+R\right|} - \frac{\mathrm{erf}(\left|\frac{L}{2}-r\right|/R_r)}{\left|\frac{L}{2}-r\right|} - \frac{\mathrm{erf}(\left|\frac{L}{2}+r\right|/R_l)}{\left|\frac{L}{2}+r\right|} - \frac{\mathrm{erf}(\left|R-r\right|/R_f)}{\left|R-r\right|}. \tag{6.1}$$

The equation uses atomic units, setting $e = Z = \hbar = 1$, in addition $m = 1$ and $M = 1836$ in the simulation. The constants R_l, R_r and R_f, as shown in Figure 6.1 are chosen to create specific adiabatic surfaces with interesting transitions, as in Figure 6.2.

The values $R_f = 5.0$, $R_l = 4.0$ and $R_r = 3.2$ are chosen, which result in an avoided crossing around $R = -1.9$ when the distance between the ions is $L = 19$. These parameters were chosen to be similar to those used in several studies of the model [230, 227]. The shape of the Born-Oppenheimer potential energy surfaces (BOPES) can be seen in Fig. 6.2

6.1.1   Ehrenfest propagation of the Model

To perform Ehrenfest propagation of the Shin-Metiu model, the system is split in two. The nucleus (p) and the electron (e^-). The electron subsystem is treated as a quantum particle described by Hamiltonian 6.1, where H_e is parameterized by the nuclear position (R). The nuclear subsystem is treated classically by tracking parameters of position (R) and velocity ($\dot{R}$).
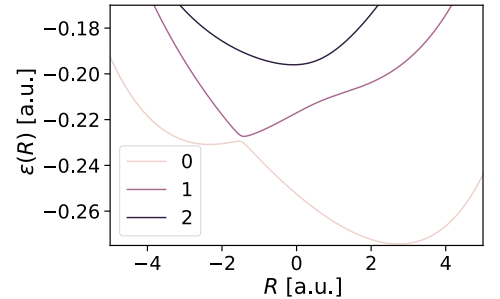

Figure 6.2: **Potential energy surfaces of the Shin-Metiu Model** for the coefficients R_f=5.0, R_l=4.0 and R_r=3.2 showing the avoided crossing around position R=−2 with L=19. Reproduced from [180] under CC4.

For initial coordinates R_0 and $\dot{R}_0$, first prepare the electronic Hamiltonian $H_e(R_0)$, which is used to compute the initial state of the electron $|\psi_0\rangle$. Thanks to the simplicity of the model, exact diagonalization can be used to compute the eigenvectors and choose any arbitrary superposition of eigenvectors as the initial state.

The nucleus is evolved using the velocity Verlet method [233] with the acceleration being computed from the Coulombic repulsion from the fixed ions and the force from the electronic state. The electronic state is evolved by unitary time evolution with the Hamiltonian at the nuclear position. Setting a timestep Δt, the system state at timestep i is denoted by under scripts i, where the time is simply $i \cdot \Delta t$. Initial conditions are defined at timestep 0, and for the ith step, compute

$$F_e(R, |\psi\rangle) = - \langle \psi | \frac{\partial H_e(R)}{\partial R} | \psi \rangle, \tag{6.2}$$

$$R_i = R_{i-1} + \dot{R}_{i-1}\Delta t + \frac{F_e(R_{i-1}, |\psi_{i-1}\rangle)}{M}\Delta t^2, \tag{6.3}$$

$$|\psi_i\rangle = e^{-iH_e(R_{i-1})\Delta t} |\psi_{i-1}\rangle, \tag{6.4}$$

$$\dot{R}_i = \dot{R}_{i-1} + \frac{F_e(R_{i-1}, |\psi_{i-1}\rangle) + F_e(R_i, |\psi_i\rangle)}{2M}\Delta t. \tag{6.5}$$

6.2   Time-Dependent Variational Quantum Propagation

The TDVQP algorithm builds on the circuit compression idea of p-VQD [225] by allowing the Hamiltonian to be time-dependent. For many large problems of interest to theoretical chemistry, especially in MD, it is impossible to fully simulate the system of interest quantum mechanically. As such, the system is subdivided into classical and quantum components. The evolution of both systems occurs in locked steps, with the classical system defining the Hamiltonian for the quantum evolution, and the quantum system then feeding back into the classical system in the way of some observable, usually the energy gradient (force). One mustn't limit themselves to molecular or even physical systems, as this algorithm would work with any set of observables that can be used to update the classical system of interest.

The algorithm begins with a parameterized circuit initialized to some desired state. This is denoted as $|\psi_0\rangle$, which will have been generated according to a Hamiltonian based on an initial vector of classical parameters $\boldsymbol{q}_0$ of the classical coordinates, denoted $H(\boldsymbol{q}_0) := H_0$. This is done by choosing some sufficiently expressive parameterized circuit ansatz $\hat{C}(\boldsymbol{\theta})$ which takes the quantum computer's initial state, denoted $|0\rangle$, to $|\psi_0\rangle$. This can be done using a VQA to find a chosen state with respect to H_0, which returns the circuit parameters $\boldsymbol{\theta}_0$. Then $|\psi_0\rangle = \hat{C}(\boldsymbol{\theta}_0) |0\rangle$. A chosen set of observables $\{\hat{O}^{(s)}(q_0)\} := \{\hat{O}_0^{(s)}\}$ are measured from $|\psi_0\rangle$, which yield a set of expectation values $\{O_0^{(s)}\}$. These observables are used to evolve the classical state of the system, generating a new vector of classical parameters $\boldsymbol{q}_1$. These can then be used to generate H_1 and

$\{\hat{O}_1^{(s)}\}$. Now, one evolves the state from $|\psi_0\rangle$ to $|\psi_1\rangle$ by applying the time evolution operator to the state, $|\psi_1\rangle = \exp\left(-i\hat{H}_0\Delta t\right)|\psi_0\rangle$.

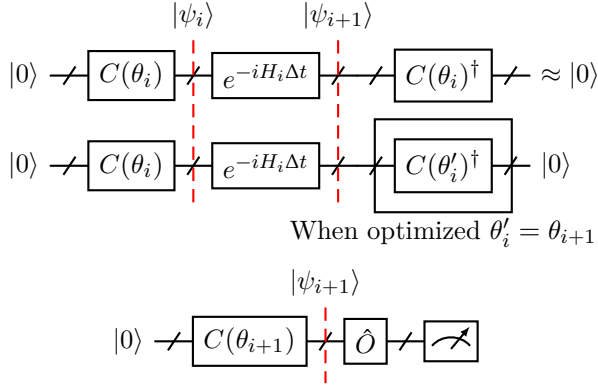When optimized $\theta_i' = \theta_{i+1}$

Figure 6.3: **Sketch of the TDVQP process with slices** showing the state after each gate at the initial condition and final condition. The initial guess for the parameter vector θ_i' is θ_i and its final value is denoted θ_{i+1}. Then observables $\hat{O}$ can be measured on $|\psi_{i+1}\rangle$ to update the Hamiltonian or the classical state of the system. Reproduced from [180] under CC4.

The physical implementation of the time evolution can either be the Trotterized form of the operator or some other approximate time evolution. This work uses $\hat{H}_0$ to evolve the state with a first-order trotter expansion. The Hamiltonian used for the time evolution could be of higher order. For example, $(\hat{H}_0 + \hat{H}_1)/2$ can be used with no extra cost in this scheme, but higher-order integrators will require an additional evolution and observable measurement for each timestep beyond H_1. In return, one gets higher-order symplectic integration. Now a single step of the p-VQD algorithm is applied which generates the new circuit parameters θ_1 such that $|\psi_1\rangle \approx \hat{C}(\theta_1)|0\rangle$ to some desired threshold. This process is repeated until the desired timestep is reached. The entire process is more precisely described in Algorithm 3, and a depiction of the quantum circuit can be seen in Fig. 6.3. The overall number of circuit evaluations is linear with respect to the number of iterations, circuit parameters, timesteps and Pauli terms of the observables. This is treated in greater depth in 6.2.2.

TDVQP should be thought of as a meta-algorithm that has re-placeable components. The most directly replaceable part is the choice of ansatz $\hat{C}(\boldsymbol{\theta})$, which at the moment is generally a heuristic choice for most problems in NISQ devices. More advanced ansatze such as the family of adaptive ansatze, which changes the ansatz throughout the evolution would work, but could not use the previous step's θ parameters as effectively. The very costly time evolution is currently a Trotterized form of the time evolution operator, as in this work, and in [225, 226]. This can be replaced by a plethora of more NISQ-friendly time evolutions as is done in [234, 208] if the form of the Hamiltonian allows this. The limit is the no-fast forwarding theorem [235, 199, 236], which states that one cannot achieve a time evolution of time t in a sublinear gate count for a general Hamiltonian, but for shorter time evolutions, limited sizes and specific cases of Hamiltonians, including the sparse Hamiltonian using short time evolutions, this likely is not the case [208].

In the classical evolution, the choice of integrator and the actual Hamiltonian used in the time evolution will depend on the type of

---

Algorithm 3: Time-Dependent Variational Quantum Propagation (TDVQP)

Input: $\hat{H}_{\text{gen}}(\boldsymbol{q})$, $C(\boldsymbol{\theta})$, $\{\hat{O}_{\text{gen}}^{(s)}(\boldsymbol{q})\}$, `EvalCost`, $\boldsymbol{q}_0$

1 Function VQE *($\hat{H}_{gen}(\boldsymbol{q})$, $C(\boldsymbol{\theta})$, $\boldsymbol{q}$)*:

2 $\boldsymbol{\theta}' = \min_\theta$ `ExpectationValue` $((\hat{H}_{\text{gen}}(\boldsymbol{q}), C(\boldsymbol{\theta})))$;

3   **return** $\boldsymbol{\theta}'$

4 Function UpdateParameters *($\boldsymbol{O}(\boldsymbol{q})$, $\boldsymbol{q}$)*:

5 $\boldsymbol{q}' = $ `UpdateFunction` $(\boldsymbol{q}, \boldsymbol{O}(\boldsymbol{q}))$;

6   **return** $\boldsymbol{q}'$

7 Function UpdateAngles *($C(\boldsymbol{\theta})$, $\hat{H}$, Δt)*:

8 $\boldsymbol{\theta} = \min_{\boldsymbol{\theta}_i'}$ `EvalCost` $(C(\boldsymbol{\theta}_\text{i}')\exp(i\hat{H}\Delta t)C(\boldsymbol{\theta}_\text{i})^\dagger)$;

9   **return** $\boldsymbol{\theta}$

10 Function MeasureObs *($\{\hat{O}_{gen}^{(s)}(\boldsymbol{q})\}$, $C(\boldsymbol{\theta})$)*:

11   for s in number of observables **repeat**

12 $O^{(s)} = $ `ExpectationValue` $(\hat{O}_{\text{gen}}^{(s)}(\boldsymbol{q}), C(\boldsymbol{\theta}))$;

13   **return** $O_i^{(s)}$

14 Function *TDVQP($H_{gen}(\boldsymbol{q})$, $C(\boldsymbol{\theta})$, $\boldsymbol{q}_0$)*:

15   store all $\boldsymbol{\theta}_i$, $\{O_i^{(s)}\}$, $\boldsymbol{q}_i$ in arrays $\boldsymbol{\theta}$, $\boldsymbol{O}$, $\boldsymbol{q}$;

16 $\boldsymbol{\theta}_0 = $ `VQE` $(\hat{H}_{\text{gen}}(\boldsymbol{q}_0), C(\boldsymbol{\theta}))$;

17 $\{O_0^{(s)}\} = $ `MeasureObs` $(\{\hat{O}_{\text{gen}}^{(s)}(\boldsymbol{q}_0)\}, C(\boldsymbol{\theta}_0))$;

18   for $i = 1$ to n_t timesteps **repeat**

19 $\boldsymbol{q}_i = $ `UpdateParameters` $(\{O_{i-1}^{(s)}\}, \boldsymbol{q}_{i-1})$;

20 $\boldsymbol{\theta}_i = $ `UpdateAngles` $(C(\theta_{i-1}), \hat{H}_{\text{gen}}(\boldsymbol{q}_{i-1}), \Delta t)$;

21 $\{O_i^{(s)}\} = $ `MeasureObs` $(\{\hat{O}_{\text{gen}}^{(s)}(\boldsymbol{q}_i)\}, C(\boldsymbol{\theta}_i))$;

22   **return** $\boldsymbol{\theta}$, $\boldsymbol{O}$, $\boldsymbol{q}$

---

problem and desired accuracy. Integrators like the Velocity Verlet algorithm require no additional resources. TDVQP becomes exact when $\hat{C}(\boldsymbol{\theta})$ can express the system perfectly for any configuration of classical parameters $\boldsymbol{q}$, given that the exact parameters θ can be found by optimization. This is only a statement of the best-case scenario. In reality, finding a good ansatz, VQA and shot-efficient optimizer is at the forefront of research in this area [216], and it is out of scope for this work.

6.2.1   Circuit Compression

A key building block of the presented algorithm and a fundamental aspect of quantum circuit optimization is the concept of circuit compression [237]. Any operation on a quantum computer must be a unitary operation $\hat{U}$, but the quantum computer only has a finite set of few qubit gates. An arbitrary $\hat{U}$ must be expressed, or compiled, into a set of native gates [1], this can always be done, but it is an NP-hard problem. If one were to implement an approximation of $\hat{U}$ within some threshold, one may find $\tilde{U}$ which might have a shorter circuit length than even an optimal decomposition of $\hat{U}$. This latter definition is what is generally known as circuit compression, although the term is sometimes used to refer to more optimal perfect decompositions [238].

   With an error-corrected quantum computer, one could use arbitrary circuit depths, but NISQ hardware benefits from using short circuits to minimize errors from occurring. But designing hardware efficient ansatze for VQAs is an unsolved problem [239]. This means that for most purposes a heuristic ansatz $\hat{C}$ parameterized by some vector θ brings the initial computational state $|0\rangle$ to a desired state $|\psi\rangle$ via $\hat{C}(\theta)|0\rangle = |\psi\rangle$. The defining property of unitary matrices, namely

$$UU^{\dagger} = U^{\dagger}U = I, \tag{6.6}$$

ensures that $C(\theta)^{\dagger}$ reverses the action of $C(\theta)$. If a unitary $\hat{U}$ is added, then it may be possible to find some parameters θ' such that $\hat{U}\hat{C}(\theta)|0\rangle \approx \hat{C}(\theta')|0\rangle$, thus compressing the action of the unitary back into the same quantum circuit, at least approximately. This is what is exploited by [224, 225, 226].

   Another approach is to approximate an initial state with such an ansatz and then perform a short-time evolution via Trotterization of the time-evolved operator as is done in [82, 232, 233]. The adjoint of the ansatz is appended to the circuit and its parameters varied such that the machine state is 'uncomputed' to its initial state. If one must assume that the chosen circuit ansatz is expressive enough to capture the entire state's time evolution, then such an approach is guaranteed to work. One then simply needs this new set of parameters and the original ansatz to express the new timestep without the time evolution operator, hence compressing the circuit. This idea has been implemented almost concurrently by Lin et al. [224], and by

Barison et al.'s "projected variational quantum dynamics" (p-VQD) algorithm [225]. Subsequent works, for example, [226], have built on the circuit compression idea.

6.2.2 Resource Cost of TDVQP

To determine the overall number of circuit evaluations required by the algorithm, the analysis is split into two parts. The first one-time cost is in finding the initial state circuit parameters. This is the same as in VQE and is $\mathcal{O}(N_H)$, where N_H is the number of Pauli strings required to express the Hamiltonian. The propagation then consists of finding the maximal overlap between the time-evolved state and its approximation, which only requires a single circuit to evaluate for the maximum number of iterations N_{iter}, and is thus $\mathcal{O}(N_{\mathrm{iter}})$. The underlying optimizer will require $\mathcal{O}((N_{\mathrm{param}}^n)$, where N_{param} is the number of circuit parameters, and n is generally small. The observable measurement requires the most circuit evaluations and is $\mathcal{O}(N_{\mathrm{obs}})$, where N_{obs} is the number of Pauli strings required to express the observable in question. The whole propagation is linear in the number of timesteps desired N_{steps}. Thus, the circuit evaluation cost is $\mathcal{O}(N_H + N_{\mathrm{steps}}(N_{\mathrm{iter}}N_{\mathrm{param}}^n + N_{\mathrm{obs}}))$. In this example, omitting the first step, there is $N_{\mathrm{obs}} = 7$ which grows either linearly with grouped observables or exponentially without, then $N_{\mathrm{param}} = 70$ and $N_{\mathrm{iter}} = 100$, which are set by choice. This gives an overall cost of around $7 \cdot 10^3$ circuits per timestep.

6.2.3 Error propagation of TDVQP

The TDVQP algorithm inherits all the errors of its constituent parts. This includes the chosen circuit compression algorithm, time evolution approximation, and in the classical propagator. Nonetheless, it is important to have an intuition of the potential pitfalls of the algorithm. This section illustrates the sources of error in the wavefunction and observables and their interaction velocity Verlet integrator. A more thorough derivation and explanation can be found in section A.3.

When running the algorithm, any coherent error on the wavefunction representation in the quantum computer $|\tilde{\psi}\rangle$ can be represented as a superposition of the desired state $|\psi\rangle$ and some combination of undesired orthogonal states $|\phi\rangle$, such that $|\tilde{\psi}\rangle = \sqrt{1 - I^2}\,|\psi\rangle + I\,|\phi\rangle$, where I is the infidelity. When the expectation value of a Hermitian observable $\mathcal{O}$ on $|\tilde{\psi}\rangle$ is measured the result is

$$\langle\tilde{\psi}|\,\mathcal{O}\,|\tilde{\psi}\rangle = \left(1 - I^2\right)\langle\psi|\,\mathcal{O}\,|\psi\rangle + I^2\,\langle\phi|\,\mathcal{O}\,|\phi\rangle. \qquad (6.7)$$

How this translates to the actual measured observable used here is completely system dependent. This will lead to an error in the observable, which in the case of the velocity Verlet integrator with a force error $F_i\epsilon \propto I^2$ in 1 dimension will give a new position $\tilde{R}_i$ of

$$\tilde{R}_i = R_i + \frac{F_{i\epsilon}}{M}\Delta t^2, \qquad (6.8)$$

shifted from the expected true position R_i. This is linear in the error of the force and quadratic with respect to the timestep Δt. The following time evolution Hamiltonian and observable operator will be based on this position with an error, which is again, system dependent. The effect is illustrated by the equations

$$|\tilde{\psi}_1\rangle = \exp(-iH_{el}(R_0)\Delta t)\,|\tilde{\psi}_0\rangle\,, \qquad (6.9)$$

$$|\tilde{\psi}_i\rangle = \exp(-iH_{el}(\tilde{R}_{i-1})\Delta t)\,|\tilde{\psi}_{i-1}\rangle\,. \qquad (6.10)$$

Even in the one-dimensional model used in this work, this effect is not analytically computable, but it is small if the timesteps are sufficiently small. This then enters the velocity ($\dot{R}$) update as

$$\tilde{\dot{R}}_{i+1} = \dot{R}_{(i+1)} + \frac{F_{i\epsilon} + F_{(i+1)\epsilon}}{2M}\Delta t, \qquad (6.11)$$

which is linear in the error and timestep. Assuming a constant error over all time of F_ϵ, that is to say, that the force deviates from the correct one by a constant offset - this is equivalent to having an additional linear term on the potential. This has the overall effect on the position at iteration i of

$$\tilde{R}_i = R_i + \dot{R}_{i-1}\Delta t + \frac{(i^2 + i)F_\epsilon}{2M}\Delta t^2. \qquad (6.12)$$

This expression is quadratic in i and quadratic in timestep. The effect on the fidelity of the TDVQP wavefunction compared to an exact propagation is nontrivial, but numerical examples are provided in section A.3.

The other main source of error inherent to the p-VQD algorithm is that the optimizer never finds a perfect representation of the time-evolved wavefunction, but rather an approximation that meets some fidelity threshold $T < 1$. If this threshold is met exactly at each p-VQD step, assuming all observable measurements are unaffected, then the decrease in the fidelity is modelled by

$$\text{Fidelity(i)} = T^i. \qquad (6.13)$$

When the algorithm is run under limited quantum resources and thus subject to finite sampling noise, both the observable and p-VQD step fidelity measurements will have some Gaussian distribution, which will feed into the errors above on a simulation by simulation basis. The effect of this has been analyzed numerically in the case of the modified Shin-Metiu Model.

6.2.4   Grid-based Mapping to Qubits

The Shin-Metiu system is treated on the quantum computer in first quantization. A finite difference method on an equidistant grid is used. For low-dimensional problems, this is an appropriate approximation, but in general discrete variable representations (DVR) are a better choice for problems in higher dimensions. In quantum computing DVRs have been used to explore first quantization simulations in [240] using the Colbert and Miller DVR [241]. Issues exist

with using DVRs on quantum computers as they generally require a full matrix Hamiltonian which is costly to measure and implement on quantum computers, and alternatives have been proposed [242].

Quantum computing in first quantization has the advantage that n_g grid points can be represented by $N = \log_2(n_g)$ qubits. Choosing $n_g = 2^N$ to maximize the use of the N available qubits. In the simplest finite differences method each position is an integer multiple L/n_g. The grid point g is by the quantum state $|g\rangle$ and mapped as

$$|g\rangle = |j_0\rangle \otimes \cdots \otimes |j_k\rangle \otimes \cdots \otimes |j_N\rangle, \tag{6.14}$$

where j_k is the k$^{\text{th}}$ bit value of the binary representation of g. The potential operator $\hat{V}$ is diagonal in this representation, simply sampling the potential at each grid point. The kinetic energy Hamiltonian is not diagonal in the position representation, and although one could use the split operator method [243] to make it diagonal in momentum space would require a quantum Fourier transform implementation, which, cannot be effectively implemented on existing quantum devices at the time of writing.

All the issues above are sidestepped by using the finite differences' method, in which the one-dimensional potential and the Laplacian form of the kinetic energy can be written as

$$V_{j,j'} = V(x_j)\delta_{j,j'} \tag{6.15}$$

$$T_{j,j'} = \frac{-\partial^2}{\partial r^2} \frac{1}{2m_{el}} f, \ f \begin{cases} -2, & \text{if } j = j'; \\ 1, & \text{if } j = j' \pm 1. \end{cases} \tag{6.16}$$

$$\tag{6.17}$$

The tridiagonal matrix that results has the same value on the off-diagonal terms, which allows them to be decomposed into fewer Pauli strings than a full matrix via an elegant recursive form that is described in 6.2.5 following from [244]. This is important because the number of non-zero entries is related to the number of terms in the Pauli decomposition, which should be kept minimal to reduce the length of the Trotterized time evolution operator and observable measurements. The finite difference method does require high grid densities to be accurate (although this depends on how oscillatory the system in question is), but this requirement will likely be met by the doubling of grid points per additional qubit.

Although not implemented in this paper, some algorithms can make the time evolution of Hamiltonians with this form more efficient on near-term devices [245]. Depending on the particular Hamiltonian one chooses to study in this way, different efficient algorithms exist to lessen the cost of the time evolution such as variational fast forwarding and qubitization [208, 234]. It is also possible to efficiently solve for the eigenstates of tridiagonal matrices on quantum computers [246].

6.2.5   Pauli String Representation of a Tridiagonal Hermitian Matrix

The real-space Hamiltonian for the Shin Metiu model is tridiagonal. There is a recursive solution to expressing tridiagonal Hermitian matrices in the Pauli basis. The off-diagonal matrices can be expressed as:

$$A_1 = X$$
$$A_n = \mathbb{I}_2 \otimes A_{n-1} + (X \otimes \mathbb{I}_{2^{n-1}}) \otimes$$
$$\left[\frac{1}{2^n} \sum_{t=0}^{\lfloor n/2 \rfloor} (-1)^t \sum_{\pi} S_{\pi} \left(X^{\otimes(n-2t)} \otimes Y^{\otimes 2t} \right) \right] \otimes$$
$$(X \otimes \mathbb{I}_{2^{n-1}}),$$

Where X and Y are the Pauli matrices, and S_{π} is the permutation function that returns a unique combination π of the Pauli string. That is to say, given the string XYY the result is the sum $XYY + YXY + YYX$. This grows exponentially, but with a qubit-wise recursive largest first commutator, it grows as $\mathcal{O}(2^{n/2})$ and if one takes at the fully commuting largest first approach, then the number of terms grows as $\mathcal{O}(n)$. The diagonal matrix is then the 2^n term weighted linear combination of all possible n length Pauli strings consisting of I and Z matrices.

Using either qubit-wise commutation relations or Pauli-word wise commutation relations, the number of Pauli strings required to implement and measure tridiagonal Hamiltonians is greatly reduced. Analyzing systems of up to 10 qubits via Qiskit and grouped the Pauli string decomposition through the above methods hints that this is indeed the case. These groupings can be used to reduce the number of measurements that have to be made, but the results are only presented here for completeness with no comment on how these measurements will be done in practice. Figure 6.4 shows the number of grouped terms using different existing techniques. These are not necessarily realizable on current machines.

Figure 6.4: **The number of grouped terms in a Hermitian tridiagonal matrix** using different grouping strategies. If no strategy is used (solid blue) then the number of terms grows exponentially, which is also the case when using qubit-wise commuting groups (dashed orange). If one uses word-wise grouping then the number of terms grows linearly (dotted green). Reproduced from [180] under CC4.

6.3   Numerical Simulations

To gauge the performance of the scheme the Shin-Metiu model as described in Section 6.1 is implemented. In the BOPES there is avoided crossing at around $R = -1.9$ a.u. The system is initialized with the nucleus at an initial position of $R = -2$ a.u. and an initial velocity of $v_0 = 1.14 \cdot 10^{-3}$ a.u., the average nuclear velocity from the Boltzmann distribution at 300K. The electronic system is initialized through the VQE with a random set of parameters and is allowed 300 iterations to approximate the ground state. The system is then evolved through the TDVQP algorithm with a timestep of $\Delta t = 0.5$ a.u. Each quantum time evolution step attempts to reach a fidelity threshold of $1 - 10^{-5}$ or up to 100 iterations of stochastic gradient descent [247]. To find the optimal circuit parameters to approximate
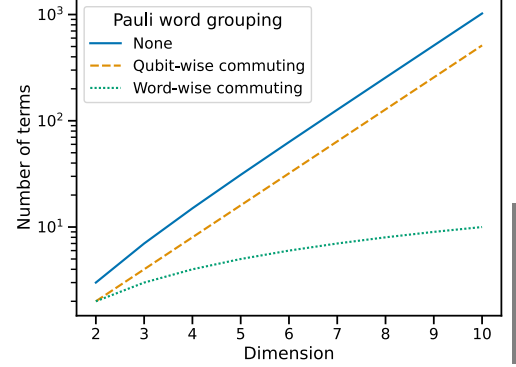
the previous time evolved state. Gradients were computed through the parameter shift rule [88]. All simulations are done on 16 grid points that can be represented by 4 qubits.

Two situations are examined. First, keep the initial conditions constant but sample different VQE ground state approximations, called the "Single Initial Condition" case. In the second case, the MD-type approach is examined in depth, where a normal distribution of initial conditions is sampled for the initial velocity of the nucleus and allow one TDVQP evolution per sample. The velocity distribution is sampled from the Boltzmann distribution, only keeping positive velocities so that the nuclei approach the avoided crossing. The results shown are 100 samples that are evolved for 1000 timesteps which bring the classical trajectory beyond the avoided crossing point.

Additional examples are provided for longer-time evolutions as well as for non-ground state evolution in Appendix A. Excited states and superpositions are prepared by using the uncomputation step of the TDVQP, but instead of starting the state with a known circuit from the VQE, the simulator is simply initialized to a desired arbitrary state, and the optimizer attempts to uncompute it with the ansatz and then those parameters are used as the initial step in place of the VQE. Various techniques to prepare excited states exist [248, 249], but are not the focus of this work.

Two different metrics to establish the accuracy of the TDVQP algorithm are used: the so-called "Ideal" evolution begins at the desired state to numerical precision and is evolved by exact diagonalization. But, precise state preparation is another area of intense study [250]. To better gauge the performance of the TDVQP in isolation, an "Exact" evolution is performed, which uses the VQE-optimized initial state for evolution via exact diagonalization. This allows the removal of any bias from a poorly optimized ground state.

The VQE uses an ansatz of the form shown in Figure 6.5, which was heuristically chosen as it can achieve ground state infidelities of up to 10^{-5} on this system with 4 layers. The same ansatz as in [225] is used, but various ansatze can be used, and for first quantization problems, in particular, there are some examples of how several different heuristic ansatze perform in [242]. The number of repetitions of the Trotterization layer is another important parameter, but as the decomposition of the Trotterized operator into native gates is deep, so a single one is used. Although for full quantum dynamics, this would be very inaccurate for larger timesteps, the interaction with a classical system necessitates the use of short time steps, so that the Hamiltonian of the system is kept up to date with the classical state of the system. This means that a single trotter step is all that is needed, and the number of layers of the ansatz can compensate as shown in 6.3.1.

The simulations were run on the Qiskit state vector simulator (version 0.28) [251] using the parameter-shift rule [252, 88] to determine the analytic gradients required for gradient-descent based
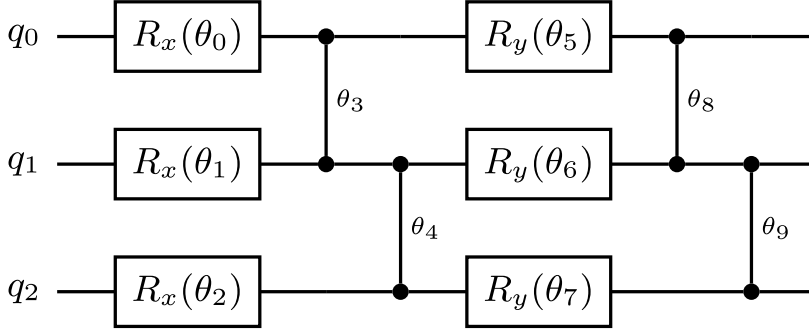
Figure 6.5: **One layer of the ansatz used for the VQE and TDVQP for three qubits.** If multiple layers are used then the above circuit is repeated. If more qubits are used then the vertical motif is continued. In both cases, more parameters can be added as needed and $\boldsymbol{\theta}$ refers to the list of all parameters. The ansatz features x rotations and the parameterized ZZ rotation. Reproduced from [180] under CC4.

optimization. Numpy [253] was used for the exact numerical simulations, to prepare the Hamiltonian and to compute the velocity Verlet steps.

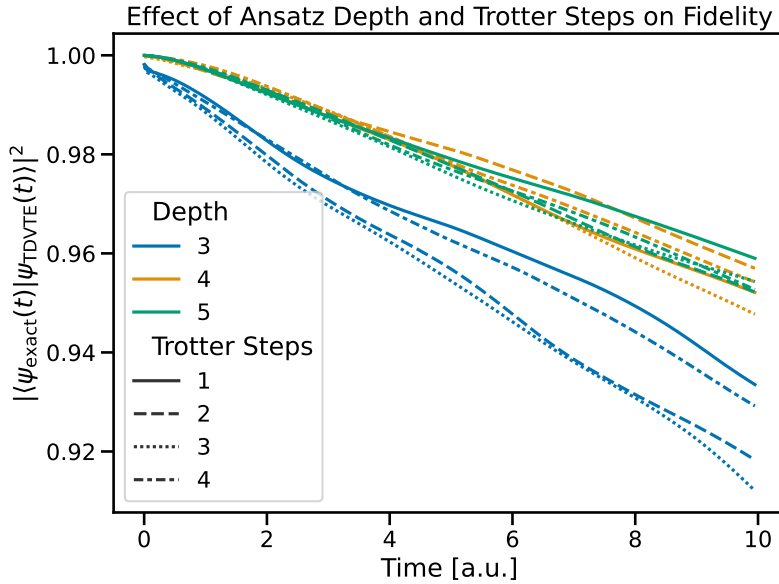### 6.3.1 Hyperparameters: Trotterization and Ansatz Depth

Figure 6.6: **Mean fidelity with respect to the depth of the ansatz and the number of trotter steps** in the evolution for 10 samples of 200 timesteps. The lowest depth is 3 (blue), while the next step is of depth 4 (orange), and the deepest is 5 (green). The different Trotter approximations are various dashed or solid lines, but since the time evolution is short, this has little effect compared to the depth of the ansatz. Reproduced from [180] under CC4.

Although one can have a near infinite amount of variability in the heuristic form of the ansatz, which has been chosen to be the that of Figure 6.5. But even for a single ansatz it is important to find what the optimal depth is for a given problem, and in this algorithm, the number of steps in the Trotter approximation is an important choice. Figure 6.6 shows the effect of that choice in the modified Shin-Metiu model with a timestep of 0.05 a.u. the order of the Trotterization has a small effect compared to the depth of the ansatz. This is not unexpected, as the timestep is very small. For the simulations, a depth of 5 and a single Trotter step was found to be a good compromise between depth and precision.

Another important point is whether it is beneficial to use longer timesteps within the limits of the chosen Trotterization depth or to

use conservatively short steps. To see the effect of this, observe the fidelity of the same simulation as the single case in the main text but with 500 steps of $\Delta t = 0.5$ [a.u.] and 5000 steps of $\Delta t = 0.05$ [a.u.]. Both have a total time of 250 [a.u.], but as should be evident from Figure 6.7 the two approaches show very different behaviors. When looking at the quality over 'simulated time', choosing larger timesteps is obvious. But looking at fidelity over the number of timesteps, the shorter timesteps do have an advantage. The reason for this is that the optimizer has a fidelity threshold of ϕ compared to its previous step; as a first approximation, one can assume this fidelity is reached exactly at each timestep T, then after T steps, would yield a fidelity of ϕ^T compared to the ideal situation.
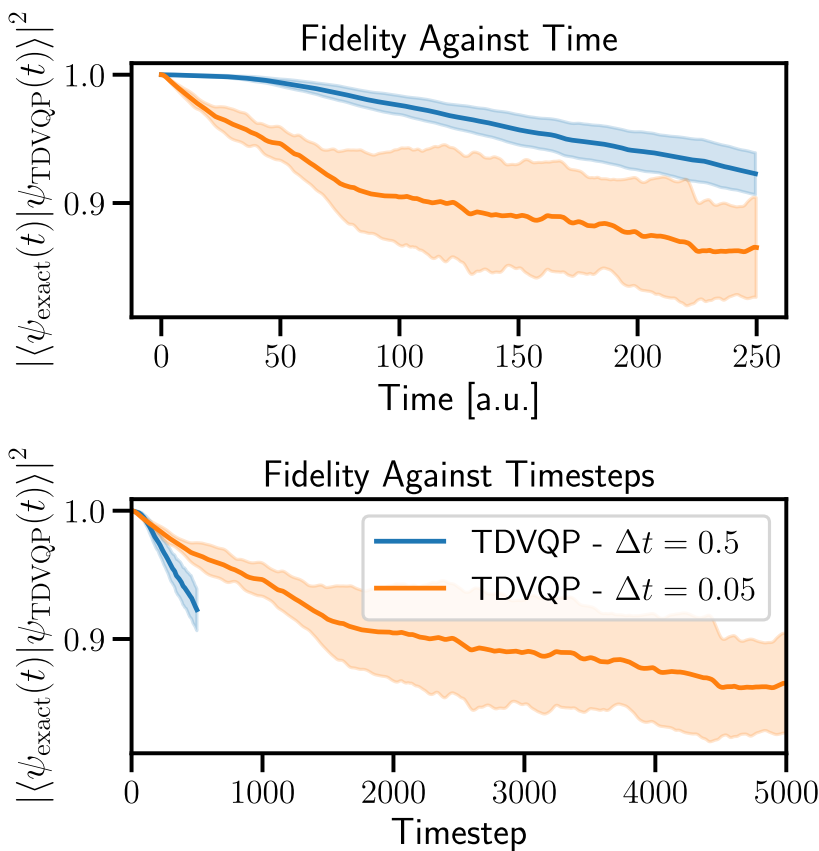
Figure 6.7: **Difference between many small timesteps and fewer large timesteps.** Using fewer larger timesteps (blue, 500 steps of 0.5) compared to many smaller timesteps (orange, 5000 steps of 0.05) leads to an interesting observation that smaller time steps lose fidelity slower, but when observing the simulated time, it can be beneficial to use fewer larger timesteps. The respective standard deviations are shown as the highlighted areas. Reproduced from [180] under CC4.

Realistically, the size of the timestep also has a large effect on the iterations the optimizer must take to converge. $1,000$ steps of $\Delta t = 0.5$ take longer than $10,000$ steps of $\Delta t = 0.05$ given the same threshold. There is also such a thing as a 'golden' initial state, which has the property that many parameters in the ansatz are initialized so that they stay constant or vary smoothly throughout the evolution. Such initializations optimize faster and retain higher fidelities than initial parameters that have much more chaotic 'spiky' evolutions. Although this is hard to quantify, it is something that could be used to filter out badly behaving initializations early on in the evolution.

6.3.2 Numerical Results

The results shown in this section are the result of two types of potential cases. The first, which is referred to as 'single' is the evolution of a single set of initial conditions meant to represent the precision of this algorithm to exactly reproduce a quantum-classical system. Although this is not the intended use case of TDVQP, it is nonetheless the most instructive to determine its behavior. The second, referred to as 'MD' is the molecular dynamics-like use case, where a single TDVQP evolution per trajectory is used. The trajectories are picked from random pairs of normal thermal distributions around the same initial state as the 'single' simulations, with the specific values written in Section 6.3. Their average behavior is taken as the approximation to the true system evolution. For NISQ devices there are some potential cases for different approaches to MD such as those mentioned in [254].
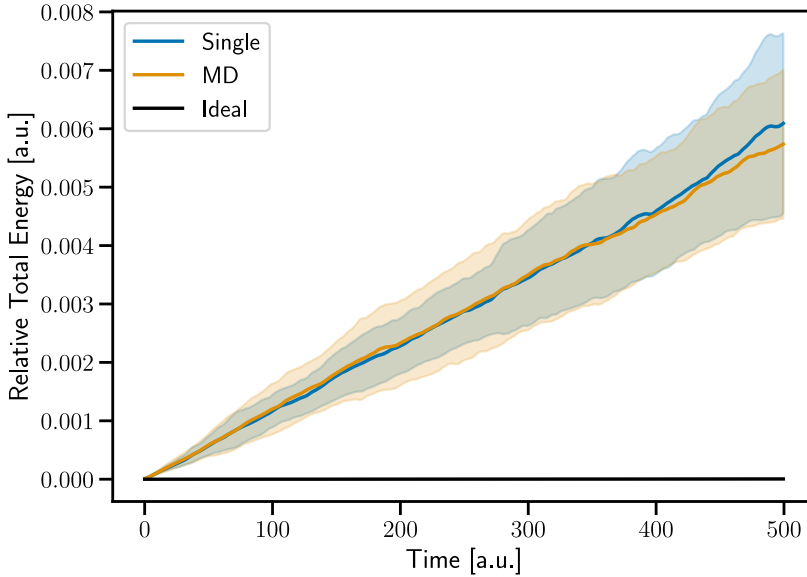

Figure 6.8: **Mean relative TDVQP energy** shown for "Single" initialization (blue) and for "MD" initialization (orange). The highlighted areas show the standard deviation of the distribution of 100 separate runs of 1000 timesteps of 0.5 a.u. at the infinite shot limit. The other lines show the ideal initial state energy evolution. The energy continually increases in the TDVQP as higher energy levels are increasingly populated through leakage. Reproduced from [180] under CC4.

An important gauge for the validity of simulations of closed systems is whether they conserve energy or not. A symplectic integrator is used in the classical system (velocity Verlet), and in the exact diagonalization case, energy conservation holds for up to 50,000 timesteps. As can be seen in Figure 6.8 the TDVQP algorithm does not conserve energy. This is because the populations are not preserved in the diagonal basis in the p-VQD step, as the optimization is limited to a finite number of iterations and the ansatz is system agnostic. The effect of this can be seen clearly in Figure 6.9, where it can be seen that the population in higher states increases much faster than in the ideal case. Although it is not shown, starting the exact evolution from the VQE state does begin with some population spread, but this does not change as the evolution progresses. The population plot is shown at the infinite shot limit for clarity, but the finite shot cases can be found in A.3.2.
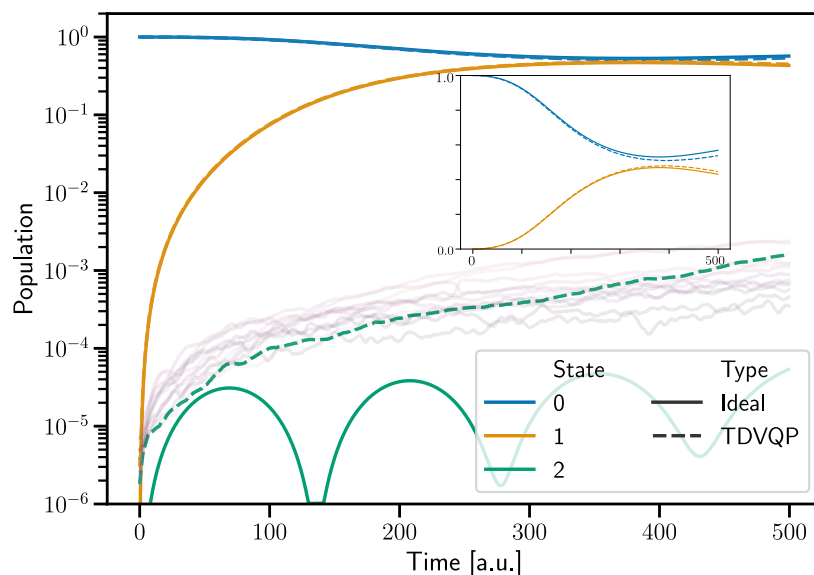
Figure 6.9: **Time evolution of state populations** between the ideal simulation (solid) and 100 instances of TDVQP evolution (dashed) over 1000 timesteps of 0.5 a.u. at the infinite shot limit for the "single" simulation. The populated states are the same as those seen in Figure 6.2. The main graph is in logarithmic scale showing faint lines for higher energy levels populated by TDVQP, with the inset showing a linear scale of the two most populated levels. Reproduced from [180] under CC4.

As a consequence of the higher energy levels being increasingly populated as the evolution progresses, it is the case that the fidelity decreases gradually as can be seen in Figure 6.11. This general degradation of quality is not optimal and strategies could be employed in the optimization to mitigate this, such as measuring the energy and allowing the cost function to penalize when the system is not conserving energy. This would require measuring the expectation value of the system Hamiltonian which would increase the cost of this algorithm.

Despite the problem with energy conservation, using such an algorithm to measure an observable such as the force exerted on the nucleus by the electron (F_{el}) can still lead to reasonable results. Figure 6.10 shows the mean of the electron force measurements from TDVQP compared to the ideal measurements at different per-circuit shot counts. It is clear that the mean value slowly deviates from the ideal evolution in even the infinite shot limit, and that one requires 10^5 shots per circuit to reach qualitatively relevant results at longer times. Efficiently estimating energy gradients is a huge undertaking, and this work does not implement some of the NISQ-friendly techniques that have been developed [255, 256], but it is expected to be a problem even in the fault-tolerant regime [257].

Figure 6.11 shows that the fidelity decays in all cases over time and that for long time evolutions, one requires more than 10^5 shots when not using any additional techniques to better measure the force or better preserve the populations when not undergoing a transition. At lower shot counts the fidelity falls quickly, following eq. 6.13 until the equal superposition is approached, which sets a higher floor than zero for the decay of the fidelity. The potential effects of other noise sources are described in more detail in A.3.

Figure 6.12 zooms into the two more reasonable fidelity lines, those of the infinite and 10^5 shot simulations. This shows that us-
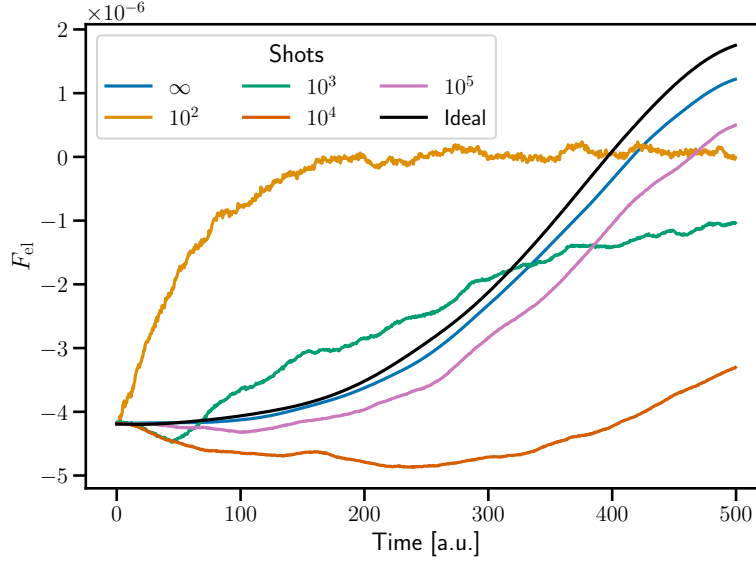
Figure 6.10: **Electron force observable for the single initialization**. The ideal simulation (black) and 100 instances of TDVQP evolutions (coloured) over 1000 timesteps of 0.5 a.u. at varying shot counts per circuit. There is qualitative agreement of TD-VQP with the ideal case at 10^5 shots and the infinite shot limit, with a shift downwards due to leakage to higher energy levels, which in this case biases the force in this direction. Reproduced from [180] under CC4.
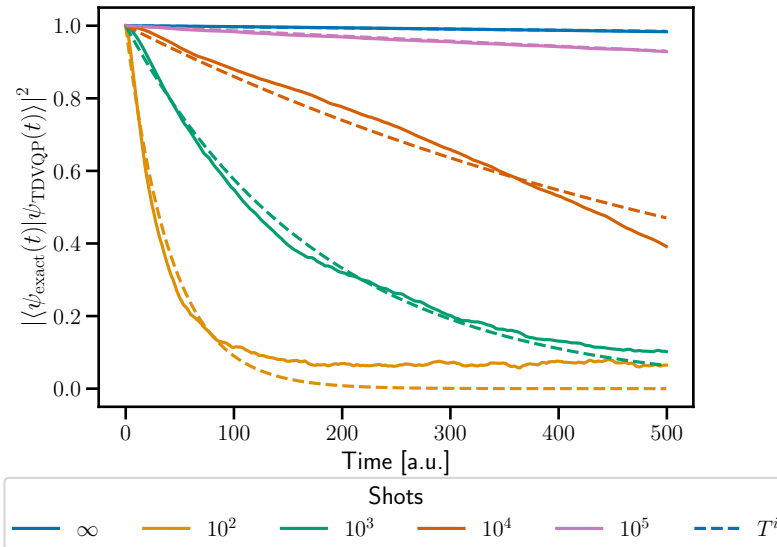
Figure 6.11: **TDVQP Fidelity over time** compared to the exact diagonalization evolution of the VQE initialized state over 1000 timesteps of 0.5 a.u. for 100 single initializations (solid, coloured) at different shot counts. The best-fit approximation of (6.13) (dashed) for each line is shown. Reproduced from [180] under CC4.

ing the multiple trajectories in an MD sense somewhat improves
the simulation fidelity compared to the single trajectory case, and
more quantum-tailored algorithms like [254] may improve this fur-
ther. In the infinite shot case, the difference is minimal - but due
to the larger variance of the MD simulations compared to the ideal
trajectory, the performance tends to be minimally worse.

Figure 6.12: **Fidelity for the MD simulation** showing the 100 instances of TDVQP evolution (orange) over 1000 timesteps of 0.5 a.u. for 100 different 'MD' trajectories and 'single' simulations. The MD simulation is of slightly higher fidelity compared to single trajectory simulations at high but finite shot counts. At lower shot counts the fidelity decreases too quickly in all cases. Reproduced from [180] under CC4.
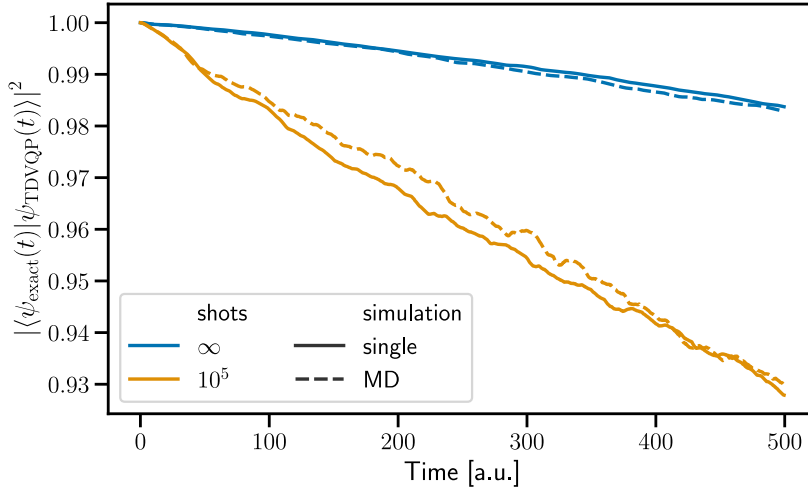
The relationship between shots and fidelity is also illustrated in
Figure 6.13, where one can more clearly see that the MD simulation
slightly improves the simulation at longer time evolutions when us-
ing finite shots. However, this improvement is not massive. It also
highlights the large jump in fidelity gained when using higher shot
counts. Following from the p-VQD result [225], the time evolution
evolves its system for 40 iterations (20 a.u. here), where there are
very high compression fidelities beyond 10^4 shots per circuit evalua-
tion.

Finally, Figure 6.14 illustrates that in the simulation the maxi-
mum number of iterations, 100, is quickly reached before the 200th
iteration at the infinite shot limit. The overall mean final infidelity
is 10^{-5}, although the fitted threshold of eq. 6.13 for the overall algo-
rithm is slightly lower at $3 \cdot 10^5$. The infidelity is $I = 1 - \mathcal{F}$, where $\mathcal{F}$
is the fidelity. This implies there is an additional error, likely due to
the drift of the exact simulation of the system from the TDVQP sim-
ulation. This is consistent with a 10^{-6} to 10^{-7} shift in the force as
described in the numerical simulations in A.3, which is also roughly
the difference in the force observable seen in Figure 6.10.

Overall the results show some interesting behavior. The number
of timesteps modelled in this work is very high, and this results in
only qualitative agreement in the region of interest where there is a
significant population transfer. Furthermore in this simple model,
the energy levels are well separated, and the simulation begins in
the ground state. This results in a strong unidirectional contribution
from population leakage to higher energy levels. In a more complex
molecule, one would begin, for example, from a thermal ensemble of
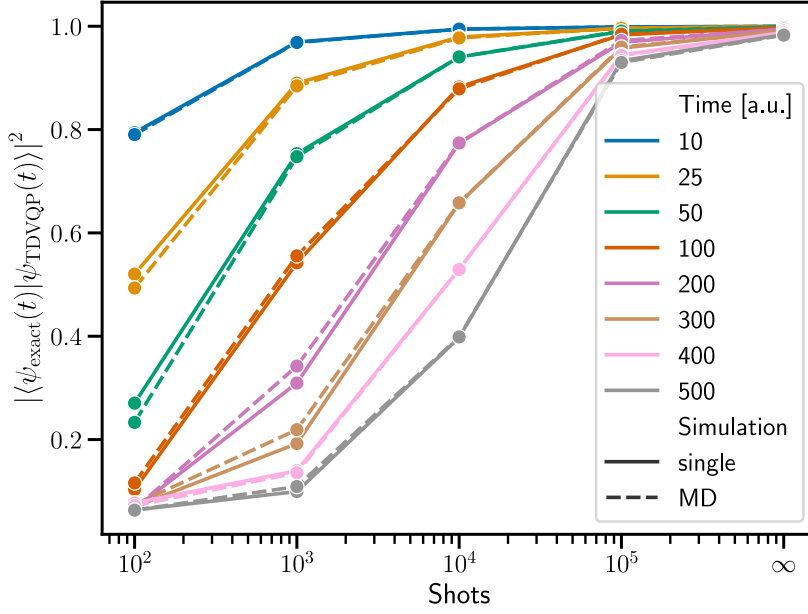
Figure 6.13: **TDVQP Fidelity over shots** of the exact diagonalization evolution of the VQE initialized state and the TDVQP evolution over 1000 timesteps of 0.5 a.u. for 100 different MD-like trajectories (dashed, coloured) and 100 single initializations (solid, coloured) at different shot counts. Reproduced from [180] under CC4.
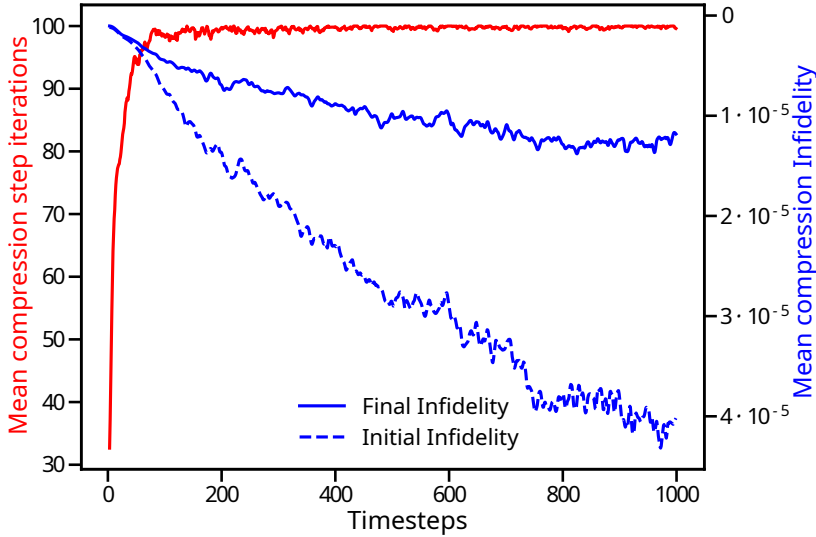
Figure 6.14: **Compression infidelity and number of iterations per step** for the single TDVQP setup at the infinite shot limit, showing the mean number of iterations in red with scale to the left. The mean initial infidelity (dashed, blue) and final infidelity (solid, blue) with the scale on the right, where infidelity (I) is $I = 1 - \text{Fidelity}$. The number of iterations taken by the optimizer quickly reaches the limit of 100, while the final and initial infidelities gradually fall with the number of timesteps into the simulation. Reproduced from [180] under CC4.

not only velocities but also states. In turn, the leakage would then result in deviations from both higher and lower energy levels and may be less detrimental to the ensemble average than here.

6.4 Discussion

Time-dependent evolution is a difficult problem that can be well explored through quantum computers. Many techniques can be used for full quantum systems [208, 240, 234, 207, 226, 225] which are suitable to both near term and fault-tolerant machines.

Algorithms that are suitable for MQC dynamics do require efficient and accurate full quantum dynamics, but the interplay between the classical and quantum systems brings a new spate of challenges. To exchange information, one must measure observables from the quantum system, which is expensive and destroys the state, requiring at minimum an efficient way to measure energy gradients, which is an area of active research [255, 256, 258]. This is a disadvantage, but it also means that one is limited to short-time evolutions between measurements. This makes it possible that a single trotter step is accurate enough [259, 260], which is beneficial to near-term devices.

Even though larger timesteps may be possible, the longer the time evolution, the longer the optimizer takes to find the time-evolved ansatz parameters. This is because the previous timestep parameters are no longer as close to the evolved ones. At the same time, most classical MQC methods do not update the parameters that govern the classical system's evolution at the small time intervals used here [218]. It would be advantageous to use the largest possible classical timestep for a given integrator. To do this, one could do multiple compression steps with short-time Trotterizations using a constant Hamiltonian and only measuring the desired observables after the quantum system has evolved for the standard timestep of the classical problem, performing updates after this point. This will leverage the underlying compression algorithm to its fullest and reduce the overall number of measurements required.

The algorithm takes advantage of the above facts and is highly modular. Although the results are shown using an algorithm like p-VQD [225] with Trotterization of the operator, there is no reason that other efficient time evolution algorithms couldn't be used. This is especially true if the time evolution operator could be efficiently represented by techniques other than the Trotterization of the Hamiltonian. The update step used here measures the Pauli string decomposition of the $\frac{dH}{dR}$ matrix to compute forces, but other techniques exist in the fault-tolerant regime [258, 257], as well as in the NISQ regime [255, 256]. The main constraint with TDVQP is the fact that throughout the time evolution, there are inevitable inaccuracies in optimization, due to the compression step not preserving the populations in the diagonal basis as would have been expected as shown in Figure 6.9. This has the direct consequence that energy is not conserved, even though in the ideal simulation this is the case

as shown in Figure 6.8. Due to this accumulated error, fidelity falls consistently, and the effect is compounded when quantum resources are finite.

These problems might be tackled by either increasing the threshold of the compression step or by measuring the energy and penalizing the optimizer when energy is not conserved. Another option that may be possible is designing an ansatz with problem-specific constraints [82]. Such an ansatz considers properties such as particle preservation within their structure, which may remove the need for expensive additional iteration steps or measurements. Furthermore, it may be possible to replace the p-VQD propagation with other compression methods [226]. Since the simulation is in the first quantization representation, it may be difficult to find what properties to conserve in the wavefunction, but the advantage comes in not needing to measure the many non-adiabatic couplings.

It was also found that the error mostly comes from the compression step or due to finite sampling effects more than from the coupling to the classical system due to the small classical timestep. Although it is always interesting to see how an algorithm behaves under noisy conditions, the performance of p-VQD under noise has been explored for full quantum dynamics in [226]. This work focuses on the interplay between the scheme under the effect of a Hamiltonian which depends on the measured observables.

Overall the TDVQP algorithm has been introduced for MQC dynamics with the quantum subsystem computed on a quantum computer and have explored it on the Shin-Metiu model as an example of Ehrenfest dynamics in first quantization. However, it is not limited to this setting. It reproduces the expected observables and state evolution qualitatively. The algorithm is modular and refinements to it may be tackled in future research. Inaccuracies of the quantum computer can also be mitigated when computing ensemble averages of the classical properties. This work shows that MQC simulations may be practically feasible on noisy quantum computers if it is proven that variational quantum algorithms can have an advantage in chemical problems.

6.5 But What About Noise?

TDVQP does work under ideal circumstances. But these simulations have all been done without the presence of errors in NISQ machines. The algorithm is, at the time of writing, currently being implemented at the Walter-Meisner Institute on a superconducting quantum processor. This process necessitated the study of the performance of this algorithm with much more stringent constraints on the depth of both the ansatz and the time evolution operator. As such, the effect of a realistic noise model on this algorithm has been computed, and the results are shown in chapter 9, compared to the noiseless case. As the work progresses, methodologies to deal with this noise are introduced and to close the thesis off, they will be

implemented on this data to show the effectiveness of the QEM techniques shown here in practice. Sadly it was not possible to fully run the algorithm on hardware at the time of writing and so the results are still numerical and will be presented in chapter 9.

Unlike on a theoretical analysis of the algorithm, in the device and time that are given, it is impossible to use a depth of more than 20 CNOTS in parallel for the entire circuit. As such, the trotter operator was approximated by a hardware efficient variational circuit at each timestep, and the ansatz was greatly simplified to have a low depth and parameter count. Due to the time evolution compression, this implementation of the algorithm is no longer scalable. Besides the change in parameterized circuit ansatz and the time evolution operator compression, the algorithm remained unchanged, except that now a noise model taken from characterization of the machine (which cannot be made public at this time) is used, and all measurements are both noisy and have a finite shot count.

In addition to this, it is essential to analyze how well this noise can be mitigated via QEM techniques. To do this for novel algorithms it may be possible to exploit certain structures in the problem. The following chapter will present how one goes about designing novel QEM, and at the final stage, this and more theoretical techniques such as the clean and dirty partial error correction model [111] will be utilized to realistically mitigate noise and perform this algorithm.

7
UNITED we Mitigate

I'd never think it possible; but I'll do what's feasible
Don Giovanni (translated), W.A. Mozart/Lorenzo Da Ponte, 1787

This chapter will go through an overview of the theoretical combination of various techniques, finally ending at unified technique for error mitigation with data (UNITED). Then, a section will be dedicated to the numerical experiments which involve the preparation of realistic noise models and the analysis of different assumptions and their effect on the various techniques. It then ends with the final results and a discussion on the implications of the results. The whole of this chapter is based on the work presented in [110].

AS SHOWN IN THE TDVQP ALGORITHM, the introduction of noise greatly impacts the performance of the algorithm. As such, beyond the implementation of a hardware efficient algorithm, an important process to designing a functional program for the NISQ era is to also design novel error mitigation approaches that can make such algorithms yield useful results with current hardware.

In section 4.2 several quantum error mitigation methods were presented. A natural question to ask is *"what improvements can be gained by combining disparate techniques?"*. This chapter presents the work carried out and published in [110] that aimed to explore that question. Here we presented a strategy to combining Clifford data regression (CDR), zero-noise extrapolation (ZNE) and virtual distillation (VD) in a framework that should allow for other techniques to be mixed and matched. It followed from the successful result of combining CDR with ZNE that was presented in a technique called variable-noise Clifford data regression (vnCDR) [132]. The addition of a more robust framework were presented under the name UNITED.

To properly compare the performance of the different QEM methods and their combination, each is benchmarked through numerical studies. In particular, they mitigated local observables of a random circuit and for a practical application on a random term of a QAOA algorithm on the Max-Cut problem of various sizes of up to 10 qubits and circuit depths of 128 sequential two-qubit gates.

Contributions: The general idea of the project was conceived by Lukasz Cincio, Patrick J. Coles, the theoretical derivations were worked on mostly by Max Hunter Gordon, Piotr Czarink, myself and corrected and expanded by Marco Cerezo. The numerical simulation code was written by Piotr and Lukasz, and all the numerical experiments were carried out by me. The writing was a joint process mostly carried out by myself, Max and Piotr, all other authors contributed to proofreading and general guidance. This particular rendition of the work was re-written by myself, reorganizing, expanding and removing sections to fit with the thesis.

To properly gauge the sampling overhead associated with combining disparate QEM that are described in section 4.2, and analyze the techniques over various fixed shot budgets ranging from 10^5 to 10^{10}. Although there are intricacies to the results which depend on the task which will be described in depth, it was found that single QEM techniques (ZNE, CDR and VD) would converge in their error mitigation ability well before the biggest shot budget was reached. The combined techniques on the other hand (vnCDR and UNITED) would be able to make use of the additional data and keep correcting up to the largest shot budgets tested, with vnCDR being the best until 10^8 and being overtaken by UNITED at 10^{10}.

7.1 Combining Quantum Error Mitigation Techniques

Combining the disparate QEM techniques requires the careful study of sub-combinations. This section builds on the foundational knowledge given in section 4.2. The first presented combination is that of Richardson extrapolation (4.14) with VD copies. Following this, CDR is combined with VD, then VD and ZNE and finally the full combination that yields UNITED. The combination of CDR and ZNE is not presented here as it was already presented in [132].

7.1.1 Virtual Distillation and Richardson Extrapolation

Given a set of VD observables for different number of copies M, denoted $\tilde{\mu}_M^{\text{VD}}$, the different noise levels in the expression can be put into a Richardson extrapolation (4.14) scheme with the different number of copies, from $1, 2 \ldots, M_{\text{max}}$ to obtain the following expression:

$$\mu'_{M_{\text{max}}} = \sum_{M=1}^{M_{\text{max}}} b_M \tilde{\mu}_M^{\text{VD}}. \tag{7.1}$$

Here b_m is an unknown coefficient that characterizes the noise level for a given number of copies. Theoretically, the larger M is, the more accurate the estimate of the noiseless observable μ will be [140]. However, in practice, the noise level of the VD observable will not decrease linearly with M, and may increase at large M, when the error from the controlled swap (CSWAP) operations exceeds the error suppression given by VD, and the estimate will become less accurate. The optimal value of M is therefore a trade-off between the accuracy of the estimate and the noise level of the observable.

Equation (4.29) tells us that there is an exponential suppression of the noise with respect to the number of copies of N qubit states, but to accurately make an extrapolation on (7.1), the values of b_M must be estimated accurately. This requires the preparation of a toy noise model.

Assuming that every gate has a probability λ of having an error. The error channel is such that every different series of errors results in orthogonal states. To ensure that there is a noise floor, it is also

possible to append a coherent error given by some unitary $U(\theta) = \exp(-i\theta H)$, where H is some Hermitian operator, this ensures that there is a noise floor when $\theta \neq 0$. Since all states are rotated equally, the orthogonality relationship between the error induced states is preserved. Given G gates, the output of some circuit can be written as ρ can be written as

$$\rho = (1-\lambda)^G \rho_0 + (1-\lambda)^{G-1}\lambda \sum_{j=1}^{G} \rho_j + (1-\lambda)^{G-2}\lambda^2 \sum_{j_1 \neq j_2} \rho_{j_1,j_2} + \cdots , \tag{7.2}$$

when λ is small, ρ_0 dominates. Now to estimate the b coefficients in (7.1). To do this, take (4.29) and substitute ρ with (7.2):

$$\tilde{\mu}_M^{\mathrm{VD}} = \frac{\mathrm{Tr}[\rho^M O]}{\mathrm{Tr}[\rho^M]} =$$

$$\frac{(1-\lambda)^{MG} O_0 + (1-\lambda)^{M(G-1)}\lambda^M \sum_{j=1}^{G} \mathrm{Tr}[\rho_j O] + \cdots}{\left[(1-\lambda)^M + \lambda^M\right]^G}$$

$$= O_0 + \lambda^M \alpha_1^{(M)} + \lambda^{M+1} \alpha_2^{(M)} + \cdots , \tag{7.3}$$

where $\alpha_1^{(M)}$, $\alpha_2^{(M)}$ are the coefficients of series expansion in λ in $\tilde{\mu}_M^{\mathrm{VD}}$. Grouping the terms via the index j in $\alpha_j^{(M)}$ gives

Note: $O_i = \langle \psi_i | O | \psi_i \rangle$, and ψ_0 refers to the dominant eigenvector of the noisy ρ.

$$\mu'_{M_{\max}} = \sum_{M=1}^{M_{\max}} b_M \tilde{\mu}_M^{\mathrm{VD}}$$

$$= O_0 \sum_{M=1}^{M_{\max}} b_M + \sum_{M=1}^{M_{\max}} b_M \sum_{j=1}^{\infty} \alpha_j^{(M)} \lambda^{j+M-1}$$

$$= O_0 \sum_{M=1}^{M_{\max}} b_M + \sum_{j=1}^{\infty} \sum_{M=1}^{M_{\max}} b_M \alpha_j^{(M)} \lambda^{j+M-1}$$

$$= O_0 \sum_{M=1}^{M_{\max}} b_M + \sum_{j=1}^{M_{\max}-1} \sum_{M=1}^{M_{\max}} b_M \alpha_j^{(M)} \lambda^{j+M-1} + b_1 \alpha_{M_{\max}}^{(1)} \lambda^{M_{\max}} + \mathcal{O}(\lambda^{M_{\max}+1}). \tag{7.4}$$

Now a system of equations for b can be written as

$$\sum_{M=1}^{M_{\max}} b_M = 1, \tag{7.5}$$

$$\sum_{M=1}^{M_{\max}} b_M \alpha_j^{(M)} \lambda^{j+M-1} = 0 , \quad \forall j = 1, ..., M_{\max}-1 , \tag{7.6}$$

which gives

$$\mu'_{M_{\max}} = O_0 + b_1 \alpha_{M_{\max}}^{(1)} \lambda^{M_{\max}} + \mathcal{O}(\lambda^{M_{\max}+1}) . \tag{7.7}$$

Taking (7.6) with small λ, the terms $b_1, b_2\lambda, b_3\lambda^2, \ldots, b_{M_{\max}}\lambda^{M_{\max}-1}$ need to have the same order, or would otherwise be overdetermined. Furthermore, (7.5), (7.6) imply that $b_{M_{\max}}$ is of order 1, which leads to $b_1 \in \mathcal{O}(\lambda^{M_{\max}-1})$. If the system of equations (7.5, 7.6) has a solution gives

Remembering that the $\alpha_j^{(M)}$ coefficients do not depend on λ.

If (7.5) and (7.6) were an overdetermined system of linear equations, it would be unlikely to be soluble.

$$|\mu'_{M_{\max}} - O_0| = |\mu'_{M_{\max}} - \mu - \varepsilon| \in \mathcal{O}(\lambda^{M_{\max}+1}). \tag{7.8}$$

Thus, linear combination of VD error-suppressed observables improves their mitigation proportional to the largest used $M = M_{\max}$. This hints at the fact that perhaps they may be useful in addition to other linear combinations, such as ZNE.

7.1.2 Virtual Distillation with Zero Noise Extrapolation

The problem with the linear combination of VD copies is that the noise floor is not cancelled out as shown in (7.8). The hypothesis is that to mitigate this, one could use a ZNE like idea and further reduce the effect of the error by taking samples with multiple VD copies and multiple noise levels.

To continue on this path, using the noise model in (7.3) and find that if all gate noise is corrected for and only retain the coherent noise given by $U(\theta) = e^{-i\theta E}$, then our extrapolated VD observable is

$$O_0 = \text{Tr}[e^{-i\theta E}|\psi_{\text{exact}}\rangle\langle\psi_{\text{exact}}|e^{i\theta E}O] \ . \tag{7.9}$$

If O and E do not commute then,

$$\tilde{\mu}_M^{\text{VD}}(\theta, \lambda) = \text{Tr}[e^{-i\theta E}|\psi_{\text{exact}}\rangle\langle\psi_{\text{exact}}|e^{i\theta E}O] + \sum_j^{M_{\max}} \lambda^{M+j-1}\alpha_j^{(M)}. \tag{7.10}$$

Rewriting (7.4) with the dependency on λ and θ gives

$$\mu'_{M_{\max}}(\theta, \lambda) = \sum_{M=1}^{M_{\max}} b_M \hat{\mu}_M^{\text{VD}}(\theta, \lambda) \ , \tag{7.11}$$

and using the coefficients computed in (7.5) and (7.6) yields a similar expression to above with the error in the order of $M_{\max+1}$,

$$\mu'_{M_{\max}}(\theta, \lambda) = \text{Tr}[e^{-i\theta E}|\psi_{\text{exact}}\rangle\langle\psi_{\text{exact}}|e^{i\theta E}O] + \mathcal{O}(\lambda^{M_{\max}+1}).$$

Since when $M_{\max}$ is large and λ is small (as previously assumed), the error term $\mathcal{O}(\lambda^{M_{\max}+1})$ is negligible and thus the dependence on λ can be removed, giving a similar expression to (7.9),

$$\mu'_{M_{\max}}(\theta) \approx \text{Tr}[e^{-i\theta E}|\psi_{\text{exact}}\rangle\langle\psi_{\text{exact}}|e^{i\theta E}O]. \tag{7.12}$$

It is now much simpler to make a MacLaurin series expansion on θ with expansion coefficients β_j to get

$$\mu'_{M_{\max}}(\theta) = \text{Tr}[|\psi_{\text{exact}}\rangle\langle\psi_{\text{exact}}|O] + \sum_j^{\infty} \beta_j\theta^j. \tag{7.13}$$

Truncating the series at $n + 1$ allows the use of Richardson extrapolation as in ZNE if one considers $c_0\theta_0 = \theta_0, c_1\theta_0, \ldots, c_n\theta_0$, then the new mitigated observable becomes

$$\mu''_{M_{\max},n} = \sum_{j=0}^n \gamma_j \mu'_{M_{\max}}(c_j\theta_0) \ , \tag{7.14}$$

Now the linear combination of VD-mitigated observables of $M_{\max}$

$\sum_{j=0}^n \gamma_j = 1$ and $\sum_{j=0}^n \gamma_j c_j^k = 0$ for $k = 1, \ldots, n$

different number of copies at n different coherent noise levels can finally be written as

$$\mu''_{M_{\max},n} = \sum_{j=0}^{n} \sum_{M=1}^{M_{\max}} \gamma_j b_M \tilde{\mu}_M^{VD}(c_j\theta_0, \lambda) . \qquad (7.15)$$

It turns out that this is the only expression required for both combining ZNE and Richardson-extrapolated VD, which turns out to be the ansatz for UNITED as well, since the training circuits would be run at different copies and noise levels, upon which a regression model is trained.

This is in effect a combination of (7.11) and (7.13)

Now the theoretical error scaling in this simplified model can also be computed as

$$\mu''_{M_{\max},n} - \mu = \mathcal{O}(\theta_0^{n+1}). \qquad (7.16)$$

If the reasonable assumption that the orthogonal errors are small, then θ and λ can be scaled. Choosing $\theta_j = c_j\theta_0$ and $\lambda_j = c_j\lambda_0$, results in

$$\mu''_{M_{\max},n} = \sum_{j=0}^{n} \sum_{M=1}^{M_{\max}} \gamma_j b_M \hat{\mu}_M^{VD}(c_j\theta_0, c_j\lambda_0), \qquad (7.17)$$

which gives the same error scaling as (7.16). This is because coherent mismatch dominates at large $M_{\max}$.

7.2   UNITED

This section presents UNIfied Technique for Error mitigation with Data UNITED, which is the combination of CDR, ZNE and VD. Chapter 4.2 and the previous sections showed that individual QEM and their combinations can be used to mitigate errors. UNITED is an extension with the property that the individual QEM can be used as a special case of UNITED. For example, using only one copy for the VD part of UNITED would result in the vnCDR technique [132], while using only one noise level would result in the technique presented in section 7.1.1. Technically all the combinations are always used on the Clifford training circuits that then train an extended CDR model, which is then used to mitigate the error on the target circuits, bypassing this steps yields the original QEM techniques, these combinations are summarized in 7.1.

The UNITED dataset is generated with a set of N_t near-Clifford circuits projected from target circuit C as in CDR in section 4.2.3, yielding $\{T_i\}_{i=1}^{N_t}$. Like in CDR, the exact expectation values μ_i of an observable O of these circuits are computed classically. The noisy and VD suppressed expectation values at n different noise levels c_0, $j = 0, \ldots, n$ and for $M = 1, \ldots, M_{\max}$ different number of copies, with $M = 1$ representing a standard run without VD by convention. This generates a training set

$$\mathcal{T}^{UNITED} = \{(\tilde{\boldsymbol{\mu}}_i^{UNITED}, \mu_i)\}_{i=1}^{N_t}, \qquad (7.18)$$

with one training value for UNITED being $\tilde{\boldsymbol{\mu}}_i^{UNITED} = (\tilde{\mu}_{i,0,1}, \ldots, \tilde{\mu}_{i,n,M_{\max}})$. This is a tensor of the c_j^{th} noise level and M-copy VD suppressed

Technique	Noise Levels	Copies	Training Circuits
ZNE [112]	✓	✗	✗
VD [115, 114]	✗	✓	✗
CDR [113]	✗	✗	✓
vnCDR [132]	✓	✗	✓
CDR+VD [110]*	✗	✓	✓
UNITED [110]*	✓	✓	✓

Table 7.1: **Resource requirements** for the different error mitigation strategies, showing the necessary resources for QEM strategies that are unified and benchmarked in this chapter. Modified from [110] under CC4.

expectation value of O of the training circuit T_i. $\mathcal{T}^{\text{UNITED}}$ trains function $f^{\text{UNITED}} : (\mathcal{E}_{\text{VD}})^{M_{\max}(n+1)} \to \mathcal{E}_{\text{noiseless}}$ which has been chosen to be

$$f^{\text{UNITED}}(\boldsymbol{x}) = \sum_{j=0}^{n} \sum_{M=1}^{M_{\max}} d_{j,M} x_{j,M} \,. \tag{7.19}$$

$\mathcal{E}_{\text{VD}}$ represents the VD-suppressed expectation values, although it does also include the standard noisy expectation values of $M = 1$.

The coefficients $d_{j,M}$ in (7.19) are determined by minimizing the loss function with respect to $\mathcal{T}^{\text{UNITED}}$, which in this work is done via least-squares regression. The trained model can then obtain a mitigated expectation value O for the target circuits C by evaluating the function f^{UNITED} on the noisy and VD suppressed expectation values of the target circuits.

7.2.1 UNITED for an idealized error model

To show that the UNITED combination is more effective than the single QEM techniques, the idealized error model presented in section 7.1.1 can be used again. In such a model, there is control of both the orthogonal noise and coherent noise. The dataset for such a model is as in the previous section denoted $\tilde{\mu}^{\text{UNITED}}$, now with $n + 1$ distinct error rates and $M_{\max}$ different number of copies. For a given noise level c_j the error rate is scaled as $\theta_j = c_j \theta_0$, where θ_0 is the inherent error coherent error in this model, parameterized by θ.

As shown in section 7.1.1, increasing the number of copies can suppress the orthogonal errors via only the combination of VD suppressed observables. Following from (7.4)

$$\mu'_{M_{\max},j} = \sum_{M=1}^{M_{\max}} b_M \hat{\mu}_{j,M} \,. \tag{7.20}$$

Then with (7.8) error scaling is

$$|\mu'_{M_{\max},j} - \mu - \varepsilon_j| \in \mathcal{O}(\lambda_j^{M_{\max}+1}) \,, \tag{7.21}$$

with ε_j representing the noise floor for a controllable error rate c_j. On these results, Richardson extrapolation has been shown to suppress the coherent noise as in section 7.1.2.

This is equivalent to section 7.1.2, and the final estimate is as before,

$$\mu''_{M_{\max},n} = \sum_{j=0}^{n} \gamma_j \mu'_{M_{\max},j} = \sum_{j=0}^{n} \sum_{M=1}^{M_{\max}} \gamma_j b_M \hat{\mu}_{j,M}, \tag{7.22}$$

where γ_j are those of (4.14), reiterating the results of section 7.1.2, at large $M_{\max}$ the error is

$$|\mu''_{M_{\max},n} - \mu| \in \mathcal{O}(\theta_0^{n+1}). \qquad (7.23)$$

So far nothing new has been shown, but the important factor that makes determining the coefficients γ and b is that the error rates for both the different number of copies and chosen noise amplifications for the machine are not known. They can only be estimated, and this is the main source of error for Richardson extrapolation when used in this context, as it is very sensitive to inaccurate coefficients [75]. UNITED now utilizes the training circuits to estimate these coefficients in a noise agnostic manner.

7.2.2   Limitations of UNITED

The main driving force of UNITED is the combination of multiple noisy observables that brought together via Richardson extrapolation, whose coefficients are determined by a regression model. This motivates the UNITED ansatz (7.19), but it is not enough to ensure its performance. The reasons for this are that any scaling to physical error rates are approximate when applied to hardware. This leads to imperfect extrapolation in the case of ZNE and vnCDR, even though both techniques are readily used to great effect [116]. Model training as in UNITED is also not identical to Richardson extrapolation, since it is completely dependent on also having a good training set. Yet, if such a training set *is* available, then fitting the coefficients during training does not require knowledge of the noise levels, which ideally minimizes the impact of imperfect control, as long as it is consistent.

Errors on current devices are very frequent, and even though hardware is continuously improving in terms of qubit and gate fidelity [261], the error rates are not negligible. This causes expectation values to concentrate around 0 since the output of any measurements tends towards a random bit string. As such, to compensate many shots are required to extract information [262]. As such, noise levels cannot be increased arbitrarily high, and even VD with numerous copies will introduce additional errors from the CSWAP operations. Of course, the use of VD data within united allows for more error rates to be included indirectly, and UNITED learns the best combination of these disparate noisy observables.

The foundational limitation of any QEM technique is the number of shots required, not only does adding noise levels and copies increase the number of experiments required, but these must evenly be distributed between all the training circuits one decides to use. Thus, any shot budget will be subdivided between all the combinations of training circuits at different copy numbers and noise levels.

7.3   Trial by Fire: Numerical Experiments

7.3.1   The Test Problems

UNITED is tested through two different metrics, one is that of mitigating random observables of random quantum circuit (RQC)s which represents a hard theoretical challenge to correct in general [263], while the other can be seen as a more practical application, the QAOA algorithm on the Max-Cut problem. To summarize, first problem is used to test the theoretical limits of the UNITED technique, while the second is used to test the practical limits. The system sizes considered range from 4 to 10 qubits Q , specifically $Q = 4, 6, 8, 10$.

RANDOM QUANTUM CIRCUITS are investigated for circuit depths defined by a g factor and number of qubits Q, where the depth, given as number of layers L is

$$L = gQ. \tag{7.24}$$

The circuits are constructed with the native gates of the chosen trapped-ion machine with a brick-like structure of nearest-neighbour Molmer-Sørensøn gates [264] with interlaced single qubit unitaries as in figure 7.1.

For such circuit structures, as L increases, RQCs are expected to give states that converge to random states sampled via a Haar measure, which is a uniform distribution over the Hilbert space [265]. This property is what makes RQCs a challenging task for error mitigation.

THE QUANTUM ALTERNATING OPERATOR ANSATZ is a stand in for a practical algorithm to solve optimization problems. The algorithm was presented in more detail in section 3.2. In this research, QAOA [56, 266] is used to find a graph bipartition (cut) that maximizes the number of edges connecting the halves, the so-called Max-Cut problem. Formally, take the graph $G = (V, E)$ with V the set of vertices and E the set of edges. The aim is to partition of the vertices into two sets S_1 and S_2 such that the number of edges connecting the two sets is maximized. In this setting, the algorithm minimizes the energy of the Hamiltonian

$$H_P = \sum_{(i,j) \in E} \sigma_Z^i \sigma_Z^j \tag{7.25}$$

defined over a number of qubits $Q = |V|$. In particular, the numerical study focuses on Erdös-Renyi graphs [267], which form an edge between vertices (i, j) with some probability p_e. The layers of the circuit are set as $L = Q$. For each Q, 28 unique graphs are constructed, with $p_e = 0.5$. This ensures that the number of graphs is consistent between the system sizes and is limited by the fact that on 4 qubits 28 is the maximum number of non-trivial graphs that can be constructed.
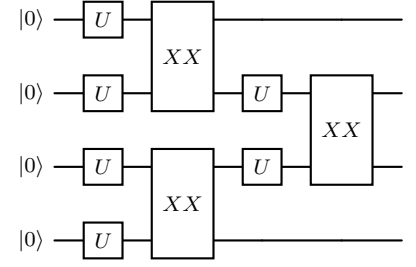
Figure 7.1: **Exemplary RQC** used in benchmark simulations for 4 qubits. A layer is built from two layers of alternating nearest-neighbor Molmer-Sørensøn gates $XX(\delta)$, around these gates the generic single-qubit unitaries $U(\alpha, \beta, \gamma)$ are inserted. The rotation angles $\alpha, \beta, \gamma, \delta$ are assigned randomly for each RQC. The parameters and indices are not shown in the figure for brevity, but each gate has a unique set of parameters.

7.3.2   Shot Budget

An essential aspect of running computations on quantum hardware is the shot budget, or how many samples one is willing to take over all experiments. This is simply some integer that will be referred to as N_{tot} and would be distributed in some way between all the experiments required to construct a dataset. Keeping table 7.1 in mind, then it should be clear that the more concurrent techniques are used, then each individual sample will have fewer shots assigned to it.

For VD, measurements are taken for M_{max} different circuits. ZNE, requires the expectation value of $n+1$ circuits built from one target circuit at $n+1$ noise levels. vnCDR needs to evaluate N_t training circuits at $n+1$ noise levels for a total of $(n+1)(N_t+1)$ circuit evaluations. UNITED is the hungriest technique requiring these observable estimates for each training circuit N_t: VD with $M_{\max}-1$ copies, each at $n+1$ noise levels. On top of this, the target circuit measurement is also required, giving a total number of circuits $(n+1)(N_t+1)(2M_{\max}-1)$ that must be evaluated.

7.3.3   The Noise Model

An ion-trap based quantum error model made by Trout *et al.* [268] and developed in [269] is used. This is because a quantum computer with all-to-all connectivity is required for VD. The gates R_X, R_Y, R_Z come with noise channels $\mathcal{S}_\Xi$ where Ξ is one of the Pauli matrices, while and XX has $\mathcal{S}_{XX}$:

$$\mathcal{S}_\Xi = \mathcal{W}^{p_d} \circ \mathcal{D}^{p_{\text{dep}}} \circ \mathcal{R}_\Xi^{p_\alpha} \,,$$
$$\mathcal{S}_{XX} = \left[\mathcal{W}_1^{p_{d_1}} \otimes \mathcal{W}_2^{p_{d_2}} \right] \circ \left[\mathcal{D}_1^{p_{\text{dep}}} \otimes \mathcal{D}_2^{p_{\text{dep}}} \right] \circ \mathcal{H}^{p_{\text{xx}}} \circ \mathcal{H}^{p_h}. \qquad (7.26)$$

Several terms have to be clarified, with $\mathcal{R}_\Xi^{p_\alpha}(\rho) = (1-p_\alpha)\rho + p_\alpha \Xi \rho \Xi$, which represent errors in the angle of rotation about the X, Y, Z axes. The various channels are defined as:

$$\mathcal{D}^p : \rho \mapsto (1-p)\rho + \frac{p}{3}(\sigma_X \rho \sigma_X + \sigma_Y \rho \sigma_Y + \sigma_Z \rho \sigma_Z), \qquad (7.27)$$

$$\mathcal{W}^{p_d} : \rho \mapsto (1-p_d)\rho + \sigma_Z \rho \sigma_Z \,, \qquad (7.28)$$

$$\mathcal{H}^p : \rho \mapsto (1-p)\rho + p(\sigma_X \otimes \sigma_X)\rho(\sigma_X \otimes \sigma_X) \qquad (7.29)$$

(7.27) is the local depolarizing channel with $p = p_{\text{dep}}$ representing depolarizing gate noise and $p = p_{\text{idle}}$ simulating idling noise [270]. (7.28) is the dephasing channel, and (7.29) is a two-qubit channel representing an error in rotation when $p = p_{\text{xx}}$ and ion heating when $p = p_h$. The error rates were chosen as

$$p_d = 1.5 \cdot 10^{-4}, \quad p_{\text{dep}} = 8 \cdot 10^{-4},$$
$$p_{d_1} = p_{d_2} = 7.5 \cdot 10^{-4}, \quad p_\alpha = 1 \cdot 10^{-4}, \qquad (7.30)$$
$$p_{\text{xx}} = 1 \cdot 10^{-3}, \quad p_h = 1.25 \cdot 10^{-3},$$
$$p_{\text{idle}} = 8 \cdot 10^{-4}.$$

No errors occur on measurement, as it is assumed that these can be well corrected separately from circuit error mitigation [271, 272].

7.3.4 Finding Optimal Hyperparameters

To maximize the use of resources, it is essential to discover what the optimal hyperparameters are for the different QEM techniques used. Specifically, it is important to know what the optimal M_{max} VD copies, the optimal number of noise levels n and the size of the near-Clifford training set N_t. It has already been determined in the literature that $N_t = 100$ [113, 132] yields heuristically good results, and 3 noise levels for ZNE is shown to be good in vnCDR [132], but this was found not to be the case here. Thus, the number of copies for VD needs to be determined. In theory, the exact distribution of the shot budget is also an important parameter, but this requires a more detailed study, so a strategy of evenly distributing the shots between the different entries in the dataset is used.

To benchmark ZNE we use two different ansatze for the function, linear and exponential. The linear ansatz can be tested with 2 and 3 noise levels, while the exponential can only be used with 3 noise levels. These are tested with at all noise budgets. Figure 7.2 shows that out of all combinations, the linear ansatz with 2 noise levels performs better in almost all cases. For clarity only 4 and 10 qubit systems are shown, but this was consistent across all system sizes. The exponential ansatz was found to always perform around an order of magnitude worse.

Similarly, for VD it is important to determine how many copies are optimal. There is an additional consideration which is whether the noise incurred in the CSWAP operations is significant. Since the net effect of additional noise will be that the results will be worse with respect to additional copies, the study on the optimal number of copies is done with this noise. The results are shown in figure 7.3, where it is found that for low depth circuits ($g = 1$) the optimal number of copies is $M = 2$, but at higher depths the situation is more complicated. For 4 qubits, $M = 4$ is optimal, and although it is not shown, the behaviors for larger systems is more complicated. As such, $M = 3$ is chosen as a compromise, since when testing the UNITED technique, it was found that the results were not significantly worse than with $M = 4$, but the computational cost was significantly lower, as VD requires $N_{qubits} = M * Q$, where Q is the size of the problem. There is a numerical trick used in the simulation that allows for the use of $N_{qubtis} = 2Q$, but at the cost of an M factor in depth, although it is outside the scope of this work, the technique is called REQUEST and can be read about in [273]. In short, instead of using M copies of the state, 2 copies are used and one of the states is reset to $|0\rangle^{\otimes Q}$ after each CSWAP operation and then prepared again. This has the same net effect, with the cost that now the depth of the circuit is multiplied by the number of copies and their respective preparation circuits.

The first is that the CSWAP operation (also known as derangement) could be noisy in certain settings, as such the effect of this must be analyzed on both VD and UNITED. Figure 7.4 shows that
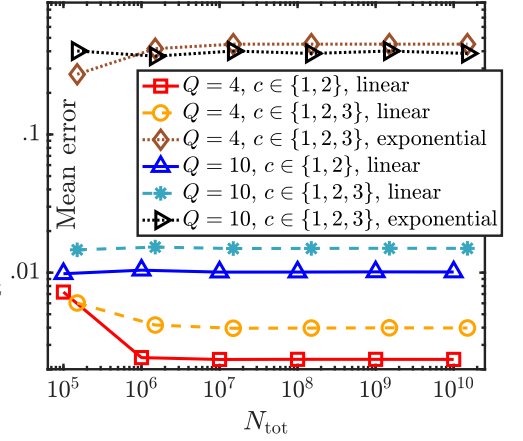
Figure 7.2: **Convergence of various ZNE implementations.** The plot shows the mean absolute error of $\langle \sigma_Z^1 \rangle$ over the shot budget. 30 RQCs with $Q = 4, 10$ and $L = Q$ are used. Please refer to the legend for the different ZNE implementations. Reproduced from [110] under CC4.

Figure 7.3: **Convergence of VD estimates with increasing copies** M. The results of the mean absolute error on the observable σ_X at $N_{tot} = 10^{10}$ against of VD copies M are shown. The $M = 1$ point is equivalent to not using VD and various depth scaling factors g and system size Q are shown (refer to legend). Reproduced from [110] under CC4.

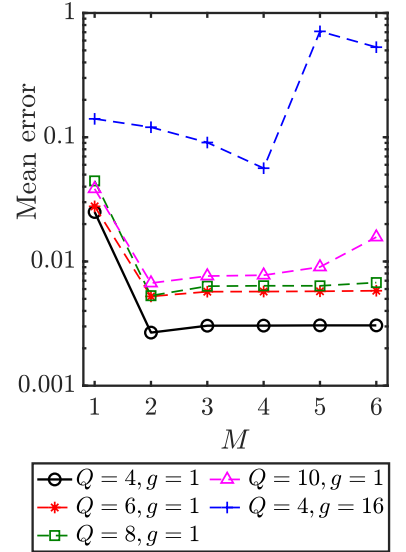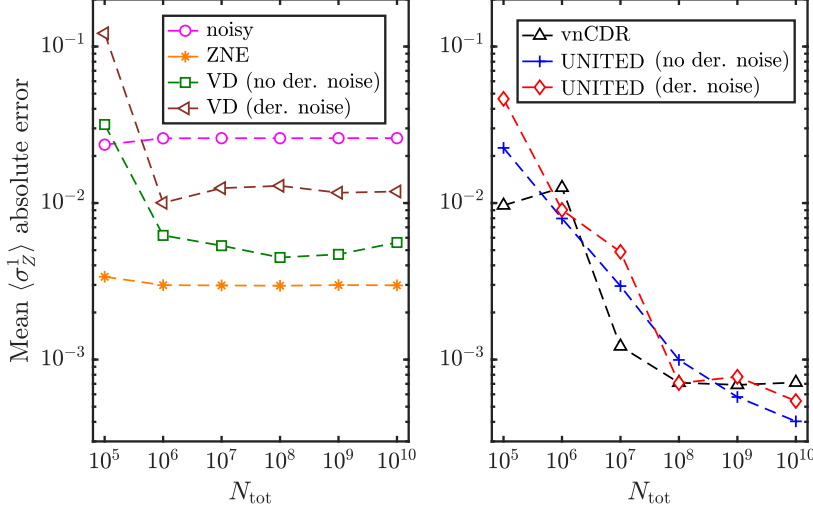Figure 7.4: **Comparing VD and UNITED** with and without CSWAP (*derangement*) noise . This was done on RQCs with $Q = 4$ and $L = 5$. Both the mitigated and noisy $\langle \sigma_Z^1 \rangle$ averaged over 30 RQCs are plotted. The CSWAP noise has an effect on both VD and UNITED, specifically harming both even at large N_{tot}. As a comparison, unmitigated, ZNE, and vnCDR results are shown. Reproduced from [110] under CC4.

there is an effect on VD and UNITED to a lesser extent when derangement noise is considered. Irrespective of this, there is a 2.5 times improvement for VD with $N_{\text{tot}} \geq 10^6$ and factor 1.2 improvement of UNITED over vnCDR at $N_{\text{tot}} = 10^{10}$. The unmitigated, ZNE, and vnCDR values are shown for comparison. Linear extrapolation ZNE is performed using $n = 2$ noise levels, which was found to be optimal previously. VD uses $M = 2$ copies which is shown to be optimal in figure 7.3, although this is not always the case, the improvement for the high depth circuits does not justify the costs. vnCDR uses 3 noise levels which was found to be optimal in [132]. UNITED also uses 3 noise levels to make the comparison to vnCDR as fair as possible. UNITED does use $M = 3$ copies, which was found to be optimal. The values are all summarized in table 7.2

Technique	Noise Levels	Copies	Training Circuits (near-Clifford)	Total Circuits
ZNE	2 (linear)	✗	✗	2
VD	✗	2	✗	2
CDR	✗	✗	100 (10)	100
vnCDR	3 (linear)	✗	100 (10)	300
UNITED	3 (linear)	3	100 (10)	900

Table 7.2: **Amount of component parts** for the different error mitigation strategies used to benchmark the various techniques.

The last hyperparameter that is unique to vnCDR and UNITED is the actual construction of the training set, which has its own subtleties. The number of training circuits is one thing, but in section 4.2.3 another problem is mentioned, specifically that the near-Clifford circuits expectation values approach 0 as Q increases. This requires a larger number of shots to resolve if nothing is done. As such it would be beneficial to increase the variance of the circuits. This is particularly important for the QAOA problem, as the structure of the ansatz is fixed and the exact expectation values are exactly 0. RQC do not have such a problem as the varying circuit structure lowers the probability that the expectation value is exactly

0.

Therefore, a strategy based on [139] is used. Here, 10^5 training circuits are constructed and post-selected fo that the exact expectation value is roughly uniformly distributed in the interval $[-0.5, 0.5]$ with 10 non-Clifford gates. Since simulating Clifford and near-Clifford circuits is classically possible, this solution is scalable. It is also possible to use more complicated techniques such as Markov-Chain Monte Carlo [273] to come to the same conclusion, but as the net effect is identical, the simpler solution is used. Figure 7.5 shows that vnCDR is not affected by post selection as much as UNITED at finite shot budgets, but at infinite shot budgets the effect is negligible, this is expected since the main issue with the concentration of expectation values around 0 is that the differences are small and need to be resolved, at the infinite shot limit they are resolved, and the issue does not exist.

7.3.5   Results

This section presents the results of the above two benchmark cases for all techniques apart from the combination of VD and CDR which is not shown. This is because the results of this in practice compared to the older vnCDR and the newly presented UNITED does not suggest that it improves greatly on CDR alone. The result is important in deriving UNITED, but including it would make already busy graphs become unnecessarily packed.

The plots are presented in two ways, one is the behaviour of the technique with increasing shot budget, and the other is the behaviour of the technique with increasing system size. The first shows when techniques start to stop being able to correct and the second shows the all important scaling with system size, which as discussed in section 4.2 is an important metric for any QEM technique. In particular, the mean and maximal absolute errors are shown. The former gives the expected performance of the techniques over many applications, while the latter gives an idea of the worst case performance.

Although another interesting comparison is the one of fixing the number of samples per database circuit instance, but the large disparity in number of shots would make this comparison not be practically useful. Since QEM is more of a practical area of application, the total number of shots is a much more important consideration than the number of shots per circuit.

FOR RANDOM CIRCUITS, There are two cases studied in RQCs, the low depth ($g = 1$) case where $L = Q$, which is studied for $Q = 4, 6, 8, 10$ and the high depth ($g = 16$) case where $L = 16Q$ for $Q = 4, 6, 8$. $Q = 10$ was not done in the high noise regime both due to the computational cost, but also the results are too noisy to perform any error mitigation. For each pair (Q, L), 30 random circuits are generated, and for each one the expectation value of σ_Z

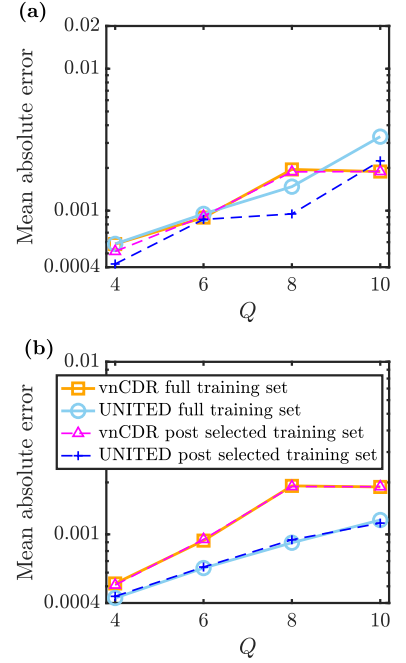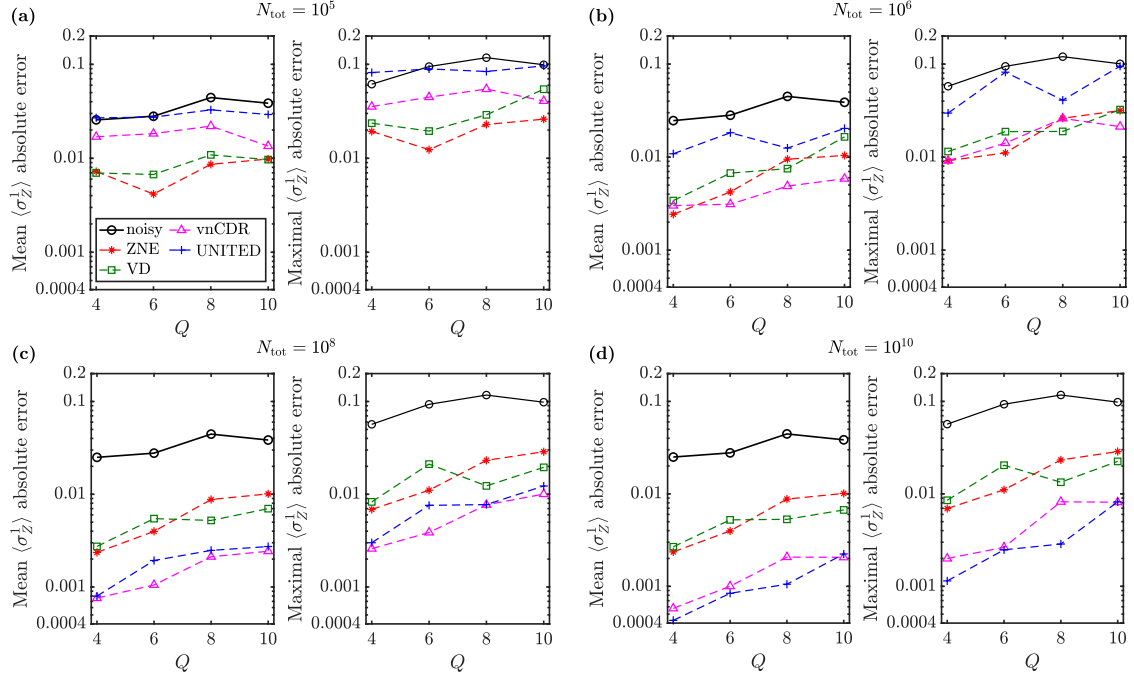Figure 7.5: **The effect of post-selection on vnCDR and UNITED**. Plots comparing the mean absolute errors of $\langle \sigma_Z^1 \rangle$ for random and post selected CDR training circuits at N_t=100. **(a)** shows the highest shot budget N_{tot}=10^{10}. While the infinite N_{tot} limit is in **(c)**. The results are the aggregate of 30 instances of g=1 RQCs. Reproduced from [110] under CC4.

Figure 7.6: **Comparing QEM methods on RQCs for layer depth** g=1**,** showing the mean and maximal absolute errors of the expectation value of σ_Z on the first qubit using the different QEM strategies, denoted $\langle\sigma_Z^1\rangle$. The title refers to the shot budget used for the given plot results for the total number of shots per mitigated and noisy expectation values. ZNE is the best at 10^5, vnCDR for N_{tot}=10^6 to 10^8 and UNITED at N_{tot}=10^{10}. Reproduced from [110] under CC4.

on the first qubit is shown. It should be noted that in when $g = 1$ the clustering problem discussed in section 4.2.3 does not occur, as this effect becomes more pronounced the deeper the circuits are.

The results for $g = 1$ are shown in figure 7.6, it is clear that the QEM performance is strongly correlated to N_{tot} and decreases as system size increases. ZNE always offers the best performance at low shot budgets, with VD performing slightly worse. Although this is more evident in figure 7.8(a) where only $Q = 4$ is shown, both VD and ZNE have in essence already converged, and their performance does not improve as N_{tot} increases, while vnCDR and UNITED converge slower but eventually overtake the single techniques at 10^7 and 10^8 respectively. Eventually vnCDR begins to reach convergence at 10^8 while UNITED keeps improving, becoming the best technique at 10^9. Looking at the relationship with system size gives a very similar picture, with the interesting caveat that at $Q = 10$ the performance does not overtake vnCDR. This is because at larger system sizes more shots are required to get a high quality expectation value per circuit.

$g = 16$ yields a very different picture. Here ZNE does not yield any improvements over the unmitigated value while VD always worsens the results except for minimal improvements at $Q = 4$. This is basically the case for all the techniques at higher shot budgets and larger system sizes. The disappointing result is due to the face that the unmitigated values, and the measured expectation values for the training circuits are probably of too low quality to be able to properly train the correction model. This is why 7.8 only compares the $Q = 4$ system, as it is the only one where the error rate is not large enough to completely wash out the mitigation ability of the techniques.
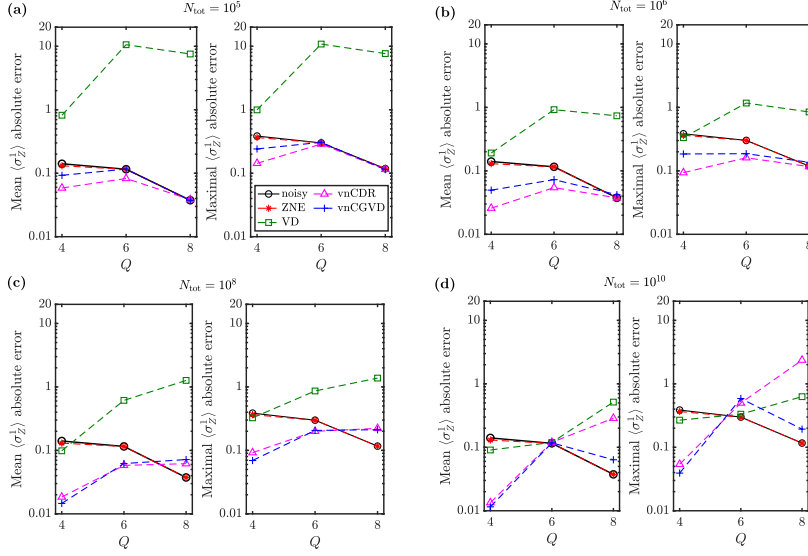
Figure 7.7: **Comparing QEM methods on RQCs for** $g=16$**,** showing the mean and maximal absolute errors of the expectation value of σ_Z on the first qubit using the different QEM strategies, deonted $\langle\sigma_Z^1\rangle$ at different shot budgets. vnCGVD was the working title of the UNITED technique, as this plot was not included in the final paper. Reproduced from [110] under CC4.

Finally, bringing all the RQC results together we can compare the performance over shot budget of $Q = 4$ at $g = 1, 16$ shown in figure 7.8. The unmitigated results are in essence already converged at 10^5. The ZNE and VD results converge at 10^6. Those of vnCDR converge slower, becoming optimal over VD and ZNE for N_{tot} from 10^7 to 10^{10}. UNITED converges slower than the others, but is the best choice at $N_{\text{tot}} = 10^9, 10^{10}$. For $g = 16$ ZNE and VD do not improve over the noisy results and are worse than vnCDR and UNITED for all considered shot budgets. UNITED improves over vnCDR for $N_{\text{tot}} = 10^9, 10^{10}$. It should be noted that these shot budgets are exceptionally generous and are only showing that at some point spending more shots using only a single mitigation method is not worth the time spent, and better strategies do exist. It is worth exploring the high noise regime a little more. It is known that here learning-based QEM is highly susceptible to shot noise [139], since just like in the training set problem, the values become clustered around the desired observable's expectation value of a random string as noise increases. Thus, as long as the number of samples is small, the measurements are highly susceptible to finite sample size effect. Indeed, this can be clearly seen in figure 7.8(b) where there is a peak in the vnCDR mitigation at $N_{\text{tot}} = 10^8$.

FOR QAOA the aim is to mitigate the expectation value of a random term of the Hamiltonian H_p (7.25) for the Max-Cut problem of the 28 Erdös-Renyi graphs. Figure 7.9 has the mean absolute errors of these expectation values, and are strongly dependent on N_{tot} as with RQCs. Unlike in the previous section, VD is much better at mitigating this family of circuits as opposed to RQCs. Indeed, it remains the best method at all system sizes until $N_{\text{tot}} = 10^8$, while ZNE behaves as before, doing quite well at the lowest budget but never improving after that. For all system sizes except for $Q = 4$
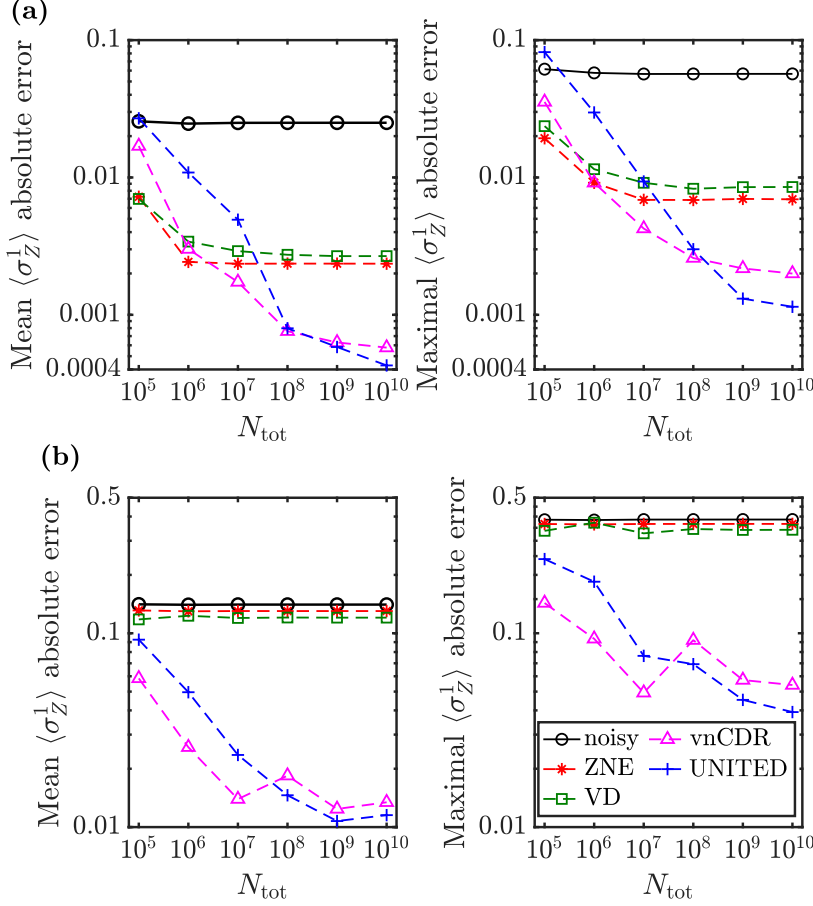
(a)

(b)

Figure 7.8: **Convergence of QEM applied to RQCs with increasing N_{tot}.** Mean and maximal absolute errors of the expectation value of $\langle \sigma_Z^1 \rangle$ for random quantum circuits plotted versus the shot budget. We use 30 instances of random quantum circuits. In (a) $Q = 4, g = 1$ is shown, while in (b) $Q = 4, g = 16$ is shown. Reproduced from [110] under CC4.

UNITED is the best at $N_{\text{tot}} = 10^{10}$ while vnCDR is the best at $Q = 4$.

Like before, it is also helpful to look at the performance of the techniques over the shot budget. Here depth is not a variable, so we can compare any system size. Since they all behave similarly, $Q = 8$ was chosen to be shown in figure 7.10. As before, VD is the best performer up to $N_{\text{tot}} = 10^8$, where it begins to converge. After which, UNITED becomes the best performer at 10^9. vnCDR also converges at 10^8.

It is quite strange to the uninitiated that VD could perform so well here and so badly in the RQC case. But a small reminder of section 4.2.4 will remind the reader that since the expectation values required of a QAOA algorithm are eigenstates of the mitigated observables and VD specifically suppresses all states that are not the dominant eigenvector (which is likely close to the eigenvector of the mitigated observable), it is expected that the noise floor of the technique is much lower than in the RQC case.

The QAOA performance of VD seems to also be a good predictor of the performance of UNITED over vnCDR. The poor VD results in the RQC circuits translated to slow improvement of UNITED with respect to shot budget, but here it clearly reaches much better results much faster. Clearly combining QEM strategies can leverage
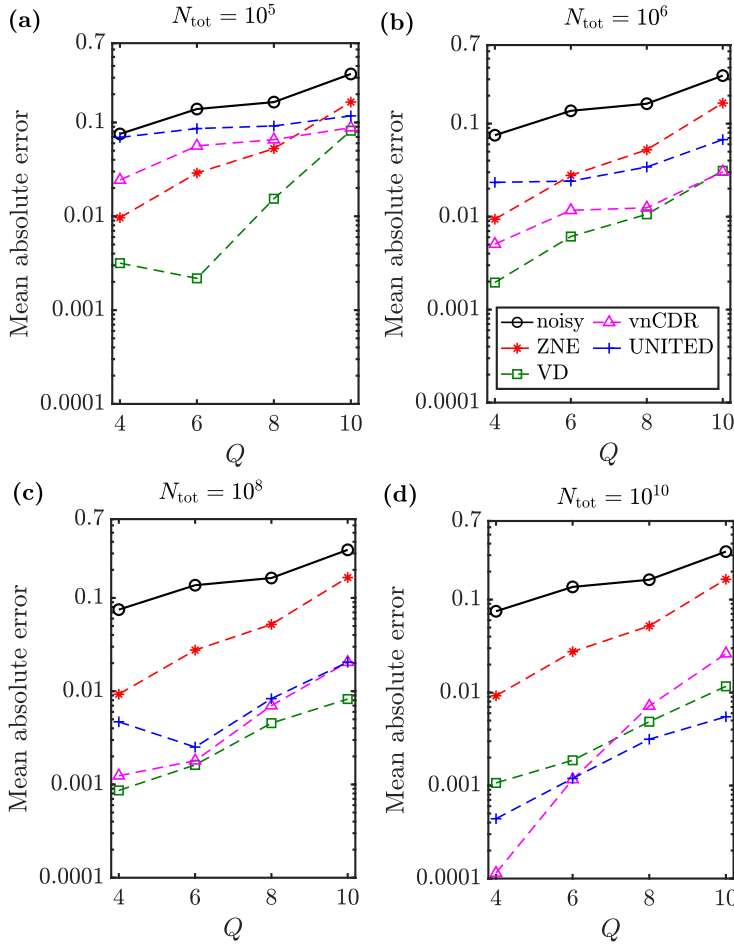
Figure 7.9: **Comparing QEM methods applied to Max-Cut QAOA.** Mean absolute errors of the expectation value of a random term of H_p averaged over 28 graphs at different shot budgets. For $N_{tot} = 10^5$, 10^6, 10^8, VD is optimal until at $N_{tot} = 10^{10}$, where UNITED overtakes it. Reproduced from [110] under CC4.
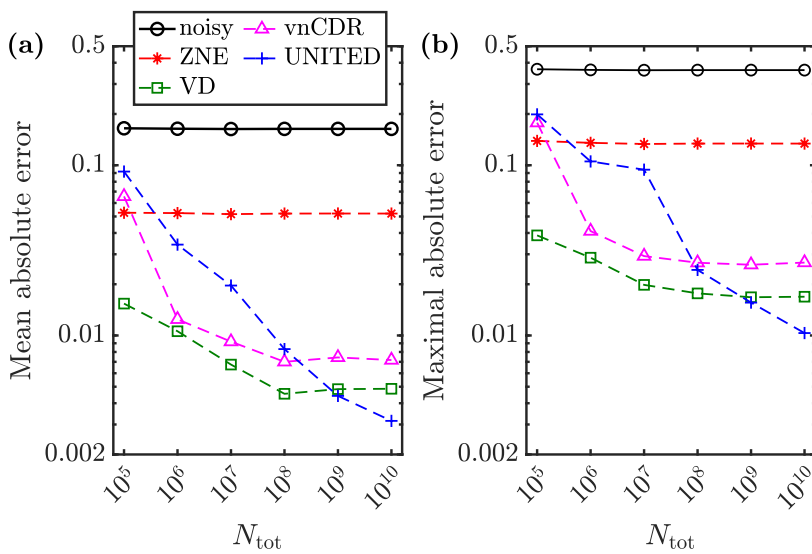
Reproduced from [110] under CC4.

Figure 7.10: **Convergence of error mitigation methods applied to the Max-Cut QAOA for $Q = 8$ with increasing N_{tot}.** Mean (a) and maximal (b) absolute errors of the expectation value of a random term of the Max-Cut Hamiltonian on 28 instances of Erdös-Renyi graphs. ZNE has already converged at $N_{tot} = 10^5$, while VD is the best performer up to its convergence at $N_{tot} = 10^8$, after which ($N_{tot} = 10^9$) UNITED becomes the best performer. vnCDR converges at $N_{tot} = 10^8$.

the strength of the individual strategies, but it is clear that potentially choosing the strategies most suited for the problem at hand, such as VD's apparent ability to correct well when the statevector is close to an eigenvector of the observable in the QAOA case.

7.4   Discussion

Although the usefulness of NISQ is not as clear as it was during this work's publication, it is still essential in this era and potentially even in the fault-tolerant era to mitigate the errors that remain, as error correction might not be completely perfect for a long time, and it might be that it may be advantageous to use a mixture of error corrected qubits and noisy qubits [111]. There have been countless approaches and combinations of various techniques that are better, although not exhaustively, explained in section 4.2. This particular work discussed some now standard techniques that were once state-of-the-art and attempted to combine them in a way that should allow other techniques to be replaced or added - of course, the way this could be done will always require some level of reasoning and consideration for how to exactly combine it as was done with ZNE, VD and CDR.

The methods and their combinations, like the older vnCDR and the newly proposed UNITED were used to mitigate errors under a realistic noise model to see how well they could correct a series of RQCs and QAOA problems for system sizes from 4 to 10 qubits. Specifically, the simulation was done for a trapped-ion device, as this is the most natural candidate for implementing VD which greatly benefits from the all-to-all connectivity.

UNITED was then derived by combining CDR with VD, forming a "Clifford-guided Virtual Distillation", this was then generalized by adding an extrapolation based on Richardson Extrapolation on the VD results at different error rates. This combination leads to first suppressing the effect of orthogonal errors from VD which are then extrapolated to the zero-noise limit. Finally, the model maps these mitigated observables to the exact observables with the hope that the same map generalizes to circuits with a similar structure.

Although the newly proposed technique did reach the lowest error in both benchmarks, this was at a very high number of shots (10^10) which in most cases would be unfeasible. The other techniques perform better with other budgets, but it is also clear from this work that the type of problem has a strong interaction with the type of error mitigation used in terms of quality of expectation values. Now, training an expensive model is not necessarily out of the question in the current experimental landscape. The issue is that the characteristics of noise in a QC can change over time, a process called *drift* [274], so an error mitigation model will not stay valid for very long, and may indeed be training on a changing error profile.

Another lesson from this work is that it is very difficult to guarantee if a QEM technique has indeed reached its maximal mitigation

ability. This is especially pertinent when there is no analytical guarantees, or if they exist when the error model is too simple. This lack of guarantees will lead to more uncertainty as quantum computers increase in size. Because although future available devices will have more, higher quality qubits, some effects of error propagation in the complex states and interactions that will be performed on these machines may be more insidious to the quality of result than we may expect at this time.

Regardless it is clear that at least for the following few years after the original paper's publication and probably that of this thesis, error mitigation has and will continue to have a large part in the experiments. This work is one of the early hybrid QEM strategies that have become quite used in the field, notably probabilistic error amplification, which was used in IBM's utility experiment [213], which uses probabilistic error correction to amplify rather than minimize, errors, and then uses something like Richardson extrapolation to extrapolate the errors to the zero noise limit with extremely well tuned error rates.

The work itself raised its own questions which perhaps a reader may have thought of already. The first is why these techniques specifically? The answer is that these that were chosen in particular all fit very well within an extrapolation strategy. VD reduces errors in a more or less known way, while controllably adding errors increases noise, and all these techniques could be blended by something akin to Richardson extrapolation. Training a model on exact-noisy pairs can sit above this approach. But of course, there is no reason to not think of other forms of data and combination that could be useful, either for specific problems or for the more challenging deep circuits. Other candidates for unification are using noise resilient algorithms [275, 269, 276, 109, 277], quasi-probabilistic error decomposition [112, 278], verified phase estimation [279], truncated Neumann series [280], and perhaps even making use of problem-specific post selection [281, 282, 283, 123].

A less complicated but more technical question is to explore the importance of dividing the shot budget unevenly between the different types of circuit. For example, it is well known that in VD the higher powers of the state lead to smaller expectation values which need higher precision/shots to measure. As such, there may be an optimal number of shots to give the $M = 1$ that is different from $M = 2$, and of course, this goes for the various different parameters.

Most importantly we see how large the effects of noise are, and especially with the consideration of controlled swap noise can quite significantly change the performance of the QEM even when trained on simulations with that noise present, as such to understand the potential of VD and UNITED in full, noisy classical simulation must also be improved.

7.5   Concluding Remarks

In retrospect, this work was quite predictive of where how the field of quantum error mitigation would evolve beyond this point. Although it was not the first to combine error mitigation techniques, it did show that the idea of combining error mitigation techniques was extensible to more than just two, as was done with vnCDR [132]. Indeed, the combination of error mitigation techniques, specifically ZNE and a hijacking of PEC to very controllably increase error rates was the basis for IBM's landmark utility experiment [213]. Another promising avenue is the use of multiple subspace expansions concurrently is another approach that is actively being explored

8

One Bad Qubit Ruins the Bunch

> The successful construction of all machinery depends on the perfection of the tools employed.
> *Industrial biography; iron-workers and tool-makers, Charles Babbage, 1864*

This chapter follows on from the previous, looking into the potential future after error mitigation but before error correction. It begins by introducing the idea that error corrected and noisy qubits could be used in conjunction. It then presents analytical derivations on the consequences of such a system, followed by numerical results and a discussion.

IT SHOULD BE CLEAR from the previous section that noise is very difficult and expensive to mitigate. Although error rates are getting lower and lower, the rate of physical error reduction will likely not outpace the effect of larger system sizes. At the same time, error correction requires a very large overhead in terms of the number of additional physical qubits. This chapter answers the question from a slightly different angle - assuming a perfect quantum computer exists, what happens if one were to begin to add physical qubits to it? Every additional physical qubit doubles the state-space that is accessible, with the ability to perform arbitrary quantum operations. This additional power and flexibility comes at the cost of introducing noise, whose effect could or could not be corrected by the QEC code.

As this work was being published it turned out that there was an emergent area of quantum computing research called partial error correction, where many approaches were being taken to try and gain some benefits from QEC without paying all the costs associated with it. AN influential approach to this are in so called partially error corrected quantum computing [284]. Outside an error corrected or mitigation context, there is an algorithmic use to partitioning a system like this. One could, for example, attach a quantum sensor to a quantum processor directly and process the incoming state, or make a more controlled open dynamics simulation where the noisy qubits act as a structured bath and the error corrected qubits are the system of interest.

One of the most important functions of quantum computing is that of estimating expectation values of an operator following the

Contributions: the paper, published under the name *the battle of clean and dirty qubits in the era of partial error correction* [111] was a joint effort split into two self contained sections, I worked on the numerics and ideation and scope of the model (with input from discussions of course) and the analytical study which has been summarized was worked on mostly by Samson Wang, it is present to bring a richer understanding to the other results. Piotr helped immensly with the numerics and code, while Lukasz and Patrick took an active suprvisory and discussionary role, and proposed the seed of the idea which was 'can one still use physical qubits in a fault tolerant era'?

preparation of a desirable state by some algorithm, be that Shor's factoring algorithm [45] in the distant future, or for more near term VQAs [57, 285]. As shown in chapter 4.2, noise damages the accessible state - in particular, the expectation values concentrate around a single value [286, 287, 262, 60]. This *concentration of expectation values* within VQA's is most felt when computing gradients for the optimizer, since this is a process that must be repeated many times throughout the runtime of these algorithms. Thus, the algorithms become *untrainable* to borrow the language of machine learning. [1]. This concentration can be mitigated with (exponentially) more resources dedicated to each measurement [110, 60, 265, 288, 289, 290, 291], but of course, this is unfeasible. This scaling has come to be known as the Noise-Induced Barren Plateau (NIBP) [60], which is distinct from general barren plateaus, which are an undesirable feature of some cost functions or of the ansatze [288]. It must be noted that QEM alone cannot solve the noise-induced barren plateau (NIBP) problem [182]. In quantum computing specifically, general BPs are studied thoroughly, and many subtypes exist [265, 288, 289, 290, 291, 292, 293, 294].

The academic aim of the work was mainly aimed at solving the great barrier to practical quantum computing that is noise. The research question was *how can we mitigate, or remove, the effect of NIBPs and exponential concentration due to noise?* [111], but in this thesis it will be presented in the tone which it was researched, which is *what happens when one starts splitting some of the logical qubits into many noisy qubits?*.

To answer these questions the *Clean and Dirty* model was proposed, after a failed attempt at calling it the Naughty and Nice model due to societal reasons. Irrespective of the name, *clean* qubits are those that are error corrected and *dirty* qubits are those that are not. In the best case, which is what is studied here, one would have the dirty qubits have some noise channels acting on them, while the clean qubits are perfectly noiseless[2]. In a more general case, clean qubits could be logical qubits encoded with high distance codes and dirty qubits could be encoded on low distance codes. Alternatively, and what has been done as a follow-up work [104], a physical qubit connected to a system with a finite distance code.

Of course, the intersection of NISQ and error correction is not new in of itself, and has been considered in contemporary works [295, 296, 297]. Indeed, the name itself is not completely original, but it is different to previous *clean qubit* models [298, 299, 300] that define clean qubits as those that are in a perfect known initial state, with the rest being in maximally mixed states. Although it is not obvious, these systems do have computational advantage [301] in areas such as factoring [302] and computing partition functions [303].

This setup represents a realizable and, at the time, new, model of quantum computation. Furthermore, with the numerical and theoretical technology at the time it was possible to come to concrete statements about the scaling of the concentration of expectation val-
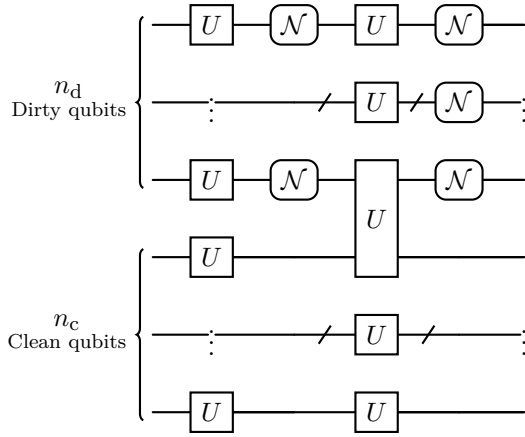
[1] Although now that the field of machine learning has received a Nobel Prize in physics and chemistry, technically it's not borrowing anymore.

[2] A completely noiseless qubit is like an infinite distance QEC.

ues for a cost function. This is because there is an analytical derivative to a parameterized quantum circuit (within some constraints) known as the parameter-shift rule [304] with respect to a cost function that can be both used to numerically compute and theoretically calculate the expected effect. How this is done will be the focus of this chapter.

8.1 Computing with Clean and Dirty qubits

The idealized clean and dirty (CnD) quantum computer consists of n qubits which are divided into n_d noisy dirty qubits and n_c noiseless clean qubits such that $n_d + n_c = n$.

Figure 8.1: **The clean and dirty quantum circuit** consists of n_c noiseless, clean qubits and n_d noisy, dirty qubits. A clean qubit gate acts by evolving the system via a perfect unitary U, while gates on dirty qubits is represented by the same U followed by the noise channel $\mathcal{N}$. When clean and dirty qubits interact together, the noise channel only acts on the dirty qubits. Reproduced from [111] under CC4.

A gate that is applied to the qubits with state ρ performs a unitary operation U, with a noise channel $\mathcal{N}$ occurring when only on the subset of qubits that are dirty as

$$\rho \underset{\text{clean}}{\to} U\rho U^\dagger, \tag{8.1}$$

$$\rho \underset{\text{dirty}}{\to} \mathcal{N}(U\rho U^\dagger). \tag{8.2}$$

$$\tag{8.3}$$

The interaction of qubits is more subtle, as the noise could act on both qubits, neither or only the dirty qubits. Since it is important to show the best case scenario in this theoretical study, the assumption that the noise exclusively acts on the dirty qubit was made. As such, figure 8.1 shows how noise is applied throughout the system. As mentioned, the setup is representative of a quantum computer comprised of high quality logical qubits and low quality uncorrected or logical qubits. At the limits of this model the standard mode of operation is recovered; ff $n_c = 0$ then the system behaves as a noisy quantum computer if the error rates are not negligible, while if $n_d = 0$ the machine acts as a fault-tolerant device or an ideal quantum computer if the noise is 0.

In the study that will be carried out many assumptions and practicalities are ignored to make an analysis of the system possible. In error correction codes, logical qubits are not perfect due to a myriad

of issues that were described in section 4.3. By making the simplifying assumption that the clean qubits are completely noiseless, the fact that implementing non-Clifford gates requires techniques such as gate teleportation [305] and magic state preparation [306], which will lead to higher costs for using such gates in a circuit that are not considered. In the worst case, the price to pay is a higher error rate for such gates, while in the best case it is a time cost, which would lead to more errors occurring in the dirty qubits. As such, this model can be used to study novel algorithms or the effect of very different error rates in a given system.

Since the time of the article's publication experimental error correction has been performed experimentally to a much larger extent than before. At the time, minimal examples of error correction were possible [307], but since then larger experiments with more complex computations have been performed [143, 144, 145, 308]. For example, [307] used partial error correction in the sense that the code implemented was a low distance one which only detected and corrected a subset of errors. Error rates of QEC have now been proven (which was only expected before) to decay exponentially with distance [144], but since the rate of improvement is related to the ratio of the physical error rate to the threshold error rate, it is still the case that many physical qubits are required for a single logical qubit [309].

There are some practical reasons to consider the CnD setup from a practical lens even with error correction on the horizon. To obtain quantum advantage, it is still necessary to have many effective qubits. It is also the case that gate times are much slower on error corrected devices, both due to the encoding overhead and the physical time it takes to implement a gate on a given architecture, which is dictated by the physics of the device [310]. As such, using the CnD model can both increase the computational space required for advantage, and potentially offloading some of the overheads in implementing non-Clifford gates to the dirty qubits instead of the clean qubits, thus a clever partition could be a good solution when these limitations are relevant.

Even given large fault-tolerant devices, the model will be valid in the case of high and low distance logical qubits. Such a subdivision could be useful to optimally use the qubits in a computer given a particular algorithm. For example, algorithms where many gates act on a subset of qubits may benefit from better correction, while temporary ancilla qubits that are reset could use no or little error correction. Such algorithms already exist, like QPE, where a multi-qubit unitary is controlled by a single ancilla [311].

8.2 Analytical study

The analytical process to deriving both the concentration of expectation values and the related gradient scaling focuses on the construction of general circuits on the depolarizing noise model, which is analytically tractable. There are three levels of model that build

on each other to show that the effect is general. Beginning with the simplest; a CNOT ladder setup. The CNOT ladder circuit is a minimal example where an arbitrary perfect initial state is prepared on all qubits appended by repeated sequential ladders of CNOT gates, with noise applied on the dirty qubits after each CNOT. This is expanded upon by preceding another CNOT construction by a perfect global unitary which could entangle the entire system and two other operations acting only on the CnD subsystems individually. To this, the ladder circuit is appended before measurement. Finally, the most general case is considered, where the previous set of gates is cast as a parameterized operation sandwiched between the CNOT ladders, which yields a direct comparison to the numerical study.

THE LADDER CONSTRUCTION shown in figure 8.2 consists of sequential CNOT gates which can be easily extended to other two qubit gates, and are used in many VQA and quantum machine learning (QML) circuits [312, 313]. Layers of linearly connected ladders are interleaved by local depolarizing error channels (4.7) acting on the first n_d qubits.

The circuit is characterized by having a probability p of the state becoming depolarized at each noise channel $\mathcal{D}_p$, which for the entire circuit of L layers forms a channel $\mathcal{W}_L^{(n_d)}$. The expectation value for the arbitrary pure input state $\rho = |\psi\rangle\langle\psi|$ for a measurement on the computational basis O is given by

$$\mathrm{Tr}[\mathcal{W}_L^{(n_d)}\rho O] \leqslant (1-p)^{\Lambda(n,n_d,L)} + \frac{1}{2^n} \tag{8.4}$$

$$\Lambda(n,n_d,L) = \begin{cases} \dfrac{Ln_d}{2^{\text{floor}(\log_2 n)}}, & \text{if } n_d = 1; \\[3mm] \dfrac{Ln_d}{2^{\text{ceil}(\log_2 n)}}, & \text{if } n_d > 1. \end{cases} \tag{8.5}$$

The proof for this can be found in [111]. The result holds for any arbitrary input state, including an unentangled product state. Thus, the net effect of the bound in (8.4) is that the expectation value which would have a maximum magnitude of 1 would be scaled by the error as $1 - p$ but potentially lifted by Γ. If the probability of error is small $p \ll 1$ then it is possible the use a Taylor expansion, and making the simplifying assumption that $n = 2^m$ for some integer m, then the scaling can be thought of as some effective number of layers $L_{\text{eff}} = \frac{Ln_d}{n}$ and to first order in p the right-hand side of (8.4), ignoring the $\frac{1}{2^n}$ system scaling term, becomes

$$(1-p)^{L_{\text{eff}}} \approx 1 - \frac{Ln_d}{n}p \approx (1 - p\frac{n_d}{n})^L. \tag{8.6}$$

This simplification is very helpful, in the rightmost expansion it tells that the system is roughly equivalent to a fully noisy system with a noise probability scaled by the number of dirty qubits to total qubits. From one perspective, the error is reduced uniformly for each additional clean qubit, from another perspective, the CNOT ladder layers, if deep enough, propagate the noise throughout the
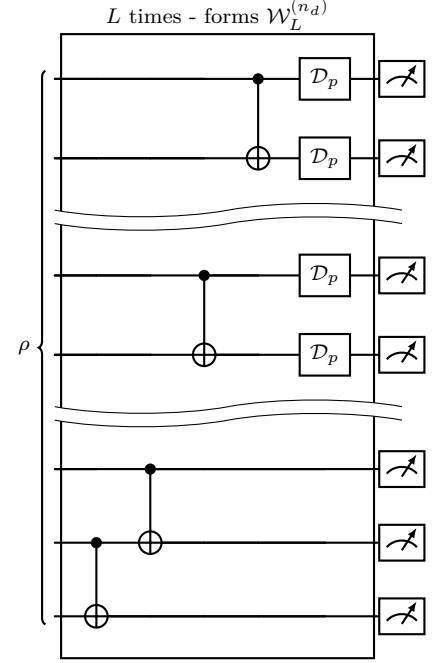

Figure 8.2: **CNOT ladder construction.** The linearly connected CNOT ladder followed by a local depolarizing noise channel $\mathcal{D}_p$ on the n_d dirty qubits, which is repeated for L repetitions, called layers followed by a projective measurement on the computational basis.

system thus leading to the tongue-in-cheek title of the chapter *one bad qubit ruins the bunch.* The numerical results preceded these findings, but it was great to see that the analytics were in line with the numerics.

THE SEPARATED OPERATOR SETTING shows that local noise in one subsystem that is entangled with the global state impacts the whole system. The construction is shown in figure 8.3 and consists of an input state ρ followed by a global perfect unitary operation W, then two arbitrary operations $\mathcal{U}, \mathcal{V}$ which define the size of the two subsystems (A, B) are applied. A subset of these are then connected via CNOTs with a Hadamard gate after the control. W is completely free perfect Unitary and as such could be even be a tensor product acting on the two subsystems separately as $W_A \otimes W_B$. The operations $\mathcal{U}, \mathcal{V}$ form some channel $\mathcal{T}$, this can be completely noiseless and unitary, or it could be noisy on one or both of the subsystems, denoting such a change as $\mathcal{U} \rightarrow \tilde{\mathcal{U}}$, the local depolarizing noise is then applied to a layered implementation as in figure 8.4, this type of decomposition is always possible, but will have a characteristic length $L_{\mathcal{U}}$ for an optimal decomposition.

For the same assumptions as the ladder construction (8.4), with the addition that there is an additional entanglement operator $\mathcal{E}_m$ representing the CNOT and Hadamard gates on $m \geqslant 1$ pairs of qubits then the expectation values concentrate as

$$\mathrm{Tr}[\mathcal{E}_m \circ \mathcal{T} \circ \mathcal{W} \rho O] \leqslant (1-p)^{L_{\mathcal{T}}} + \frac{1}{2^n}, \tag{8.7}$$

$$L_{\mathcal{T}} = \begin{cases} L_{\mathcal{U}}, & \text{if } \mathcal{T} = \tilde{\mathcal{U}} \otimes \mathcal{V}; \\ L_{\mathcal{U}} + L_{\mathcal{V}}, & \text{if } \mathcal{T} = \tilde{\mathcal{U}} \otimes \tilde{\mathcal{V}}. \end{cases} \tag{8.8}$$

This result is surprising there is no dependence on m, which means even a single CNOT after the noisy operation is enough to concentrate the expectation values. Intuitively this is not clear, so an example is in order. If n_d is small, then the operator on the noisy subsystem $\tilde{\mathcal{U}}$ will have a small $L_{\mathcal{U}}$. If the systems are of similar size, then the low quality measurements on the dirty subsystem A will be enough to deteriorate the whole system's expectation values. Nonetheless, compared to the fully noisy system which scales as $L_{\mathcal{U}} + L_{\mathcal{V}}$, the scaling compared to this is exponentially suppressed
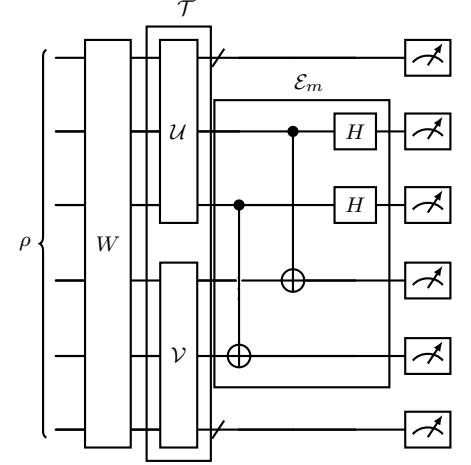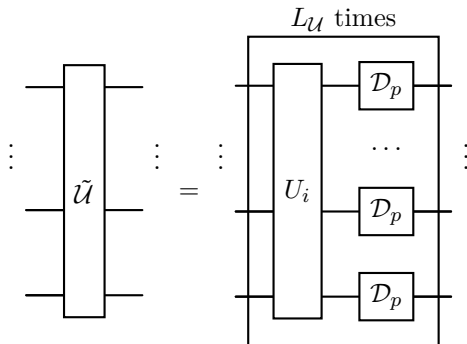

Figure 8.3: **The separated operator construction** showing a circuit where a general input state ρ is acted upon by a perfect global operator W followed by a channel that acts on either n_d or n_c qubits, which are defined by whether $\mathcal{U}$ or $\mathcal{V}$ is noisy. This is then followed by a CNOT ladder on a subset of the two subsystems with H gates after the controls. These are then measured in the computational basis.

The choice of $\mathcal{V}$ or $\mathcal{U}$ as noisy is arbitrary, and the same result holds with a change of index.


Figure 8.4: U **is made noisy** by first decomposing U into a number of non-parallelizabale layers formed of factors U_i, $i = 1, \ldots, L$ which are then interleaved by local depolarizing channels $\mathcal{D}_p$, defining the depth of U to be L. U is recovered when $p = 0$.

since the exponent in (8.7) is $\frac{L_{\mathcal{U}}}{L_{\mathcal{U}}+L_{\mathcal{V}}}$. This can be further strengthened if one assumes a proportional relationship between the number of qubits in the two subsystems and the layers L. For a linear relationship this is

$$L_{\mathcal{U}} \propto n_d, \quad L_{\mathcal{V}} \propto n_c \implies \frac{L_{\mathcal{U}}}{L_{\mathcal{U}}+L_{\mathcal{V}}} \propto \frac{n_d}{n_d+n_c} = \frac{n_d}{n}, \qquad (8.9)$$

which is similar to the findings in the ladder construction.

THE PARAMETERIZED CIRCUIT SETTING is the final setting which brings the results of the previous two cases together and is the most directly comparable to the numerical results. Both the previous results have the undesirable feature that with any noise there are exponentially vanishing expectation values. If one were to use parameterized unitaries this would still hold, thus barren plateaus exist for circuits of linear depth if n_d scales linearly with n. It also follows that any higher polynomial depth circuits will have the same if $n_d > 0$.
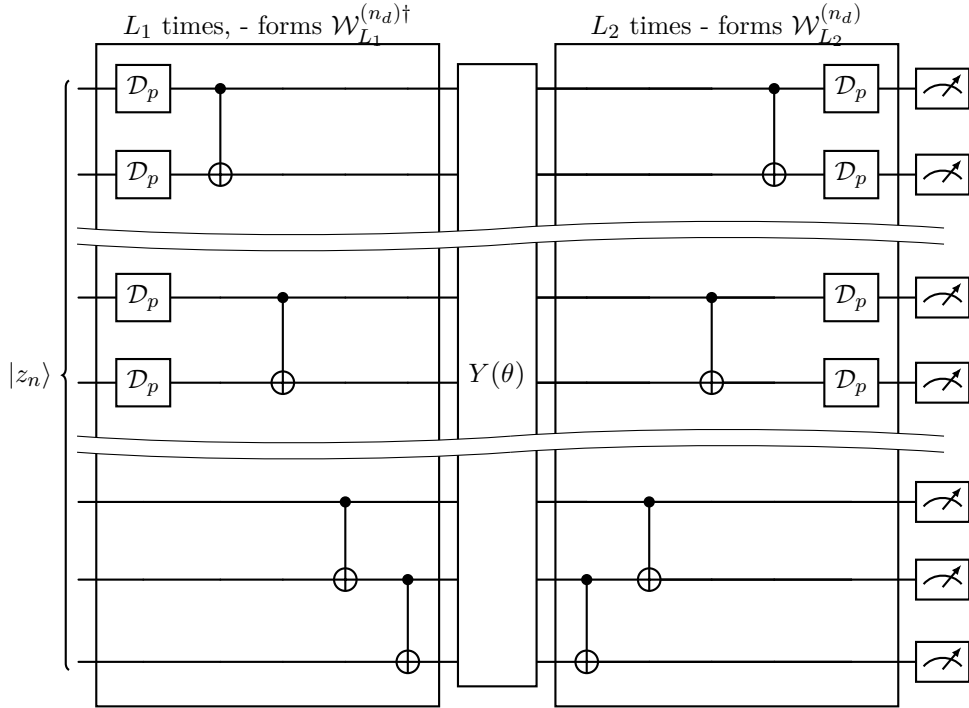
A trainable unitary is given as

$$Y(\boldsymbol{\theta}) = W_0 \prod_{k=1}^{K} e^{-i\theta_k H_k} W_k, \qquad (8.10)$$

where $\{W_k\}_{k=0}^{K}$ are some fixed unitary operators, and $\{H_k\}_{k=0}^{K}$ are Hermitian operators. This forms a channel $\mathcal{Y}(\boldsymbol{\theta})$ which acts on the input state ρ, which now will be a computational basis state. This is now inserted into the two previous constructions. Figure 8.5 shows this for the CNOT construction, but an equivalent 'sandwiching' can be done for the separated operator construction.

Figure 8.5: **Parameterized ladder construction.** An analytical circuit consisting of the adjoint CNOT ladder followed by a parameterized unitary $Y(\theta)$ and finished by the CNOT ladder and measurement.

Here, instead of measuring an observable in the computational basis, some cost function $C(\boldsymbol{\theta}) = \text{Tr}[\mathcal{W}_{L_2}^{(n_d)} \circ \mathcal{Y}(\boldsymbol{\theta}) \circ \mathcal{W}_{L_1}^{(n_d)\dagger} \rho O]$ that comprises multiple measurements in the computational basis O. Here, the observables of interest are partial derivatives of C, as these are what allow the optimization of parameters in all VQAs. Thus the expectation value for a given parameter is

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \text{Tr}[\mathcal{W}_{L_2}^{(n_d)} \circ \frac{\partial \mathcal{Y}(\boldsymbol{\theta})}{\partial \theta_k} \circ \mathcal{W}_{L_1}^{(n_d)\dagger} \rho O]. \qquad (8.11)$$

The channels $\mathcal{W}$ can be replaced by $\mathcal{T} \circ \mathcal{E}_m$ to give the expression for the other setting. Irrespective of the setting, this replacement yields the following bounds on the absolute value of the gradient scaling

$$\left| \frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} \right| \leqslant (1-p)^\Gamma \, ||H_k||_\infty \qquad (8.12)$$

$$\Gamma = \begin{cases} \dfrac{n_d}{2^{\text{ceil}(\log_2 n)}}(L_1 + L_2), & \text{if CNOT ladder and } n_d > 1; \\ 2L_\mathcal{U}, & \text{if separable construction with } \mathcal{T} = \tilde{\mathcal{U}} \otimes \mathcal{V}. \end{cases} \qquad (8.13)$$

Again, there is an exponential suppression of the magnitude of the measured gradients, which leads to the NIBP effect. This is expected from the more trivial intuitive example based on the previous construction where each individual unitary can be thought of as a circuit giving the partial derivative of the cost function. The big difference is with the fact that the infinitum of the Hermitian operators $||H_k||_\infty$ multiplies the previous expressions, which means that if the largest singular values of $\{H_k\}_{k=1}^K$ grow faster than the scaling of the concentration with system size, it will outpace the suppression. But since the rate of concentration is exponential if n_d scales linearly with n, this is unlikely to be the case.

8.3 Numerical Investigation

The numerical aspect of this work follows the same principle as the parameterized circuit setting, but for a realistic VQA type circuit. Specifically, the behavior of a cost function of a Hamiltonian variational ansatz (HVA) [56, 266] is used, beginning with a presentation of the problem and numerical computation of the gradients. Then the results of the simulations under the same depolarizing noise model as the analytical investigation and a realistic noise model are presented.

8.3.1 Hamiltonian Variational Ansatz

The HVA is a specific form of ansatz to a VQE problem, which in this case is to find the ground state of a periodic transverse field quantum Ising model in one dimension. The Hamiltonian of the model used for n spin sites is

$$H = -\sum_i^n X_i X_{i+1} - \sum_{i=1}^n Z_i, \qquad (8.14)$$
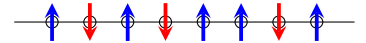

Figure 8.6: **Spin chain** with 8 spin sites in an arbitrary state, depending on the interactions between spins and boundary conditions many different 1D models can be studied.
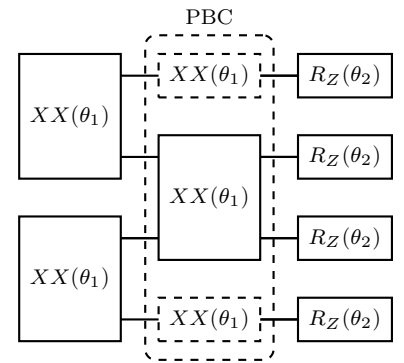

Figure 8.7: **The HVA ansatz** showing a single layer of (8.15) on 4 qubits with periodic boundary conditions (dotted gate). The ansatz is decomposed to native gates of a trapped-ion quantum computer.

where X_i, Z_i are the Pauli matrices acting on the ith site. The reason why such a simple system is chosen is because it is known that the ansatz with the cost function has no barren plateau [314], thus any concentration of expectation values is due to NIBP. For completeness, the ansatz used with L layers is parameterized by the vector $\boldsymbol{\theta}$ with $2L$ entries, which form a state

$$|\psi(\boldsymbol{\theta})\rangle = \left(\prod_{i=1}^{L} e^{-i\theta_{2i} H_Z} e^{-i\theta_{2i-1} H_{XX}} \right) |0\rangle^{\otimes n}, \qquad (8.15)$$

Each H_{XX}, H_Z are the terms in the Hamiltonian, which in the trapped ion gateset have direct translations to Molmer-Sørensøn gates (XX) and single qubit rotations about the Z axis (R_Z). A layer of the ansatz for 4 qubits is shown in figure 8.7. The variational principle is then used to find the ground state energy by minimizing the energy as

$$C(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle. \qquad (8.16)$$

This has a gradient

$$\nabla C(\boldsymbol{\theta}) = (\partial_{\theta_1} C(\boldsymbol{\theta}), \partial_{\theta_2} C(\boldsymbol{\theta}), \dots, \partial_{\theta_{2L}} C(\boldsymbol{\theta})), \qquad (8.17)$$

which is numerically computed through a parameter shift rule for Pauli strings [304][3].

[3] See section [88] for a description of the parameter shift rule. In short the parameter shift rule allows for analytical gradients for certain classes of circuits

8.3.2   Numerical Clean and Dirty Model

To make the results comparable to the analytical results, particularly those in (8.12), the mean of the absolute values of the cost function's partial derivatives for all *theta*s are calculated for systems of 4, 6 and 8 qubits. The ansatze tested is comprised from 1 to 600 layers for $n_d = 0, 1, \dots, n$. This is done both with a local depolarizing noise model and a scaled trapped ion noise model [268, 269] described in section 7.3.3. The implementation of the noise channels on the HVA ansatz is shown in figure 8.8.

Figure 8.8: **Numerical noise placement** for both the depolarizing case (inner channels labelled $\mathcal{D}_p$) and the trapped ion case (outer gates labelled $\mathcal{C}$ and $\mathcal{S}_Z$) for one HVA ansatz layer on 4 qubits. The dashed gates represent a single XX gate which can act like this due to the periodic boundary conditions.
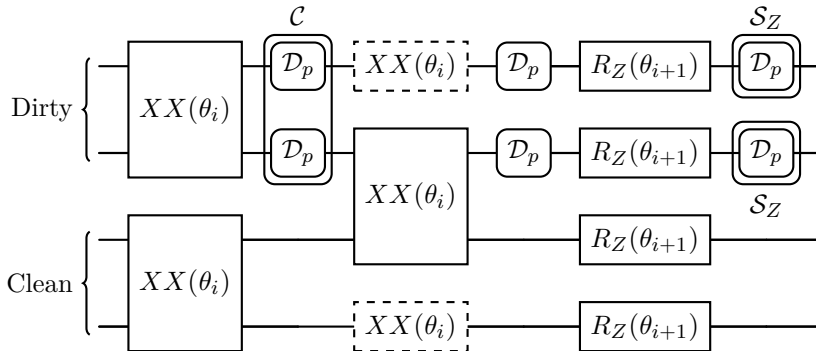
The way the noise channels are placed on the dirty qubits demands some explanation. Referring to figure 8.8 it can be seen that the depolarizing case follows the analytics and has a depolarizing channel $\mathcal{D}_p$ placed after each operation on a dirty qubit. The trapped ion case includes two channels - the correlated 2 qubit error

$\mathcal{C}$, and the single qubit error channel $\mathcal{S}_Z$ is placed after each R_Z gate on a dirty qubit. The subtlety comes in when to add $\mathcal{C}$, since the character of a two qubit gate on hardware is to have more complex error channels the choice was to use an arbitrary single qubit channel on the dirty qubit, have the correlated error occur on the clean qubit as well or to skip the noise completely on a clean-dirty connection. The latter was chosen with the idea that the results would represent the best case scenario - or some upper bound to performance. To further minimize the amount of connections, the dirty qubits are set up in a continuous block so that there are minimal connections between the CnD subsystems as in figure 8.1.

To show that it is indeed the case that the overall effect of the dirty qubits is the same as having an overall noisy machine with a proportionally lower error rate per qubit on all qubits as in (8.6). This is done for both the depolarizing and realistic noise model, by simply altering the probabilities by some factor f to replace the dirty qubit model with the reduced fully noisy realization as

$$p \to p f_{n_d}, \quad \text{with } f_{n_d} = n_d/n. \tag{8.18}$$

For the depolarizing noise, this is merely the probability of the depolarizing channel, and for the trapped ion, all probabilities in section 7.3.3 are scaled by f_{n_d}.

8.3.3 Results

The results focus on the most important marker of NIBP; the scaling of the mean of the absolute value of the cost function's partial derivative for each parameter

$$\overline{|\partial_{\theta_k} C|} = \frac{1}{2L} \sum_{i=1}^{2L} |\partial_{\theta_i} C|, \tag{8.19}$$

for the cost function C in (8.16). The numerical simulations are discussed for the 8 qubit case, but the trends are identical for the other tested system sizes.

Figure 8.9 shows the outcome for both the depolarizing and realistic cases. The relationship is very clear, the high depth of the circuits tested clearly show the expected characteristic exponential decay of $\overline{|\partial_{\theta_k} C|}$, which is a straight line on a log linear scale. This is the case for any number of dirty qubits, i.e. $n_d \geqslant 1$, with the average gradient decreasing exponentially as more dirty qubits are added. As such the theoretical results predicted this outcome correctly, and they extend to the realistic noise model as well.

The situation becomes more interesting when considering the first order approximation (8.6), which says that the system is roughly equivalent to the fully noisy machine with rescaled error rates as (8.18). In the depolarizing noise model (figure 8.9(a)) this is a very good approximation, although the fully noisy machine performs a little bit worse than the equivalent CnD model. The realistic noise model (figure 8.9(b)) shows a similar trend, but the fully noisy machine appears to perform significantly worse (as though there was
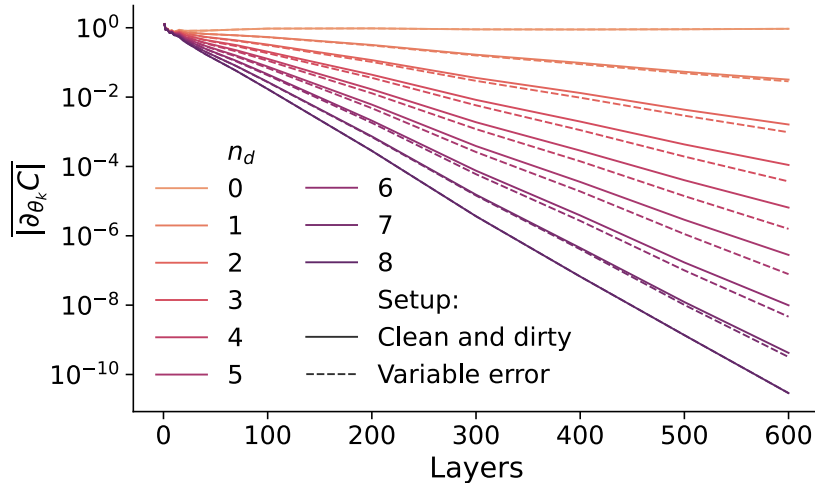
(a) Depolarizing noise model - gradient versus number of layers
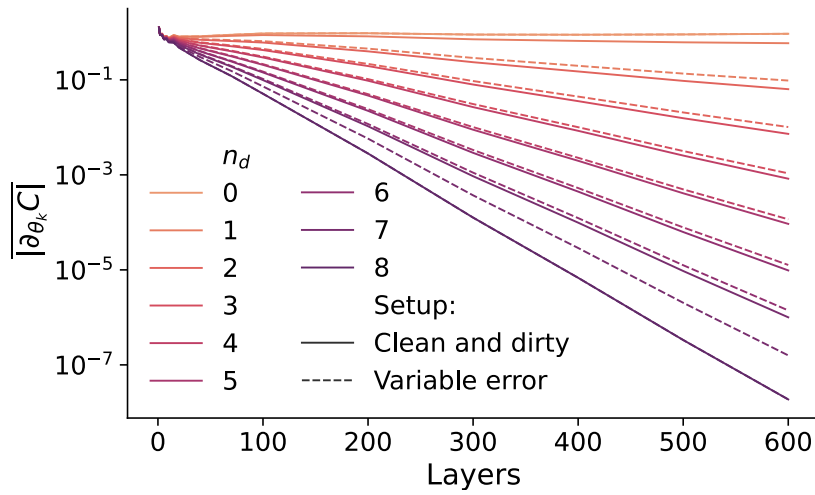
(b) Realistic noise model - gradient versus number of layers

Figure 8.9: **8 qubit HVA gradient versus HVA layers.** (a) shows results for local depolarizing noise while (b) shows results for the realistic model. Solid lines show the clean and dirty model for $n_d = 0, 1, \ldots, 8$, dashed lines show a fully noisy machine with error rates scaled by a factor $f = n_d/8$. The rescaled result are always below the dirty qubit results, this is because the coloring scheme makes it a little difficult to distinguish the two. Reproduced from [103] under CC4.

an additional dirty qubit). The results above may not appear to be
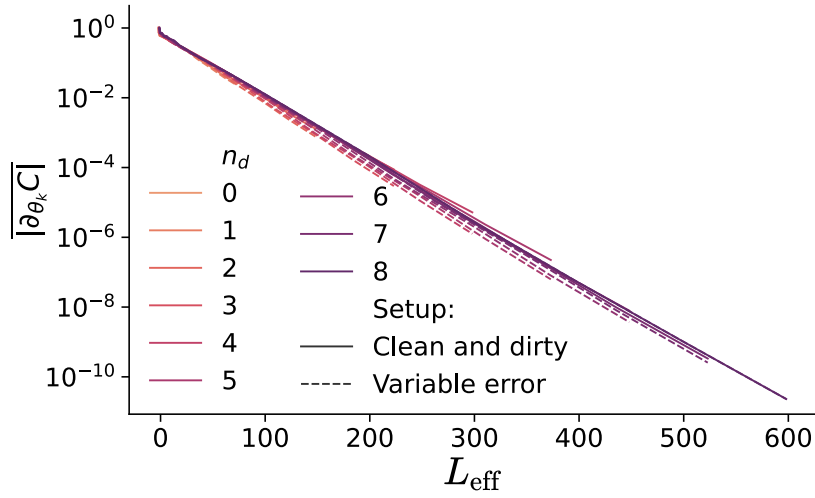
(a) Depolarizing noise model

(b) Realistic noise model

Figure 8.10: **HVA gradient versus total error rate.** Both the depolarizing noise model (a) and the realistic noise (b) results show an exponential decay of the mean absolute value of the partial derivative with the total error rate up to small deviations. The inset in (b) shows the error rate for the two setups and the realistic noise model in relation to n_d demonstrating that it is approximately linear. Figure (a) is a new figure from the data in [111] and (b) is reproduced from [111] under CC4.

good evidence that the fully noisy approximation is correct, so it is good to attempt to rescale the results and observe if there is a form in which the qualitative differences between the two models can be made more quantitative. This was done through a lot of trial and error, but it turned out that plotting $\overline{|\partial_{\theta_k} C|}$ against the total error rate [4]. In the depolarizing noise case, it is easier to find an analytic expression for the transformation, which is to define an effective layer $L_{\text{eff}} = L\frac{n_d}{n}$, but for the realistic noise model this is just a sum of the existing error rates. Under this transformation (figure ref-fig:numericalResultScaled) the results apprriximately 'collapse' onto a single line, which indicates that the truly important factor is the total error rate. For the realistic noise model specifically, there is an inset in figure 8.10(b) which shows the total error for the fully noisy and CnD model, which explains the differences in figure 8.9(b) a little better. In the depolarizing noise model the approximation is also qualitatively good, with a slighly worse performance for the fully noisy machine.

One important consideration in all quantum computing research is the scaling with respect to system size, figures 8.11(a-d) show the

[4] Defined as the sum of all error rates of all gates in the circuit

same plots for the other tested systems with 4 and 6 qubits. The re-
sults show the same trends with smaller deviations of the scaled fully
noisy machine from the CnD model. This implies one of two things,
the first being that as the state becomes bigger the approximation
worsens or that the larger the system size, the gradients are more
sensitive to finite sample artifacts. This is likely as the deviations
(not shown for clarity) are within a standard deviation of the mean
gradient.

Most circuits on so few qubits are generally not 600 layers deep,
so it is interesting to look at what the results would be in a practi-
cal setting. That is to say, is this a scaling that is only relevant in
the deep circuit regime or is it something that would occur at any
depth? At large circuit depths, simulations for many noisy qubits are
expensive, but at lower depths it is feasible to examine the effect on
more qubits. As such, a system of 10 qubits is shown in figure 8.12.
This exhibits similar scaling to previous examples, but with spikier
gradients due to the stronger dependence on the specific problem
instance. Nonetheless, the different values of n_d do not cross each
other.

Finally, the numerical results can be summarized as showing that
indeed, the theoretical prediction is quite accurate in the first order.
To fully understand the origin of the differences between the fully
noisy and CnD model there would need to be further investigation,
especially for the depolarizing case. In the realistic case the collapse
is almost perfect. The scaling of the total error is directly propor-
tional to n_d/n for the depolarizing case, while it is approximately
linear in the case of the realistic noise model (as shown in the in-
set of figure 8.10(b)). The collapses clearly show the exponential
suppression of the gradient by the predicted ratio of n_d/n.

8.4   Discussion

The battle of CnD qubits has shown that a single dirty qubit will
ruin the bunch. By splitting a quantum computer into error cor-
rected *clean* qubits and physical *dirty* qubits one essentially creates
a noisy quantum computer with an error rate lowered by $\frac{n_d}{n}$. The
consequence of this result is that this model exhibits the same issues
that all noisy machines exhibit, namingly exponential scaling symp-
toms like barren plateaus. Even QEM cannot truly mitigate such ef-
fects [182, 141], and this work added to the growing body of evidence
that a noiseless system might be the only way to avoid these issues.
The results are extensible to more general statements about the
concentration of arbitrary expectation values rather than just gra-
dients, as the results in both domains are proven to be transferrable
[289]. Nonetheless, an exponential suppression of the effective error
rate does mean that at finite depths and a favourable $\frac{n_d}{n}$, the error
rate may be so low as to be inconsequential for many algorithms,
although this error rate will need to be on the order of 10^{-15} [149]
for the most demanding algorithms. In this work the error rates used
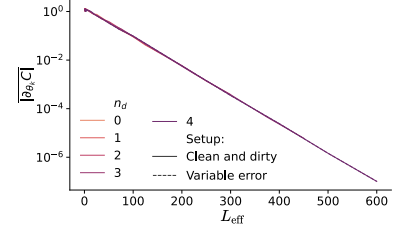
(a) Depolarizing noise model - 4
Qubits
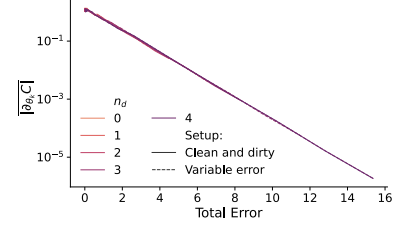
(b) Realistic noise model - 4 Qubits
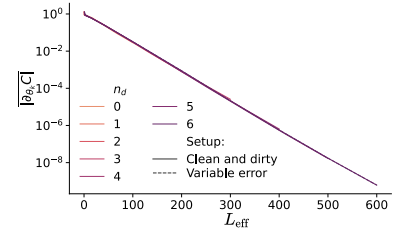
(c) Depolarizing noise model - 6
Qubits
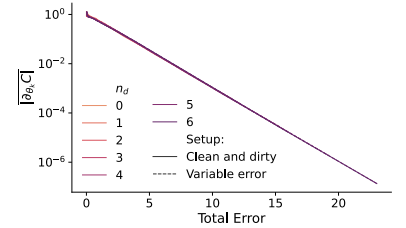
(d) Realistic noise model - 6 Qubits

Figure 8.11: **HVA** $\overline{|\partial_{\theta_k} C|}$
against the total error rate
for the depolarizing noise model
(a,c) and the realistic noise
model (b,d) for 4 and 6 qubits.
The results show the same
trends as for 8 qubits, but the
fully noisy machine is more in-
distinguishable to the clean and
dirty model. Reproduced from
[103] under CC4.
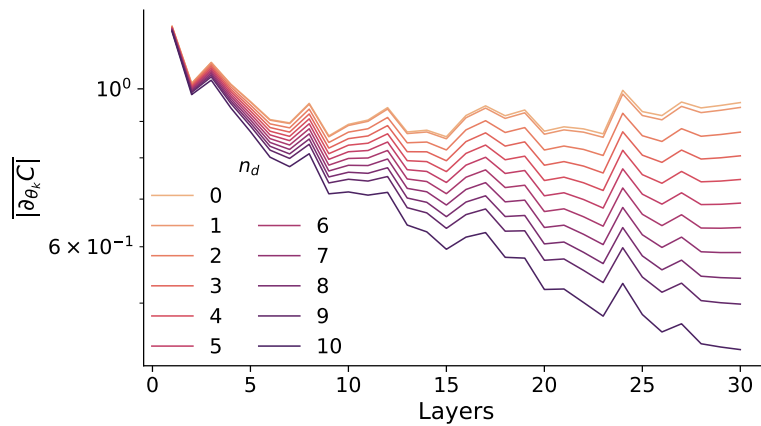
Figure 8.12: **HVA** $\overline{|\partial_{\theta_k} C|}$ **for 10 qubits at low layer depths.** The clean and dirty model with realistic noise is plotted at depths of up to 30 layers, with results given for each layer count. The same pattern as the high depth is also seen here, with high n_d systems' gradients decaying faster.

were in line with error rates of the time, but to implement such a scheme it is necessary that the error rates will be below the threshold required for fault tolerance, which would further favour such a scheme.

Whilst the numerical analysis is within the realm of VQAs, the analytical results are likely a general phenomenon in partially noisy circuits. However, the setup overlooks the fact that one of the biggest issues in quantum error correction is performing non-Clifford gates and that the gate-dependent error rate is more relevant than the qubit-dependent error rate. Furthermore, VQAs are not necessarily the ideal scenario for the CnD setup. Algorithms can be designed to specifically take full advantage of such a setup or scenarios where noise in a subsystem is unavoidable. For example, a quantum machine learning algorithm that is using noisy experimental data [315] from an attached qubit-based quantum sensor [316, 317]. Another class of candidates is algorithms where some qubits are acted upon much more so than others, such as in QPE [311]. Finally, algorithms could be custom-built to suit such a setup, here the scope can be in studying open system dynamics to as yet unexplored territory.

8.5 Concluding remarks

This work was based on very simplifying assumptions that may either aid or hinder the real world performance of such a scheme. Actual logical qubits are not noiseless, but are characterized by logical error rates. The error correction codes might also be able to act as *entropy sinks* that could actually reduce the effect of the noisy qubits on the whole machine, or not. The temptation to do this exact study was too great, and indeed, has lead to a direct follow-up work [104].

The first problem is that this approach was so novel that the idea of a gate between a physical or noisy qubit was not even explored. This question was solved by Nikolaos Kuokoulekidis [104]

who showed that all that is required is a series of CNOT gates between the physical qubits and the relevant qubits of the error correction code as shown in figure 8.13 for the Steane code. Other error correction codes have specific constructions, but it is always possible to create such a gate. This immediately shows additional issues for such a scheme - each additional dirty-clean CNOT could introduce errors which could destroy the encoding of the logical qubit, and the larger the code, the more CNOTS are necessary. Also, for a fully noisy machine with error correction, it is necessary to perform syndrome measurements and then apply corrections as necessary, something that cannot be done while the dirty-clean CNOT is applied.

The second problem is numerical, since to simulate this setup with a fully noisy machine is completely out of the realm of numerical feasibility with current hardware with a classic density matrix approach. As such, for this follow-up work it was necessary to limit the simulation to Clifford circuits with noise models derived from IBM Ourense process matrices [269] which were. This is a significant limitation, which limited the analysis to using randomized benchmarking [318] and a novel mirrored random Clifford circuit approach to determine the process fidelity. This amazing numerical work was carried out by Piotr Czarnik and supported by Lukasz Cincio. The work itself is an excellent extension of the original work, and has shown one very striking difference between the theoretical model presented and the more realistic implementation. This is that now, instead of having a simple $\frac{n_d}{n}$ scaling, the fully noisy machine outperforms unfavourable $\frac{n_d}{n}$, specifically, there exists some pseudo-fidelity function $f(n_c, n)$ with the property

Figure 8.13: **Logical and dirty CNOT on the Steane code,** the dirty qubit q is connected via a series of CNOT gates to the relevant qubits of the steane code $s_{1,2,3}$, the qubits that are not acted upon are relevant for the logical encoding, but needn't be flipped to change state.

$$f(n_d, n) = \begin{cases} 1, & \text{if } n_c = n; \\ 0, & \text{if } n_c = 0; \\ < 0, & \text{if } n_c < n_{\text{threshhold}}; \\ > 0 & \text{if } n_c > n_{\text{threshhold}}. \end{cases} \quad (8.20)$$

That is to say that there is some threshold requirement $n_{\text{threshhold}}$ for the number of error corrected 'clean' qubits n_c before which the completely uncorrected machine outperforms the system with too many dirty qubits and too few clean qubits. Here 1 represents the highest possible pseudo-fidelity (since some error can still occur in the fully corrected noisy machine) and 0 is set as the uncorrected machine with n dirty qubits. In the case of the Steane code implemented in the paper [104], a single n_c consists of 7 noisy qubits, so a machine with $n_c = n$ has 7n physical qubits and one with $n_d = n$ has only n qubits, not counting some auxiliary qubits that are required for syndrome measurement. The results for this are shown in figure 8.14.

Although much can be said about this follow-up work, the field of partially error corrected research has gained traction. In the first place there are many ideas on splitting quantum circuits between many quantum processors, which can be roughly classed as *circuit*

(a) Randomized benchmarking fidelity

(b) $n_{\text{threshhold}}$ scaling for various error ratios.

Figure 8.14: **Numerics for the clean and dirty model with logical qubits.** (a) shows the fidelity decay for various numbers of clean qubits in a $n=30$ qubit systems, the thick blue line is the performance of the fully noisy machine which is better than a system with few clean qubits. (b) shows $n_{\text{threshhold}}$ for various error ratios between a constant CNOT error $\mathcal{E}_{\text{CNOT}}$ and varying idling error $\mathcal{E}_I$. The inset shows the change of this over the error ratio. Unpublished results based on reviewer comments for [104], under CC4.

division protocols [319]. This includes techniques such as in circuit knitting [320], lattice surgery [321], amongst many others. Some of these allow for the use of noisy systems connected to a logical system, and may have some practical advantages [319]. In the second place, there is now the exploration of using the dirty qubits to generate non-Clifford operations on logical qubits [322] with high fidelity. As yet no algorithms have been tailored to suit such an esoteric setup, the fact that such systems are being researched might lead to Clean and Dirty tailored algorithms coming into existence.

9
All Together Now

'Take these broken wings and learn to fly'
The Beatles, Blackbird, 1968

This chapter brings together the algorithm and error mitigation research alongside other considerations that need to be taken when performing actual quantum hardware experiments. It begins with a presentation of how one can go about getting time on a quantum computer, and goes on to describe the hardware limitations of the machine, followed by the alterations and choices made to the ansatz of TDVQP as well as further approximations made for the time evolution to make the algorithm work. A simplified noise model is used to run noisy simulations that are then mitigated.

THE RESEARCH PRESENTED in this thesis has had a strong focus on either analytical or numerical processes and algorithmic design. There is a chasm between these simulations and equations and running an algorithm on actual hardware. Although different platforms have their own limits and peculiarities, it is difficult to implement even conceptually simple algorithms and output a reasonable result. This work is in progress as a collaboration between the author and Malay Singh and Federico Roy of the Walter Meisner Institute (WMI) who want to implement the TDVQP algorithm introduced in chapter 6 on their experimental 4 qubit quantum computer.

This machine has much higher noise levels than commercial hardware, but it has one advantage, which is time. There is comparably unlimited access to every facet of the machine that is completely impossible to purchase from anywhere, and this allows for the implementation of every trick in the book to make an algorithm work on real hardware. As of the time of writing, there are no final results so most of the plots presented will be of simulations, but taking into consideration the very harsh limitations of the machine.

The first thing that one must do when attempting to run a computational experiment is to justify that the time invested would lead to some valuable scientific outcome. In the original paper [180] it was enough to present a near term friendly algorithm for dynamics on an interesting model. This is not quite enough for an actual

Contributions: this work is the result of a collaboration with Malay Singh and Federico Roy who are working on the experimental side, although no experimental results are present yet, their input and running of the simulation code from the TDVQP paper has been invaluable to finding the optimal parameters. Dominik Tonne also has helped run a lot of the simulations required to find the optimal VQE ansatz and reworked the original code to better function with the experimental setup.

experiment, so the first goal was to find an interesting physics or chemistry phenomena that could be simulated within reasonable constraints and in 4 qubits. Fortunately, the Shin-Metiu model[1] is often used as an initial model for many interesting phenomena, but simulation for most of these require the full quantum description and sometimes even additional interactions [323] or particles [324]. Any increase in dimensions would make the 4 qubit machine be unable to reasonably simulate it without major changes to the algorithm [325], nor would an increase in complexity in the Hamiltonian. As discussed in subsection 5.4.2 the ability to perform time evolution efficiently is harmed by an increase of Pauli terms. What was left if one couldn't actually simulate a more complex system than the one quantum electron and classical proton in a potential?

9.1 Proton Coupled Electron Transfer

Figure 9.1: **The Shin-Metiu model** with labelled components for comparison with the PCET process shown in figure 9.2.

Figure 9.2: **Donor-acceptor pathways** showing a simplified version of the potential reaction pathways. Assuming that the acceptor with the proton and electron is stable, there are no backward reactions from there, in the half filled case there can be a backward reaction to the donor with a proton and electron. Since PCET is a one-step reaction it can be faster if the system has a pathway for it to occur.

There is an important phenomenon in both chemical and biological mechanisms known as the Proton Coupled Electron Transfer (PCET) which is present in the Shin-Metiu model for certain parameterization [323]. It appears in several redox processes especially, where it describes the coupling of an electron to the nuclear motion. This is energetically favourable to a concerted proton transfer followed by a electron transfer (or vice versa). As mentioned, this mechanism is present in a wide variety of systems which involve some X-H bond, where X is most likely a C, N, O, S or metal [326]. The net effect in these reactions is that the reaction rate is faster than if the concerted (or some other) process was taking place. Like all things in chemistry there are many flavours of this phenomenon in different systems [327].

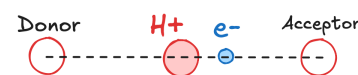The abstract picture of the reaction can be thought of as a donor on one side and a acceptor on the other side. In reality these can be as complicated as they might be, but in the Shin-Metiu model they would be assigned as the fixed right or left ions. To illustrate this a reminder of the model is given in figure 9.1, which is labelled so that figure 9.2 can be mapped onto locations of the electron and proton in the model. As such the goal of the quantum simulation is to somehow capture this transfer within the limitations imposed by formulating it in first quantization and within the number of iterations that are feasible to do within a reasonable timeframe.

9.2 Hardware Constraints and How to Work Around Them

Two qubit error rates in commercial hardware have fallen to the 0.1% range for superconducting devices [328], but for the experimental device at the WMI which is using a tunable coupling qubit which is not as mature as other technologies [329], but with a high potential. The machine available for this experiment has 4 qubits with a single gate effective error rate of 0.1% and a two qubit gate error rate of 1%. This limits the simulation depth quite drastically. The qubit topology is on the other hand quite friendly, being a periodic chain as shown in figure 9.3

Since the TDVQP algorithm at its deepest requirements 2 copies of the ansatz circuit and the time evolution circuit, this means that one must try to use as few layers as possible at each stage. For the VQE ansatz this is both simple and hard, since one can choose any circuit and test to see if it is appropriate, which is what was done. For the time evolution, this is more subtle, since if one were to use a traditional technique like Trotterization the depths are high since the Hamiltonian determines the complexity. Of course, due to the tridiagonal nature and commuting groups this is not a huge issue as it could be, but it still leads to circuits of depth 60 on 4 qubits. This is too much, and the whole algorithm should ideally fit to within depth 15 at most.

To do this it is of course possible to use alternative time evolution strategies as described in section 5.4.2. This will be important for larger experiments, but for small machines like this it is possible to do something even better. This is the area of approximate unitary implementations which do not scale fantastically with the size of the unitary, but for a 4 qubit device they are more than enough. Such systems can be used to simplify unitary sub-blocks in larger circuits as well - but here they can be used for the whole circuit. In particular, approximate quantum compiling with tensor networks was used, the so called AQCtensor algorithm [330]. The specifics of this are beyond scope, but the traditional method of approximating a unitary gate was to have a parameterized circuit and alter the parameters so that the effective unitary was as close as possible to the desired unitary. This is different from standard compilation where the operations are optimized, but there is a guarantee that the
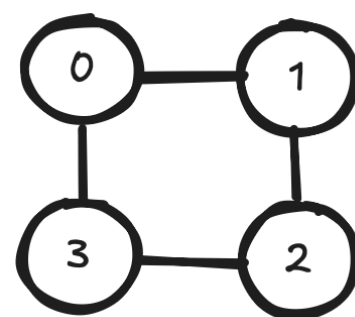

Figure 9.3: **Topology of the chip** to be used in this work. Qubits are the numbered circles and lines represent the connections on which 2 qubit gates can be performed.

output unitary is mathematically identical to the input unitary.

The integration of error mitigation techniques is essential for the functioning of the algorithm on hardware. Due to the higher error rates (although it likely would be important regardless) the use of error mitigation becomes essential. To this end, it is thought that it is a perfect test case for the UNITED technique presented in chapter 7. This is in addition to the generation of a readout mitigation routine which can be the full-inverted confusion matrix[2] due to the small number of qubits. This will be essential as otherwise it is very likely that the noise would wash out any reasonable results especially when going through the different timesteps.

[2] See section 4.2.1 for details

Underpinning the whole experiment is the parameters chosen for the model. Since it was decided that PCET would be the phenomenon that should be observed, this means that the transfer of the electron from the left side of the model (the donor) to the right side (the acceptor) would need to be observed. For this it is necessary to measure a PCET witness (W_{PCET}), which is simply the expectation value of the position ($\hat{x}$) of the electron wavefunction ($|\psi\rangle$) as

$$W_{\mathrm{PCET}} = \langle\psi|\,\hat{x}\,|\psi\rangle. \tag{9.1}$$

This all has to be done within 20 timesteps, so it is necessary to tweak all the parameters so that the effect would occur within these timesteps. Of course this means that the model parameters would be quite nonphysical, but since the model doesn't actually represent a real system this was a compromise that was deemed acceptable given the constraints.

9.3 Finding Optimal Parameters

9.3.1 Model Parameters

As mentioned, to find the actual model parameters it is necessary to sweep through most parameters. Although (6.1) is shown, for convenience the model equation is repeated here:

$$\begin{aligned}
H_e = -\,\frac{1}{2m}\frac{\partial^2}{\partial r^2} + \frac{1}{\left|\frac{L}{2}-R\right|} + \frac{1}{\left|\frac{L}{2}+R\right|} - \\
-\,\frac{\mathrm{erf}(|\frac{L}{2}-r|/R_r)}{\left|\frac{L}{2}-r\right|} - \frac{\mathrm{erf}(|\frac{L}{2}+r|/R_l)}{\left|\frac{L}{2}+r\right|} - \frac{\mathrm{erf}(|R-r|/R_f)}{|R-r|},
\end{aligned} \tag{9.2}$$

the margin note[3] contains a reminder of the physical meaning of the parameters. It is the case that some parameters are dependent on others, for example, the length of the box determines the energy scale and thus the timescale, these are somewhat arbitrary and so were fixed to have (in atomic units) a R parameters of $R_l = R_r = 1.71$, $R_p = 1.818$, the padding beyond the fixed nuclei was fixed at pad $= 5.7$, The length was then changed at some fixed initial position of the proton $x_p0 = -3.9$ and initial velocity $v_p = 0.004$ to see if the transition would occur. The potential energy surfaces of the various tested settings are shown in figure 9.4.

[3] The Shin-Metiu model consists of two stationary ions separated by a distance of L, specifically located at $\frac{L}{2}$ and $\frac{-L}{2}$. These enclose a mobile ion p of mass M at distance R from the origin and an electron e^- at distance r. The modified Coulomb potential is parameterized by the constants R_l, R_r and R_f, as shown in (9.2).
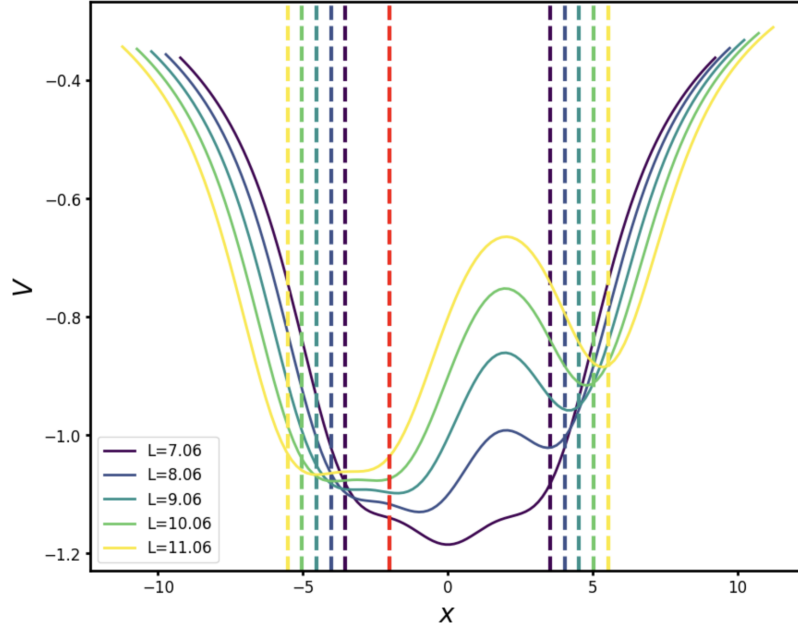
Figure 9.4: **Sweep of the PES** for the Shin-Metiu model. Credit to Malay Singh for the visualization code.

The W_{PCET} is hard to quantify for an optimizer to do the sweep automatically, since it is desirable that not only does the electron move to the right but also that there is no abrupt change in the electronic wavefunction when the proton moves, as the electron can stay stationary and be in an excited state as a fast proton moves away. By analyzing the trajectories and eigenstate evolution for the chosen lengths it was found that a length of $L = 8.6$ was optimal at a time step of $\tau = 17.857$ and yielded an evolution shown in figure 9.5. This was done on the exact ground state and with simple linear algebra for the evolution, and is hence numerically exact. Looking at it through the lens of the witness gives a very clear transition from left to right, which is shown in figure 9.6. One interesting feature is that the electron jumps from one side of the other as the proton (not shown) is crossing the midpoint of the system.
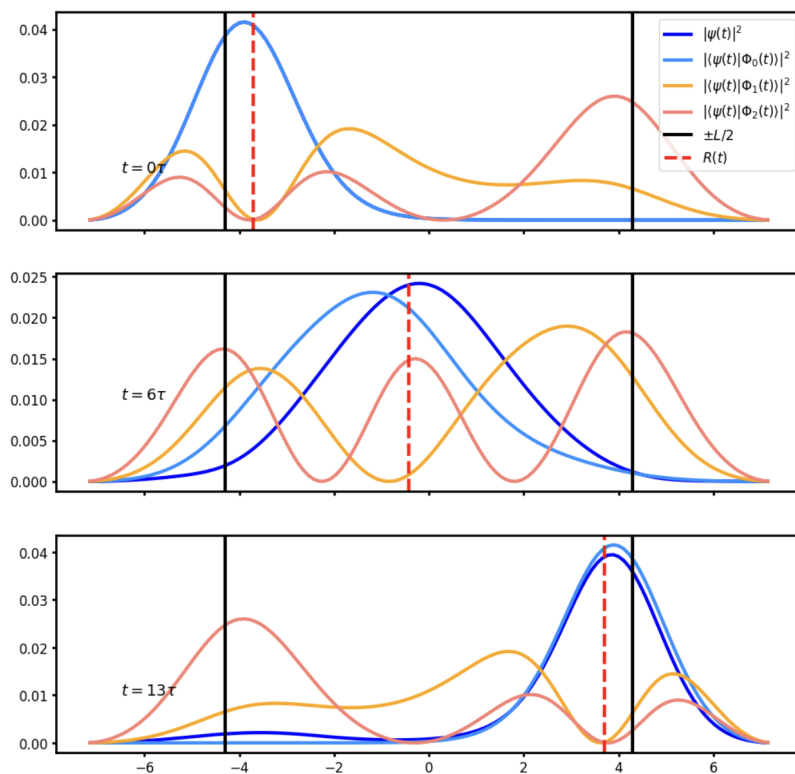
Figure 9.5: **Evolution of the wavefunction and eigenstates** for the Shin-Metiu model with the candidate parameters. $|\phi\rangle$ represent the (time dependent) eigenstates of electronic system while the dark blue line is the wavefunction $|\psi\rangle$. The red dashed line represents the position of the free proton. Credit to Malay Singh For the visualization code.
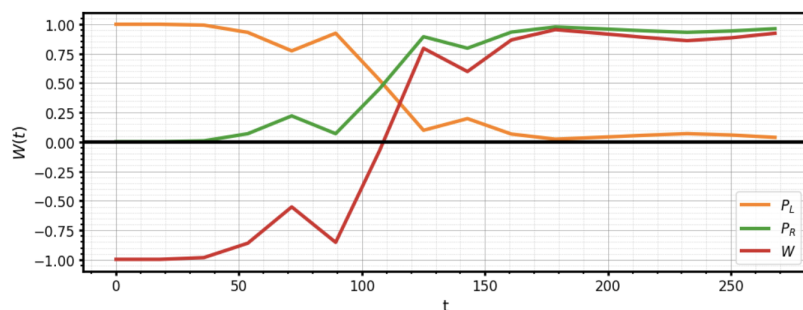
Figure 9.6: **Evolution of the PCET witness** for the discovered parameters over 13 timesteps (a little more than 250 atomic time units). The red line is the electron's average normalized position, the green and yellow lines measure the amount of amplitude in the right and left halves of the model respectively, quantifying the completeness of the electron transition. Within this time the proton smoothly transitions from left to right almost linearly, but this is not shown. Credit to Malay Singh For the visualization code.

9.3.2 VQE Ansatz

To find the optimal VQE ansatz with as low a depth as practically possible, a heuristic search over 4 qubit ansatze with 2 to 4 layers of CNOTs was carried out with a variety of connectivities and single qubit gate types. It turned out that the best ansatz that could be found was the one shown in figure 9.7, although many others were tested they either achieved comparable performance with more parameters or were worse. The ansatz has 24 parameters which is a reasonable number, although even this requires 48 different circuits to measure the analytical gradients required for the optimizer if that strategy is the one that is chosen.

9.4 Preliminary Results and Conclusion

Sadly it is not possible at this time to show any of the actual experimental results at the time of writing, but it has been possible to simulate the system with good estimates of the hardware noise. This is still preliminary work required to show to the experimental committee such that it will be possible to run the algorithm on real hardware, as everything should ideally be working by the time the machine is made available for this work, and it should be readily comparable to the results.

To achieve this goal it was necessary to run a noisy simulation. To do this, a noise model was prepared with local depolarizing error after every single qubit gate with a probability of 0.003 and a two-qubit depolarizing channel with an error probability of 0.01. The depolarizing channel is a bit of a worst case scenario, but better noise models for the machine were not available at the time of this simulation. It was again performed on 4 simulated qubits through Qiskit [251] with 10^5 shots per expectation value.

To perform error mitigation UNITED[4] was implemented with again 10^5 shots per circuit, but only at two noise levels, two copies and 10 training circuits which lead to a total shot budget of $4 \cdot 10^6$. It should be noted that it is unlikely that such an expensive error mitigation strategy will be used in the experiment as it likely will result in excessive measurements and fine-tuning for not such high rewards. At 4 qubits the system is still small enough that other techniques should work more effectively. The results from the noisy, shot limited simulation and the mitigated results are shown in figure 9.8. These results are still preliminary, and it is possible that there are errors in the implementation that have yet to be solved, however they do look plausible.

The results do not look extremely promising as the number of timesteps is actually quite small, at 14 iterations. The ideal results, which although subject to shot noise, actually are very good, but here the algorithm is likely only limited by the shot noise and the optimizer limits, and in the original TDVQP paper a decay can be
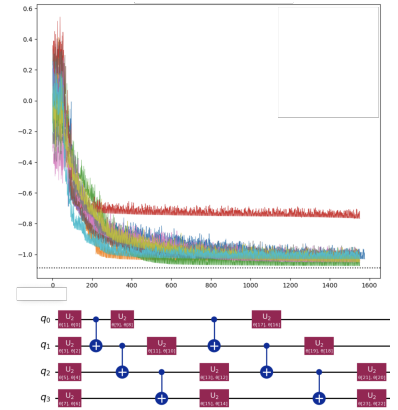
Figure 9.7: **Various optimizations of the candidate ansatz** each colored line represents a different optimization trial, with the y-axis being energy and the x-axis being iteration. The target energy is the dashed line. Underneath is the circuit structure where the red boxes are SU_2 gates with 2 parameters each for 24 in total. Credit to Dominik Tonne for the heuristic search.

[4] Please see section 7.2 for details on UNITED.
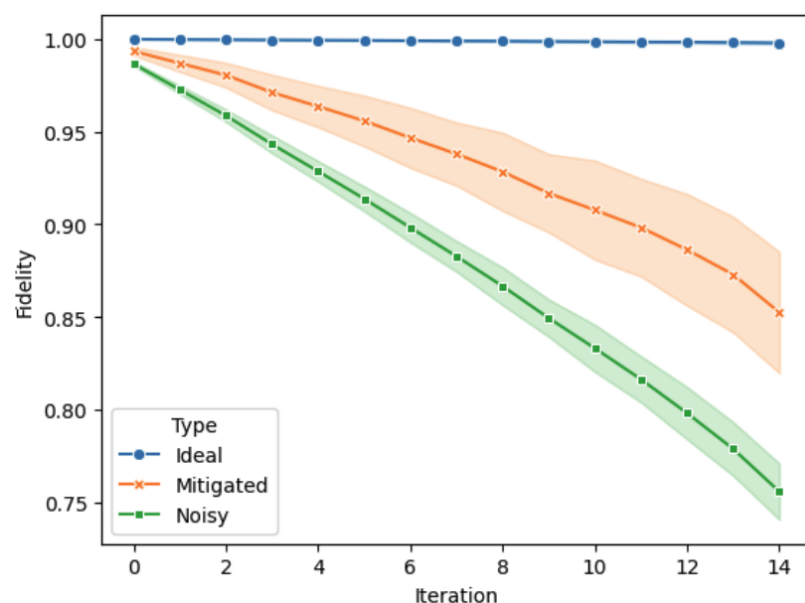
Figure 9.8: **Noisy and Mitigated fidelities of TDVQP** with optimal parameters showing 14 iterations of the algorithm. The blue line is the ideal performance, the green line shows the raw results from the machine and the orange line is the UNITED mitigated results. This all used 10^5 shots per required expectation value and was repeated 100 times from the best starting initial VQE state. The highlighted areas show the standard deviation of the respective means. One particularly fun thing to see is the increase in the uncertainty due to the error mitigation, leading to a larger standard deviation of the results. This is exactly the bias-variance tradeoff mentioned in section 4.2.

noticed, but it is worth remembering that it had 1000 iterations. The fidelities for the noisy and mitigated results are both promising and not promising - they are good enough to see some qualitative results especially after mitigation. At the same time, it would be good if something could be improved. A factor that may have come into play is the large size of the timestep, which at ~ 18 a.u. might make the approximately compiled trotter expansion too approximate. This is exactly why such studies are essential to perform before an experiment.

It is hoped that it will be possible to perfect the simulation and improve the results by tweaking various parameters and eventually run the full experiment on actual quantum hardware. It is hoped that the reader has gotten a feeling for the considerations and work that goes into translating an algorithm that attempts to be efficient to one that can actually be run on hardware.

Part III

Closing Matters

10
Conclusion and Outlook

"Which way ought I to walk from here?" – "That depends a good deal on where you want to get to."
Alice's Adventures in Wonderland, Lewis Carroll, 1865

THIS THESIS has presented a broad introduction to quantum computing with a strong focus on error mitigation and algorithms. This was to lay the foundation for the research that followed. The research can be thought of in three steps, first an almost ideal algorithm with potential to be used in current devices, then presenting research in dealing with the many errors presenting a new method of error mitigation, finally looking forward into a future where there is some error correction. Finally, the algorithmic and practical sides were brought together in a still in progress work where one of the first experiments in quantum computer-simulated first quantization based dynamics will be performed.

To achieve this goal, the reader and author have gone through a journey beginning with a short overview of the fundamentals of quantum computing, beginning with a mathematical underpinning and the foundational postulates of quantum computing. This allowed the introduction of gate based quantum computing concepts and some fundamental theorems of quantum information and computing.

From here it was possible to discuss various quantum algorithms and the kinds of advantages they have over classical algorithms. In particular a strong focus was given to the variational quantum eigensolver as it was one of the most promising candidates for these of current machines should error correction not work out. Of course, mentioning error correction means a discussion on errors, of which their characterization and foundational error mitigation techniques and considerations were presented. A brief discussion on error correction and its recent advances were shown for context.

This laid the foundations to discuss a short introduction and history to quantum chemistry on quantum computers, looking at both the second quantized approach which is more popular for electronic structure, and then first quantization approaches that may be more promising for dynamics. Then foundational algorithms for Hamiltonian simulations were presented. This knowledge and context is hopefully of help for the true meat of the work, which is the research.

Here the TDVQP algorithm for mixed quantum-classical dynamics is presented, which is used to solve the classical nuclear and quantum electron dynamics of the Shin-Metiu model in first quantization. Many simulations considering finite shot effects were run, yielding qualitatively good results. The simulations themselves were noiseless, but just to be sure that noise won't be a problem, a novel error mitigation technique is developed that combines the previously introduced techniques and combined them into one called UNITED[1]. It turns out that doing so, with enough measurements, allows one to correct better than any one technique alone.

[1] Because it is a method to combine different techniques in harmony.

Around this time though it was becoming clear that error mitigation would probably never truly bring about a useful quantum algorithm, so instead the question of what happens when one has a perfect quantum computer but then attempt to add noisy subsystems to it came up. So instead of making a computer better, what happens when it begins perfect and make it worse? It turned out that even a single qubit subject to noise would destroy the state eventually.

Undeterred and armed with all the tools and techniques required to write an algorithm and think about how it could be made to work on noisy hardware it is time to actually do so. The last chapter of the research presents a currently incomplete work that brings together the algorithm and error mitigation into a soon-to-be experiment, and discusses all the practical considerations that must be taken, hopefully instructing all who read this about the various complexities of this era of quantum computing, from how to justify time on a quantum computer to the compromises and tricks that have to be played.

As such the thesis has explored a large range of methodologies with a focus on the chemical use case for NISQ computers. Many times it is mentioned that the main obstacle to quantum computing is the presence of noise, and that when it is there most hope of advantage is pushed away. At the time of writing there is no hint of a clear advantage, even though techniques today do go beyond any exact computations [214], but not beyond very good approximations by the arch nemesis and sometimes fundamental friend of quantum computing, tensor networks.

As such at the moment noisy quantum computers are approaching state of the art classical techniques, but for a very small subset of problems that happen to be Hamiltonian simulation, which is exactly what Feynman said they would be good for. Whether this will extend to more complicated systems like those in chemistry, that is likely not the case as simulating chemical systems require extremely deep circuits. Variational circuits in general have been falling out of favour, but things like geometric deep learning and their translation into quantum machine learning may still be fruitful.

However, the major advances happening in quantum error correction make the prospect of fault-tolerant machines much more tangible. If there is any prediction to be made it is that going back to the

ideal algorithms and making them resource efficient in terms of gate counts or discovering new constructions[2] for important operations for fault-tolerant machines is the future of the craft.

At the same time, it is the case that to keep this research exciting it would be very beneficial to the field if even existing machines yield useful results, so such a pursuit is beneficial to both the quantum computer providers and the lucky research groups that achieve the first hints at advantage. If they can cut through the noise!

[2] A construction is something like an algorithm to generate a specific type of state or some N-qubit unitary.

11
A Reflection

'Roses have thornes, and silver fountaines mud, Cloudes and eclipses staine both Moone and Sunne'
No more be grieved at that which thou hast done, Sonnet 35, William Shakespeare, c.a.1592-1595

Quantum computing is a truly odd field. It is full of promise and problems, like many others, but it also has unrealistic expectations brought upon it from an extreme amount of investment that has not truly been seen often in the history of science. This has lead to many claims that at best stretch the truth and due to this the research has become applied before it truly has a chance at being realistic. In my opinion this is a bit of a pity, and I fear this work has somewhat fallen into that race of trying to apply something too soon when it was still time to explore the waters.

When I started working in QC in late 2018 Preskill had only just written his prediction of the potential of the so called NISQ era [168]. Indeed, in that year was the first implementation of what is now known as zero noise extrapolation [75] from theoretical work only done in the previous year [112]. As such it was an exciting time to be in the field. As I have mentioned several times throughout the work, there was a veritable explosion in the exploration of NISQ ready algorithms.

Indeed, for better or for worse the work I did through this thesis fits fairly nicely within this scheme. My first paper of the PhD was a result of the LANL summer school I had done just before and throughout the first month or two of the PhD which was the UNITED error mitigation technique. This then lead to some discussions where the idea of combining perfect and imperfect qubits came along, which was the second paper of my PhD. Concurrently to this Oriol Vendrell and me had been working on various ideas on non-adiabatic dynamics on QCs, a lot of which were researched and published in the early stages of our own work which was disheartening to me, but such is research, and we found a nice topic in mixed quantum classical dynamics in first quantization. This was based on an algorithm based on VQE, and so very appropriate for this era of research. This is being picked up and being slowly but surely adapted to run on an actual QC. Indeed, now the very same model is being studied by at least one other group [325], under a very different approach, surprisingly working off different results from the

Vendrell group.

Now that it's 2025 there are many more doubts on the validity or scalability of the VQE and to a lesser extent other VQAs, for after all the variational principle and parameterized models will always be a compass to numerical and theoretical studies, leaving quantum computing in a split between two paths. One is chasing the road pitched by IBM which is that of *quantum utility* [213], whether it's utility or futility, only time will tell. The other is slightly more far term, and it is a return to *fault-tolerant algorithms*, which are a little different from the early 'idealistic' algorithms of the before times[1]. These algorithms go back to not worrying about noise, but now take into account some of the remaining limitations imposed by error correction. In particular, the amount of time it takes to run an algorithm and how many measurements are required to estimate important expectation values, since the overhead to error correction is likely to be quite large. This is also not helped by the fact that the clock speed of QCs are limited by the physical speed of the interactions required to perform gates, and as such there will be a plateau.

As such there is a resurgence of creativity beyond variational algorithms, which I think will be a great benefit to the potential usefulness to quantum computing, be it in the actual discovery of an actionable quantum advantage or in the discovery of techniques that may be useful in classical computing. I call this potential phenomena *scientific backporting*[2], and it has happened a few times, usually due to the fact that noise removes quantum advantage from many classes of algorithms [331], but some observables are easier to simulate than others even in noiseless cases [62]. Once in a while a supposed algorithmic advantage is disproven [332].

There is also the question, now that machine learning has effectively solved[3] some long-standing problems such as protein folding [333], if QCs are truly necessary if one can learn the right approach to specific problems that do not need to explore an exponentially large Hilbert space. This is part of the race of being able to numerically compute or simulate problems with an ever-growing number of components, but you shouldn't forget that sometimes mathematical or purely theoretical insights can shed much more light on a problem rather than more accurately computing yet another test case. Nonetheless, for problems of practical importance if quantum computing or machine learning are the ones that will solve ever larger problems, it will be a net gain for all research.

[1] Before 2017

[2] In the same way programs and games from newer machines back in the early days of home computing would be rewritten for a particularly popular older machine.

[3] By this I mean that for many practical purposes excellent solutions have been found

Appendix

A

Time-Dependent Variational Quantum Propagation

A.1 Multiple transitions

The simulation parameters for these examples are synthetic, with a $\Delta t = 0.05$ and an initial velocity of $v = 0.2$, which tests if the algorithm can deal with more complex dynamics within 700 timesteps. All other parameters are kept as in the main text. Figure A.1 shows the dynamics of the populations with 4 population crossings that are well described between states 0,1 and 2.
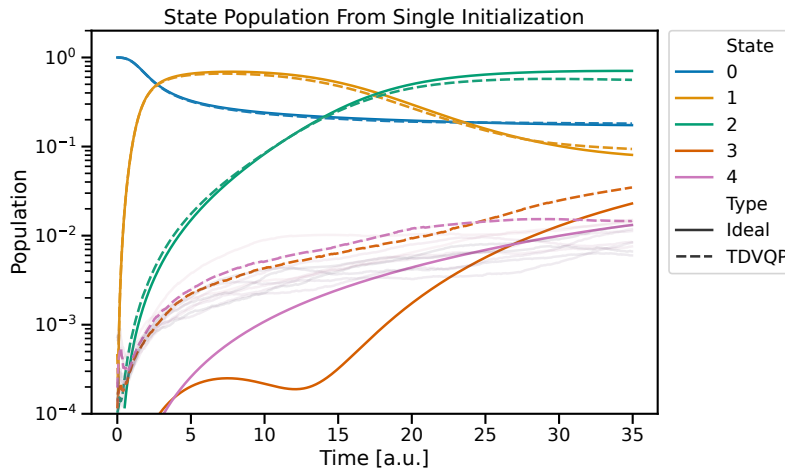
Figure A.1: **State population evolution for multiple transitions.** The simulation of the Shin-Metiu model is initialized in the ground state and evolved for 700 steps of $\Delta t = 0.05$ with a fast-moving proton ($v = 0.2$). This induces 4 transitions. The lines show the mean BOPES state populations (solid for ideal and dashed for TDVQP) with faint lines representing higher energy levels of the TDVQP.

What is particularly interesting is that around the end of the simulation ($t > 17$), there is a fall in occupation of the second state and an increase in the first state. This is nicely matched by what the energy in Figure A.2. Now that the state is highly excited, population loss to lower energy states causes a drop in the energy rather than the increase in the other examples.

This fall in energy is not accompanied by an increase in fidelity, and as Figure A.3 shows, the fidelity keeps decreasing at a steady rate, although likely that at very long times it would begin to oscillate at around 0.5.

A.2 Arbitrary state evolution

Total Energy
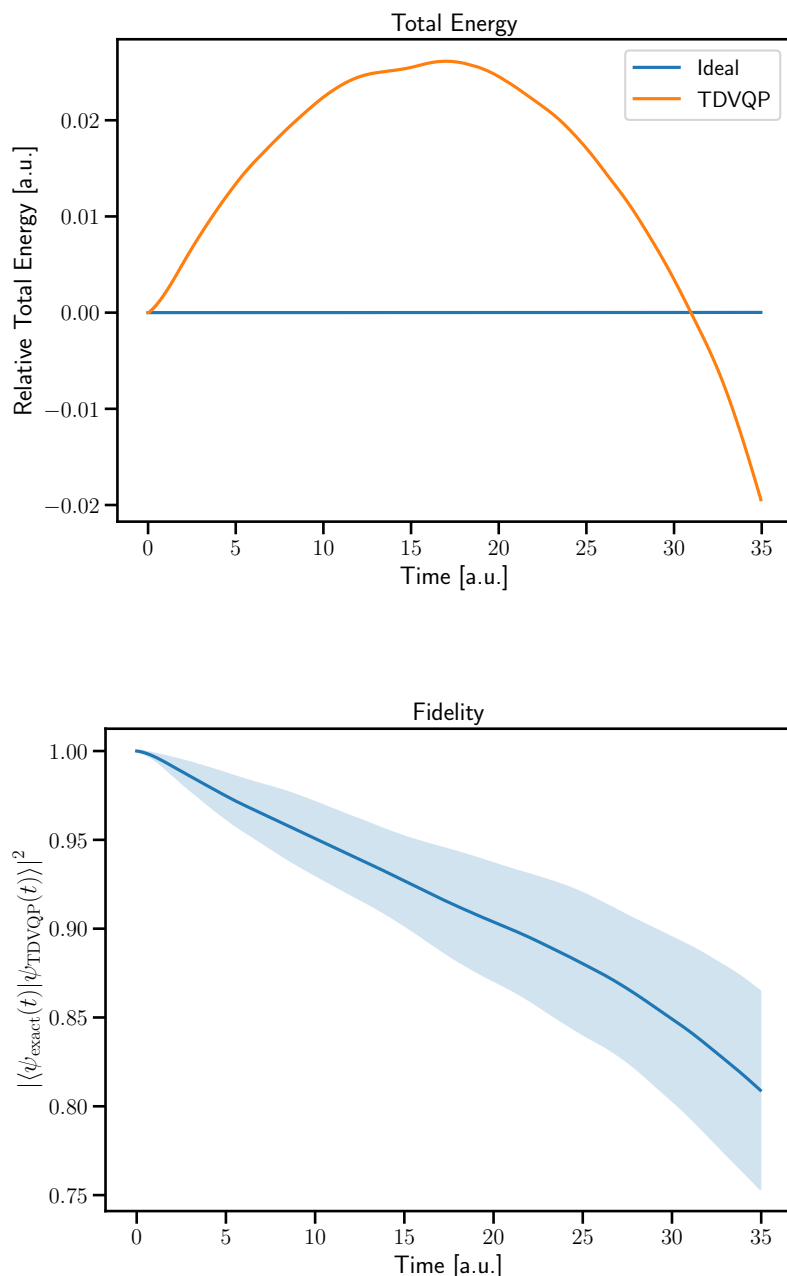
Figure A.2: **Change in energy for the multiple transition simulation.** The plot shows the energy behaviour of the TDVQP algorithm (orange solid line) as it progresses over 700 timesteps of $\Delta t = 0.05$ starting in the ground state, but with a fast-moving proton (with an initial velocity of $v = 0.2$). The ideal simulation is energy-conserving. The rise in TDVQP energy is due to leakage to higher energy levels, which then transitions to leakage in lower energy levels as the higher energy level is populated.

Fidelity

Figure A.3: **Fidelity for the multiple transition simulation.** The fidelity (solid blue line) of the TDVQP wavefunctions compared to the exact evolution of the initial state for 700 timesteps of $\Delta t = 0.05$. It can be seen that it decreases over time more quickly than the ground state evolution.

To prepare an arbitrary state it is possible to find some unitary U that can operate on a state $|\psi\rangle$ such that $U|\psi\rangle = |0\rangle$. This circuit can be made to have the form of the desired ansatz and only it is only allowed to vary the parameters θ. Some threshold or number of iterations is set and the expression $\min_\theta(1 - |\langle\psi|U(\theta)|0\rangle|^2)$ is optimized to said threshold. This is done by running the circuit in Figure A.4, where the quantum computer is initialized int the excited state. For this simulation, the starting state is set to be the desired state, and an initialization in the first excited state is shown in Figure A.5 and a superposition of the two lowest BOPES in Figure A.6.

$$|\psi_{\text{desired}}\rangle \;\text{/}\!\!\!-\; \boxed{U(\theta)^\dagger} \;\text{/}\!\!\!-\; \approx |0\rangle$$

Figure A.4: **Computing the ansatz parameters for arbitrary state preparation.** The quantum simulator is initialized in the desired state, and the ansatz U parameters θ are varied until the final state is close to $|0\rangle$.
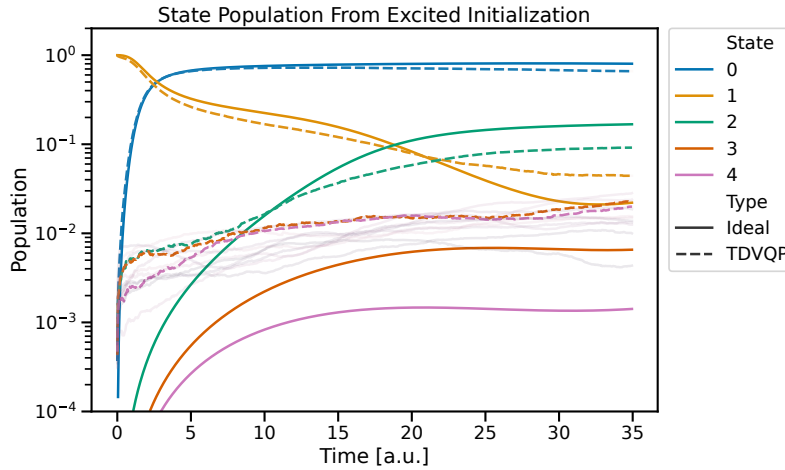
Figure A.5: **State population evolution of the first excited state.** The simulation of the Shin-Metiu model initialized in the first excited state and evolved for 700 steps of $\Delta t = 0.05$. Initial conditions are the same as in Section 6.3 and the lines show the mean BOPES state populations. The faint lines represent higher energy levels of the TDVQP.
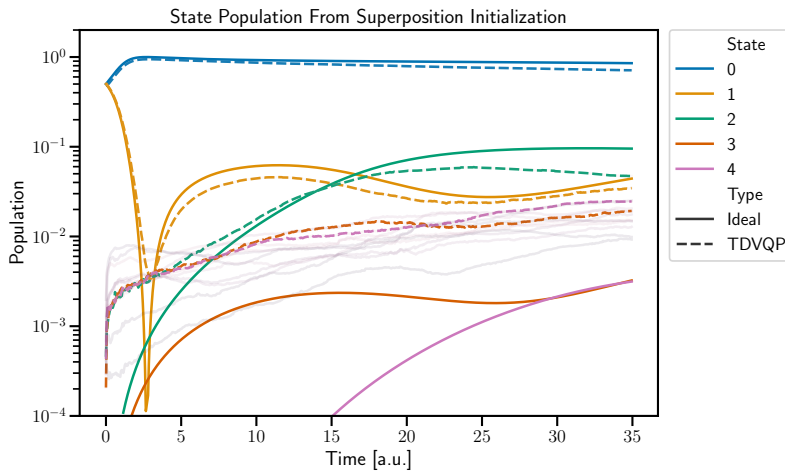
Figure A.6: **State population evolution of the superposition state.** The simulation of the Shin-Metiu model initialized in an equal superposition of the 0 and 1 states showing the mean BOPES state populations and evolved for 700 steps of $\Delta t = 0.05$. Initial conditions are the same as in Section 6.3 and the lines show the mean BOPES state populations. The faint lines represent higher energy levels of the TDVQP.
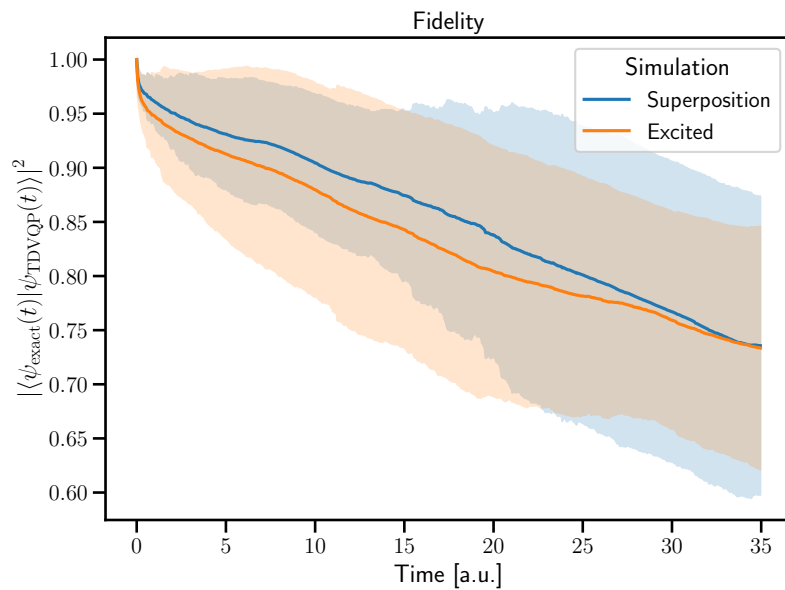
Figure A.7: **Mean fidelity** over the evolution of the equal superposition showing the lowest BOPES states (solid, blue) and first excited state (orange, dashed). The highlighted areas show the standard deviation of their respective values.
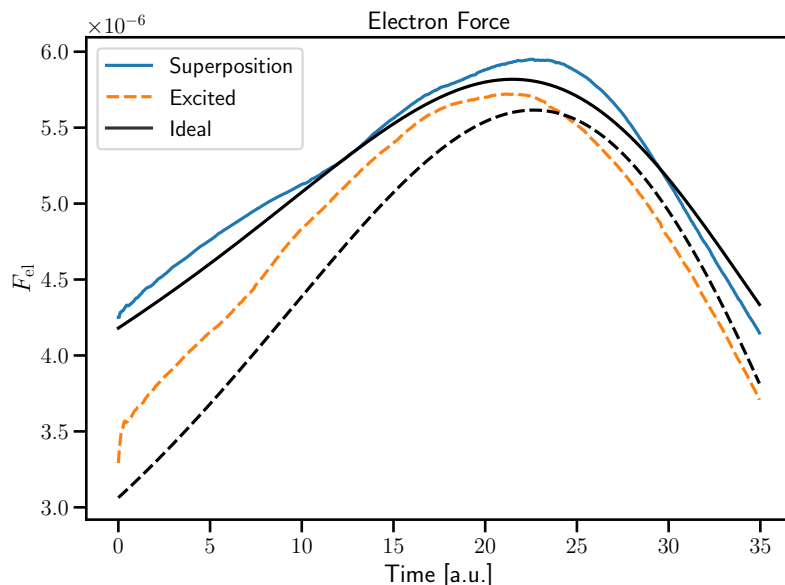
Figure A.8: **Mean force measurements** for an equal superposition of the lowest BOPES states (solid, blue) and first excited state (orange, dashed) compared to their respective ideal simulations (black).

Figures A.4 and A.5 both show that the states of interest are initially well represented by the algorithm. As the evolution continues the evolution remains qualitatively similar, but degrades, especially when populations approach the 'noise floor' of the algorithm around 10^{-3} where the higher energy levels are populated. Figure A.7 shows a very sharp decrease in fidelity in the first timesteps and quite a large standard deviation compared to ground state results shown in Figure A.3. This may either be due to not being able to initialize the ansatz as well in non-ground state settings with this approach. Forces are still well followed qualitatively as in Figure A.8.

A.3   Inherent Errors in Ideal TDVQP

The TDVQP algorithm inherits all of the errors of its constituent parts. This includes the chosen circuit compression algorithm, time evolution approximation, and in the classical propagator. Nonetheless, it is important to have an intuition of the potential pitfalls of the algorithm. The first subsection will begin with a short derivation of the effect of the magnitude of the infidelity of the wave function on an observable. This is followed by an analysis of the propagation through the velocity Verlet integrator, and finally, numerical experiments comparing the effect of various potential errors on the Shin-Metiu model. The second subsection will show the effect of finite shots on the algorithm

A.3.1   Errors Due to Incorrect Force Expecation Values

The error can be thought of as a superposition of the desired state $|\psi\rangle$ some combination of undesired orthogonal states $|\phi\rangle$ are in a superposition of $|\tilde{\psi}\rangle = \sqrt{1 - I^2}\,|\psi\rangle + I\,|\phi\rangle$, where I is the infidelity.

When a Hermitian observable $\mathcal{O}$ is measured, the result is

$$\langle \tilde{\psi} | \mathcal{O} | \tilde{\psi} \rangle = \left(1 - I^2\right) \langle \psi | \mathcal{O} | \psi \rangle + I^2 \langle \phi | \mathcal{O} | \phi \rangle . \tag{A.1}$$

The actual measured observable is completely system dependent, so the effect of its magnitude on the rest of the algorithm cannot be estimated at this stage. Nonetheless, an idea of the effect of this on the integrator can be formed by assuming that this directly translates to a worst-case error in the force observable, so that for an error of some magnitude, F_e is replaced in eq. 6.2 with $F_e(R_i, |\psi_i\rangle) = F_i + F_{i\epsilon}$ and propagate the first timestep.

$$\tilde{R}_1 = R_0 + \dot{R}_0 \Delta t + \frac{F_0 + F_{0\epsilon}}{M} \Delta t^2 \tag{A.2}$$

$$= R_1 + \frac{F_\epsilon}{M} \Delta t^2, \tag{A.3}$$

$$\tag{A.4}$$

This shows the error in position is linear in the force error and quadratic in the timestep. This now enters the generation of the new position-dependent Hamiltonian such that $H_{el}(R_1 + \frac{F_\epsilon}{M} \Delta t^2)$ which is used on the subsequent step. The first evolution begins at a known position and the second is already affected by the previous error as

$$|\tilde{\psi}_1\rangle = \exp(-i H_{el}(R_0) \Delta t) |\tilde{\psi}_0\rangle , \tag{A.5}$$

$$|\tilde{\psi}_2\rangle = \exp(-i H_{el}(\tilde{R}_1) \Delta t) |\tilde{\psi}_1\rangle . \tag{A.6}$$

Sadly even in this simple 1-dimensional model, it is difficult to analytically determine the effect on the evolution of the subsequent wavefunction, so assume this effect is negligible within one timestep. Now the effect on the velocity update can be computed as

$$\dot{\tilde{R}}_1 = \dot{R}_0 + \frac{F_0 + F_{0\epsilon} + F_1 + F_{1\epsilon}}{2M} \Delta t \tag{A.7}$$

$$\dot{\tilde{R}}_1 = \dot{R}_0 + \frac{F_0 + F_1}{2M} \Delta t + \frac{F_{0\epsilon} + F_{1\epsilon}}{2M} \Delta t \tag{A.8}$$

$$\dot{\tilde{R}}_1 = \dot{R}_1 + \frac{F_{0\epsilon} + F_{1\epsilon}}{2M} \Delta t \tag{A.9}$$

The update can always be separated into the contributions of the ideal integration and the integration of the error. This shows that the velocity estimation error is linear in the error in the force and linear in time. From the second timestep onwards, the error will accumulate leading to behaviour like

$$\tilde{R}_i = \tilde{R}_{i-1} + \dot{\tilde{R}}_{i-1} \Delta t + \frac{F_i + F_{i\epsilon}}{2M} \Delta t^2 . \tag{A.10}$$

This expression can then be expressed as the ideal contributions and the contributions from the force error as

$$\tilde{R}_i = R_i + \dot{R}_{i-1} \Delta t + \sum_{j=0}^{i} \left(\frac{F_{(j-2)\epsilon} + F_{(j-1)\epsilon}}{2M} + \frac{F_{j\epsilon}}{M} \right) \Delta t^2 . \tag{A.11}$$

$$\tag{A.12}$$

And if the error in the force is constant the expression simplifies to

$$\tilde{R}_i = R_i + \dot{R}_{i-1}\Delta t + \frac{(i^2 + i)F_\epsilon}{2M}\Delta t^2, \qquad (A.13)$$

which is quadratic in the timestep, linear in the magnitude of the error, and quadratic with respect to the number of time steps. The error is also likely to be proportional to I^2, which will initially be small but may increase unexpectedly as the desired position and subsequent error-prone time evolution will increasingly diverge from the ideal time evolution.

The above section illustrates that there is an effect due to the inherent interplay between the observables and the classical propagation which increases through time. But The fidelity of the wavefunction is always affected by the set optimizer threshold for fidelity. Assuming this is always met within the maximum allowed number of iterations per timestep, with a threshold of T and assuming no other errors, the fidelity with the number of iterations i should fall as

$$|\langle\psi_{\text{exact}}(i)|\psi_{\text{TDVQP}}(i)\rangle|^2 \approx T^i. \qquad (A.14)$$

The overall effect of the threshold error is already illustrated in Figure 6.11, but the effect of the observable deserves numerical simulations, so two different types of errors have been simulated. The first is a constant additive offset ($\tilde{F} = F + F_\epsilon$) that simply adds a force of the stated magnitude at each timestep and is shown in Figure A.9. The second is a multiplicative factor $\tilde{F} = F + F \cdot \epsilon$, which is force-dependent and its effect is shown in Figure A.10. The effect of both types of errors is quantified by comparing the fidelity of the ideal evolution to the one in which this error is injected at each timestep, but otherwise, the quantum evolution and integrator are the same as in the ideal case. The effect in both is a straight line in the log-log plot of $1 - \text{fidelity}$ over time, which hints at there being a power law as in eq. A.13. The gradient of the lines is identical in all cases within one error group (additive or multiplicative), but it is larger than 2. This may be due to the change in the Hamiltonian as the positions diverge which is not taken into account in eq. A.13.

Another potential error that should be disentangled from the rest is the effect of the trajectory on the TDVQP. To do this, the nucleus was set to follow the path it would have followed on the exact simulation irrespective of the measured observable. In effect, an evolution under an external time-dependent Hamiltonian. Figure A.11 shows the effect of the parameterization, which is insignificant at the infinite shot limit, but quite noticeable in the finite shot case. The difference is particularly pronounced near the transition point (which is beginning to be approached around 200 a.u.), where the speed of the nucleus has a large effect on the transition probability, as expected by the Landau-Zener formula. The lower measured force over the simulation as seen in Figure 6.10 means that the speed of approach differs enough for the transition to cause a divergence in fidelity of the two.
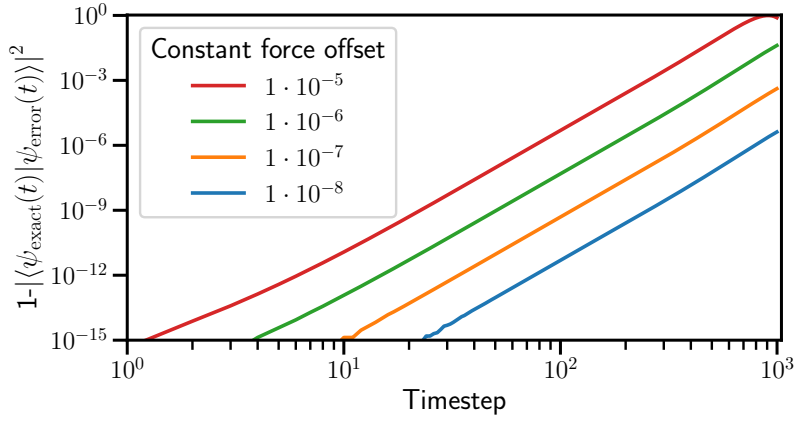
Figure A.9: **Effect of an additive force measurement error on ideal propagation.** The force is a constant addition of F_ϵ at each step such that $\tilde{F}_i = F_i + F_\epsilon$. Different colours represent different error factors. Reproduced from [180] under CC4.
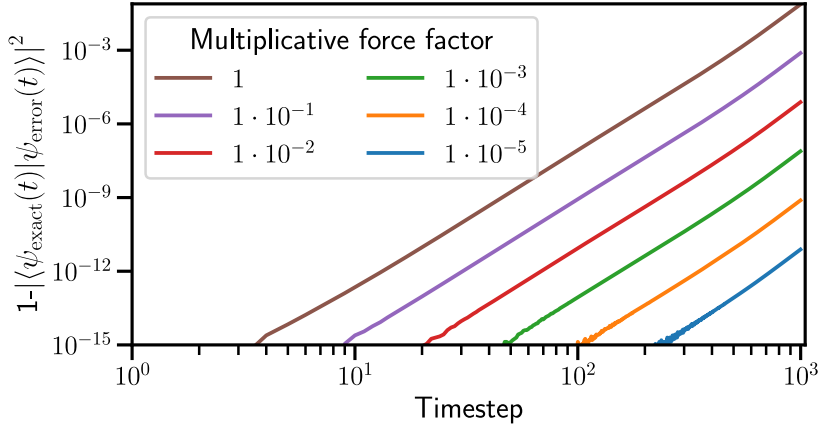
Figure A.10: **Effect of a multiplicative force measurement error on ideal propagation.** Here the force is a constant addition of a magnitude (ϵ) dependent error such that $\tilde{F}_i = F_i + F_i \cdot \epsilon$. Different colours represent different error magnitudes. Reproduced from [180] under CC4.
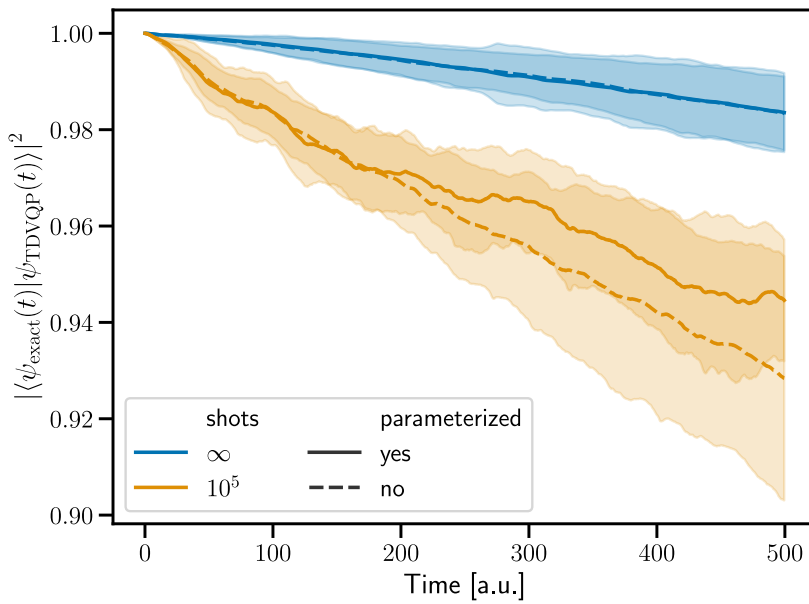
Figure A.11: **Effect of using a parameterized trajectory** on the p-VQD algorithm, showing the infinite shot limit in blue and 10^5 shots in orange, with highlights showing the standard deviation. Solid lines are unparameterized while dashed lines are. In the infinite shot limit, there is little difference between the two, but in the finite shot case, fidelity does not fall as quickly nearing the transition point, likely due to a difference in the speed of approach of the nucleus. Reproduced from [180] under CC4.

A.3.2   Finite Shot Effects
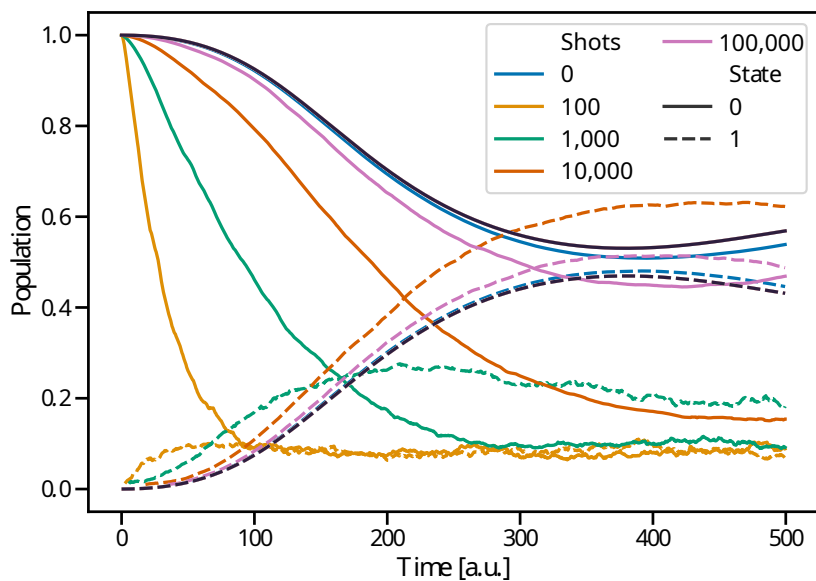
Figure A.12: **Populations with finite sampling effect** for the MD initialization simulation, showing different colours for the various shot counts, 0 being the infinite shot limit. Dashed lines represent the excited state while solid lines represent the ground state. Reproduced from [180] under CC4.

Figure A.12 shows the effect of finite sampling error on the state populations. When shot noise is taken into account, the system tends to move towards the equal superposition state. Interestingly, the population transfer between states 0 and 1 is enhanced, likely due to the faster decrease of the 0 state to other states.

B

UNITED

B.1 Correcting Depolarizing Noise with UNITED

Consider a global depolarizing noise channel that acts on all Q qubits, this channel takes the form:

$$\rho \longrightarrow (1-p)\rho + p\frac{\mathbb{I}}{d} \tag{B.1}$$

where d is the dimension of the Hilbert space and $0 \leqslant p \leqslant 1$. If this channel acts i times throughout this can be represented as:

$$\rho_{\psi,i} = (1-p)^i \rho_\psi + (1 - (1-p)^i)\frac{\mathbb{I}}{d} \tag{B.2}$$

Consider the effect of the above channel on some observable of interest X:

$$X_\psi^{noisy,i} = \mathrm{Tr}(\rho_{\psi,i}X) \tag{B.3}$$

$$= (1-p)^i X_\psi^{exact} + (1 - (1-p)^i)\frac{\mathrm{Tr}(X)}{d} \tag{B.4}$$

Using an ansatz which combines observables evaluated at multiple state preparation noise levels an several numbers of copies in the VD circuit structure. For simplicity swap noise increases are omitted in this treatment.

$$\hat{X}_\psi^{mit} = \sum_{i=1,m=1}^{n,M} a_{i,m} X_\psi^{noisy,i,m} \tag{B.5}$$

where $\hat{X}_\psi^{mit}$ is the estimate of the observable of interest, n is the number of noise levels and $a_{i,m}$ are the parameters fitted using least squares regression on the data generated by the near-Clifford VD training circuits with m copies. Therefore,

$$X_\psi^{noisy,i,m} = \frac{\mathrm{Tr}(\rho_{\psi,i}^m X)}{\mathrm{Tr}(\rho_{\psi,i}^m)} \tag{B.6}$$

This expression can be evaluated. First focusing on the denominator:

$$\mathrm{Tr}(\rho_{\psi,i}^m) = \mathrm{Tr}\left(\sum_{k=0}^m \binom{m}{k}\left((1-p)^i\rho_\psi\right)^{(m-k)}\left(\frac{1-(1-p)^i}{d}\right)^k\right) \tag{B.7}$$

$$= \sum_{k=0}^m \binom{m}{k}(1-p)^{i(m-k)}\mathrm{Tr}(\rho_\psi{}^{m-k})\left(\frac{1-(1-p)^i}{d}\right)^k \tag{B.8}$$

If it is assumed that ρ_ψ is pure then $\rho_\psi^2 = \rho_\psi$ and $\mathrm{Tr}\rho_\psi = 1$. Noting that when $m = k$ it is a trace over the identity, giving:

$$\mathrm{Tr}(\rho_{\psi,i}^m) = \sum_{k=0}^{m-1} \binom{m}{k} \left((1-p)^{i(m-k)} \left(\frac{1-(1-p)^i}{d} \right)^k \right) + d \left(\frac{1-(1-p)^i}{d} \right)^m \tag{B.9}$$

$$= (1-p)^{im} \left(\frac{(1-p)^{-i}-1}{d} + 1 \right)^m + (d-1) \left(\frac{1-(1-p)^i}{d} \right)^m \tag{B.10}$$

$$= \left(\frac{1+(d-1)(1-p)^i}{d} \right)^m + (d-1) \left(\frac{1-(1-p)^i}{d} \right)^m \tag{B.11}$$

Now consider the numerator of Eq.B.6

$$\mathrm{Tr}(\rho_{\psi,i}^m X) = \mathrm{Tr} \left(\sum_{k=0}^{m} \binom{m}{k} ((1-p)^i \rho_\psi)^{(m-k)} \left(\frac{(1-(1-p)^i)}{d} \right)^k X \right) \tag{B.12}$$

$$= \left(\frac{1+(d-1)(1-p)^i}{d} \right)^m \mathrm{Tr}(\rho_\psi X) + (\mathrm{Tr}(X) - \mathrm{Tr}(\rho_\psi X)) \left(\frac{1-(1-p)^i}{d} \right)^m \tag{B.13}$$

$$= \mathrm{Tr}(\rho_\psi X) \left(\left(\frac{1+(d-1)(1-p)^i}{d} \right)^m - \left(\frac{1-(1-p)^i}{d} \right)^m \right) + \mathrm{Tr}(X) \left(\frac{1-(1-p)^i}{d} \right)^m \tag{B.14}$$

Combining the above yields the following expression:

$$X_\psi^{noisy,i,m} = f_{i,m} \mathrm{Tr}(\rho_\psi X) + g_{i,m} \mathrm{Tr}(X) \tag{B.15}$$

where

$$f_{i,m} = 1 - \frac{d}{((1-(1-p)^i)^{-m}(1+(d-1)(1-p)^i)^m + d - 1)} \tag{B.16}$$

$$g_{i,m} = \frac{1}{((1-(1-p)^i)^{-m}(1+(d-1)(1-p)^i)^m + d - 1)} \tag{B.17}$$

This can be simplified further as $f_{i,m} = 1 - dg_{i,m}$. Therefore,

$$X_\psi^{noisy,i,m} = (1 - dg_{i,m})\mathrm{Tr}(\rho_\psi X) + g_{i,m} \mathrm{Tr}(X) \tag{B.18}$$

It is good to check that $g_{1,1} = p/d$, $f_{1,1} = 1 - p$ leads to the expected result for one copy and one noise level, $g_{2,1} = \frac{(2-p)p}{d} = \frac{1-(1-p)^2}{d}$, $f_{2,1} = (1-p)^2$ also as expected.

Take the case where the ansatz is formed from one noise level $i = 1$ and two copies $m = 1, 2$.

$$\hat{X}_\psi^{mit} = a_{1,1} X_\psi^{noisy,1,1} + a_{1,2} X_\psi^{noisy,1,2} \tag{B.19}$$

with this anstaz is it possible to completely remove the effects of global depolarizing noise?

$$\hat{X}_\psi^{mit} = a_{1,1}((1-p)X_\psi^{exact} + p\frac{\mathrm{Tr}(X)}{d}) + a_{1,2}((1 - dg_{1,2})X_\psi^{exact} + g_{1,2}\mathrm{Tr}(X)) \tag{B.20}$$

$$= X_\psi^{exact}(a_{1,1}((1-p) + a_{1,2}(1 - dg_{1,2})) + \mathrm{Tr}(X)(\frac{p}{d}a_{1,1} + a_{1,2}g_{1,2}) \tag{B.21}$$

For $\hat{X}_\psi^{mit} = X_\psi^{exact}$, this gives:

$$a_{1,1} = -g_{1,2}\frac{d}{p - dg_{1,2}}, \quad a_{1,2} = \frac{p}{p - dg_{1,2}} \tag{B.22}$$

Considering $m = 2$ gives

$$g_{1,2} = \frac{p^2}{d(d(p-1)^2 - (p-2)p)} \tag{B.23}$$

So it is shown that this ansatz can correct global depolarizing noise for one noise level. This extends to any number of noise levels.

Bibliography

[1] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Higher Education from Cambridge University Press. Dec. 9, 2010. DOI: `10.1017/CBO9780511976667`. URL: `https://www.cambridge.org/highereducation/books/quantum-computation-and-quantum-information/01E10196D0A682A6AEFFEA52D53BE9AE` (visited on 03/08/2023).

[2] Sheldon Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Cham: Springer International Publishing, 2024. ISBN: 978-3-031-41025-3 978-3-031-41026-0. DOI: `10.1007/978-3-031-41026-0`. URL: `https://link.springer.com/10.1007/978-3-031-41026-0` (visited on 10/01/2024).

[3] Yuri Manin. *Mathematics as Metaphor: Selected Essays of Yuri I. Manin*. American Mathematical Soc., 2007. 258 pp. ISBN: 978-0-8218-4331-4. Google Books: `D8t0rWSX1DEC`.

[4] Richard P. Feynman. "Quantum Mechanical Computers". In: *Optics News* 11.2 (Feb. 1985), pp. 11–11. DOI: `10.1364/ON.11.000011`. URL: `http://www.opticsinfobase.org/abstract.cfm?URI=ON-11-2-11`.

[5] Gordon E Moore. "Cramming More Components onto Integrated Circuits". In: *Electronics* 38.8 (1965).

[6] Charles E. Leiserson et al. "There's Plenty of Room at the Top: What Will Drive Computer Performance after Moore's Law?" In: *Science* 368.6495 (June 5, 2020), eaam9744. DOI: `10.1126/science.aam9744`. URL: `https://www.science.org/doi/full/10.1126/science.aam9744` (visited on 09/03/2024).

[7] Zryan Najat Rashid et al. "Distributed Cloud Computing and Distributed Parallel Computing: A Review". In: *2018 International Conference on Advanced Science and Engineering (ICOASE)*. 2018 International Conference on Advanced Science and Engineering (ICOASE). Oct. 2018, pp. 167–172. DOI: `10.1109/ICOASE.2018.8548937`. URL: `https://ieeexplore.ieee.org/abstract/document/8548937` (visited on 09/03/2024).

[8] John D. Owens et al. "GPU Computing". In: *Proceedings of the IEEE* 96.5 (May 2008), pp. 879–899. ISSN: 1558-2256. DOI: `10.1109/JPROC.2008.917757`. URL: `https://ieeexplore.ieee.org/abstract/document/4490127` (visited on 09/03/2024).

[9] Stefano Markidis et al. "NVIDIA Tensor Core Programmability, Performance & Precision". In: *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. 2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). May 2018, pp. 522–531. DOI: `10.1109/IPDPSW.2018.00091`. URL: `https://ieeexplore.ieee.org/abstract/document/8425458` (visited on 09/03/2024).

[10] Christopher De Sa et al. *High-Accuracy Low-Precision Training*. Mar. 8, 2018. DOI: `10.48550/arXiv.1803.03383`. arXiv: 1803.03383 [cs, stat]. URL: `http://arxiv.org/abs/1803.03383` (visited on 09/03/2024). Pre-published.

[11]   Xiao Sun et al. "Ultra-Low Precision 4-Bit Training of Deep Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 1796–1807. URL: https://papers.nips.cc/paper_files/paper/2020/hash/13b919438259814cd5be8cb45877d577-Abstract.html (visited on 09/03/2024).

[12]   Tom Conte et al. *Thermodynamic Computing*. Nov. 14, 2019. DOI: 10.48550/arXiv.1911.01968. arXiv: 1911.01968 [cs]. URL: http://arxiv.org/abs/1911.01968 (visited on 01/31/2024). Pre-published.

[13]   Denis Melanson et al. *Thermodynamic Computing System for AI Applications*. Dec. 8, 2023. DOI: 10.48550/arXiv.2312.04836. arXiv: 2312.04836 [cond-mat]. URL: http://arxiv.org/abs/2312.04836 (visited on 01/31/2024). Pre-published.

[14]   Maya Gokhale et al. *Report of the DOE/NSF Workshop on Correctness in Scientific Computing, June 2023, Orlando, FL*. Version 2. Dec. 28, 2023. DOI: 10.48550/arXiv.2312.15640. arXiv: 2312.15640. URL: http://arxiv.org/abs/2312.15640 (visited on 12/08/2024). Pre-published.

[15]   M. A. Nielsen and Isaac L. Chuang. "Programmable Quantum Gate Arrays". In: *Physical Review Letters* 79.2 (July 14, 1997), pp. 321–324. DOI: 10.1103/PhysRevLett.79.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.79.321 (visited on 01/31/2024).

[16]   Jan Kaiser and Supriyo Datta. "Probabilistic Computing with P-Bits". In: *Applied Physics Letters* 119.15 (Oct. 12, 2021), p. 150503. ISSN: 0003-6951. DOI: 10.1063/5.0067927. URL: https://doi.org/10.1063/5.0067927 (visited on 01/31/2024).

[17]   S. Somaroo et al. "Quantum Simulations on a Quantum Computer". In: *Physical Review Letters* 82.26 (June 28, 1999), pp. 5381–5384. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.82.5381. URL: https://link.aps.org/doi/10.1103/PhysRevLett.82.5381 (visited on 12/03/2024).

[18]   Jiangfeng Du et al. "NMR Implementation of a Molecular Hydrogen Quantum Simulation with Adiabatic State Preparation". In: *Physical Review Letters* 104.3 (Jan. 2010), pp. 030502–030502. ISSN: 0031-9007\n1079-7114. DOI: 10.1103/PhysRevLett.104.030502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.030502.

[19]   Kenneth R. Brown, Robert J. Clark, and Isaac L. Chuang. "Limitations of Quantum Simulation Examined by Simulating a Pairing Hamiltonian Using Nuclear Magnetic Resonance". In: *Physical Review Letters* 97.5 (Aug. 4, 2006), p. 050504. DOI: 10.1103/PhysRevLett.97.050504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.050504 (visited on 12/04/2024).

[20]   P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136 (3B 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864. URL: https://link.aps.org/doi/10.1103/PhysRev.136.B864.

[21]   W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140 (4A Nov. 15, 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133. URL: https://link.aps.org/doi/10.1103/PhysRev.140.A1133 (visited on 12/08/2024).

[22]   Oriol Vendrell and Hans-Dieter Meyer. "Multilayer Multiconfiguration Time-Dependent Hartree Method: Implementation and Applications to a Henon–Heiles Hamiltonian and to Pyrazine". In: *The Journal of Chemical Physics* 134.4 (Jan. 28, 2011), p. 044135. ISSN: 0021-9606. DOI: 10.1063/1.3535541. URL: https://aip.scitation.org/doi/full/10.1063/1.3535541 (visited on 03/20/2023).

[23]   H. -D. Meyer, U. Manthe, and L. S. Cederbaum. "The Multi-Configurational Time-Dependent Hartree Approach". In: *Chemical Physics Letters* 165.1 (Jan. 5, 1990), pp. 73–78. ISSN: 0009-2614. DOI: 10.1016/0009-2614(90)87014-I. URL: https://www.sciencedirect.com/science/article/pii/000926149087014I (visited on 03/08/2023).

[24] Dominic W. Berry et al. "Qubitization of Arbitrary Basis Quantum Chemistry Leveraging Sparsity and Low Rank Factorization". In: *Quantum* 3 (Dec. 2, 2019), p. 208. DOI: 10.22331/q-2019-12-02-208. URL: https://quantum-journal.org/papers/q-2019-12-02-208/ (visited on 12/08/2024).

[25] Matthew Otten et al. "QREChem: Quantum Resource Estimation Software for Chemistry Applications". In: *Frontiers in Quantum Science and Technology* 2 (Nov. 10, 2023). ISSN: 2813-2181. DOI: 10.3389/frqst.2023.1232624. URL: https://www.frontiersin.org/journals/quantum-science-and-technology/articles/10.3389/frqst.2023.1232624/full (visited on 12/08/2024).

[26] Seokmin Shin and Horia Metiu. "Nonadiabatic Effects on the Charge Transfer Rate Constant: A Numerical Study of a Simple Model System". In: *The Journal of Chemical Physics* 102.23 (June 15, 1995), pp. 9285–9295. ISSN: 0021-9606. DOI: 10.1063/1.468795. URL: https://aip.scitation.org/doi/10.1063/1.468795 (visited on 09/08/2021).

[27] Adrián Pérez-Salinas et al. "Data Re-Uploading for a Universal Quantum Classifier". In: *Quantum* 4 (Feb. 6, 2020), p. 226. DOI: 10.22331/q-2020-02-06-226. URL: https://quantum-journal.org/papers/q-2020-02-06-226/ (visited on 03/11/2024).

[28] Tarun Dutta et al. "Single-Qubit Universal Classifier Implemented on an Ion-Trap Quantum Device". In: *Physical Review A* 106.1 (July 7, 2022), p. 012411. DOI: 10.1103/PhysRevA.106.012411. URL: https://link.aps.org/doi/10.1103/PhysRevA.106.012411 (visited on 03/11/2024).

[29] Adrián Pérez-Salinas et al. "One Qubit as a Universal Approximant". In: *Physical Review A* 104.1 (July 2, 2021), p. 012405. DOI: 10.1103/PhysRevA.104.012405. URL: https://link.aps.org/doi/10.1103/PhysRevA.104.012405 (visited on 03/11/2024).

[30] Roy J. Garcia et al. *On the Hardness of Measuring Magic.* Aug. 3, 2024. DOI: 10.48550/arXiv.2408.01663. arXiv: 2408.01663. URL: http://arxiv.org/abs/2408.01663 (visited on 12/08/2024). Pre-published.

[31] Farrokh Vatan and Colin Williams. "Optimal Quantum Circuits for General Two-Qubit Gates". In: *Physical Review A* 69.3 (Mar. 22, 2004), p. 032315. DOI: 10.1103/PhysRevA.69.032315. URL: https://link.aps.org/doi/10.1103/PhysRevA.69.032315 (visited on 12/08/2024).

[32] Christopher M. Dawson and Michael A. Nielsen. *The Solovay-Kitaev Algorithm.* Aug. 23, 2005. DOI: 10.48550/arXiv.quant-ph/0505030. arXiv: quant-ph/0505030. URL: http://arxiv.org/abs/quant-ph/0505030 (visited on 11/26/2024). Pre-published.

[33] Marco Maronese et al. *Quantum Compiling.* Dec. 1, 2021. DOI: 10.48550/arXiv.2112.00187. arXiv: 2112.00187. URL: http://arxiv.org/abs/2112.00187 (visited on 11/11/2024). Pre-published.

[34] J. S. Bell. "On the Einstein Podolsky Rosen Paradox". In: *Physics Physique Fizika* 1.3 (Nov. 1, 1964), pp. 195–200. DOI: 10.1103/PhysicsPhysiqueFizika.1.195. URL: https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195 (visited on 09/10/2024).

[35] W. K. Wootters and W. H. Zurek. "A Single Quantum Cannot Be Cloned". In: *Nature* 299.5886 (Oct. 1982), pp. 802–803. DOI: 10.1038/299802a0. URL: http://www.nature.com/articles/299802a0.

[36] Stephen Wiesner. "Conjugate Coding". In: *ACM SIGACT News* (1983). DOI: 10.1145/1008908.1008920.

[37] Nick Herbert. "FLASH-A Superluminal Communicator Based upon a New Kind of Quantum Measurement". In: *Foundations of Physics* (1982). DOI: 10.1007/BF00729622.

[38]   Howard Barnum et al. "Noncommuting Mixed States Cannot Be Broadcast". In: *Physical Review Letters* 76.15 (Apr. 8, 1996), pp. 2818–2821. DOI: 10.1103/PhysRevLett.76.2818. URL: https://link.aps.org/doi/10.1103/PhysRevLett.76.2818 (visited on 09/10/2024).

[39]   Giacomo Mauro D'Ariano, Chiara Macchiavello, and Paolo Perinotti. "Superbroadcasting of Mixed States". In: *Physical Review Letters* 95.6 (Aug. 5, 2005), p. 060503. DOI: 10.1103/PhysRevLett.95.060503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.060503 (visited on 10/01/2024).

[40]   D. Dieks. "Communication by EPR Devices". In: *Physics Letters A* 92.6 (Nov. 22, 1982), pp. 271–272. ISSN: 0375-9601. DOI: 10.1016/0375-9601(82)90084-6. URL: https://www.sciencedirect.com/science/article/pii/0375960182900846 (visited on 10/01/2024).

[41]   Sergey Bravyi and Alexei Kitaev. "Universal Quantum Computation with Ideal Clifford Gates and Noisy Ancillas". In: *Physical Review A: Atomic, Molecular, and Optical Physics* 71.2 (Feb. 2005), p. 022316. DOI: 10.1103/PhysRevA.71.022316. URL: https://link.aps.org/doi/10.1103/PhysRevA.71.022316.

[42]   M Cerezo et al. "Does Provable Absence of Barren Plateaus Imply Classical Simulability? Or, Why We Need to Rethink Variational Quantum Computing". 2023. arXiv: 2312.09121. URL: https://arxiv.org/abs/2312.09121.

[43]   Martin Larocca et al. "A Review of Barren Plateaus in Variational Quantum Computing". 2024. arXiv: 2405.00781. URL: https://arxiv.org/abs/2405.00781.

[44]   David Deutsch and Roger Penrose. "Quantum Theory, the Church–Turing Principle and the Universal Quantum Computer". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (July 8, 1985), pp. 97–117. DOI: 10.1098/rspa.1985.0070. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1985.0070 (visited on 03/14/2022).

[45]   Peter W Shor. "Scheme for Reducing Decoherence in Quantum Computer Memory". In: *Physical review A* 52.4 (1995), R2493. DOI: 10.1103/PhysRevA.52.R2493.

[46]   Wim van Dam and Yoshitaka Sasaki. *Quantum Algorithms for Problems in Number Theory, Algebraic Geometry, and Group Theory.* June 26, 2012. DOI: 10.48550/arXiv.1206.6126. arXiv: 1206.6126. URL: http://arxiv.org/abs/1206.6126 (visited on 11/08/2024). Pre-published.

[47]   Andrew M. Childs and Wim van Dam. "Quantum Algorithms for Algebraic Problems". In: *Reviews of Modern Physics* 82.1 (Jan. 2010), pp. 1–52. DOI: 10.1103/revmodphys.82.1. URL: https://doi.org/10.1103%2Frevmodphys.82.1.

[48]   Ashley Montanaro. "Quantum Algorithms: An Overview". In: *npj Quantum Information* 2.1 (Jan. 12, 2016), pp. 1–8. ISSN: 2056-6387. DOI: 10.1038/npjqi.2015.23. URL: https://www.nature.com/articles/npjqi201523 (visited on 11/08/2024).

[49]   Alexander M. Dalzell et al. *Quantum Algorithms: A Survey of Applications and End-to-End Complexities.* Oct. 4, 2023. DOI: 10.48550/arXiv.2310.03011. arXiv: 2310.03011. URL: http://arxiv.org/abs/2310.03011 (visited on 11/08/2024). Pre-published.

[50]   Guang Song and Andreas Klappenecker. *The Simplified Toffoli Gate Implementation by Margolus Is Optimal.* Dec. 31, 2003. arXiv: quant-ph/0312225. URL: http://arxiv.org/abs/quant-ph/0312225 (visited on 11/11/2024). Pre-published.

[51]   Lov K Grover. "A Fast Quantum Mechanical Algorithm for Database Search". In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing.* 1996, pp. 212–219. DOI: 10.1145/237814.237866. URL: https://doi.org/10.1145/237814.237866.

[52]   D. Coppersmith. *An Approximate Fourier Transform Useful in Quantum Factoring.* Jan. 16, 2002. DOI: 10.48550/arXiv.quant-ph/0201067. arXiv: quant-ph/0201067. URL: http://arxiv.org/abs/quant-ph/0201067 (visited on 04/23/2024). Pre-published.

[53]   A. Yu Kitaev. *Quantum Measurements and the Abelian Stabilizer Problem.* Nov. 20, 1995. DOI: 10.48550/arXiv.quant-ph/9511026. arXiv: `quant-ph/9511026`. URL: `http://arxiv.org/abs/quant-ph/9511026` (visited on 11/11/2024). Pre-published.

[54]   Berend Klaver et al. *SWAP-less Implementation of Quantum Algorithms.* Aug. 20, 2024. DOI: 10.48550/arXiv.2408.10907. arXiv: `2408.10907`. URL: `http://arxiv.org/abs/2408.10907` (visited on 10/30/2024). Pre-published.

[55]   Alberto Peruzzo et al. "A Variational Eigenvalue Solver on a Photonic Quantum Processor". In: *Nature Communications* 5.1 (1 Dec. 2014), pp. 4213–4213. DOI: `10.1038/ncomms5213`. URL: `http://www.nature.com/articles/ncomms5213`.

[56]   Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A Quantum Approximate Optimization Algorithm". 2014. arXiv: `1411.4028`. URL: `https://arxiv.org/abs/1411.4028`.

[57]   M. Cerezo et al. "Variational Quantum State Eigensolver". In: *npj Quantum Information* 8.1 (2022), pp. 1–11. DOI: `10.1038/s41534-022-00611-6`. URL: `https://doi.org/10.1038/s41534-022-00611-6`.

[58]   Junhan Qin. "Review of Ansatz Designing Techniques for Variational Quantum Algorithms". In: *Journal of Physics: Conference Series* 2634.1 (Nov. 1, 2023), p. 012043. ISSN: 1742-6596. DOI: `10.1088/1742-6596/2634/1/012043`. URL: `https://iopscience.iop.org/article/10.1088/1742-6596/2634/1/012043/meta` (visited on 11/18/2024).

[59]   Yuguo Shao et al. "Simulating Quantum Mean Values in Noisy Variational Quantum Algorithms: A Polynomial-Scale Approach". 2023. arXiv: `2306.05804`. URL: `https://doi.org/10.48550/arXiv.2306.05804`.

[60]   Samson Wang et al. "Noise-Induced Barren Plateaus in Variational Quantum Algorithms". Feb. 8, 2021. arXiv: `2007.14384 [quant-ph]`. URL: `http://arxiv.org/abs/2007.14384` (visited on 09/14/2021).

[61]   Phattharaporn Singkanipa and Daniel A Lidar. "Beyond Unital Noise in Variational Quantum Algorithms: Noise-Induced Barren Plateaus and Fixed Points". 2024. arXiv: `2402.08721`. URL: `https://arxiv.org/abs/2402.08721`.

[62]   Armando Angrisani et al. *Classically Estimating Observables of Noiseless Quantum Circuits.* Sept. 3, 2024. DOI: `10.48550/arXiv.2409.01706`. arXiv: `2409.01706`. URL: `http://arxiv.org/abs/2409.01706` (visited on 11/18/2024). Pre-published.

[63]   David J. Griffiths. *Introduction to Quantum Mechanics.* 2. ed. Pearson International Edition. Upper Saddle River, NJ London: Pearson Prentice Hall, 2005. 468 pp. ISBN: 978-0-13-111892-8 978-0-13-191175-8.

[64]   Cornelius Hempel et al. "Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator". In: (Mar. 2018). DOI: `10.1103/PhysRevX.8.031022`. URL: `http://arxiv.org/abs/1803.10238`.

[65]   Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. "Quantum Autoencoders for Efficient Compression of Quantum Data". In: *Quantum Science and Technology* 2.4 (2017), p. 045001. DOI: `10.1088/2058-9565/aa8072`. URL: `https://iopscience.iop.org/article/10.1088/2058-9565/aa8072`.

[66]   Sergey B. Bravyi and Alexei Yu. Kitaev. "Fermionic Quantum Computation". In: *Annals of Physics* 298.1 (May 2002), pp. 210–226. DOI: `10.1006/APHY.2002.6254`. URL: `https://www.sciencedirect.com/science/article/pii/S0003491602962548`.

[67]   Atharv Joshi, Kyungjoo Noh, and Yvonne Y. Gao. "Quantum Information Processing with Bosonic Qubits in Circuit QED". In: *Quantum Science and Technology* 6.3 (Apr. 2021), p. 033001. ISSN: 2058-9565. DOI: 10.1088/2058-9565/abe989. URL: https://dx.doi.org/10.1088/2058-9565/abe989 (visited on 11/19/2024).

[68]   Thomas E. Baker. "Density Functionals and Kohn-Sham Potentials with Minimal Wavefunction Preparations on a Quantum Computer". In: *Physical Review Research* 2.4 (2020). DOI: 10.1103/PhysRevResearch.2.043238.

[69]   P. J. J. O'Malley et al. "Scalable Quantum Simulation of Molecular Energies". In: *Physical Review X* 6.3 (July 2016), pp. 031007–031007. DOI: 10.1103/PhysRevX.6.031007. URL: https://link.aps.org/doi/10.1103/PhysRevX.6.031007.

[70]   P. Jordan and E. Wigner. "Über Das Paulische Äquivalenzverbot". In: *Zeitschrift für Physik* 47.9 (Sept. 1, 1928), pp. 631–651. ISSN: 0044-3328. DOI: 10.1007/BF01331938. URL: https://doi.org/10.1007/BF01331938.

[71]   Yu-An Chen and Yijia Xu. "Equivalence between Fermion-to-Qubit Mappings in Two Spatial Dimensions". In: *PRX Quantum* 4.1 (Mar. 13, 2023), p. 010326. DOI: 10.1103/PRXQuantum.4.010326. URL: https://link.aps.org/doi/10.1103/PRXQuantum.4.010326 (visited on 11/20/2024).

[72]   Oliver O'Brien. "Ultrafast Hybrid Fermion-to-Qubit Mapping". In: *Physical Review B* 109.11 (2024). DOI: 10.1103/PhysRevB.109.115149.

[73]   Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Courier Corporation, July 2, 1996. 484 pp. ISBN: 978-0-486-69186-2.

[74]   Pranav Gokhale et al. *Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families.* July 31, 2019. DOI: 10.48550/arXiv.1907.13623. arXiv: 1907.13623. URL: http://arxiv.org/abs/1907.13623 (visited on 11/20/2024). Pre-published.

[75]   Abhinav Kandala et al. "Extending the Computational Reach of a Noisy Superconducting Quantum Processor". In: *Nature* 567.7749 (Mar. 2019), pp. 491–495. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-1040-7. arXiv: 1805.04492. URL: http://arxiv.org/abs/1805.04492 (visited on 11/07/2020).

[76]   Jules Tilly et al. "The Variational Quantum Eigensolver: A Review of Methods and Best Practices". In: *Physics Reports* 986 (2022), pp. 1–128. URL: https://www.sciencedirect.com/science/article/pii/S0370157322003118.

[77]   Abhinav Anand et al. "A Quantum Computing View on Unitary Coupled Cluster Theory". Sept. 30, 2021. arXiv: 2109.15176 [physics, physics:quant-ph]. URL: http://arxiv.org/abs/2109.15176 (visited on 01/26/2022).

[78]   Harper R Grimsley et al. "An Adaptive Variational Algorithm for Exact Molecular Simulations on a Quantum Computer". In: *Nature Communications* 10.1 (2019), pp. 1–9. DOI: 10.1038/s41467-019-10988-2. URL: https://www.nature.com/articles/s41467-019-10988-2.

[79]   Ho Lun Tang et al. "Qubit-Adapt-Vqe: An Adaptive Algorithm for Constructing Hardware-Efficient Ansätze on a Quantum Processor". In: *PRX Quantum* 2.2 (2021), p. 020310. DOI: 10.1103/PRXQuantum.2.020310.

[80]   Davide Materia et al. "Quantum Information Driven Ansatz (QIDA): Shallow-Depth Empirical Quantum Circuits from Quantum Chemistry". In: *The Journal of Physical Chemistry A* 128.39 (Oct. 3, 2024), pp. 8533–8543. ISSN: 1089-5639. DOI: 10.1021/acs.jpca.4c03756. URL: https://doi.org/10.1021/acs.jpca.4c03756 (visited on 11/22/2024).

[81]   Qiskit contributors. *Qiskit: An Open-Source Framework for Quantum Computing.* 2023. DOI: 10.5281/zenodo.2573505.

[82] Bryan T. Gard et al. "Efficient Symmetry-Preserving State Preparation Circuits for the Vari-
      ational Quantum Eigensolver Algorithm". In: *npj Quantum Information* 6.1 (1 Jan. 28, 2020),
      pp. 1–9. ISSN: 2056-6387. DOI: 10.1038/s41534-019-0240-1. URL: https://www.nature.com/
      articles/s41534-019-0240-1 (visited on 03/10/2023).

[83] Yuxuan Du et al. "Efficient Measure for the Expressivity of Variational Quantum Algorithms". In:
      Physical Review Letters 128.8 (2022), p. 080506. DOI: 10.1103/PhysRevLett.128.080506. URL:
      https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.128.080506 (visited on
      11/25/2024).

[84] W. J. Hehre, R. F. Stewart, and J. A. Pople. "Self-Consistent Molecular-Orbital Methods. I. Use
      of Gaussian Expansions of Slater-Type Atomic Orbitals". In: *The Journal of Chemical Physics*
      51.6 (Sept. 15, 1969), pp. 2657–2664. ISSN: 0021-9606. DOI: 10.1063/1.1672392. URL: https:
      //doi.org/10.1063/1.1672392 (visited on 11/25/2024).

[85] Caio M. Porto, Rene Alfonso Nome, and Nelson H. Morgon. "The Influence of Basis Sets and
      Ansatze Building to Quantum Computing in Chemistry". In: *Journal of Molecular Modeling* 30.8
      (July 19, 2024), p. 275. ISSN: 0948-5023. DOI: 10.1007/s00894-024-06072-2. URL: https:
      //doi.org/10.1007/s00894-024-06072-2 (visited on 11/25/2024).

[86] Sebastian Ruder. *An Overview of Gradient Descent Optimization Algorithms*. June 15, 2017. DOI:
      10.48550/arXiv.1609.04747. arXiv: 1609.04747. URL: http://arxiv.org/abs/1609.04747
      (visited on 11/25/2024). Pre-published.

[87] Diederik P Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *Proceed-
      ings of the 3rd International Conference on Learning Representations (ICLR)*. 2015. URL: http:
      //arxiv.org/abs/1412.6980.

[88] David Wierichs et al. "General Parameter-Shift Rules for Quantum Gradients". In: *Quantum* 6
      (Mar. 30, 2022), p. 677. DOI: 10.22331/q-2022-03-30-677. URL: https://quantum-journal.
      org/papers/q-2022-03-30-677/ (visited on 02/24/2023).

[89] Tatsuya Hirokami, Yutaka Maeda, and Hiroyuki Tsukada. "Parameter Estimation Using Simulta-
      neous Perturbation Stochastic Approximation". In: *Electrical Engineering in Japan* 154.2 (2006),
      pp. 30–39. ISSN: 1520-6416. DOI: 10.1002/eej.20239. URL: https://onlinelibrary.wiley.
      com/doi/abs/10.1002/eej.20239 (visited on 11/25/2024).

[90] Michael JD Powell. "A Direct Search Optimization Method That Models the Objective and Con-
      straint Functions by Linear Interpolation". In: *Advances in Optimization and Numerical Analysis*.
      Dordrecht: Springer, 1994, pp. 51–67. ISBN: 978-94-015-8330-5. DOI: 10.1007/978-94-015-8330-
      5_4. URL: https://link.springer.com/chapter/10.1007/978-94-015-8330-5_4 (visited on
      11/25/2024).

[91] Harshdeep Singh, Sonjoy Majumder, and Sabyashachi Mishra. "Benchmarking of Different Opti-
      mizers in the Variational Quantum Algorithms for Applications in Quantum Chemistry". In: *The
      Journal of Chemical Physics* 159.4 (July 31, 2023), p. 044117. ISSN: 0021-9606. DOI: 10.1063/5.
      0161057. URL: https://doi.org/10.1063/5.0161057 (visited on 11/25/2024).

[92] Xia Liu et al. "Mitigating Barren Plateaus of Variational Quantum Eigensolvers". In: *IEEE Trans-
      actions on Quantum Engineering* (2024), pp. 1–19. DOI: 10.1109/TQE.2024.3383050. URL:
      https://ieeexplore.ieee.org/abstract/document/10485449.

[93] Marco Schumann, Frank K Wilhelm, and Alessandro Ciani. "Emergence of Noise-Induced Barren
      Plateaus in Arbitrary Layered Noise Models". 2023. arXiv: 2310.08405. URL: https://arxiv.
      org/abs/2310.08405v2.

[94] Eric R Anschuetz and Bobak T Kiani. "Beyond Barren Plateaus: Quantum Variational Algorithms Are Swamped with Traps". In: *Nature Communications* 13.1 (2022), p. 7760. DOI: 10.1038/s41467-022-35364-5. URL: https://doi.org/10.1038/s41467-022-35364-5.

[95] Tameem Albash and Daniel A. Lidar. "Adiabatic Quantum Computation". In: *Reviews of Modern Physics* 90.1 (Jan. 29, 2018), p. 015002. DOI: 10.1103/RevModPhys.90.015002. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.015002 (visited on 11/25/2024).

[96] Elizabeth Crosson et al. *Different Strategies for Optimization Using the Quantum Adiabatic Algorithm.* Jan. 28, 2014. DOI: 10.48550/arXiv.1401.7320. arXiv: 1401.7320. URL: http://arxiv.org/abs/1401.7320 (visited on 11/25/2024). Pre-published.

[97] Kostas Blekos et al. *A Review on Quantum Approximate Optimization Algorithm and Its Variants.* June 26, 2023. DOI: 10.48550/arXiv.2306.09198. arXiv: 2306.09198. URL: http://arxiv.org/abs/2306.09198 (visited on 11/25/2024). Pre-published.

[98] Rebekah Herrman et al. "Multi-Angle Quantum Approximate Optimization Algorithm". In: *Scientific Reports* 12.1 (2022), pp. 1–10. ISSN: 2045-2322. DOI: 10.1038/s41598-022-10555-8. URL: https://www.nature.com/articles/s41598-022-10555-8 (visited on 11/25/2024).

[99] David C. McKay et al. "Efficient Z Gates for Quantum Computing". In: *Physical Review A* 96.2 (Aug. 2017), pp. 022330–022330. DOI: 10.1103/PhysRevA.96.022330. URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.96.022330.

[100] Benjamin Schumacher. "Quantum Coding". In: *PHYSICAL REVIEW A* 51 (1995). URL: https://journals.aps.org/pra/pdf/10.1103/PhysRevA.51.2738.

[101] Daniel Bultrini et al. "Simple Mitigation Strategy for a Systematic Gate Error in IBMQ". In: *Journal of Applied Mathematics and Physics* 9.6 (6 June 9, 2021), pp. 1215–1229. DOI: 10.4236/jamp.2021.96083. URL: https://www.scirp.org/journal/paperinformation?paperid=109835 (visited on 10/24/2024).

[102] Eric Huang, Andrew C. Doherty, and Steven Flammia. "Performance of Quantum Error Correction with Coherent Errors". In: *Physical Review A* 99.2 (Feb. 12, 2019), p. 022313. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.99.022313. URL: https://link.aps.org/doi/10.1103/PhysRevA.99.022313 (visited on 10/24/2024).

[103] Daniel Bultrini et al. *The Battle of Clean and Dirty Qubits in the Era of Partial Error Correction.* July 28, 2022. DOI: 10.48550/arXiv.2205.13454. arXiv: 2205.13454 [quant-ph]. URL: http://arxiv.org/abs/2205.13454 (visited on 05/19/2023). Pre-published.

[104] Nikolaos Koukoulekidis et al. "A Framework of Partial Error Correction for Intermediate-Scale Quantum Computers". 2023. arXiv: 2306.15531. URL: https://arxiv.org/abs/2306.15531.

[105] Zhenyu Cai, Xiaosi Xu, and Simon C. Benjamin. "Mitigating Coherent Noise Using Pauli Conjugation". In: *npj Quantum Information* 6.1 (Feb. 5, 2020), pp. 1–9. ISSN: 2056-6387. DOI: 10.1038/s41534-019-0233-0. URL: https://www.nature.com/articles/s41534-019-0233-0 (visited on 10/24/2024).

[106] Sergey Bravyi et al. "Correcting Coherent Errors with Surface Codes". In: *npj Quantum Information* 4.1 (Oct. 31, 2018), pp. 1–6. ISSN: 2056-6387. DOI: 10.1038/s41534-018-0106-y. URL: https://www.nature.com/articles/s41534-018-0106-y (visited on 10/24/2024).

[107] Karl Kraus et al., eds. *States, Effects, and Operations Fundamental Notions of Quantum Theory.* Vol. 190. Lecture Notes in Physics. Berlin, Heidelberg: Springer, 1983. ISBN: 978-3-540-12732-1 978-3-540-38725-1. DOI: 10.1007/3-540-12732-1. URL: http://link.springer.com/10.1007/3-540-12732-1 (visited on 10/25/2024).

[108]   Sumeet Khatri and Mark M. Wilde. *Principles of Quantum Communication Theory: A Modern Approach*. Feb. 11, 2024. DOI: 10.48550/arXiv.2011.04672. arXiv: 2011.04672. URL: http://arxiv.org/abs/2011.04672 (visited on 10/26/2024). Pre-published.

[109]   Sumeet Khatri et al. "Quantum-Assisted Quantum Compiling". In: *Quantum* 3 (2019), p. 140. DOI: 10.22331/q-2019-05-13-140. URL: https://quantum-journal.org/papers/q-2019-05-13-140/.

[110]   Daniel Bultrini et al. "Unifying and Benchmarking State-of-the-Art Quantum Error Mitigation Techniques". In: *Quantum* 7 (2023), p. 1034. DOI: 10.22331/q-2023-06-06-1034. URL: https://quantum-journal.org/papers/q-2023-06-06-1034.

[111]   Daniel Bultrini et al. "The Battle of Clean and Dirty Qubits in the Era of Partial Error Correction". 2022. arXiv: 2205.13454. URL: https://arxiv.org/abs/2205.13454.

[112]   Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. "Error Mitigation for Short-Depth Quantum Circuits". In: *Physical Review Letters* 119.18 (Nov. 3, 2017), p. 180509. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.119.180509. arXiv: 1612.02058. URL: http://arxiv.org/abs/1612.02058 (visited on 11/07/2020).

[113]   Piotr Czarnik et al. "Error Mitigation with Clifford Quantum-Circuit Data". Feb. 22, 2021. arXiv: 2005.10189 [quant-ph]. URL: http://arxiv.org/abs/2005.10189 (visited on 06/10/2021).

[114]   William J Huggins et al. "Virtual Distillation for Quantum Error Mitigation". In: *Physical Review X* 11.4 (2021), p. 041036. DOI: 10.1103/PhysRevX.11.041036.

[115]   Bálint Koczor. "Exponential Error Suppression for Near-Term Quantum Devices". June 3, 2021. arXiv: 2011.05942 [quant-ph]. URL: http://arxiv.org/abs/2011.05942 (visited on 06/23/2021).

[116]   Zhenyu Cai et al. "Quantum Error Mitigation". In: *Reviews of Modern Physics* 95.4 (2023), p. 045005. DOI: 10.1103/RevModPhys.95.045005. URL: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.95.045005.

[117]   Daniel A Lidar. "Review of Decoherence-Free Subspaces, Noiseless Subsystems, and Dynamical Decoupling". In: *Quantum information and computation for chemistry* (2014), pp. 295–354. DOI: 10.1002/9781118742631. URL: https://doi.org/10.1002/9781118742631.

[118]   Nobuyuki Yoshioka et al. "Generalized Quantum Subspace Expansion". In: *Physical Review Letters* 129.2 (July 2022), p. 020502. DOI: 10.1103/PhysRevLett.129.020502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.129.020502.

[119]   Scott Aaronson. *Quantum Computing, Postselection, and Probabilistic Polynomial-Time*. Dec. 23, 2004. DOI: 10.48550/arXiv.quant-ph/0412187. arXiv: quant-ph/0412187. URL: http://arxiv.org/abs/quant-ph/0412187 (visited on 10/09/2024). Pre-published.

[120]   Alex Linden and Betül Gül. *Optimization of Postselection in Quantum Algorithms: A Two-Way Quantum Computing Approach*. Aug. 30, 2024. DOI: 10.48550/arXiv.2409.03785. arXiv: 2409.03785 [physics]. URL: http://arxiv.org/abs/2409.03785 (visited on 10/09/2024). Pre-published.

[121]   Ewout Van Den Berg et al. "Probabilistic Error Cancellation with Sparse Pauli–Lindblad Models on Noisy Quantum Processors". In: *Nature Physics* (2023), pp. 1–6. DOI: 10.1038/s41567-023-02042-2. URL: https://doi.org/10.1038/s41567-023-02042-2.

[122]   Sergei Filippov et al. *Scalable Tensor-Network Error Mitigation for near-Term Quantum Computing*. arXiv.org. July 21, 2023. URL: https://arxiv.org/abs/2307.11740v2 (visited on 10/09/2024).

[123] Zhenyu Cai. "Quantum Error Mitigation Using Symmetry Expansion". In: *Quantum* 5 (2021), p. 548. DOI: 10.22331/q-2021-09-21-548.

[124] P. Krantz et al. "A Quantum Engineer's Guide to Superconducting Qubits". In: *Applied Physics Reviews* 6.2 (June 17, 2019), p. 021318. ISSN: 1931-9401. DOI: 10.1063/1.5089550. URL: https://doi.org/10.1063/1.5089550 (visited on 10/02/2024).

[125] Changwon Lee and Daniel K. Park. "Scalable Quantum Measurement Error Mitigation via Conditional Independence and Transfer Learning". In: *Machine Learning: Science and Technology* 4.4 (Dec. 2023), p. 045051. ISSN: 2632-2153. DOI: 10.1088/2632-2153/ad1007. URL: https://dx.doi.org/10.1088/2632-2153/ad1007 (visited on 10/08/2024).

[126] Paul D. Nation et al. "Scalable Mitigation of Measurement Errors on Quantum Computers". In: *PRX Quantum* 2.4 (Nov. 2021), p. 040326. DOI: 10.1103/PRXQuantum.2.040326. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.040326.

[127] Adrian Skasberg Aasen et al. "Readout Error Mitigated Quantum State Tomography Tested on Superconducting Qubits". In: *Communications Physics* 7.1 (Sept. 6, 2024), pp. 1–11. ISSN: 2399-3650. DOI: 10.1038/s42005-024-01790-8. URL: https://www.nature.com/articles/s42005-024-01790-8 (visited on 10/08/2024).

[128] Yanzhu Chen et al. "Detector Tomography on IBM 5-Qubit Quantum Computers and Mitigation of Imperfect Measurement". In: (Apr. 2019). DOI: 10.1103/PhysRevA.100.052315. URL: http://arxiv.org/abs/1904.11935.

[129] Morten Kjaergaard et al. *Superconducting Qubits: Current State of Play*. 2020.

[130] Andre He et al. "Zero-Noise Extrapolation for Quantum-Gate Error Mitigation with Identity Insertions". In: *Physical Review A* 102.1 (July 2020), p. 012426. DOI: 10.1103/PhysRevA.102.012426. URL: https://link.aps.org/doi/10.1103/PhysRevA.102.012426.

[131] Lewis Fry Richardson and J. Arthur Gaunt. "VIII. The Deferred Approach to the Limit". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 226.636–646 (Jan. 1927), pp. 299–361. DOI: 10.1098/rsta.1927.0008. URL: https://doi.org/10.1098/rsta.1927.0008.

[132] Angus Lowe et al. "Unified Approach to Data-Driven Quantum Error Mitigation". Nov. 2, 2020. arXiv: 2011.01157 [quant-ph]. URL: http://arxiv.org/abs/2011.01157 (visited on 06/10/2021).

[133] Ritajit Majumdar et al. *Best Practices for Quantum Error Mitigation with Digital Zero-Noise Extrapolation*. July 20, 2023. DOI: 10.48550/arXiv.2307.05203. arXiv: 2307.05203. URL: http://arxiv.org/abs/2307.05203 (visited on 10/10/2024). Pre-published.

[134] Andrea Mari, Nathan Shammah, and William J Zeng. "Extending Quantum Probabilistic Error Cancellation by Noise Scaling". In: *Physical Review A* 104.5 (2021), p. 052607. DOI: 10.1103/PhysRevA.104.052607.

[135] Benjamin McDonough et al. "Automated Quantum Error Mitigation Based on Probabilistic Error Reduction". In: *2022 IEEE/ACM third international workshop on quantum computing software (QCS)* (2022), pp. 83–93. DOI: 10.1109/QCS56647.2022.00015.

[136] Tian-Ren Jin et al. "Purity-Assisted Zero-Noise Extrapolation for Quantum Error Mitigation". In: *Advanced Quantum Technologies* n/a.n/a (), p. 2400150. ISSN: 2511-9044. DOI: 10.1002/qute.202400150. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.202400150 (visited on 10/10/2024).

[137] Ao Chen and Markus Heyl. "Efficient Optimization of Deep Neural Quantum States toward Machine Precision". 2023. DOI: 10.48550/arXiv.2302.01941. arXiv: 2302.01941. URL: https://doi.org/10.48550/arXiv.2302.01941.

[138]   Daniel Gottesman. *The Heisenberg Representation of Quantum Computers*. arXiv.org. July 1, 1998. URL: https://arxiv.org/abs/quant-ph/9807006v1 (visited on 10/05/2024).

[139]   Piotr Czarnik et al. "Improving the Efficiency of Learning-Based Error Mitigation". 2022. arXiv: 2204.07109. URL: https://arxiv.org/abs/2204.07109.

[140]   Bálint Koczor. "The Dominant Eigenvector of a Noisy Quantum State". Apr. 1, 2021. arXiv: 2104.00608 [quant-ph]. URL: http://arxiv.org/abs/2104.00608 (visited on 06/23/2021).

[141]   Ryuji Takagi et al. "Fundamental Limits of Quantum Error Mitigation". In: *npj Quantum Information* 8.1 (2022), p. 114. DOI: 10.1038/s41534-022-00618-z. URL: https://www.nature.com/articles/s41534-022-00618-z.

[142]   Yihui Quek et al. "Exponentially Tighter Bounds on Limitations of Quantum Error Mitigation". In: *Nature Physics* (July 25, 2024), pp. 1–11. ISSN: 1745-2481. DOI: 10.1038/s41567-024-02536-7. URL: https://www.nature.com/articles/s41567-024-02536-7 (visited on 10/09/2024).

[143]   Benjamin L. Brock et al. *Quantum Error Correction of Qudits Beyond Break-even*. arXiv.org. Sept. 23, 2024. URL: https://arxiv.org/abs/2409.15065v1 (visited on 10/02/2024).

[144]   Rajeev Acharya et al. *Quantum Error Correction below the Surface Code Threshold*. Aug. 24, 2024. DOI: 10.48550/arXiv.2408.13687. arXiv: 2408.13687 [quant-ph]. URL: http://arxiv.org/abs/2408.13687 (visited on 10/02/2024). Pre-published.

[145]   Harald Putterman et al. *Hardware-Efficient Quantum Error Correction Using Concatenated Bosonic Qubits*. arXiv.org. Sept. 19, 2024. URL: https://arxiv.org/abs/2409.13025v1 (visited on 10/02/2024).

[146]   Riddhi S. Gupta et al. "Encoding a Magic State with beyond Break-Even Fidelity". In: *Nature* 625.7994 (Jan. 2024), pp. 259–263. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06846-3. URL: https://www.nature.com/articles/s41586-023-06846-3 (visited on 11/20/2024).

[147]   Daniel Gottesman. "An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation". In: *Quantum information science and its contributions to mathematics, Proceedings of Symposia in Applied Mathematics* 63 (2010), pp. 13–58. DOI: 10.1090/psapm/068/2762145. URL: https://arxiv.org/abs/0904.2557.

[148]   Joschka Roffe. "Quantum Error Correction: An Introductory Guide". In: *Contemporary Physics* 60.3 (July 3, 2019), pp. 226–245. ISSN: 0010-7514. DOI: 10.1080/00107514.2019.1667078. URL: https://doi.org/10.1080/00107514.2019.1667078 (visited on 06/14/2021).

[149]   Zijun Chen et al. "Exponential Suppression of Bit or Phase Errors with Cyclic Error Correction". In: *Nature* 595.7867 (7867 July 2021), pp. 383–387. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03588-y. URL: https://www.nature.com/articles/s41586-021-03588-y (visited on 09/07/2021).

[150]   Colin D. Bruzewicz et al. "Trapped-Ion Quantum Computing: Progress and Challenges". In: *Applied Physics Reviews* 6.2 (May 29, 2019), p. 021314. ISSN: 1931-9401. DOI: 10.1063/1.5088164. URL: https://doi.org/10.1063/1.5088164 (visited on 11/26/2024).

[151]   J. von Neumann. "Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components". In: *Automata Studies*. Ed. by C. E. Shannon and J. McCarthy. Princeton University Press, Mar. 2, 2016, pp. 43–98. ISBN: 978-1-4008-8261-8. DOI: 10.1515/9781400882618-003. URL: https://www.degruyter.com/document/doi/10.1515/9781400882618-003/html (visited on 11/26/2024).

[152]   P.W. Shor. "Fault-Tolerant Quantum Computation". In: *Proceedings of 37th Conference on Foundations of Computer Science*. Proceedings of 37th Conference on Foundations of Computer Science. Oct. 1996, pp. 56–65. DOI: 10.1109/SFCS.1996.548464. URL: https://ieeexplore.ieee.org/document/548464 (visited on 11/26/2024).

[153]   Emanuel Knill, Raymond Laflamme, and Wojciech H. Zurek. "Resilient Quantum Computation". In: *Science* 279.5349 (Jan. 16, 1998), pp. 342–345. DOI: 10.1126/science.279.5349.342. URL: https://www.science.org/doi/10.1126/science.279.5349.342 (visited on 11/26/2024).

[154]   Bryan Eastin and Emanuel Knill. "Restrictions on Transversal Encoded Quantum Gate Sets". In: *Physical review letters* 102.11 (2009), p. 110502. DOI: 10.1103/PhysRevLett.102.110502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.110502.

[155]   Scott Aaronson and Daniel Gottesman. "Improved Simulation of Stabilizer Circuits". In: *Physical Review A* 70.5 (2004), p. 052328. DOI: 10.1103/PhysRevA.70.052328. URL: https://link.aps.org/doi/10.1103/PhysRevA.70.052328.

[156]   Daniel S. Abrams and Seth Lloyd. "Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors". In: *Physical Review Letters* 83.24 (Dec. 1999), pp. 5162–5165. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.83.5162. URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.5162 (visited on 12/03/2024).

[157]   Daniel A. Lidar and Haobin Wang. "Calculating the Thermal Rate Constant with Exponential Speedup on a Quantum Computer". In: *Physical Review E* 59.2 (Feb. 1, 1999), pp. 2429–2438. DOI: 10.1103/PhysRevE.59.2429. URL: https://link.aps.org/doi/10.1103/PhysRevE.59.2429 (visited on 12/04/2024).

[158]   L.-A. Wu, M. S. Byrd, and D. A. Lidar. "Polynomial-Time Simulation of Pairing Models on a Quantum Computer". In: *Physical Review Letters* 89.5 (July 16, 2002), p. 057904. DOI: 10.1103/PhysRevLett.89.057904. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.057904 (visited on 12/04/2024).

[159]   B. P. Lanyon et al. "Towards Quantum Chemistry on a Quantum Computer". In: *Nature Chemistry* 2.2 (Feb. 2010), pp. 106–111. ISSN: 1755-4349. DOI: 10.1038/nchem.483. URL: http://www.nature.com/articles/nchem.483 (visited on 12/03/2024).

[160]   Alán Aspuru-Guzik et al. "Simulated Quantum Computation of Molecular Energies". In: *Science* 309.5741 (2005), pp. 1704–1707. DOI: 10.1126/science.1113479. URL: https://science.sciencemag.org/content/309/5741/1704.

[161]   Ivan Kassal et al. "Polynomial-Time Quantum Algorithm for the Simulation of Chemical Dynamics". In: *Proceedings of the National Academy of Sciences* 105.48 (Dec. 2, 2008), pp. 18681–18686. DOI: 10.1073/pnas.0808245105. URL: https://www.pnas.org/doi/10.1073/pnas.0808245105 (visited on 03/29/2022).

[162]   Ivan Kassal et al. "Simulating Chemistry Using Quantum Computers". In: *Annual Review of Physical Chemistry* 62 (Volume 62, 2011 May 5, 2011), pp. 185–207. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev-physchem-032210-103512. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-physchem-032210-103512 (visited on 12/03/2024).

[163]   Dawei Lu et al. "Quantum Chemistry Simulation on Quantum Computers: Theories and Experiments". In: *Physical Chemistry Chemical Physics* 14.26 (June 13, 2012), pp. 9411–9420. ISSN: 1463-9084. DOI: 10.1039/C2CP23700H. URL: https://pubs.rsc.org/en/content/articlelanding/2012/cp/c2cp23700h (visited on 12/03/2024).

[164]   Hefeng Wang, S. Ashhab, and Franco Nori. "Efficient Quantum Algorithm for Preparing Molecular-System-like States on a Quantum Computer". In: *Physical Review A* 79.4 (Apr. 28, 2009), p. 042335. DOI: 10.1103/PhysRevA.79.042335. URL: https://link.aps.org/doi/10.1103/PhysRevA.79.042335 (visited on 12/04/2024).

[165] Hefeng Wang, S. Ashhab, and Franco Nori. "Quantum Algorithm for Simulating the Dynamics of an Open Quantum System". In: *Physical Review A* 83.6 (June 14, 2011), p. 062317. DOI: `10.1103/PhysRevA.83.062317`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.83.062317` (visited on 12/04/2024).

[166] Xiaoyu Xie et al. "Time-Dependent Density Matrix Renormalization Group Quantum Dynamics for Realistic Chemical Systems". In: *The Journal of Chemical Physics* 151.22 (Dec. 14, 2019), p. 224101. ISSN: 0021-9606. DOI: `10.1063/1.5125945`. URL: `https://aip.scitation.org/doi/10.1063/1.5125945` (visited on 05/24/2022).

[167] Stephen R Clark. "Unifying Neural-Network Quantum States and Correlator Product States via Tensor Networks". In: *Journal of Physics A: Mathematical and Theoretical* 51.13 (Apr. 3, 2018), p. 135301. ISSN: 1751-8113, 1751-8121. DOI: `10.1088/1751-8121/aaaaf2`. URL: `https://iopscience.iop.org/article/10.1088/1751-8121/aaaaf2` (visited on 12/04/2024).

[168] John Preskill. "Quantum Computing in the NISQ Era and Beyond". In: *Quantum* 2 (Jan. 2018), pp. 79–79. DOI: `10.22331/q-2018-08-06-79`. URL: `http://arxiv.org/abs/1801.00862`.

[169] Jared D. Weidman et al. "Quantum Computing and Chemistry". In: *Cell Reports Physical Science* 5.9 (Sept. 18, 2024), p. 102105. ISSN: 2666-3864. DOI: `10.1016/j.xcrp.2024.102105`. URL: `https://www.sciencedirect.com/science/article/pii/S2666386424003837` (visited on 12/03/2024).

[170] Gautam Kumar et al. "Recent Advances in Quantum Computing for Drug Discovery and Development". In: *IEEE Access* 12 (2024), pp. 64491–64509. ISSN: 2169-3536. DOI: `10.1109/ACCESS.2024.3376408`. URL: `https://ieeexplore.ieee.org/abstract/document/10466774` (visited on 12/03/2024).

[171] Seunghoon Lee et al. "Evaluating the Evidence for Exponential Quantum Advantage in Ground-State Quantum Chemistry". In: *Nature Communications* 14.1 (2023), p. 1952. DOI: `10.1038/s41467-023-37587-6`. URL: `https://www.nature.com/articles/s41467-023-37587-6`.

[172] Mario Motta and Julia Rice. "Emerging Quantum Computing Algorithms for Quantum Chemistry". Sept. 7, 2021. DOI: `10.1002/wcms.1580`. arXiv: `2109.02873 [quant-ph]`. URL: `http://arxiv.org/abs/2109.02873` (visited on 09/10/2021).

[173] Xiaoxiao Xiao et al. "Physics-Constrained Hardware-Efficient Ansatz on Quantum Computers That Is Universal, Systematically Improvable, and Size-Consistent". In: *Journal of Chemical Theory and Computation* 20.5 (Mar. 12, 2024), pp. 1912–1922. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.3c00966`. URL: `https://doi.org/10.1021/acs.jctc.3c00966` (visited on 12/03/2024).

[174] Panagiotis G Anastasiou et al. "TETRIS-ADAPT-VQE: An Adaptive Algorithm That Yields Shallower, Denser Circuit Ansätze". In: *Physical Review Research* 6.1 (2024), p. 013254. DOI: `10.1103/PhysRevResearch.6.013254`. URL: `https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.6.013254`.

[175] Nonia Vaquero-Sabater et al. "Physically Motivated Improvements of Variational Quantum Eigensolvers". In: *Journal of Chemical Theory and Computation* 20.12 (June 25, 2024), pp. 5133–5144. ISSN: 1549-9618. DOI: `10.1021/acs.jctc.4c00329`. URL: `https://doi.org/10.1021/acs.jctc.4c00329` (visited on 11/22/2024).

[176] Oscar Higgott, Daochen Wang, and Stephen Brierley. "Variational Quantum Computation of Excited States". In: *Quantum* 3 (2019), p. 156. DOI: `10.22331/q-2019-07-01-156`. URL: `https://quantum-journal.org/papers/q-2019-07-01-156/`.

[177] Lila Cadi Tazi and Alex J. W. Thom. "Folded Spectrum VQE: A Quantum Computing Method for the Calculation of Molecular Excited States". In: *Journal of Chemical Theory and Computation* 20.6 (Mar. 26, 2024), pp. 2491–2504. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.3c01378. URL: https://doi.org/10.1021/acs.jctc.3c01378 (visited on 12/03/2024).

[178] Eduarda Sangiogo Gil et al. "SHARC Meets TEQUILA: Mixed Quantum-Classical Dynamics on a Quantum Computer Using a Hybrid Quantum-Classical Algorithm". In: *Chemical Science* (Nov. 28, 2024). ISSN: 2041-6539. DOI: 10.1039/D4SC04987J. URL: https://pubs.rsc.org/en/content/articlelanding/2024/sc/d4sc04987j (visited on 12/04/2024).

[179] Stefano Barison et al. "Variational Dynamics as a Ground-State Problem on a Quantum Computer". In: *Physical Review Research* 4.4 (2022), p. 043161. URL: https://doi.org/10.1103/PhysRevResearch.4.043161.

[180] Daniel Bultrini and Oriol Vendrell. *Towards Mixed Quantum-Classical Dynamics on Quantum Computers*. Mar. 20, 2023. DOI: 10.48550/arXiv.2303.11375. arXiv: 2303.11375 [physics, physics:quant-ph]. URL: http://arxiv.org/abs/2303.11375 (visited on 05/19/2023). Pre-published.

[181] Sebastian Krinner et al. "Realizing Repeated Quantum Error Correction in a Distance-Three Surface Code". Dec. 7, 2021. arXiv: 2112.03708 [cond-mat, physics:quant-ph]. URL: http://arxiv.org/abs/2112.03708 (visited on 03/04/2022).

[182] Samson Wang et al. "Can Error Mitigation Improve Trainability of Noisy Variational Quantum Algorithms?" In: *Quantum* 8 (2024), p. 1287. DOI: 10.22331/q-2024-03-14-1287. URL: https://quantum-journal.org/papers/q-2024-03-14-1287/.

[183] Iulia M Georgescu, Sahel Ashhab, and Franco Nori. "Quantum Simulation". In: *Reviews of Modern Physics* 86.1 (2014), p. 153. DOI: 10.1103/RevModPhys.86.153. URL: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.86.153.

[184] Trygve Helgaker et al. "Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations". In: *Chemical Reviews* 112.1 (Jan. 11, 2012), pp. 543–631. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr2002239. URL: https://pubs.acs.org/doi/10.1021/cr2002239 (visited on 12/03/2024).

[185] Nikolaj Moll et al. "Optimizing Qubit Resources for Quantum Chemistry Simulations in Second Quantization on a Quantum Computer". In: *Journal of Physics A: Mathematical and Theoretical* 49.29 (July 2016), pp. 295301–295301. DOI: 10.1088/1751-8113/49/29/295301. URL: http://stacks.iop.org/1751-8121/49/i=29/a=295301?key=crossref.42584c866176798a2347025a2d912fce.

[186] J. I. Colless et al. "Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm". In: *Physical Review X* 8.1 (Feb. 2018), pp. 011021–011021. DOI: 10.1103/PhysRevX.8.011021. URL: https://link.aps.org/doi/10.1103/PhysRevX.8.011021.

[187] Jacob T. Seeley, Martin J. Richard, and Peter J. Love. "The Bravyi-Kitaev Transformation for Quantum Computation of Electronic Structure". In: *The Journal of Chemical Physics* 137.22 (Dec. 2012), pp. 224109–224109. DOI: 10.1063/1.4768229. URL: http://aip.scitation.org/doi/10.1063/1.4768229.

[188] Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and Quantum Computation.* 47. American Mathematical Soc., 2002.

[189] Andrew Tranter et al. "A Comparison of the Bravyi–Kitaev and Jordan–Wigner Transformations for the Quantum Simulation of Quantum Chemistry". In: *Journal of Chemical Theory and Computation* 14.11 (Nov. 13, 2018), pp. 5617–5630. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.8b00450. URL: https://doi.org/10.1021/acs.jctc.8b00450 (visited on 12/03/2024).

[190]   Peter M. Fenwick. "A New Data Structure for Cumulative Frequency Tables". In: *Software: Practice and Experience* 24.3 (1994), pp. 327–336. ISSN: 1097-024X. DOI: 10.1002/spe.4380240306. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.4380240306 (visited on 12/03/2024).

[191]   Mitchell Chiew and Sergii Strelchuk. "Optimal Fermion-Qubit Mappings". Oct. 25, 2021. arXiv: 2110.12792 [quant-ph]. URL: http://arxiv.org/abs/2110.12792 (visited on 01/26/2022).

[192]   Bochen Tan and Jason Cong. "Optimal Qubit Mapping with Simultaneous Gate Absorption". Sept. 14, 2021. arXiv: 2109.06445 [quant-ph]. URL: http://arxiv.org/abs/2109.06445 (visited on 09/15/2021).

[193]   John C. Light and Tucker Carrington Jr. "Discrete-Variable Representations and Their Utilization". In: *Advances in Chemical Physics.* John Wiley & Sons, Ltd, 2000, pp. 263–310. ISBN: 978-0-470-14173-1. DOI: 10.1002/9780470141731.ch4. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470141731.ch4 (visited on 03/25/2022).

[194]   Yuan Su et al. "Fault-Tolerant Quantum Simulations of Chemistry in First Quantization". In: (May 28, 2021). URL: https://scirate.com/arxiv/2105.12767 (visited on 05/30/2022).

[195]   Dominic W. Berry et al. "Improved Techniques for Preparing Eigenstates of Fermionic Hamiltonians". In: *npj Quantum Information* 4.1 (May 2, 2018), pp. 1–7. ISSN: 2056-6387. DOI: 10.1038/s41534-018-0071-5. URL: https://www.nature.com/articles/s41534-018-0071-5 (visited on 12/05/2024).

[196]   Benedikt Fauseweh. "Quantum Many-Body Simulations on Digital Quantum Computers: State-of-the-art and Future Challenges". In: *Nature Communications* 15.1 (Mar. 8, 2024), p. 2123. ISSN: 2041-1723. DOI: 10.1038/s41467-024-46402-9. URL: https://www.nature.com/articles/s41467-024-46402-9 (visited on 12/03/2024).

[197]   Stefano Barison, Javier Robledo Moreno, and Mario Motta. *Quantum-Centric Computation of Molecular Excited States with Extended Sample-Based Quantum Diagonalization.* Nov. 1, 2024. DOI: 10.48550/arXiv.2411.00468. arXiv: 2411.00468. URL: http://arxiv.org/abs/2411.00468 (visited on 12/05/2024). Pre-published.

[198]   Keita Kanno et al. *Quantum-Selected Configuration Interaction: Classical Diagonalization of Hamiltonians in Subspaces Selected by Quantum Computers.* Feb. 22, 2023. DOI: 10.48550/arXiv.2302.11320. arXiv: 2302.11320. URL: http://arxiv.org/abs/2302.11320 (visited on 12/05/2024). Pre-published.

[199]   Dominic W. Berry et al. "Efficient Quantum Algorithms for Simulating Sparse Hamiltonians". In: *Communications in Mathematical Physics* 270.2 (Mar. 1, 2007), pp. 359–371. ISSN: 1432-0916. DOI: 10.1007/s00220-006-0150-x. URL: https://doi.org/10.1007/s00220-006-0150-x (visited on 07/24/2023).

[200]   César R. De Oliveira. *Intermediate Spectral Theory and Quantum Dynamics.* Basel: Birkhäuser, 2009. ISBN: 978-3-7643-8794-5 978-3-7643-8795-2. DOI: 10.1007/978-3-7643-8795-2. URL: http://link.springer.com/10.1007/978-3-7643-8795-2 (visited on 12/06/2024).

[201]   James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. "Simulation of Electronic Structure Hamiltonians Using Quantum Computers". In: *Molecular Physics* 109.5 (Mar. 2011), pp. 735–750. DOI: 10.1080/00268976.2011.552441. URL: http://www.tandfonline.com/doi/abs/10.1080/00268976.2011.552441.

[202]   Masuo Suzuki. "Decomposition Formulas of Exponential Operators and Lie Exponentials with Some Applications to Quantum Mechanics and Statistical Physics". In: *Journal of mathematical physics* 26.4 (1985), pp. 601–612.

[203]   Andrew Tranter et al. "Ordering of Trotterization: Impact on Errors in Quantum Simulation of Electronic Structure". In: *Entropy* 21.12 (12 Dec. 2019), p. 1218. ISSN: 1099-4300. DOI: 10.3390/e21121218. URL: https://www.mdpi.com/1099-4300/21/12/1218 (visited on 12/06/2024).

[204]   Conor Mc Keever and Michael Lubasch. "Classically Optimized Hamiltonian Simulation". May 23, 2022. DOI: 10.48550/arXiv.2205.11427. arXiv: 2205.11427 [quant-ph]. URL: http://arxiv.org/abs/2205.11427 (visited on 05/30/2022).

[205]   Earl Campbell. "Random Compiler for Fast Hamiltonian Simulation". In: *Physical Review Letters* 123.7 (Aug. 2019), p. 070503. DOI: 10.1103/PhysRevLett.123.070503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.070503.

[206]   Kouhei Nakaji, Mohsen Bagherimehrab, and Alán Aspuru-Guzik. "High-Order Randomized Compiler for Hamiltonian Simulation". In: *PRX Quantum* 5.2 (May 8, 2024), p. 020330. DOI: 10.1103/PRXQuantum.5.020330. URL: https://link.aps.org/doi/10.1103/PRXQuantum.5.020330 (visited on 12/06/2024).

[207]   Yong-Xin Yao et al. "Adaptive Variational Quantum Dynamics Simulations". In: *PRX Quantum* 2.3 (July 12, 2021), p. 030307. DOI: 10.1103/PRXQuantum.2.030307. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.030307 (visited on 04/25/2022).

[208]   Cristina Cîrstoiu et al. "Variational Fast Forwarding for Quantum Simulation beyond the Coherence Time". In: *npj Quantum Information* 6.1 (1 Sept. 18, 2020), pp. 1–10. ISSN: 2056-6387. DOI: 10.1038/s41534-020-00302-0. URL: https://www.nature.com/articles/s41534-020-00302-0 (visited on 04/25/2022).

[209]   Benjamin Commeau et al. "Variational Hamiltonian Diagonalization for Dynamical Quantum Simulation". Sept. 5, 2020. arXiv: 2009.02559 [quant-ph]. URL: http://arxiv.org/abs/2009.02559 (visited on 07/21/2021).

[210]   Konstantinos D Vogiatzis et al. "Pushing Configuration-Interaction to the Limit: Towards Massively Parallel MCSCF Calculations". In: *Journal of Chemical Physics* 147.18 (2017). ISSN: 9788578110796. DOI: 10.1063/1.4989858. URL: https://arxiv.org/pdf/1707.04346.pdf.

[211]   Yangchao Shen et al. "Quantum Implementation of the Unitary Coupled Cluster for Simulating Molecular Electronic Structure". In: *Physical Review A* 95.2 (Feb. 2017), pp. 020501–020501. DOI: 10.1103/PhysRevA.95.020501. URL: https://link.aps.org/doi/10.1103/PhysRevA.95.020501.

[212]   Jarrod R. McClean et al. "Exploiting Locality in Quantum Computation for Quantum Chemistry". In: *The Journal of Physical Chemistry Letters* 5.24 (Dec. 2014), pp. 4368–4380. DOI: 10.1021/jz501649m. URL: http://pubs.acs.org/doi/10.1021/jz501649m.

[213]   Youngseok Kim et al. "Evidence for the Utility of Quantum Computing before Fault Tolerance". In: *Nature* 618.7965 (2023), pp. 500–505. DOI: 10.1038/s41586-023-06096-3. URL: https://www.nature.com/articles/s41586-023-06096-3.

[214]   Laurin E. Fischer et al. *Dynamical Simulations of Many-Body Quantum Chaos on a Quantum Computer*. Nov. 1, 2024. DOI: 10.48550/arXiv.2411.00765. arXiv: 2411.00765. URL: http://arxiv.org/abs/2411.00765 (visited on 12/06/2024). Pre-published.

[215]   Uwe Manthe. "A Multilayer Multiconfigurational Time-Dependent Hartree Approach for Quantum Dynamics on General Potential Energy Surfaces". In: *The Journal of Chemical Physics* 128.16 (Apr. 28, 2008), p. 164116. ISSN: 0021-9606. DOI: 10.1063/1.2902982. URL: https://aip.scitation.org/doi/full/10.1063/1.2902982 (visited on 03/20/2023).

[216]   M. Cerezo et al. "Variational Quantum Algorithms". Dec. 16, 2020. arXiv: 2012.09265 [quant-ph, stat]. URL: http://arxiv.org/abs/2012.09265 (visited on 09/10/2021).

[217]   Pauline J. Ollitrault, Alexander Miessen, and Ivano Tavernelli. "Molecular Quantum Dynamics: A Quantum Computing Perspective". In: *Accounts of Chemical Research* (Nov. 17, 2021). ISSN: 0001-4842. DOI: 10.1021/acs.accounts.1c00514. URL: https://doi.org/10.1021/acs.accounts.1c00514 (visited on 12/01/2021).

[218]   Basile F. E. Curchod and Todd J. Martínez. "Ab Initio Nonadiabatic Quantum Molecular Dynamics". In: *Chemical Reviews* 118.7 (Apr. 11, 2018), pp. 3305–3336. ISSN: 0009-2665. DOI: 10.1021/acs.chemrev.7b00423. URL: https://doi.org/10.1021/acs.chemrev.7b00423 (visited on 03/09/2023).

[219]   Adam Kirrander and Morgane Vacher. "Ehrenfest Methods for Electron and Nuclear Dynamics". In: *Quantum Chemistry and Dynamics of Excited States*. John Wiley & Sons, Ltd, 2020, pp. 469–497. ISBN: 978-1-119-41777-4. DOI: 10.1002/9781119417774.ch15. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119417774.ch15 (visited on 03/08/2023).

[220]   Pauline J. Ollitrault, Guglielmo Mazzola, and Ivano Tavernelli. "Nonadiabatic Molecular Quantum Dynamics with Quantum Computers". In: *Physical Review Letters* 125.26 (Dec. 31, 2020), p. 260511. DOI: 10.1103/PhysRevLett.125.260511. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.260511 (visited on 09/06/2021).

[221]   Igor O. Sokolov et al. "Microcanonical and Finite-Temperature Ab Initio Molecular Dynamics Simulations on Quantum Computers". In: *Physical Review Research* 3.1 (Feb. 10, 2021), p. 013125. DOI: 10.1103/PhysRevResearch.3.013125. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.3.013125 (visited on 09/06/2021).

[222]   Max Rossmannek et al. "Quantum HF/DFT-Embedding Algorithms for Electronic Structure Calculations: Scaling up to Complex Molecular Systems". Sept. 3, 2020. arXiv: 2009.01872 [physics, physics:quant-ph]. URL: http://arxiv.org/abs/2009.01872 (visited on 11/06/2020).

[223]   Daniel S. Levine et al. "CASSCF with Extremely Large Active Spaces Using the Adaptive Sampling Configuration Interaction Method". In: *Journal of Chemical Theory and Computation* 16.4 (Apr. 14, 2020), pp. 2340–2354. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.9b01255. URL: https://doi.org/10.1021/acs.jctc.9b01255 (visited on 07/24/2023).

[224]   Sheng-Hsuan Lin et al. "Real- and Imaginary-Time Evolution with Compressed Quantum Circuits". In: *PRX Quantum* 2.1 (Mar. 15, 2021), p. 010342. DOI: 10.1103/PRXQuantum.2.010342. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.010342 (visited on 01/25/2023).

[225]   Stefano Barison, Filippo Vicentini, and Giuseppe Carleo. "An Efficient Quantum Algorithm for the Time Evolution of Parameterized Circuits". In: *Quantum* 5 (July 28, 2021), p. 512. DOI: 10.22331/q-2021-07-28-512. URL: https://quantum-journal.org/papers/q-2021-07-28-512/ (visited on 01/25/2023).

[226]   Noah F. Berthusen et al. "Quantum Dynamics Simulations beyond the Coherence Time on Noisy Intermediate-Scale Quantum Hardware by Variational Trotter Compression". In: *Physical Review Research* 4.2 (May 4, 2022), p. 023097. DOI: 10.1103/PhysRevResearch.4.023097. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.023097 (visited on 01/25/2023).

[227]   Guillermo Albareda et al. "Universal Steps in Quantum Dynamics with Time-Dependent Potential-Energy Surfaces: Beyond the Born-Oppenheimer Picture". In: *Physical Review A* 94.6 (Dec. 22, 2016), p. 062511. DOI: 10.1103/PhysRevA.94.062511. URL: https://link.aps.org/doi/10.1103/PhysRevA.94.062511 (visited on 12/09/2022).

[228] M. Erdmann, P. Marquetand, and V. Engel. "Combined Electronic and Nuclear Dynamics in a Simple Model System". In: *The Journal of Chemical Physics* 119.2 (July 8, 2003), pp. 672–679. ISSN: 0021-9606. DOI: 10.1063/1.1578618. URL: https://aip.scitation.org/doi/abs/10.1063/1.1578618 (visited on 12/08/2022).

[229] Mirjam Falge et al. "Quantum Wave-Packet Dynamics in Spin-Coupled Vibronic States". In: *The Journal of Physical Chemistry A* 116.46 (Nov. 26, 2012), pp. 11427–11433. ISSN: 1089-5639. DOI: 10.1021/jp306566x. URL: https://doi.org/10.1021/jp306566x (visited on 12/08/2022).

[230] Graeme H. Gossel, Lionel Lacombe, and Neepa T. Maitra. "On the Numerical Solution of the Exact Factorization Equations". In: *The Journal of Chemical Physics* 150.15 (Apr. 21, 2019), p. 154112. ISSN: 0021-9606. DOI: 10.1063/1.5090802. URL: https://aip.scitation.org/doi/full/10.1063/1.5090802 (visited on 09/20/2021).

[231] Kishor Bharti et al. "Noisy Intermediate-Scale Quantum (NISQ) Algorithms". Jan. 21, 2021. arXiv: 2101.08448 [cond-mat, physics:quant-ph]. URL: http://arxiv.org/abs/2101.08448 (visited on 09/10/2021).

[232] Johannes Flick et al. "Cavity Born–Oppenheimer Approximation for Correlated Electron–Nuclear-Photon Systems". In: *Journal of Chemical Theory and Computation* 13.4 (Apr. 11, 2017), pp. 1616–1625. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b01126. URL: https://doi.org/10.1021/acs.jctc.6b01126 (visited on 12/08/2022).

[233] William C. Swope et al. "A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters". In: *The Journal of Chemical Physics* 76.1 (Jan. 1982), pp. 637–649. ISSN: 0021-9606. DOI: 10.1063/1.442716. URL: https://aip.scitation.org/doi/10.1063/1.442716 (visited on 03/08/2023).

[234] Guang Hao Low and Isaac L. Chuang. "Hamiltonian Simulation by Qubitization". In: *Quantum* 3 (July 12, 2019), p. 163. DOI: 10.22331/q-2019-07-12-163. URL: https://quantum-journal.org/papers/q-2019-07-12-163/ (visited on 02/20/2023).

[235] Yosi Atia and Dorit Aharonov. "Fast-Forwarding of Hamiltonians and Exponentially Precise Measurements". In: *Nature Communications* 8.1 (1 Nov. 17, 2017), p. 1572. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01637-7. URL: https://www.nature.com/articles/s41467-017-01637-7 (visited on 04/26/2022).

[236] Andrew M. Childs and Robin Kothari. "Limitations on the Simulation of Non-Sparse Hamiltonians". In: *Quantum Information and Computation* 10 (7&8). ISSN: 15337146, 15337146. DOI: 10.26421/QIC10.7-8. arXiv: 0908.4398 [quant-ph]. URL: http://arxiv.org/abs/0908.4398 (visited on 07/24/2023).

[237] Péter Rakyta and Zoltán Zimborás. *Efficient Quantum Gate Decomposition via Adaptive Circuit Compression*. Nov. 15, 2022. DOI: 10.48550/arXiv.2203.04426. arXiv: 2203.04426 [quant-ph]. URL: http://arxiv.org/abs/2203.04426 (visited on 03/06/2023). Pre-published.

[238] Efekan Kökcü et al. "Algebraic Compression of Quantum Circuits for Hamiltonian Evolution". In: *Physical Review A* 105.3 (Mar. 10, 2022), p. 032420. DOI: 10.1103/PhysRevA.105.032420. URL: https://link.aps.org/doi/10.1103/PhysRevA.105.032420 (visited on 03/06/2023).

[239] Dmitry A. Fedorov et al. "VQE Method: A Short Survey and Recent Developments". In: *Materials Theory* 6.1 (Jan. 6, 2022), p. 2. ISSN: 2509-8012. DOI: 10.1186/s41313-021-00032-6. URL: https://doi.org/10.1186/s41313-021-00032-6 (visited on 04/07/2022).

[240]   Chee-Kong Lee et al. "Variational Quantum Simulation of Chemical Dynamics with Quantum Computers". In: *Journal of Chemical Theory and Computation* (Mar. 16, 2022). ISSN: 1549-9618. DOI: 10.1021/acs.jctc.1c01176. URL: https://doi.org/10.1021/acs.jctc.1c01176 (visited on 03/29/2022).

[241]   Daniel T. Colbert and William H. Miller. "A Novel Discrete Variable Representation for Quantum Mechanical Reactive Scattering via the S-matrix Kohn Method". In: *The Journal of Chemical Physics* 96.3 (Feb. 1992), pp. 1982–1991. ISSN: 0021-9606. DOI: 10.1063/1.462100. URL: https://aip.scitation.org/doi/10.1063/1.462100 (visited on 03/28/2022).

[242]   Pauline J. Ollitrault et al. "Quantum Algorithms for Grid-Based Variational Time Evolution". Mar. 4, 2022. arXiv: 2203.02521 [quant-ph]. URL: http://arxiv.org/abs/2203.02521 (visited on 03/16/2022).

[243]   Mark R. Hermann and J. A. Fleck. "Split-Operator Spectral Method for Solving the Time-Dependent Schr\"odinger Equation in Spherical Coordinates". In: *Physical Review A* 38.12 (Dec. 1, 1988), pp. 6000–6012. DOI: 10.1103/PhysRevA.38.6000. URL: https://link.aps.org/doi/10.1103/PhysRevA.38.6000 (visited on 02/15/2023).

[244]   Otfried Gühne et al. "Toolbox for Entanglement Detection and Fidelity Estimation". In: *Physical Review A* 76.3 (Sept. 18, 2007), p. 030305. DOI: 10.1103/PhysRevA.76.030305. URL: https://link.aps.org/doi/10.1103/PhysRevA.76.030305 (visited on 07/25/2023).

[245]   Amir Kalev and Itay Hen. "Quantum Algorithm for Simulating Hamiltonian Dynamics with an Off-diagonal Series Expansion". In: *Quantum* 5 (Apr. 8, 2021), p. 426. DOI: 10.22331/q-2021-04-08-426. URL: https://quantum-journal.org/papers/q-2021-04-08-426/ (visited on 02/20/2023).

[246]   Hefeng Wang and Hua Xiang. "A Quantum Eigensolver for Symmetric Tridiagonal Matrices". In: *Quantum Information Processing* 18 (Mar. 1, 2019), p. 93. DOI: 10.1007/s11128-019-2211-z. URL: https://ui.adsabs.harvard.edu/abs/2019QuIP...18...93W (visited on 07/25/2023).

[247]   Herbert Robbins and Sutton Monro. "A Stochastic Approximation Method". In: *The Annals of Mathematical Statistics* 22.3 (Sept. 1951), pp. 400–407. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729586. URL: http://projecteuclid.org/euclid.aoms/1177729586 (visited on 07/25/2023).

[248]   Shigeki Gocho et al. "Excited State Calculations Using Variational Quantum Eigensolver with Spin-Restricted Ansätze and Automatically-Adjusted Constraints". In: *npj Computational Materials* 9.1 (1 Jan. 21, 2023), pp. 1–9. ISSN: 2057-3960. DOI: 10.1038/s41524-023-00965-1. URL: https://www.nature.com/articles/s41524-023-00965-1 (visited on 03/08/2023).

[249]   Jarrod R. McClean et al. "Hybrid Quantum-Classical Hierarchy for Mitigation of Decoherence and Determination of Excited States". In: *Physical Review A* 95.4 (Apr. 2017), pp. 042308–042308. DOI: 10.1103/PhysRevA.95.042308. URL: http://link.aps.org/doi/10.1103/PhysRevA.95.042308.

[250]   Joseph C. Aulicino, Trevor Keen, and Bo Peng. "State Preparation and Evolution in Quantum Computing: A Perspective from Hamiltonian Moments". In: *International Journal of Quantum Chemistry* 122.5 (2022), e26853. ISSN: 1097-461X. DOI: 10.1002/qua.26853. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.26853 (visited on 03/20/2023).

[251]   Gadi Aleksandrowicz et al. *Qiskit: An Open-source Framework for Quantum Computing.* Zenodo, Jan. 23, 2019. DOI: 10.5281/zenodo.2562111. URL: https://zenodo.org/record/2562111 (visited on 07/25/2023).

[252]   Gavin E. Crooks. *Gradients of Parameterized Quantum Gates Using the Parameter-Shift Rule and Gate Decomposition.* May 30, 2019. DOI: 10.48550/arXiv.1905.13311. arXiv: 1905.13311 [quant-ph]. URL: http://arxiv.org/abs/1905.13311 (visited on 02/24/2023). Pre-published.

[253]   Charles R. Harris et al. "Array Programming with NumPy". In: *Nature* 585.7825 (7825 Sept. 2020), pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: https://www.nature.com/articles/s41586-020-2649-2 (visited on 07/25/2023).

[254]   Kohdai Kuroiwa et al. *Quantum Car-Parrinello Molecular Dynamics: A Cost-Efficient Molecular Simulation Method on Near-Term Quantum Computers.* Dec. 22, 2022. DOI: 10.48550/arXiv.2212.11921. arXiv: 2212.11921 [quant-ph]. URL: http://arxiv.org/abs/2212.11921 (visited on 07/24/2023). Pre-published.

[255]   Utkarsh Azad and Harjinder Singh. "Quantum Chemistry Calculations Using Energy Derivatives on Quantum Computers". In: *Chemical Physics* 558 (June 1, 2022), p. 111506. ISSN: 0301-0104. DOI: 10.1016/j.chemphys.2022.111506. URL: https://www.sciencedirect.com/science/article/pii/S0301010422000611 (visited on 07/24/2023).

[256]   Jack Ceroni et al. *Tailgating Quantum Circuits for High-Order Energy Derivatives.* July 22, 2022. DOI: 10.48550/arXiv.2207.11274. arXiv: 2207.11274 [quant-ph]. URL: http://arxiv.org/abs/2207.11274 (visited on 07/24/2023). Pre-published.

[257]   Thomas E. O'Brien et al. "Efficient Quantum Computation of Molecular Forces and Other Energy Gradients". Dec. 16, 2021. arXiv: 2111.12437 [quant-ph]. URL: http://arxiv.org/abs/2111.12437 (visited on 01/26/2022).

[258]   Thomas E. O'Brien et al. "Calculating Energy Derivatives for Quantum Chemistry on a Quantum Computer". In: *npj Quantum Information* 5.1 (1 Dec. 11, 2019), pp. 1–12. ISSN: 2056-6387. DOI: 10.1038/s41534-019-0213-4. URL: https://www.nature.com/articles/s41534-019-0213-4 (visited on 09/06/2021).

[259]   Ryan Babbush et al. "Chemical Basis of Trotter-Suzuki Errors in Quantum Chemistry Simulation". In: *Physical Review A* 91.2 (Feb. 2015), pp. 022311–022311. DOI: 10.1103/PhysRevA.91.022311. URL: https://link.aps.org/doi/10.1103/PhysRevA.91.022311.

[260]   H. F. Trotter. "On the Product of Semi-Groups of Operators". In: *Proceedings of the American Mathematical Society* 10.4 (Aug. 1959), pp. 545–545. DOI: 10.2307/2033649. JSTOR: 2033649. URL: https://www.jstor.org/stable/2033649?origin=crossref.

[261]   Justin Kalloor et al. "Quantum Hardware Roofline: Evaluating the Impact of Gate Expressivity on Quantum Processor Design". In: (Feb. 2024).

[262]   Daniel Stilck França and Raul Garcia-Patron. "Limitations of Optimization Algorithms on Noisy Quantum Devices". In: *Nature Physics* 17.11 (2021), pp. 1221–1227. DOI: 10.1038/s41567-021-01356-3.

[263]   Matthew PA Fisher et al. "Random Quantum Circuits". In: *Annual Review of Condensed Matter Physics* 14 (2023), pp. 335–379. DOI: 10.1146/annurev-conmatphys-031720-030658. URL: https://www.annualreviews.org/doi/10.1146/annurev-conmatphys-031720-030658.

[264]   Anupam Mitra et al. "Robust M\o{}lmer-S\o{}rensen Gate for Neutral Atoms Using Rapid Adiabatic Rydberg Dressing". In: *Physical Review A* 101.3 (Mar. 20, 2020), p. 030301. DOI: 10.1103/PhysRevA.101.030301. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.030301 (visited on 10/14/2024).

[265]   Jarrod R McClean et al. "Barren Plateaus in Quantum Neural Network Training Landscapes". In: *Nature Communications* 9.1 (2018), pp. 1–6. DOI: 10.1038/s41467-018-07090-4. URL: https://doi.org/10.1038/s41467-018-07090-4.

[266] Stuart Hadfield et al. "From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz". In: *Algorithms* 12.2 (2019), p. 34. DOI: 10.3390/a12020034. URL: https://www.mdpi.com/1999-4893/12/2/34.

[267] P Erdos and A Renyi. "On Random Graphs I". In: *Universitatis Debreceniensis* 6.290–297 (1959), p. 18. URL: http://snap.stanford.edu/class/cs224w-readings/erdos59random.pdf.

[268] Colin J Trout et al. "Simulating the Performance of a Distance-3 Surface Code in a Linear Ion Trap". In: *New Journal of Physics* 20.4 (2018), p. 043038. DOI: 10.1088/1367-2630/aab341. URL: https://doi.org/10.1088/1367-2630/aab341.

[269] Lukasz Cincio et al. "Machine Learning of Noise-Resilient Quantum Circuits". In: *PRX Quantum* 2.1 (Feb. 15, 2021), p. 010324. DOI: 10.1103/PRXQuantum.2.010324. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.010324 (visited on 07/02/2021).

[270] Elijah Pelofske, Georg Hahn, and Hristo N. Djidjev. "Noise Dynamics of Quantum Annealers: Estimating the Effective Noise Using Idle Qubits". In: *Quantum Science and Technology* 8.3 (Apr. 2023), p. 035005. ISSN: 2058-9565. DOI: 10.1088/2058-9565/accbe6. URL: https://dx.doi.org/10.1088/2058-9565/accbe6 (visited on 10/14/2024).

[271] Kathleen E Hamilton et al. "Scalable Quantum Processor Noise Characterization". In: *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2020, pp. 430–440. DOI: 10.1109/QCE49297.2020.00060. URL: https://doi.org/10.1109/QCE49297.2020.00060.

[272] Sergey Bravyi et al. "Mitigating Measurement Errors in Multiqubit Experiments". In: *Physical Review A* 103.4 (2021), p. 042605. DOI: 10.1103/PhysRevA.103.042605. URL: https://doi.org/10.1103/PhysRevA.103.042605.

[273] Piotr Czarnik et al. "Qubit-Efficient Exponential Suppression of Errors". Mar. 24, 2021. arXiv: 2102.06056 [quant-ph]. URL: http://arxiv.org/abs/2102.06056 (visited on 06/10/2021).

[274] Akel Hashim et al. "Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor". In: *Physical Review X* 11.4 (Nov. 24, 2021), p. 041039. DOI: 10.1103/PhysRevX.11.041039. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.041039 (visited on 10/02/2024).

[275] Lukasz Cincio et al. "Learning the Quantum Algorithm for State Overlap". In: *New Journal of Physics* 20.11 (Nov. 15, 2018), p. 113022. ISSN: 1367-2630. DOI: 10.1088/1367-2630/aae94a. arXiv: 1803.04114. URL: http://arxiv.org/abs/1803.04114 (visited on 11/23/2020).

[276] Prakash Murali et al. "Noise-Adaptive Compiler Mappings for Noisy Intermediate-Scale Quantum Computers". In: Asplos '19 (2019), pp. 1015–1029. DOI: 10.1145/3297858.3304075. URL: https://doi.org/10.1145/3297858.3304075.

[277] Kunal Sharma et al. "Noise Resilience of Variational Quantum Compiling". In: *New Journal of Physics* 22.4 (2020), p. 043006. DOI: 10.1088/1367-2630/ab784c. URL: https://iopscience.iop.org/article/10.1088/1367-2630/ab784c.

[278] Ryuji Takagi. "Optimal Resource Cost for Error Mitigation". In: *Phys. Rev. Res.* 3.3 (Aug. 2021), p. 033178. DOI: 10.1103/PhysRevResearch.3.033178. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.3.033178.

[279] Thomas E. O'Brien et al. "Error Mitigation via Verified Phase Estimation". In: *PRX Quantum* 2.2 (May 2021), p. 020317. DOI: 10.1103/PRXQuantum.2.020317. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.020317.

[280] Kun Wang, Yu-Ao Chen, and Xin Wang. "Mitigating Quantum Errors via Truncated Neumann Series". In: *Science China Information Sciences* 66.8 (2023), p. 180508. DOI: 10.1007/s11432-023-3786-1. URL: https://link.springer.com/article/10.1007/s11432-023-3786-1.

[281] Sam McArdle, Xiao Yuan, and Simon Benjamin. "Error-Mitigated Digital Quantum Simulation". In: *Physical Review Letters* 122.18 (May 2019), p. 180501. DOI: 10.1103/PhysRevLett.122.180501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.180501.

[282] Xavi Bonet-Monroig et al. "Low-Cost Error Mitigation by Symmetry Verification". In: *Physical Review A* 98.6 (2018), p. 062339. DOI: 10.1103/PhysRevA.98.062339. URL: https://doi.org/10.1103/PhysRevA.98.062339.

[283] Matthew Otten, Cristian L Cortes, and Stephen K Gray. "Noise-Resilient Quantum Dynamics Using Symmetry-Preserving Ansatzes". 2019. arXiv: 1910.06284. URL: https://arxiv.org/abs/1910.06284.

[284] Yutaro Akahoshi et al. "Partially Fault-Tolerant Quantum Computing Architecture with Error-Corrected Clifford Gates and Space-Time Efficient Analog Rotations". In: *PRX Quantum* 5.1 (Mar. 5, 2024), p. 010337. DOI: 10.1103/PRXQuantum.5.010337. URL: https://link.aps.org/doi/10.1103/PRXQuantum.5.010337 (visited on 11/07/2024).

[285] Kishor Bharti et al. "Noisy Intermediate-Scale Quantum Algorithms". In: *Reviews of Modern Physics* 94.1 (2022), p. 015004. DOI: 10.1103/RevModPhys.94.015004. URL: https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.94.015004.

[286] Dorit Aharonov et al. "Limitations of Noisy Reversible Computation". 1996. arXiv: quant-ph/9611028. URL: https://arxiv.org/abs/quant-ph/9611028.

[287] Michael Ben-Or, Daniel Gottesman, and Avinatan Hassidim. "Quantum Refrigerator". 2013. arXiv: 1301.1995. URL: https://arxiv.org/abs/1301.1995.

[288] M. Cerezo et al. "Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits". In: *Nature Communications* 12.1 (Dec. 2021), p. 1791. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21728-w. arXiv: 2001.00550. URL: http://arxiv.org/abs/2001.00550 (visited on 07/01/2021).

[289] Andrew Arrasmith et al. "Equivalence of Quantum Barren Plateaus to Cost Concentration and Narrow Gorges". In: *Quantum Science and Technology* 7.4 (2022), p. 045015. DOI: 10.1088/2058-9565/ac7d06. URL: https://doi.org/10.1088/2058-9565/ac7d06.

[290] Andrew Arrasmith et al. "Effect of Barren Plateaus on Gradient-Free Optimization". In: *Quantum* 5 (2021), p. 558. DOI: 10.22331/q-2021-10-05-558. URL: https://quantum-journal.org/papers/q-2021-10-05-558/.

[291] M. Cerezo and Patrick J Coles. "Higher Order Derivatives of Quantum Neural Networks with Barren Plateaus". In: *Quantum Science and Technology* 6.2 (2021), p. 035006. DOI: 10.1088/2058-9565/abf51a. URL: https://iopscience.iop.org/article/10.1088/2058-9565/abf51a.

[292] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. "Entanglement-Induced Barren Plateaus". In: *PRX Quantum* 2.4 (2021), p. 040316. DOI: 10.1103/PRXQuantum.2.040316. URL: https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.040316.

[293] Martin Larocca et al. "Diagnosing Barren Plateaus with Tools from Quantum Optimal Control". May 29, 2021. arXiv: 2105.14377 [quant-ph]. URL: http://arxiv.org/abs/2105.14377 (visited on 07/01/2021).

[294] Zoë Holmes et al. "Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus". In: *PRX Quantum* 3.1 (Jan. 2022), p. 010313. DOI: 10.1103/PRXQuantum.3.010313. URL: https://doi.org/10.1103/PRXQuantum.3.010313.

[295] Ningping Cao et al. "NISQ: Error Correction, Mitigation, and Noise Simulation". 2021. arXiv: 2111.02345. URL: https://arxiv.org/abs/2111.02345.

[296]   Adam Holmes et al. "NISQ+: Boosting Quantum Computing Power by Approximating Quantum Error Correction". In: *Proceedings of the ACM/IEEE 47th Annual International Symposium on Computer Architecture*. ISCA '20. Virtual Event: IEEE Press, May 30, 2020, pp. 556–569. ISBN: 978-1-72814-661-4. DOI: `10.1109/ISCA45697.2020.00053`. URL: `https://doi.org/10.1109/ISCA45697.2020.00053` (visited on 05/23/2022).

[297]   Yasunari Suzuki et al. "Quantum Error Mitigation as a Universal Error Reduction Technique: Applications from the NISQ to the Fault-Tolerant Quantum Computing Eras". In: *PRX Quantum* 3.1 (Mar. 18, 2022), p. 010345. DOI: `10.1103/PRXQuantum.3.010345`. URL: `https://link.aps.org/doi/10.1103/PRXQuantum.3.010345` (visited on 05/23/2022).

[298]   Emanuel Knill and Raymond Laflamme. "Power of One Bit of Quantum Information". In: *Physical Review Letters* 81.25 (1998), p. 5672. DOI: `10.1103/PhysRevLett.81.5672`. URL: `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.81.5672`.

[299]   Keisuke Fujii et al. "Power of Quantum Computation with Few Clean Qubits". 2016. DOI: `10.4230/LIPIcs.ICALP.2016.13`. arXiv: `1509.07276 [quant-ph]`. URL: `http://arxiv.org/abs/1509.07276` (visited on 07/30/2021).

[300]   Tomoyuki Morimae, Keisuke Fujii, and Harumichi Nishimura. "Power of One Non-Clean Qubit". In: *Physical Review A* 95.4 (Apr. 25, 2017), p. 042336. ISSN: 2469-9926, 2469-9934. DOI: `10.1103/PhysRevA.95.042336`. arXiv: `1610.07244`. URL: `http://arxiv.org/abs/1610.07244` (visited on 07/30/2021).

[301]   Keisuke Fujii et al. "Impossibility of Classically Simulating One-Clean-Qubit Model with Multiplicative Error". In: *Physical Review Letters* 120.20 (May 17, 2018), p. 200502. DOI: `10.1103/PhysRevLett.120.200502`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.120.200502` (visited on 07/30/2021).

[302]   Craig Gidney. "QFactoring with N+2 Clean Qubits and N-1 Dirty Qubits". 2017. arXiv: `1706.07884`. URL: `https://arxiv.org/abs/706.07884`.

[303]   Anirban N. Chowdhury, Rolando D. Somma, and Yiğit Subaşı. "Computing Partition Functions in the One-Clean-Qubit Model". In: *Physical Review A* 103.3 (Mar. 17, 2021), p. 032422. DOI: `10.1103/PhysRevA.103.032422`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.103.032422` (visited on 07/30/2021).

[304]   Maria Schuld et al. "Evaluating Analytic Gradients on Quantum Hardware". In: *Physical Review A* 99.3 (2019), p. 032331. DOI: `10.1103/PhysRevA.99.032331`. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.99.032331`.

[305]   Daniel Gottesman and Isaac L. Chuang. "Demonstrating the Viability of Universal Quantum Computation Using Teleportation and Single-Qubit Operations". In: *Nature* 402.6760 (Nov. 1999), pp. 390–393. DOI: `10.1038/46503`. URL: `https://www.nature.com/articles/46503`.

[306]   Laird Egan et al. "Fault-Tolerant Operation of a Quantum Error-Correction Code". Jan. 7, 2021. arXiv: `2009.11482 [quant-ph]`. URL: `http://arxiv.org/abs/2009.11482` (visited on 09/10/2021).

[307]   Laird Egan et al. "Fault-Tolerant Control of an Error-Corrected Qubit". In: *Nature* 598.7880 (2021), pp. 281–286. DOI: `10.1038/s41586-021-03928-y`. URL: `https://www.nature.com/articles/s41586-021-03928-y`.

[308]   Ben W. Reichardt et al. *Demonstration of Quantum Computation and Error Correction with a Tesseract Code*. Sept. 6, 2024. URL: `https://arxiv.org/abs/2409.04628v1` (visited on 10/02/2024). Pre-published.

[309] Austin G. Fowler et al. "Surface Codes: Towards Practical Large-Scale Quantum Computation". In: *Physical Review A - Atomic, Molecular, and Optical Physics* 86.3 (Sept. 2012), pp. 032324–032324. DOI: `10.1103/PhysRevA.86.032324`. URL: `https://journals.aps.org/pra/abstract/10.1103/PhysRevA.86.032324`.

[310] Martin Suchara et al. *Comparing the Overhead of Topological and Concatenated Quantum Error Correction*. Dec. 9, 2013. DOI: `10.48550/arXiv.1312.2316`. arXiv: `1312.2316`. URL: `http://arxiv.org/abs/1312.2316` (visited on 10/28/2024). Pre-published.

[311] Rolando D Somma. "Quantum Eigenvalue Estimation via Time Series Analysis". In: *New Journal of Physics* 21.12 (2019), p. 123025. DOI: `10.1088/1367-2630/ab5c60`. URL: `https://iopscience.iop.org/article/10.1088/1367-2630/ab5c60`.

[312] Vojtěch Havlíček et al. "Supervised Learning with Quantum-Enhanced Feature Spaces". In: *Nature* 567.7747 (2019), pp. 209–212. DOI: `10.1038/s41586-019-0980-2`. URL: `https://www.nature.com/articles/s41586-019-0980-2`.

[313] Andrew G Taube and Rodney J Bartlett. "New Perspectives on Unitary Coupled-Cluster Theory". In: *International journal of quantum chemistry* 106.15 (2006), pp. 3393–3401. DOI: `10.1002/qua.21198`. URL: `https://onlinelibrary.wiley.com/doi/full/10.1002/qua.21198`.

[314] Chae-Yeun Park and Nathan Killoran. *Hamiltonian Variational Ansatz without Barren Plateaus*. Jan. 24, 2024. DOI: `10.48550/arXiv.2302.08529`. arXiv: `2302.08529`. URL: `http://arxiv.org/abs/2302.08529` (visited on 10/31/2024). Pre-published.

[315] Hsin-Yuan Huang et al. "Quantum Advantage in Learning from Experiments". In: *Science* 376.6598 (2022), pp. 1182–1186. DOI: `10.1126/science.abn7293`. URL: `https://www.science.org/doi/10.1126/science.abn7293`.

[316] Sergey Danilin, Nicholas Nugent, and Martin Weides. "Quantum Sensing with Tuneable Superconducting Qubits: Optimization and Speed-Up". 2022. arXiv: `2211.08344`. URL: `https://arxiv.org/abs/2211.08344`.

[317] Nikolai Lauk et al. "Perspectives on Quantum Transduction". In: *Quantum Science and Technology* 5.2 (Mar. 2020), p. 020501. DOI: `10.1088/2058-9565/ab788a`. URL: `https://dx.doi.org/10.1088/2058-9565/ab788a`.

[318] Easwar Magesan et al. "Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking". In: *Physical review letters* 109.8 (2012), p. 080505. DOI: `10.1103/PhysRevLett.109.080505`. URL: `https://doi.org/10.1103/PhysRevLett.109.080505`.

[319] Alicja Dutkiewicz et al. *Error Mitigation and Circuit Division for Early Fault-Tolerant Quantum Phase Estimation*. Oct. 7, 2024. DOI: `10.48550/arXiv.2410.05369`. arXiv: `2410.05369`. URL: `http://arxiv.org/abs/2410.05369` (visited on 10/10/2024). Pre-published.

[320] Christophe Piveteau and David Sutter. "Circuit Knitting With Classical Communication". In: *IEEE Transactions on Information Theory* 70.4 (Apr. 2024), pp. 2734–2745. ISSN: 1557-9654. DOI: `10.1109/TIT.2023.3310797`. URL: `https://ieeexplore.ieee.org/document/10236453` (visited on 11/07/2024).

[321] Clare Horsman et al. "Surface Code Quantum Computing by Lattice Surgery". In: *New Journal of Physics* 14.12 (2012), p. 123011. ISSN: 1367-2630. DOI: `10.1088/1367-2630/14/12/123011`. URL: `https://dx.doi.org/10.1088/1367-2630/14/12/123011` (visited on 11/07/2024).

[322] Riki Toshio et al. *Practical Quantum Advantage on Partially Fault-Tolerant Quantum Computer*. Aug. 27, 2024. DOI: `10.48550/arXiv.2408.14848`. arXiv: `2408.14848`. URL: `http://arxiv.org/abs/2408.14848` (visited on 11/07/2024). Pre-published.

[323]   Arkajit Mandal, Michael A. D. Taylor, and Pengfei Huo. "Theory for Cavity-Modified Ground-State Reactivities via Electron–Photon Interactions". In: *The Journal of Physical Chemistry A* 127.32 (Aug. 17, 2023), pp. 6830–6841. ISSN: 1089-5639. DOI: 10.1021/acs.jpca.3c01421. URL: https://doi.org/10.1021/acs.jpca.3c01421 (visited on 12/06/2024).

[324]   Karl Michael Ziems et al. "Nuclear–Electron Correlation Effects and Their Photoelectron Imprint in Molecular XUV Ionisation". In: *Frontiers in Chemistry* 10 (Aug. 5, 2022). ISSN: 2296-2646. DOI: 10.3389/fchem.2022.942633. URL: https://www.frontiersin.org/journals/chemistry/articles/10.3389/fchem.2022.942633/full (visited on 12/06/2024).

[325]   Jong-Kwon Ha and Ryan J. MacDonell. *Analog Quantum Simulation of Coupled Electron-Nuclear Dynamics in Molecules.* Version 1. Sept. 6, 2024. DOI: 10.48550/arXiv.2409.04427. arXiv: 2409.04427. URL: http://arxiv.org/abs/2409.04427 (visited on 12/06/2024). Pre-published.

[326]   Robin Tyburski et al. "Proton-Coupled Electron Transfer Guidelines, Fair and Square". In: *Journal of the American Chemical Society* 143.2 (Jan. 20, 2021), pp. 560–576. ISSN: 0002-7863. DOI: 10.1021/jacs.0c09106. URL: https://doi.org/10.1021/jacs.0c09106 (visited on 12/06/2024).

[327]   Noah B. Lewis et al. "A Molecular-Level Mechanistic Framework for Interfacial Proton-Coupled Electron Transfer Kinetics". In: *Nature Chemistry* 16.3 (Mar. 2024), pp. 343–352. ISSN: 1755-4349. DOI: 10.1038/s41557-023-01400-0. URL: https://www.nature.com/articles/s41557-023-01400-0 (visited on 12/06/2024).

[328]   *IBM Quantum Has Achieved Its Highest Quantum Volume yet | IBM Quantum Computing Blog.* URL: https://www.ibm.com/quantum/blog/quantum-volume-256 (visited on 12/07/2024).

[329]   F. Pfeiffer et al. "Efficient Decoupling of a Nonlinear Qubit Mode from Its Environment". In: *Physical Review X* 14.4 (Oct. 8, 2024), p. 041007. DOI: 10.1103/PhysRevX.14.041007. URL: https://link.aps.org/doi/10.1103/PhysRevX.14.041007 (visited on 12/07/2024).

[330]   Niall F. Robertson et al. *Approximate Quantum Compiling for Quantum Simulation: A Tensor Network Based Approach.* June 15, 2024. DOI: 10.48550/arXiv.2301.08609. arXiv: 2301.08609. URL: http://arxiv.org/abs/2301.08609 (visited on 12/07/2024). Pre-published.

[331]   Weixiao Sun et al. "Sudden Death of Quantum Advantage in Correlation Generations". In: *Science Advances* 10.47 (Nov. 22, 2024), eadr5002. DOI: 10.1126/sciadv.adr5002. URL: https://www.science.org/doi/10.1126/sciadv.adr5002 (visited on 12/07/2024).

[332]   Ewin Tang. "A Quantum-Inspired Classical Algorithm for Recommendation Systems". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing.* 2019, pp. 217–228. DOI: 10.1145/3313276.3316310. URL: https://dl.acm.org/doi/10.1145/3313276.3316310.

[333]   Josh Abramson et al. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3". In: *Nature* 630.8016 (June 2024), pp. 493–500. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07487-w. URL: https://www.nature.com/articles/s41586-024-07487-w (visited on 12/07/2024).

Acronyms

BK Bravyi-Kitaev. 44, 73, 76

BOA Born-Oppenheimer approximation. 72, 78

BOPES Born-Oppenheimer potential energy surfaces. 85, 93

CDR Clifford data regression. 57–59, 63, 64, 105, 106, 109, 116, 121

CPTP completely positive, trace preserving. 53

CSWAP controlled swap. 106, 111, 114, 115

FCI full configuration interaction. 71, 79

GPU graphical processing unit. 20, 21

HVA Hamiltonian variational ansatz. 132, 133

JW Jordan-Wigner. 44, 73, 75, 76

MCTDH multi-configurational time-dependent Hartree. 77

MQC mixed quantum-classical. 83, 102, 103

NAMD non-adiabatic molecular dynamics. 84

NIBP noise-induced barren plateau. 126, 132–134

NISQ noisy intermediate-scale Quantum. 39, 41, 43, 46, 50, 57, 63, 65, 69–72, 83, 84, 87, 89, 97, 98, 102, 105, 121, 126, 152

NMR nuclear magnetic resonance. 22, 70

p-VQD projected variational quantum dynamics. 84, 86, 87, 90, 91, 97, 100, 102, 103, 165

PEC probabilistic error cancellation. 59, 123

POVM positive operator-valued measure. 60

QAOA quantum approximate optimization algorithm. 41, 42, 47, 57, 105, 112, 115, 119, 121

QC quantum computer. 20, 21, 32, 34, 39, 41, 50, 57, 59, 61, 70, 71, 77, 121, 155

QEC quantum error correction. 58, 61, 67, 68, 125, 126, 128

QEM quantum error mitigation. 43, 57–59, 61, 63, 66, 104–106, 109–111, 114, 116–119, 121, 122, 126, 137

QFT quantum Fourier transform. 40, 41

GESAMTFAKULTÄT FÜR MATHEMATIK,
INGENIEUR- UND NATURWISSENSCHAFTEN

COMBINED FACULTY OF MATHEMATICS,
ENGINEERING AND NATURAL SCIENCES

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften der Universität Heidelberg / Sworn Affidavit according to § 8 of the doctoral degree regulations of the Combined Faculty of Mathematics, Engineering and Natural Sciences at the Heidelberg University

1. Bei der eingereichten Dissertation zu dem Thema / The thesis I have submitted entitled

   ## Quantum chemistry on quantum computers

   handelt es sich um meine eigenständig erbrachte Leistung / is my own work.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. / I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

3. Die Arbeit oder Teile davon habe ich wie folgt/bislang nicht[1] an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. / I have not yet/have already[1] presented this thesis or parts thereof to a university as part of an examination or degree.

   Titel der Arbeit / Title of the thesis: Quantum chemistry on quantum computers

   Hochschule und Jahr / University and year: **University of Heidelberg, 2025**

   Art der Prüfungs- oder Qualifikationsleistung / Type of examination or degree: **Doctorate**

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. / I confirm that the declarations made above are correct.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. / I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe. / I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.

25.01.2025 Heidelberg

....
Ort und Datum / Place and date

....
Unterschrift / Signature

[1] Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung. / Please cross out what is not applicable. If applicable, please provide: the title of the thesis that was presented elsewhere, the name of the university, the year of presentation and the type of examination or degree.

The German text is legally binding.