A Framework for Approximate Quantum Circuit Optimisation

Owen Agnel

University of Oxford

A thesis submitted for the degree of

Masters in Computer Science and Philosophy

Trinity 2024

Word count: 9,656

Statement of Originality

This thesis and the work to which it refers are the results of my own efforts, unless indicated otherwise. Any ideas, proofs, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification.

Acknowledgements

I would like to thank my supervisors Professor Thomas Melham and Professor Aleks Kissinger for all their help and advice with this master's thesis. I'm also indebted to Professor Noah Linden who provided the invaluable proof structure for Theorem 3.2.5.

I'm immensely grateful to my brother and parents for all the encouragement I received during the completion of this thesis.

Finally, I'd like to thank my partner Minshu Gupta and friends Edward Yarrow, Alex Ogleby, Alyssa Stein, Jan Gradon, Thomas Feerick, Natalia Puczek, Temea Turjaka, and Ghizlane Slaoui and for their unwavering friendship and support.

Abstract

We are currently in the noisy intermediate-scale quantum (NISQ) era, where the practical implementation of most quantum algorithms remains infeasible due to hardware limitations. Overcoming these challenges necessitates, in part, the effective optimisation of quantum circuits. This thesis reports on an extensive study of approximate Quantum Circuit Optimisation (QCO). Our contributions include the development of a robust mathematical framework established through systematic investigation, resulting in a physically motivated metric for quantifying the distance between circuits. Additionally, we offer a novel proof of an unproven result concerning the diamond norm, yielding a markedly more efficient algorithm for computing diamond distances. Using these theoretical advancements, we conduct an empirical investigation of the space of approximate circuit identities. Our findings suggest that phase-squashing up to π emerges as the predominant strategy for local approximate circuit compilation, providing guidance for future research endeavours and technological advancements in this domain.

Contents

Chapter 1

Introduction

Quantum computing is a revolutionary technology that holds the potential to bring about disruptive progress in a wide range of scientific research and technology-related domains [1]. From fundamental results in physics [2] and chemistry [3] to decryption [4], database search [5], and optimisation problems [6].

Building large-scale fault-tolerant quantum computers is a challenging task that requires isolating a large number of qubits, enabling fine-grain control over their states, and performing highly accurate measurements [7]. Despite considerable advances in the development of these devices, with quantum hardware reaching a state where it can compete with classical supercomputers in specific tasks [7, 8, 9], we remain in the noisy intermediate-scale quantum (NISQ [10]) era of quantum computing. Available devices are relatively small and prone to decoherence noise resulting from undesirable environmental interactions. The resource requirements of many quantum algorithms make their implementation infeasible in the short term.

To overcome these limitations and achieve quantum advantage, the research community has proposed several strategies [11]. One approach is the design of novel *noise resilient* hybrid quantum-classical algorithms that can deliver valuable results on NISQ devices [12, 13]. Another strategy involves designing efficient quantum circuits tailored to specific devices [14]. As each gate in a quantum circuit introduces

error, a general heuristic in circuit design is to minimise the circuit depth [15]. The ability to construct shallow circuits using hardware-specific gate sets constitutes an essential step for running quantum algorithms on NISQ devices [15].

The general problem of generating quantum circuits is referred to as *quantum circuit synthesis* (QCS) [14]. We divide circuit synthesis into two more specific problems: *quantum circuit compilation* (QCC) and *optimisation* (QCO). QCC constructs circuits that meet a set of constraints, such as device connectivity and native gate sets, using a high-level description [16, 17]. On the other hand, QCO takes an existing circuit and produces a new one that minimises the impact of hardware error, usually by reducing depth or T-count [18, 19]. Although there is a distinction between these two problems, methods can be adapted from one to the other.

In this thesis, we concentrate on the problem of QCO, and more specifically on *approximate* quantum circuit optimisation. Unlike *exact* QCO, which aims to find logically equivalent circuits maximising a heuristic, *approximate* QCO relaxes the problem and aims to produce circuits which are merely close to the target with respect to a distance metric.

To understand this subtlety, it's important to distinguish two types of error. On the one hand, quantum devices are subject to decoherence noise and gate infidelities, leading to inaccuracies that worsen with circuit depth; we call this *hardware error*. On the other hand, in approximate QCO, the circuits we construct are not logically equivalent to the input circuits. This also results in inaccuracy; we call this approximation error. Approximate QCO aims to achieve a favourable trade-off between these errors [20].

Significant progress has been made in the field of exact QCO in the past two decades [18] resulting in methods such as T-par [19, 21], TOpt [22], T-Optimizer [23], QCO based on ZX calculus [24, 25], and ML methods [18, 26, 14]. Many of these methods

are based on the compression and parallelisation of quantum gates, often using hard-coded rewrite rules [18].

The problem of approximate QCO has not received the same level of attention. Adapted from approximate QCC, a general strategy has been to construct a parametrised circuit (ansatz) and tune the gate parameters to minimise the distance between the output and target unitaries [27]. As these techniques operate globally, we will call them *global* optimisation methods. Depending on the hardware used, we can further classify these techniques into classically-assisted and quantum-assisted [28, 29].

Both strategies face difficulties. Classically assisted methods quickly become infeasible as the numbers of qubits grow [28, 1] and many distance metrics are intractable to compute classically [29]. Quantum-assisted methods, on the other hand, face noise-related convergence issues [30] even when the approach is noise resilient [31]. Indeed, both techniques are vulnerable to barren plateaus in the optimisation landscapes. These are large regions of the parameter space where the cost function gradient vanishes exponentially [28, 32, 33]. This phenomenon makes training an ansatz infeasible as the number of qubits and ansatz expressibility grow.

As global optimisation methods for approximate QCO face several limitations, local optimisation strategies are incentivised. We propose a hybrid approach to approximate QCO, identifying approximate *local rewrite rules*, which can be combined with existing exact QCO libraries to maximise a heuristic.

Although relatively unexplored, this approach has been employed previously in limited applications. Approximations of quantum algorithms relying on local rewrites, such as the approximate quantum Fourier transform (AQFT), have been studied extensively in the literature [34, 35, 20, 36, 37]. In the case of the AQFT, an approximate rewrite rule called *phase-squashing* is used to reduce the circuit depth: single qubit

gates implementing small rotations are deleted, resulting in better performance in the presence of noise [35, 20].

A thorough exploration of the space of approximate circuit identities and rewrite rules beyond phase-squashing is notably lacking. This thesis reports on an extensive study of the feasibility of *local approximate QCO*. The main contributions of this work include:

- The elaboration of a rigorous mathematical framework to quantify the distance between two circuits and assess the quality of optimised circuits.
- A novel proof of an unproved result in [38] resulting in a substantially more efficient algorithm for calculating the diamond distance of two unitary channels.
- Development of a novel algorithm for computing the diamond distance between two unitary channels, exhibiting significant speedup compared to existing methods. We contribute our implementation to popular quantum computing libraries such as QuTiP [39] and Qiskit [40] (pending future release).
- Substantial empirical exploration of the space of approximate circuit equivalences corroborating previous theoretical results and leading to practical insights on the feasibility of local approximate QCO.

These contributions collectively serve to examine the feasibility of approximate QCO and open new avenues for future research.

Chapter 2

Background

This chapter briefly introduces quantum information theory (QIT), including the mathematical preliminaries required to elaborate our framework for approximate QCO. Notation differs substantially from author to author, and definitions do not always coincide. In this chapter, we set out the relevant notation and concepts we will use throughout this thesis. Section 2.1 will focus on the linear algebra needed to describe quantum systems. This will be followed by Section 2.2 briefly describing the standard model for quantum computing. We end with Section 2.3 introducing operator and superoperator norms.

2.1 Mathematical Preliminaries

The most common mathematical formulation of QIT (see [41] for an alternative) relies heavily on linear algebra to describe the state of systems and the operations which can be performed on these systems. This section reviews the concepts in linear algebra that are relevant to QIT. Much of the notation is borrowed from the book "Quantum Information Theory" by John Watrous [42].

2.1.1 Euclidean Spaces

The most fundamental mathematical object in QIT is the *complex Euclidean space*. Quantum states, measurements, observables, and operations can all be expressed in mathematical terms related to these spaces. As such, they form the bedrock of QIT.

A complex Euclidean space $\mathcal{X}$ is a finite-dimensional complex vector space with an *inner product* satisfying three properties: *conjugate symmetry*, *linearity*, and *positive definiteness*. In general, we will refer to these spaces using scripted capital letters. Vectors belonging to these spaces are written $\mathbf{u}$ or $|u\rangle$ interchangeably. The Bra-Ket notation $|u\rangle$ is particularly useful in QIT, as we often take the conjugate transpose of vectors, written $\langle u| = \overline{(|u\rangle)^T}$, and will often be preferred.

A basis for a Euclidean space is a linearly independent set which spans the space. One notable basis in QIT, which is also orthonormal, is the *computational basis* $\{|i\rangle\}$. For the Euclidean space $\mathbb{C}^4$ the computational basis is

$$|0\rangle = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \qquad |1\rangle \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \qquad |2\rangle = \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \qquad |3\rangle = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

The inner product between vectors $\mathbf{u}, \mathbf{v}$ belonging to $\mathcal{X}$ a d-dimensional Euclidean space is defined as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle u|v\rangle = \sum_{i=1}^{d} \overline{u_i} v_i \tag{2.1}$$

The Bra-Ket notation makes this definition rather intuitive. Note that this definition satisfies the three properties cited above. Using the inner product, we can derive the *Euclidean norm* for vectors, defined as

$$\|\mathbf{u}\| = \sqrt{\langle u|u\rangle} \tag{2.2}$$

Like for inner products, norms must satisfy several properties, namely:

- Positive definiteness: if $\|\mathbf{u}\| = 0$ then $|u\rangle = \mathbf{0}$
- Absolute homogeneity: $\|a\ \mathbf{u}\| = |a|\|\mathbf{u}\|$ for all $a \in \mathbb{C}$
- Triangle inequality: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ for all $\mathbf{u}, \mathbf{v} \in \mathcal{X}$

Although the Euclidean norm is perhaps the most familiar, many other norms exist. Indeed, the *p-norm* of a vector $\mathbf{u} \in \mathbb{C}^n$ for $p \in \mathbb{N}^* \cup \{\infty\}$ is defined as

$$\|\mathbf{u}|_p = \left(\sum_{i=1}^{d} |u_i|^p\right)^{\frac{1}{p}} \tag{2.3}$$

for $p < \infty$, and

$$\|\mathbf{u}\|_\infty = \max\{|u_i| : i \in \{1, \cdots, n\}\} \tag{2.4}$$

for $p = \infty$.

Earlier, we suggested that Euclidean spaces are used to represent quantum systems. To represent *composite* quantum systems, we need a way to "combine" Euclidean spaces into composite spaces. There are several ways to do this; however, in QIT specifically, we use the *tensor product* $\otimes$. Let $\mathcal{X}$ and $\mathcal{Y}$ be two complex Euclidean spaces such that $\dim(\mathcal{X}) = n$ and $\dim(\mathcal{Y}) = m$. Their tensor product $\mathcal{X} \otimes \mathcal{Y}$, is the $n \times m$-dimensional complex Euclidean space containing all vectors of the form $\mathbf{u} \otimes \mathbf{v}$ for $\mathbf{u} \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{Y}$.

Notably, there are vectors $\mathbf{w} \in \mathcal{X} \otimes \mathcal{Y}$ which cannot be written $\mathbf{w} = \mathbf{u} \otimes \mathbf{v}$ for vectors $\mathbf{u} \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{Y}$. For our purposes, we define the tensor product of two vectors explicitly as $(\mathbf{u} \otimes \mathbf{v})_i = \mathbf{u}_{\lceil \frac{i}{m} \rceil} \mathbf{v}_{i \pmod m + 1}$, although this definition will rarely be used in practice.

Finally, we identify a subset of $\mathcal{X}$ which is particularly relevant to QIT, the set of unit vectors. Pure states of a quantum system can be represented as outer products of vectors belonging to this set. $\mathcal{S}(\mathcal{X})$ is defined as

$$\mathcal{S}(\mathcal{X}) = \{\mathbf{u} \in \mathcal{X} : \|\mathbf{u}\| = 1\}. \tag{2.5}$$

2.1.2 Linear Operators

Linear operators are linear mappings of the form $A : \mathcal{X} \to \mathcal{Y}$. We write $L(\mathcal{X}, \mathcal{Y})$ to denote the collection of all linear operators between spaces $\mathcal{X}$ and $\mathcal{Y}$ where $L(\mathcal{X})$ is shorthand for the collection of *square* operators $L(\mathcal{X}, \mathcal{X})$. Every linear operator $A \in L(\mathcal{X}, \mathcal{Y})$ can be represented as a matrix $A \in \mathbb{C}^{m \times n}$ where $\dim(\mathcal{X}) = n$ and $\dim(\mathcal{Y}) = m$ (in the standard basis). In other words, there is a bijective correspondence between $\mathbb{C}^{m \times n}$ and $L(\mathcal{X}, \mathcal{X})$. We thus write $A\mathbf{u}$ and $A(\mathbf{u})$ interchangeably.

For any operator $A \in L(\mathcal{X}, \mathcal{Y})$, there exists a unique adjoint operator $A^\dagger \in L(\mathcal{Y}, \mathcal{X})$ satisfying:

$$\langle \mathbf{v}, A\mathbf{u} \rangle = \langle A^\dagger \mathbf{v}, \mathbf{u} \rangle \tag{2.6}$$

for all $\mathbf{u} \in \mathcal{X}$ and $\mathbf{v} \in \mathcal{Y}$. The adjoint can also be understood in terms of its effect on the matrix A, i.e., $A^\dagger = \overline{A^T}$.

Interestingly, the collection of operators $L(\mathcal{X}, \mathcal{Y})$ also forms a Euclidean space with the following inner product:

$$\langle A, B \rangle = \mathrm{Tr}(A^\dagger B) \tag{2.7}$$

This realisation enables us to define norms over operators in section 2.3.

There are several classes of operators which are particularly relevant to QIT. We list a few of these here.

- *Hermitian Operators* : An operator $X \in L(\mathcal{X})$ is Hermitian if it holds that $X = X^\dagger$. The set of Hermitian operators for a Euclidean space $\mathcal{X}$ is

$$\mathrm{Herm}(\mathcal{X}) = \left\{ X \in L(\mathcal{X}) : X = X^\dagger \right\} \tag{2.8}$$

- *Positive semidefinite operators* : An operator $X \in L(\mathcal{X})$ is positive semidefinite if $X = Y^\dagger Y$ for some $Y \in L(\mathcal{X})$. It follows that any positive semidefinite

operator is also Hermitian. The set of positive semidefinite operators is

$$\mathrm{Pos}(\mathcal{X}) = \left\{Y^\dagger Y : Y \in L(\mathcal{X})\right\} \tag{2.9}$$

- *Positive definite operators* : An operator $X \in L(\mathcal{X})$ is positive definite if, in addition to being positive semidefinite, it is invertible. The collection of positive definite operators over $\mathcal{X}$

$$\mathrm{Pd}(\mathcal{X}) = \{P \in \mathrm{Pos}(\mathcal{X}) : \mathrm{Det}(P) \neq 0\} \tag{2.10}$$

- *Density operators* : An operator $X \in L(\mathcal{X})$ is a density operator if it is both positive definite and has a trace equal to 1. Density operators are often written ρ. The collection of all such operators is

$$\mathrm{D}(\mathcal{X}) = \{\rho \in \mathrm{Pos}(\mathcal{X}) : \mathrm{Tr}(\rho) = 1\} \tag{2.11}$$

- *Isometries* : An operator $A \in L(\mathcal{X}, \mathcal{Y})$ is an isometry if it holds that $A^\dagger A = \mathbb{1}_\mathcal{X}$. Equivalently, isometries preserve the Euclidean norm and the inner product of vectors on which they act. The set of isometries is

$$\mathrm{U}(\mathcal{X}, \mathcal{Y}) = \left\{A \in L(\mathcal{X}, \mathcal{Y}) : A^\dagger A = \mathbb{1}_\mathcal{X}\right\} \tag{2.12}$$

- *Unitary operators* : The set of isometries mapping from a space $\mathcal{X}$ to itself are called unitary operators. We denote the set of such operators $U(\mathcal{X})$. Note that for $U \in U(\mathcal{X})$ it holds that $U^\dagger = U^{-1}$.

A fundamental theorem in operator theory is the *spectral theorem*. The theorem, when applied to Hermitian matrices, is:

Theorem 2.1.1 (Spectral Theorem for Hermitian matrices)**.** *Let $\mathcal{X}$ be a complex Euclidean space of dimension n and let $X \in L(\mathcal{X})$ be a Hermitian operator. There exist an orthonormal basis $\{|u_1\rangle, \cdots, |u_n\rangle\}$ of $\mathcal{X}$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ such that*

$$X = \sum_{i=1}^{m} \lambda_i |u_i\rangle \langle u_i|$$

where each vector $|u_i\rangle$ is an eigenvector of X and λ_i is its associated eigenvalue.

The equivalent theorem for unitary matrices is:

Theorem 2.1.2 (Spectral Theorem for Unitary matrices)**.** *Let $\mathcal{X}$ be a complex Euclidean space of dimension n and let $U \in L(\mathcal{X})$ be a unitary operator. There exist an orthonormal basis $\{|u_1\rangle, \cdots, |u_n\rangle\}$ of $\mathcal{X}$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{C}$ such that*

$$U = \sum_{i=1}^{m} \lambda_i |u_i\rangle \langle u_i|$$

where each vector $|u_i\rangle$ is an eigenvector of U and λ_i is its associated eigenvalue. Moreover $|\lambda_i| = 1$ for all i.

We can define the tensor product of operators. Let $\mathcal{X}$ and $\mathcal{Y}$ be two complex Euclidean spaces and $\mathcal{L}(\mathcal{X})$ and $\mathcal{L}(\mathcal{Y})$ their associated operator spaces. For $X \in \mathcal{L}(\mathcal{X})$ and $Y \in \mathcal{L}(\mathcal{Y})$ we define $X \otimes Y$ to be the unique linear operator acting on the space $\mathcal{X} \otimes \mathcal{Y}$ such that for all $|u\rangle \in \mathcal{X}$, and $|v\rangle \in \mathcal{Y}$:

$$(X \otimes Y)(|u\rangle \otimes |v\rangle) = X|u\rangle \otimes Y|v\rangle \tag{2.13}$$

In general, for $X, X' \in \mathcal{L}(\mathcal{X})$ and $Y, Y' \in \mathcal{L}(\mathcal{Y})$ the following identities hold:

$$(X \otimes Y)(X' \otimes Y') = XX' \otimes YY' \tag{2.14}$$

$$(X \otimes Y)^\dagger = X^\dagger \otimes Y^\dagger \tag{2.15}$$

2.1.3 Superoperators

We have now defined Euclidean spaces and operators over these spaces (the set of which also forms an Euclidean space). Hence, it is natural to define linear operators over spaces of operators, i.e., *superoperators*. These maps are crucial in QIT and form the theoretical foundations for approximate circuit optimisation.

Superoperators (or *maps* as we will often call them) are linear mappings of the form $\Phi : L(\mathcal{X}) \to L(\mathcal{Y})$ for two square operator spaces $L(\mathcal{X})$ and $L(\mathcal{Y})$. We write $T(\mathcal{X}, \mathcal{Y})$

to denote the collection of all linear operators between spaces $L(\mathcal{X})$ and $L(\mathcal{Y})$ and, like for operators, use $T(\mathcal{X})$ as shorthand for $T(\mathcal{X}, \mathcal{X})$.

Similar to operators, for every map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ we define a unique adjoint $\Phi^\dagger \in T(\mathcal{Y}, \mathcal{X})$ which satisfies

$$\langle \Phi^\dagger(Y), X \rangle = \langle Y, \Phi(X) \rangle \tag{2.16}$$

for all $X \in L(\mathcal{X})$ and $Y \in L(\mathcal{Y})$.

Two common maps in QIT are the *identity map* and the *partial trace* map. The identity map $\mathbb{1}_{L(\mathcal{X})} \in T(\mathcal{X})$ is the unique map which satisfies

$$\mathbb{1}_{L(\mathcal{X})}(X) = X \tag{2.17}$$

for all $X \in L(\mathcal{X})$.

The partial trace is $\mathrm{Tr}_{\mathcal{X}} \in T(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Y})$ is the map $\mathrm{Tr} \otimes \mathbb{1}_{L(\mathcal{Y})}$. Alternatively, we can define it as

$$\mathrm{Tr}_{\mathcal{X}}(A) = \sum_{k=1}^{n} (\langle k| \otimes \mathbb{1}_{\mathcal{Y}}) A (|k\rangle \otimes \mathbb{1}_{\mathcal{Y}}) \tag{2.18}$$

where $|k\rangle$ are the computational basis states. Setting $\mathcal{Y}$ to $\mathbb{C}$ gives us the usual trace function. Semantically, this map has the effect of *discarding* the quantum system represented by $\mathcal{X}$ leaving only $\mathcal{Y}$. The following theorem illustrates this property.

Theorem 2.1.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Euclidean spaces and operators $A \in L(X \otimes \mathcal{Y})$ and $B \in L(\mathcal{X})$. It holds that*

$$\mathrm{Tr}((B \otimes \mathbb{1}_{\mathcal{Y}})A) = \mathrm{Tr}(B \,\mathrm{Tr}_{\mathcal{Y}}(A))$$

Proof. We show the equality holds for all operators $A \in L(X \otimes \mathcal{Y})$ and $B \in L(\mathcal{X})$

using definition 2.18 of the partial trace.

$$\begin{aligned}
\operatorname{Tr}(B \operatorname{Tr}_{\mathcal{Y}}(A)) &= \operatorname{Tr}(B \sum_{k=1}^{n} (\mathbb{1}_{\mathcal{X}} \otimes \langle k|) A (\mathbb{1}_{\mathcal{X}} \otimes |k\rangle)) && \text{by definition 2.18} \\
&= \sum_{k=1}^{n} \operatorname{Tr}((\mathbb{1}_{\mathcal{X}} \otimes \langle k|) A (\mathbb{1}_{\mathcal{X}} \otimes |k\rangle) B) && \text{by linearity of Tr} \\
&= \operatorname{Tr}(A \sum_{k=1}^{n} (\mathbb{1}_{\mathcal{X}} \otimes |k\rangle) B (\mathbb{1}_{\mathcal{X}} \otimes \langle k|)) && \text{by cyclic property}
\end{aligned}$$

All that remains to be shown is that:

$$\sum_{k=1}^{n} (\mathbb{1}_{\mathcal{X}} \otimes |k\rangle) B (\mathbb{1}_{\mathcal{X}} \otimes \langle k|) = B \otimes \mathbb{1}_{\mathcal{Y}}$$

We start by noting that $\sum_{k=1}^{n} |k\rangle \langle k| = \mathbb{1}_{\mathcal{Y}}$ as $\{|k\rangle\}$ is an orthonormal basis. Moreover, for $X \in L(\mathcal{X})$ and $|u\rangle \in L(\mathbb{C}, \mathcal{Y})$, we have $(\mathbb{1}_{\mathcal{X}} \otimes |u\rangle) X = X \otimes |u\rangle$. Using these identities:

$$\begin{aligned}
\sum_{k=1}^{n} (\mathbb{1}_{\mathcal{X}} \otimes |k\rangle) B (\mathbb{1}_{\mathcal{X}} \otimes \langle k|) &= \sum_{k=1}^{n} (B \otimes |k\rangle)(\mathbb{1}_{\mathcal{X}} \otimes \langle k|) \\
&= \sum_{k=1}^{n} B \otimes (|k\rangle \langle k|) \\
&= B \otimes (\sum_{k=1}^{n} |k\rangle \langle k|) \\
&= B \otimes \mathbb{1}_{\mathcal{Y}}
\end{aligned}$$

□

There are four classes of maps which will have a special status in QIT.

- *Hermitian-preserving maps* : A map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is Hermitian-preserving if it holds that

  $$\Phi(X) \in \operatorname{Herm}(\mathcal{Y}) \tag{2.19}$$

  for every Hermitian operator $X \in \operatorname{Herm}(\mathcal{X})$.

- *Positive maps* : A map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is positive if it holds that

$$\Phi(X) \in \text{Pos}(\mathcal{Y}) \tag{2.20}$$

for every Hermitian operator $X \in \text{Pos}(\mathcal{X})$.

- *Completely positive maps* : A map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is completely positive if it holds that

$$\Phi \otimes \mathbb{1}_{L(\mathcal{Z})} \tag{2.21}$$

is a positive map for all complex Euclidean spaces $\mathcal{Z}$.

- *Trace-preserving maps* : A map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is trace-preserving if it holds that:

$$\text{Tr}(\Phi(X)) = \text{Tr}(X) \tag{2.22}$$

for all $X \in L(\mathcal{X})$

- *Unitary maps* : A map $\Phi \in T(\mathcal{X})$ is unitary if we can write it as:

$$\Phi(\cdot) = UXU^\dagger \tag{2.23}$$

for some unitary operator $U \in U(\mathcal{X})$.

2.2 Quantum Information Theory

Now that we have reviewed the relevant mathematics, we can move on to QIT. In this section, we define the notion of a quantum system, describe what operations we can perform on it, and formalise how we can extract information from it. This mathematical formalism is the basis for the standard model of quantum computing.

2.2.1 Quantum Systems

For a given d-dimensional quantum system S, a quantum state is any operator $\rho \in L(\mathcal{X})$ where $\dim(\mathcal{X}) = d$ such that:

- $\rho \in \text{Herm}(\mathcal{X})$
- $\rho \in \text{Pos}(\mathcal{X})$
- $\text{Tr}(\rho) = 1$

We know, by equation (2.9) that every positive semidefinite operator is Hermitian ($\text{Pos}(\mathcal{X}) \subseteq \text{Herm}(\mathcal{X})$). Thus, the collection of operators in $L(\mathcal{X})$, which represents quantum states, is precisely $D(\mathcal{X})$.

If a quantum state $\rho \in D(\mathcal{X})$ satisfies the additional condition that $\rho^2 = \rho$, we call ρ a pure state. This is equivalent to saying $\rho = |\psi\rangle \langle\psi|$ for $|\psi\rangle \in \mathcal{S}(\mathcal{X})$.

Now suppose we have two quantum subsystems S_A and S_B with associated Euclidean spaces $\mathcal{X}_A$ and $\mathcal{X}_B$, and states ρ_A and ρ_B respectively. We represent the state of this composite system by taking the tensor product of the two states $\rho_{AB} = \rho_A \otimes \rho_B$. Note that this composite state belongs to the composite Euclidean space $\mathcal{X}_{AB} = \mathcal{X}_A \otimes \mathcal{X}_B$.

If the state of a composite system S_{AB} is of the form $\rho_{AB} = \rho_A \otimes \rho_B$ then we call it a *product state*. However, as we saw in 2.1.1, not all operators in a composite space can be written in this form. The states of this form are called *entangled states*. Entanglement is a concept at the heart of quantum theory and provides the basis for several quantum protocols.

If we have a composite system S_{AB} and we wish to know the state of system S_A, we can use the partial trace Tr_B to "discard" or "trace out" system S_B leaving the state of S_A. A consequence of this is the ability to express a mixed quantum state ρ_A by a global pure state ρ_{AB} such that $\text{Tr}_B(\rho_{AB}) = \rho_A$. This is called purifying states.

A notable quantum system is the *qubit*. Qubits are quantum systems with dimension 2, making $D(\mathbb{C}^2)$ the set of possible states of a qubit. They are the fundamental unit of information in quantum computing and will be discussed repeatedly in this thesis.

2.2.2 Operations on Systems

The most general evolution a quantum system can undergo is called a quantum channel. For two Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ a channel is any superoperators $\mathcal{E} \in T(\mathcal{X}, \mathcal{Y})$ such that $\mathcal{E}(\rho) \in D(\mathcal{Y})$ for all $\rho \in D(\mathcal{X})$. In other words, channels are maps that map density operators to other density operators. Quantum channels are also referred to as *completely-positive trace-preserving* (CPTP) maps. Note that the partial trace is a valid quantum channel.

Multiple decompositions of channels exist, but one of the most common is the Kraus decomposition. In general, for $\mathcal{E} \in T(\mathcal{X}, \mathcal{Y})$ a quantum channel, $\mathcal{E}(\rho)$ can be written

$$\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger \tag{2.24}$$

for a set of complex matrices $\{A_i\} \subset L(\mathcal{X}, \mathcal{Y})$ satisfying:

$$\sum_i A_i^\dagger A_i = \mathbb{1}_\mathcal{X} \tag{2.25}$$

The set of operators $\{A_i\}$ are called the Kraus operators.

When the set of Kraus operators is a singleton set $\{U\}$ with $U \in U(\mathcal{X})$, the channel is called *unitary*. The standard quantum computing model comprises unitary channels called *gates* acting on one or more qubits.

2.2.3 Measurements

The final concept we discuss is quantum measurements. Intuitively, if we can describe the state of a quantum system with perfect accuracy, we should be able to predict the outcomes of a measurement. This is the case in classical theory, after all.

Unfortunately, things are more complicated in QIT. Unlike classical measurements, quantum measurements are distinctly nondeterministic. Identical measurements of

an identical state sometimes return different results. To make things worse, measurements change the state of the system we are measuring. Quantum systems always jump to the state we measure, even if they were not in that state initially.

The type of measurement we will be concerned with is called *projective measurements*. These are described by sets of projectors $\{\Pi_i\}$ whose eigenvectors form an orthogonal basis of the Euclidean space being measured. Equivalently, let $\{|\phi_1\rangle \cdots |\phi_d\rangle\}$ be any *orthonormal basis* (ONB) for $\mathcal{X}_A$, then the set $\{|\phi_i\rangle \langle\phi_i|\}$ is a projective measurement for $\mathcal{X}_A$.

For a given quantum system S_A in state ρ, the probability of getting outcome m for projective measurement $\{\Pi_i\}$ is given by the born rule:

$$p(m) = \mathrm{Tr}(\Pi_m \rho) \tag{2.26}$$

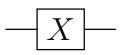
And the state of the system ρ' after measurement has been performed is updated to

$$\rho' = \frac{\Pi_m \rho \Pi_m}{\mathrm{Tr}(\Pi_m \rho)} \tag{2.27}$$

2.2.4 The Quantum Circuit Model

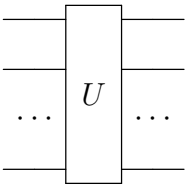
The previous sections provided us with the foundations of QIT. In quantum computing, we aim to harness the properties of this model to achieve computational advantage in specific tasks. We use the quantum circuit model to formalise computational tasks on quantum hardware. Like in classical computing, we use circuits to describe the algorithm we are implementing–but the similarities stop there. Rather than bits, wires represent qubits to which we apply quantum gates (unitary channels).

We represent quantum gates using the following standard notation:



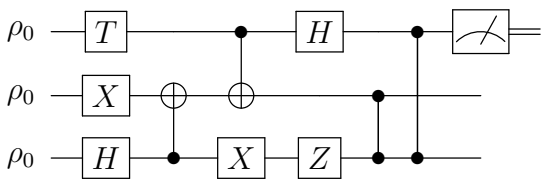
However, quantum gates can operate on more than one qubit. For instance, the controlled-NOT gate (CNOT), which acts on two qubits, is written as

More generally, the gate implementing unitary U acting on any number of qubits can be written as



Measurements in the circuit model are always in the computational basis (i.e. $\Pi_i = |i\rangle\langle i|$) or *computational measurements*. Importantly, this does not result in any loss of generality. Any projective measurement can be reduced to a measurement in the computational basis, assuming we can implement the corresponding change of the basis unitary. Computational measurements are written:

where the doubled wire represents classical information. Generally, the *input* state of a quantum circuit is the zero state $\rho_{\text{init}} = \rho_0^{\otimes n}$ where $\rho_0 = |0\rangle\langle 0|$ and n is the number of qubits. Combining these three elements – gates, measurements, and input state – here is an example of a full quantum circuit:



Translating the circuit model to and from the mathematical model we developed earlier is straightforward. The sequential composition of two unitary channels is done using multiplication. Let $\mathcal{E}(\cdot) = U \cdot U^\dagger$ and $\mathcal{F}(\cdot) = V \cdot V^\dagger$. Then $\mathcal{F} \circ \mathcal{E}(\cdot) = VU \cdot (VU)^\dagger$.

Parallel composition is done using the tensor product. These two composition rules, in addition to the identity channel, are sufficient to translate any circuit into a global unitary operation.

In practice, quantum computers can only implement a finite subset of the aforementioned unitary channels. Therefore, this *gate set* must be *universal* in a relevant sense. We call a gate set universal if any unitary channel can be approximated by a finite sequence of gates from the gate set. The Solovay–Kitaev theorem [43] guarantees this approximation can be done efficiently. Some universal quantum gate sets are:

- The Clifford set $\{CNOT, H, S\}$ + T gate $\{T\}$
- The 1-qubit rotation gates + controlled-NOT gate $\{CNOT\}$
- Toffoli + Hadamard $\{H\}$

2.3 Operator and Superoperator Norms

At this stage, we still require an element essential to elaborating our framework for approximate QCO, namely a way to quantify the distance between quantum states and quantum channels. As discussed in section 2.2, states and channels are represented using operators and maps acting on complex Euclidean spaces. In this section, we define norms for these objects and discuss some of their properties.

2.3.1 Operator Norms

Norms can be defined over any Euclidean space. In section 2.1.1, we defined the norm of a vector $|u\rangle$. Analogously, we can define norms over spaces of operators. Like the norm of a vector, the norm of an operator is a function $\|\cdot\| : L(\mathcal{X}, \mathcal{Y}) \to \mathbb{R}$ which satisfies:

1. *Positive-valued* : $\|A\| \geq 0$

2. *Definite* : $\|A\| = 0 \Leftrightarrow A = \mathbf{0}_{L(\mathcal{X},\mathcal{Y})}$

3. *Absolutely homogeneous* : $\|\alpha A\| = |\alpha|\|A\|$ for all $\alpha \in \mathbb{C}$

4. *Triangle inequality* : $\|A + B\| \leq \|A\| + \|B\|$

for all $A, B \in L(\mathcal{X}, \mathcal{Y})$

There are infinitely many norms one can define over an operator space. However, in quantum information, we will stick to a particular family of norms $\|\cdot\|_p$ called *Schatten p-norms*. The definition of a Schatten p-norm is

$$\|A\|_p = \left(\mathrm{Tr}\left((A^\dagger A)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}} \tag{2.28}$$

for $p \neq \infty$ and

$$\|A\|_\infty = \max\left\{\|A|u\rangle\| : |u\rangle \in \mathcal{X}, \||u\rangle]\| \leq 1\right\} \tag{2.29}$$

for $p = \infty$.

Setting p to 1, 2, and ∞ yields the three most commonly used norms in quantum information: the *trace norm*, *Frobenius norm*, and *spectral norm* respectively. The spectral norm $\|\cdot\|_\infty$ is also called the *operator norm* or *infinity norm* and is sometimes written $\|\cdot\|_{\mathrm{op}}$ accordingly.

In addition to the properties above, the Schatten norms satisfy a few additional properties:

1. Schatten p-norms are non-increasing, i.e., for all $A \in L(\mathcal{X}, \mathcal{Y})$ and $1 \leq p \leq q \leq \infty$

$$\|A\|_p \geq \|A\|_q \tag{2.30}$$

2. Schatten p-norms are isometrically invariant. For every $p \in [1, \infty]$, $A \in L(\mathcal{X}, \mathcal{Y})$, $U \in U(\mathcal{Y}, \mathcal{Z})$, and $V \in U(\mathcal{X}, \mathcal{W})$

$$\|A\|_p = \|UAV^\dagger\|_p \tag{2.31}$$

3. Schatten p-norms are dual in the sense that for every $p \in [1, \infty]$, we can define $p^\dagger$ as

$$\frac{1}{p} + \frac{1}{p^\dagger} = 1 \tag{2.32}$$

such that for every $A \in L(\mathcal{X}, \mathcal{Y})$, it holds that

$$\|A\|_p = \max\{|\langle B, A\rangle| : B \in L(\mathcal{X}, \mathcal{Y}), \|B\|_{p^\dagger} \leq 1\} \tag{2.33}$$

A consequence of this is the *Hölder inequality*

$$|\langle B, A\rangle| \leq \|A\|_p \|B\|_{p^\dagger} \tag{2.34}$$

4. The Schatten p-norms are sub-multiplicative

$$\|AB\|_p \leq \|A\|_p \|B\|_p \tag{2.35}$$

In the remainder of this thesis, we will interchangeably use the terms *norm* and *distance*. For any norm $\|\phi\|$ we can easily derive a distance measure $d(\cdot, \cdot)$ called the *induced metric* such that $d(\phi, \psi) = \|\phi - \psi\|$. An example is the *trace distance* between quantum states. It is derived from the trace (see above). This leads to a convenient identity we will make use of later:

$$\||\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|\|_1 = 2\sqrt{1 - |\langle\psi|\phi\rangle|^2} \tag{2.36}$$

2.3.2 Superoperator Norms

In section 2.2, we saw that the general representation of a quantum operation is a superoperator (c.f. section 2.1.3). To assess the effectiveness of approximate QCO, we require a distance measure on quantum channels, which in turn requires defining norms over superoperators. This section introduces the *induced trace norm*.

Superoperator norms are defined on the space of superoperators $\Phi \in T(\mathcal{X}, \mathcal{Y})$. The most straightforward superoperator norms belong to the family of *induced norms*. In

general, for a suitable operator norm $\|\cdot\|$ defined over spaces $\mathcal{L}(\mathcal{X})$ and $\mathcal{L}(\mathcal{Y})$ the induced superoperator norm $\|\cdot\| : T(\mathcal{X}, \mathcal{Y}) \to \mathbb{R}$ is:

$$\|\Phi\| = \max\{\|\Phi(A)\| : A \in L(\mathcal{X}), \|A\| \leq 1\} \tag{2.37}$$

Using this, we can define the *induced trace norm* as:

$$\|\Phi\|_1 = \max\{\|\Phi(A)\|_1 : A \in L(\mathcal{X}), \|A\|_1 \leq 1\} \tag{2.38}$$

Likewise, we could define the *induced spectral norm*:

$$\|\Phi\|_\infty = \max\{\|\Phi(A)\|_\infty : A \in L(\mathcal{X}), \|A\|_\infty \leq 1\} \tag{2.39}$$

Conveniently, induced norms inherit the properties of the operator norm from which they are derived. Hence, any induced norm will satisfy the usual "metric" properties. In addition, the induced trace norm has several desirable properties for approximate QCO.

1. *Submultiplicativity*: For all maps $\Phi \in T(\mathcal{X}, \mathcal{Y})$ and $\Psi \in T(\mathcal{Y}, \mathcal{Z})$ it holds that

$$\|\Psi\Phi\|_1 \leq \|\Phi\|_1 \|\Psi\|_1 \tag{2.40}$$

2. *Additivity of channel differences*: For all channels $\Phi_0, \Phi_1 \in C(\mathcal{X}, \mathcal{Y})$ and $\Psi_0, \Psi_1 \in C(\mathcal{Y}, \mathcal{Z})$ it holds that

$$\|\Psi_0\Phi_0 - \Psi_1\Phi_1\|_1 \leq \|\Psi_0 - \Psi_1\|_1 + \|\Phi_0 - \Phi_1\|_1 \tag{2.41}$$

3. *Unitary invariance*: Let $\Phi \in T(\mathcal{X}, \mathcal{Y})$ be a map and let $U_0, U_1 \in U(\mathcal{X})$ and $V_0, V_1 \in U(\mathcal{Y})$ be unitary operators. Define $\Psi \in T(\mathcal{X}, \mathcal{Y})$ as

$$\Psi(X) = V_0\Phi(U_0 X U_1)V_1 \tag{2.42}$$

for all $X \in L(\mathcal{X})$. Then $\|\Psi\|_1 = \|\Phi\|_1$

Chapter 3

Quantifying Distance Between Quantum Channels

The main focus of this chapter is defining a quantifiable distance metric for quantum channels. Section 3.1 sets out the general criteria a metric must comply with and defines our metric of choice: the *diamond norm*. In section 3.2, we discuss the practicalities of calculating the diamond norm numerically and provide a novel proof leading to numerical optimisation. In sections 3.3 and 3.4.1, we formalise the algorithm we use to compute the diamond norm and evaluate its efficiency empirically with respect to alternatives.

3.1 An Operationally Meaningful Distance Measure

3.1.1 Desirable Properties for a Distance Measure

In section 2.3.2, we introduced the induced trace norm $\|\cdot\|_1$ and discussed several desirable properties, including submultiplicativity, additivity of channel differences, and unitary invariance. Nonetheless, it generally fails to provide a physically well-motivated measure of distance between quantum channels [42]. This is mainly because it is not stable with respect to tensoring with the identity. This means that, in general, $\|\Phi\|_1 \neq \|\Phi \otimes \mathbb{1}_{L(\mathcal{Z})}\|_1$ for a Euclidean space $\mathcal{Z}$.

Physically, this means that unrelated ancillary quantum systems sometimes affect the value of the induced trace norm. As quantum algorithms are usually broken down into sequences of smaller operations (gates or sub-circuits, c.f. section 2.2.4) and we wish to derive bounds on the total error by analysing individual components, the *stability* criterion is essential for our purposes.

Gilchrist et al. compile a list of criteria that should be satisfied by a good distance measure in order to identify a *gold standard* metric for QIT [44]. Their criteria include:

1. *Metric* : The measure should be a metric satisfying three properties: positive definiteness, symmetry, and satisfaction of the triangle inequality.

2. *Easy to calculate* : It should be possible to evaluate the measure numerically and efficiently.

3. *Physical Interpretation* : The measure should have a clear physical interpretation.

4. *Stability* : The measure should be stable with respect to tensoring with the identity.

5. *Additive channel difference* : Let $d(\cdot, \cdot)$ be our metric, then $d(\mathcal{E}_1 \circ \mathcal{F}_1, \mathcal{E}_2 \circ \mathcal{F}_2) \leq d(\mathcal{E}_1, \mathcal{E}_2) + d(\mathcal{F}_1, \mathcal{F}_2)$

After evaluating an extensive collection of candidate norms using these criteria, Gilchrist et al. recommend the *diamond norm* as a gold standard.

3.1.2 The Diamond Norm

We now define an alternative to the induced trace norm which satisfies the criteria set out above: the *diamond norm*[1] $\|\cdot\|_\diamond$. Let Φ be a mapping $\Phi \in T(\mathcal{X}, \mathcal{Y})$, we define

[1]The relevance of the diamond norm for QIT was first realised by Kitaev in 1997 [43]. The norm gets its name from the notation $\|\cdot\|_\diamond$ used in that paper. It is also referred to as the *completely bounded trace norm* (written $\|\|\cdot\|\|_1$) in [42] or the *stabilised process distance* in [44].

the diamond norm to be:

$$\|\Phi\|_\diamond = \|\Phi \otimes \mathbb{1}_{L(\mathcal{X})}\|_1 \tag{3.1}$$

The diamond norm inherits all the properties of the induced trace norm. Hence, when used as a distance, it is both a *metric* and satisfies the *additive channel difference* property. In addition, the diamond norm is stable when tensored with the identity.

Proposition 3.1.1 (Watrous 3.47)**.** *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be Euclidean spacs, and let $\Phi \in T(\mathcal{X}, \mathcal{Y})$ be a superoperator between $\mathcal{X}$ and $\mathcal{Y}$. Then, it holds that*

$$\|\Phi \otimes \mathbb{1}_{L(\mathcal{Z})}\|_\diamond = \|\Phi\|_\diamond \tag{3.2}$$

This is a consequence of the following two propositions. First, the diamond norm is multiplicative with respect to tensor products.

Proposition 3.1.2 (Watrous 3.49)**.** *Let $\mathcal{W}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be Euclidean spaces, and let $\Phi \in T(\mathcal{W}, \mathcal{X})$ and $\Psi \in T(\mathcal{Y}, \mathcal{Z})$ be superoperators. Then, it holds that*

$$\|\Phi \otimes \Psi\|_\diamond = \|\Phi\|_\diamond \|\Psi\|_\diamond \tag{3.3}$$

Moreover, the diamond norm of a CPTP channel is always 1.

Proposition 3.1.3 (Watrous 3.40)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Euclidean spaces, and let $\Phi \in T(\mathcal{X}, \mathcal{Y})$ be a CPTP channel. Then, it holds that*

$$\|\Phi\|_\diamond = 1 \tag{3.4}$$

As the identity channel is CPTP and the diamond norm is multiplicative with respect to tensor products, it trivially follows that it is stable in the sense described above. Moreover, when we take the difference of two CPTP channels, the diamond norm of this map is always between 0 and 2.

The diamond norm also has the virtue of being physically meaningful. Historically, it was used to derive probabilistic bounds for the channel discrimination task. When evaluating the approximation error between a pair of channels, the diamond norm bounds the worst-case error. Gilchrist et al. argue this is an advantage as we are often more interested in the worst-case error than the average case [44].

The only remaining criterion, especially relevant for our purposes, is that the diamond norm is easy to calculate numerically.

3.2 Computing the Diamond Norm

Given its formulation, it is surprising that the diamond norm admits a fairly efficient numerical calculation. Watrous and Johnston et al. derive semidefinite programs, and Ben-Aroya et al. provide a convex optimisation problem, which compute the norm in the general case of a CPTP quantum channel [45, 46, 47, 48].

Although these methods are generally quite efficient, the problem they solve quickly becomes intractable as the number of qubits grows (see section 3.4.1 for details). These methods make use of the Choi–Jamiołkowski representation of a quantum channel, which has space complexity $\mathcal{O}(2^{4n})$ in the number of qubits. This is less optimal than the $\mathcal{O}(2^{2n})$ space required to represent unitary channels. Moreover, they require solving an optimisation problem, which is significantly more resource-intensive than evaluating a closed-form equation.

As our empirical exploration of the space of approximate circuit equivalences relies on the rapid evaluation of the diamond distance between a pair of unitary channels, we propose an alternative algorithm which computes the diamond norm in the specific case of a difference of unitary channels. The method relies on an unproved result first discussed in [38] about the diamond norm of a difference of two unitary channels.

3.2.1 Preliminary Results

Our novel proof of the result in [38] relies on several properties of the diamond and induced trace norms stated in this section. To begin, Watrous shows that the diamond norm of a map Φ is always larger than or equal to the induced trace norm of that map.

Theorem 3.2.1 (Watrous 3.46)**.** *Let $\Phi \in T(\mathcal{X}, \mathcal{Y})$ be a map between Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, then*

$$\|\Phi\|_1 \leq \|\Phi\|_\diamond$$

Although we defined the diamond norm as a maximisation over all operators with trace norm less than one, Watrous shows that if the map $\Phi \in T(\mathcal{X}, \mathcal{Y})$ is Hermitian preserving, this isn't necessary. For Φ a Hermitian preserving map, there exists a unit vector $|u\rangle \in \mathcal{X} \otimes \mathcal{X}$ such that the rank one projection $|u\rangle \langle u|$ maximises the diamond norm (equation 3.1). This result is stated in the following theorem:

Theorem 3.2.2 (Watrous 3.51)**.** *For $\Phi \in T(\mathcal{X}, \mathcal{Y})$ a Hermitian preserving map, it holds that*

$$\|\Phi\|_\diamond = \max_{|u\rangle \in \mathcal{S}(\mathcal{X} \otimes \mathcal{X})} \|(\Phi \otimes \mathbb{1}_{L(\mathcal{X})})(|u\rangle \langle u|)\|_1$$

Additionally, we can prove that in the case of a difference of unitary maps, the diamond norm is equivalent to the induced trace norm. This result is one of the main reasons we can simplify the methods in [47, 45]. We begin with the following lemma.

Lemma 3.2.3. *Let $\Phi, \Psi \in T(\mathcal{X})$ be unitary maps for $\mathcal{X}$ a Euclidean space,*

$$\Phi(\cdot) := U \cdot U^\dagger$$

$$\Psi(\cdot) := V \cdot V^\dagger$$

then for every vector $|\phi\rangle \in \mathcal{S}(\mathcal{X} \otimes \mathcal{X})$ there exists a vector $|\psi\rangle \in \mathcal{S}(\mathcal{X})$ such that

$$\|((\Phi - \Psi) \otimes \mathbb{1}_{L(\mathcal{X})})(|\phi\rangle \langle \phi|)\|_1 = \|(\Phi - \Psi)(|\psi\rangle \langle \psi|)\|_1$$

Proof. We define $\mathcal{Y}$, the Euclidean space equal to $\mathcal{X}$ for notational convenience. Let $|\phi\rangle \in \mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$, using equation 2.36 we have:

$$\begin{aligned}\|((\Phi - \Psi) \otimes \mathbb{1}_{L(\mathcal{Y})})(|\phi\rangle \langle\phi|)\|_1 &= \|(U \otimes \mathbb{1}_{\mathcal{Y}}) |\phi\rangle \langle\phi| (U^\dagger \otimes \mathbb{1}_{\mathcal{Y}}) \\ &\quad - (V \otimes \mathbb{1}_{\mathcal{Y}}) |\phi\rangle \langle\phi| (V^\dagger \otimes \mathbb{1}_{\mathcal{Y}})\|_1 \\ &= 2\sqrt{1 - |\langle\phi| (U^\dagger \otimes \mathbb{1}_{\mathcal{Y}})(V \otimes \mathbb{1}_{\mathcal{Y}}) |\phi\rangle|^2} \\ &= 2\sqrt{1 - |\langle\phi| (U^\dagger V \otimes \mathbb{1}_{\mathcal{Y}}) |\phi\rangle|^2}\end{aligned}$$

Notice that:

$$\begin{aligned}\langle\phi| \left(U^\dagger V \otimes \mathbb{1}_{\mathcal{Y}}\right) |\phi\rangle &= \mathrm{Tr}\left(\langle\phi| \left(U^\dagger V \otimes \mathbb{1}_{\mathcal{Y}}\right) |\phi\rangle\right) \\ &= \mathrm{Tr}\left(\left(U^\dagger V \otimes \mathbb{1}_{\mathcal{Y}}\right) |\phi\rangle \langle\phi|\right) \\ &= \mathrm{Tr}\left(U^\dagger V \rho\right) \qquad \text{by Theorem 2.1.3} \\ &= \langle\rho, U^\dagger V\rangle\end{aligned}$$

for $\rho = \mathrm{Tr}_{\mathcal{Y}}(|\phi\rangle \langle\phi|)$. Applying the spectral theorem (Theorem 2.1.1) to ρ, it follows that $\langle\rho, U^\dagger V\rangle$ is a convex combination of complex values

$$\langle\chi| U^\dagger V |\chi\rangle$$

where $\{|\chi\rangle\}$ are the eigenvectors of ρ. Each of these values is contained in the numerical range of $U^\dagger V$ written $\mathcal{N}(U^\dagger V)$. The Toeplitz–Hausdorff theorem tells us that the numerical range of any operator is convex. Hence, $\langle\rho, U^\dagger V\rangle \in \mathcal{N}(U^\dagger V)$. Therefore, there must exist a vector $|\psi\rangle \in \mathcal{S}(\mathcal{X})$ such that

$$\langle\psi| U^\dagger V |\psi\rangle = \langle\rho, U^\dagger V\rangle \tag{3.5}$$

$$= \langle\phi| U^\dagger V \otimes \mathbb{1}_{\mathcal{X}} |\phi\rangle \tag{3.6}$$

plugging this into equation 2.36 gives the desired result. □

Theorem 3.2.4. *Let $\Phi, \Psi \in T(\mathcal{X})$ be unitary maps for $\mathcal{X}$ a Euclidean space, then*

$$\|\Phi - \Psi\|_\diamond = \|\Phi - \Psi\|_1$$

Proof. We know that the sum of two Hermitian preserving maps is also Hermitian preserving. As Φ and Ψ are unitary maps, it follows that $\Phi - \Psi$ is also Hermitian preserving. Theorem 3.2.2 tells us there exists a unit vector $|\phi\rangle \in \mathcal{S}(\mathcal{X} \otimes \mathcal{X})$ such that:

$$\|\Phi - \Psi\|_\diamond = \|((\Phi - \Psi) \otimes \mathbb{1}_{L(\mathcal{X})}) |\phi\rangle \langle\phi| \|_1$$

We also know by Lemma 3.2.3 that there exists a vector $|\psi\rangle \in \mathcal{S}(\mathcal{X})$ such that:

$$\|((\Phi - \Psi) \otimes \mathbb{1}_{L(\mathcal{X})})(|\phi\rangle \langle\phi|)\|_1 = \|(\Phi - \Psi)(|\psi\rangle \langle\psi|)\|_1$$

Moreover the vector $|\psi\rangle$ maximises $\|(\Phi - \Psi)(|\psi\rangle \langle\psi|)\|_1$ (by Theorem 3.2.1) and thus $\|\Phi - \Psi\|_\diamond = \|\Phi - \Psi\|_1$ □

3.2.2 Novel Proof of Diamond Norm Optimisation

This section gives an original proof of the result on page 29 of [38]. The proof relies on two intermediate results which we discussed in section 3.2.1. We are indebted to Professor Noah Linden[2] for providing the rough structure of this proof.

Theorem 3.2.5. *Let $\Phi, \Psi \in T(\mathcal{X})$ be unitary channels for $\mathcal{X}$ a Euclidean space with*

$$\Phi(\cdot) := U \cdot U^\dagger$$
$$\Psi(\cdot) := V \cdot V^\dagger$$

then the diamond metric $\|\Phi - \Psi\|_\diamond$ is equal to $2\sqrt{1 - d^2}$ where d is the shortest distance between the origin and the surface formed by the convex hull of the eigenvalues of $U^\dagger V$.

Proof. For $\mathcal{X}$ be a Euclidean space with $\dim(\mathcal{X}) = n$ let $\Phi \in T(\mathcal{X})$ and $\Psi \in T(\mathcal{X})$ be unitary channels

$$\Phi(\cdot) := U \cdot U^\dagger$$
$$\Psi(\cdot) := V \cdot V^\dagger$$

[2]Professor of Theoretical Physics, University of Bristol

with $U, V \in U(\mathcal{X})$.

$$\begin{aligned}
\|\Phi - \Psi\|_\diamond &= \max_{|\phi\rangle \in \mathcal{S}(\mathcal{X}\otimes\mathcal{X})} \|((\Phi - \Psi) \otimes \mathbb{1}_{L(\mathcal{X})})(|\phi\rangle\langle\phi|)\|_1 && \text{by Theorem 3.2.2} \\
&= \max_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} \|(\Phi - \Psi)(|\psi\rangle\langle\psi|)\|_1 && \text{by Theorem 3.2.1 and Lemma 3.2.3} \\
&= \max_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} \|U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger\|_1 \\
&= \max_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} 2\sqrt{1 - |\langle\psi|U^\dagger V|\psi\rangle|^2} && \text{by Equation 2.36}
\end{aligned}$$

Hence, we can evaluate the diamond norm by calculating

$$\min_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} \left|\langle\psi|U^\dagger V\,|\psi\rangle\right|^2 \tag{3.7}$$

We know $U^\dagger V$ is unitary. Therefore, by Theorem 2.1.2, $U^\dagger V$ admits a decomposition of the form $WDW^\dagger$ with $W \in U(\mathcal{X})$ and D a diagonal matrix. Let $D = \text{diag}(e^{i\theta_1}, e^{i\theta_2}, \cdots, e^{i\theta_n})$. Since W is an isometry (unitary operator) and isometries preserve the dot product

$$\begin{aligned}
\min_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} \left|\langle\psi|U^\dagger V|\psi\rangle\right|^2 &= \min_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} \left|\langle\psi|WDW^\dagger|\psi\rangle\right|^2 \\
&= \min_{|\psi\rangle \in \mathcal{S}(\mathcal{X})} |\langle\psi|D|\psi\rangle|^2 \\
&= \min_{|\alpha\rangle \in \mathbb{C}^n} \left|\sum_j |\alpha_j|^2 e^{i\theta_j}\right|^2 && \text{such that } \||\alpha\rangle\| = 1
\end{aligned}$$

We know that $\||\alpha\rangle\| = 1$ implies that $\sum_j |\alpha_j|^2 = 1$. Hence

$$\left\{\sum_j |\alpha_j|^2 e^{i\theta_j} : \||\alpha\rangle\| = 1\right\}$$

is the convex hull of the eigenvalues of $U^\dagger V$. Thus, the optimal value of the minimisation problem above is the square of the shortest distance between the origin and the surface formed by the convex hull, which completes the proof.

□

Notice that the diamond norm presents a very vivid geometric interpretation in the case of a difference of unitary channels. To calculate the diamond norm, it suffices to calculate the eigenvalues of the unitary $U^\dagger V$ (all of which lie on the unit circle in the complex plane) and determine the distance between the origin and the convex hull formed by these eigenvalues. We denote this distance as d henceforth. We explore this geometrical interpretation further in section 3.4.1.

3.3 The Algorithm

Using the theoretical result of Theorem 3.2.5, we can develop a more efficient alternative to the methods developed by Watrous and Johnston et al. [46, 45, 47]. It is important to emphasise that this implementation is not as general as Watrous and Johnston's, as our method only works on differences of unitary channels.

Despite this restriction, the novel approach to calculating the diamond norm developed in this thesis represents a significant contribution to quantum information research. The standard model of quantum computing restricts itself to unitary gates. Thus, in quantum circuit optimisation, an efficient method for computing the diamond distance between two circuits is extremely valuable. We have contributed the algorithm to two popular quantum computing libraries, Qiskit [40] and QuTiP [39], providing further evidence of this.

We develop the hyper-efficient diamond norm algorithm (denoted DiamondNorm) in three steps: (1) calculate the eigenvalues of $U^\dagger V$; (2) find the distance d; (3) plug d into $2\sqrt{1-d^2}$. Algorithms 1 and 2 formalise our approach in pseudo-code.

We start by calculating the eigenvalues of $U^\dagger V$ using the procedure GetEigenVals. These are passed into FindPolyDistance, which calculates the distance d by either finding the edge of the polygon closest to the origin or determining that the origin is contained within the polygon.

This edge is found by first calculating the phase of each eigenvalue and partitioning them into two sets *neg* and *pos* (Where *neg* contains phases in $(-\pi, 0]$ and *pos* contains phases in $(0, \pi]$). We call the procedure which does this PartitionPhases. We then find the largest and smallest phase in both these sets using GetExtremePhases. Observe that the edge closest to the origin will necessarily connect a pair of eigenvalues whose phases are one of these extremes.

Verifying whether either set is empty and calculating the angle between the extreme eigenvalues enables us to determine the distance between the polygon and the origin quickly.

Algorithm 1 Hyper-efficient Diamond Norm

```
1: procedure DiamondNorm(U, V)                ▷ For U, V unitary matrices
2:     eigs ← GetEigenVals(U†V)
3:     d ← FindPolyDistance(eigs)
4:     return 2√(1 − d²)
5: end procedure
```

For empirical exploration in later chapters, we implement this algorithm in Python. We use numpy's native `linalg.eigvals` to determine the unitary's eigenvalues. Computing the distance d is done straightforwardly using native numpy and python functions.

3.4 Empirical Results

This section reports on the two main advantages of this approach to evaluating the diamond norm: the intuitive geometric interpretation it offers and the impressive computational speed-up. We use empirical results to illustrate these claims and shed light on the operation of the diamond norm.

Algorithm 2 Find Polygon Distance

```
1: procedure FindPolyDistance(eigs)
2:     pos, neg ← PartitionPhases(eigs)          ▷ Partitions the phases of eigs
3:     pmin, pmax ← GetExtremePhases(pos)        ▷ Returns extreme phases
4:     nmin, nmax ← GetExtremePhases(neg)
5:     if neg is empty then
6:         return cos((pmax−pmin)/2)
7:     end if
8:     if pos is empty then
9:         return cos((nmax−nmin)/2)
10:    end if
11:    angle_big ← pmax − nmin
12:    angle_small ← pmax − nmin
13:    if angle_big ≥ π then
14:        if angle_small ≤ π then
15:            return 0                          ▷ Origin in polygon
16:        else
17:            return cos((2π−angle_small)/2)
18:        end if
19:    else
20:        return cos(angle_big/2)
21:    end if
22: end procedure
```

3.4.1 Performance

Our method provides significant speed-up over existing methods to compute the diamond norm. Indeed, all prior implementations of the diamond norm investigated by the author use the semidefinite programs of [46, 45, 47]. As discussed above, these methods are generally less efficient than the closed-form solution we provide in Section 3.3. Moreover, our method does not use the map's Choi–Jamiołkowski representation, which results in better computational complexity.

To illustrate the improvements driven by our novel approach, we run empirical testing for each method. We generate random unitaries $U, V \in U(\mathcal{X})$ and calculate $\|U \cdot U^\dagger - V \cdot V^\dagger\|_\diamond$ where $\dim(\mathcal{X}) = 2^1, 2^2, \cdots, 2^10$ (i.e. for $1, 2, \cdots, 10$ qubits). We stop testing at four qubits for the semidefinite methods as the calculation lasted too long. We compare our algorithm (Hyper-efficient) with the Qiskit semidefinite implementation (Qiskit) [40], and an alternative implementation of the semidefinite program (Semidefinite) [45] as a sanity check.

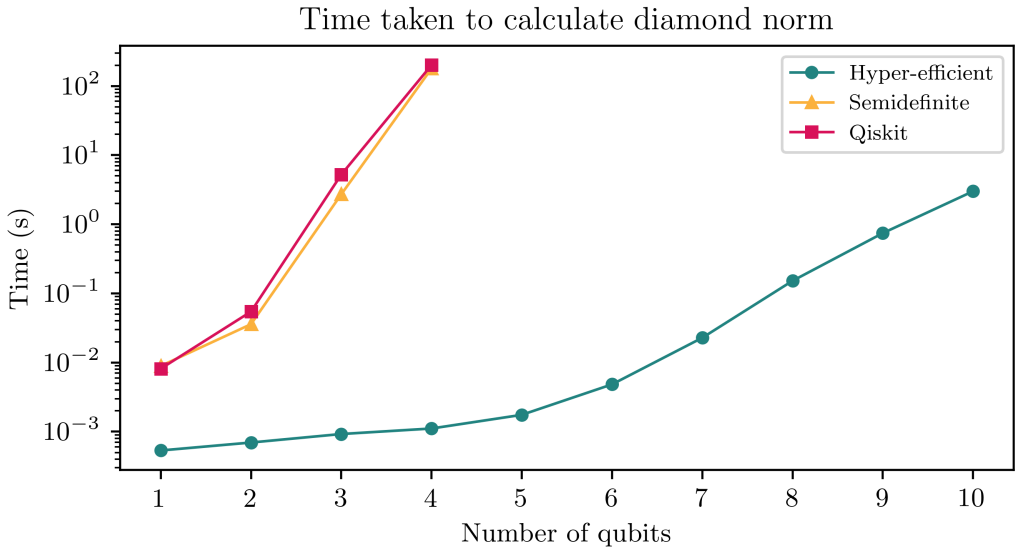


Figure 3.1: Time comparison of different numerical calculations of the diamond norm

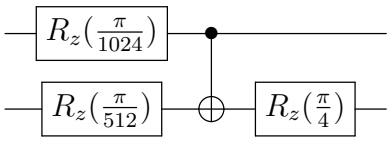
The execution time of the norm was calculated using the Python Standard Library package `timeit`. We run the tests on a 2022 Macbook Air with M2 processor (8 core 2.4-3.5 GHz) and 8 GB LPDDR5 RAM. The `timeit` function performs seven runs of the computation and returns the mean and standard deviation of the execution time. The number of loops per run is determined at runtime. In Figure 3.1 we plot the mean execution times of the seven runs using a logarithmic scale to improve readability. The standard deviation is omitted as it is negligible.

Although the hyper-efficient implementation still has exponential time complexity with respect to the number of qubits, it provides a tremendous gain in performance compared to existing implementations. Indeed, even on single-qubit channels, our method is around 15 times faster on average than Qiskit's method. This jumps to an 181,081-fold improvement on four-qubit channels.

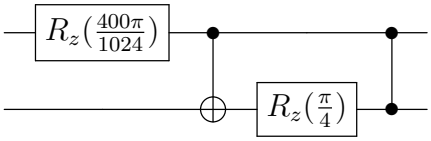
3.4.2 Visualising the Diamond Norm

In addition to providing significant speed-up, our approach enables an intuitive geometric interpretation of the diamond norm. In this section, we illustrate the norm's *modus operandi* on two arbitrary circuits, described below.

Circuit 1:



Circuit 2:



Let U be the 4×4 unitary matrix describing Circuit 1 and V the unitary describing Circuit 2. We plot the eigenvalues of $U^\dagger V$ in Figure 3.2.

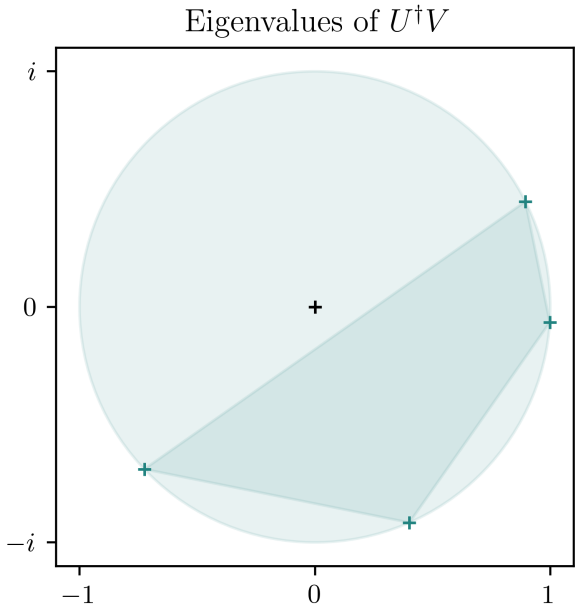


Figure 3.2: Plot of the eigenvalues of $U^\dagger V$ and the polygon formed by their convex hull.

As we can see, the polygon of eigenvalues does not include the origin. This allows us to infer that the diamond distance between the two circuits is not 2 (i.e., they are not maximally distant). In this specific case, the distance d between the origin and the polygon is around 0.1475. This results in a diamond norm of around 1.9781 when plugging it into $2\sqrt{1-d^2}$.

This geometrical interpretation also intuitively illustrates how the diamond norm holds up to a global phase. Applying a global phase to either of our unitary channels amounts to rotating the polygon around the origin. This rotation does not affect the distance d and, hence, the norm's value.

3.4.3 Case Study: Phase-Squashing in Quantum Circuits

In this section, we explore a practical application of our framework by examining phase-squashing in quantum circuits. As previously discussed, the Approximate Quantum Fourier Transform (AQFT) employs phase-squashing to reduce circuit

depth [36]. The underlying concept of this approach is that single-qubit rotation with angles below a certain threshold can be ignored without compromising the calculation. This results in an overall better circuit. We propose to analyse phase-squashing using the mathematical framework we have elaborated.

The phase gate $R_z(\alpha)$, which represents a rotation of α around the z-axis of the Bloch sphere, is represented by the following unitary matrix:

$$U(\alpha) := \left[\!\!\left[-\boxed{R_z(\alpha)}- \right]\!\!\right] = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{bmatrix} \tag{3.8}$$

Phase-squashing replaces a phase gate by the identity. Calculating the resulting approximation error boils down to finding the eigenvalues of $U(\alpha)$ ($\mathbb{1}^\dagger U(\alpha) = U(\alpha)$). Since $U(\alpha)$ is diagonal for the phase gate, the eigenvalues are 1 and $e^{i\alpha}$.

With this in mind, finding an analytical solution to the diamond norm is easy. The distance d is $\cos(\frac{\alpha}{2})$ by a simple geometric argument. Plugging this into the usual identity yields:

$$\|\mathbb{1}_{L(\mathcal{X})} - \Phi\|_\diamond = 2\sqrt{1 - \cos^2\left(\frac{\alpha}{2}\right)} \tag{3.9}$$

Where $\Phi(\cdot) = U(\alpha) \cdot U(\alpha)^\dagger$ is the unitary channel representing the phase gate and $\mathbb{1}_{L(\mathcal{X})}$ is the identity channel on the same operator space. It is clear that the smaller the angle α, the better the approximation error.

In Figure 3.3, we plot both the eigenvalues of $\mathbb{1}^\dagger U(\alpha)$ for different values of α as well as the value of the diamond norm for $\alpha \in [0, 2\pi)$. This illustrates that smaller values of α (i.e., those closest to zero modulo 2π) shrink the distance d, resulting in a better approximation error.

This analysis generalises well to any quantum circuit where the set of eigenvalues can be determined analytically. In general, computing the diamond norm for circuits with many qubits ($n \geq 20$) is intractable, even using our novel approach, as this

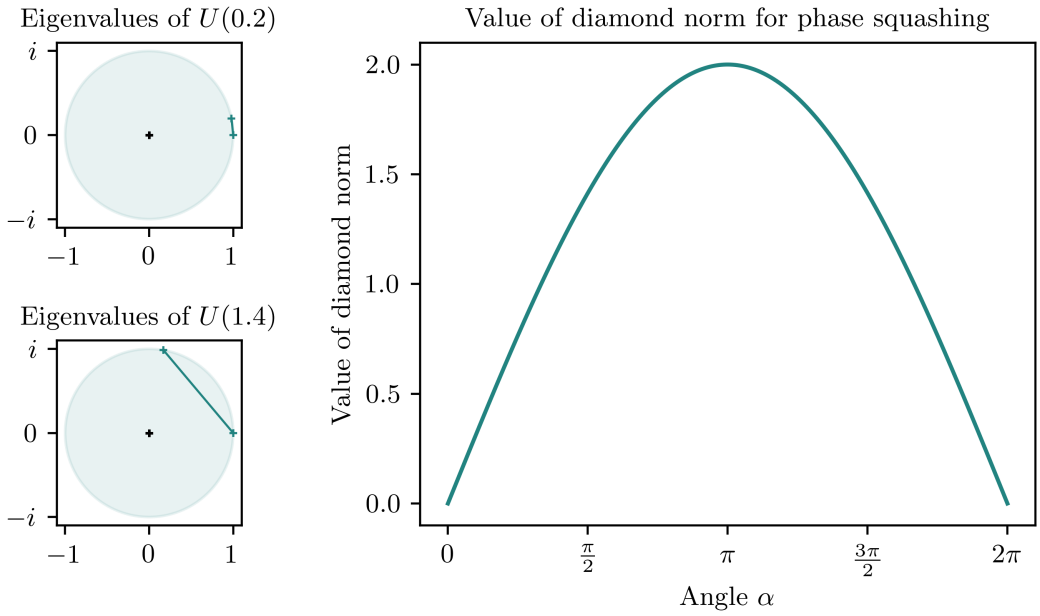


Figure 3.3: Eigenvalues of $U(\alpha)$ for $\alpha = 0.2$ and $\alpha = 1.4$ and value of the phase-squashing diamond approximation error as a function of α.

involves diagonalising an exponentially large matrix. However, some circuits, such as Z-phase gadgets [49], admit analytical solutions for an arbitrary number of qubits (the eigenvalues are $e^{i\frac{\alpha}{2}}$ and $e^{-i\frac{\alpha}{2}}$ in the case of a phase gadget). Finding such circuits is a potential avenue for future research in circuit optimisation, allowing for more global classical optimisation techniques.

Chapter 4

Random Search for Approximate Identities

In chapter 3, we settle on the diamond norm $\|\cdot\|_\diamond$ as a useful metric for quantifying the distance between two quantum channels. After developing an efficient numerical method for evaluating the norm, we can undertake a more pragmatic investigation of the space of quantum circuits up to ϵ-equivalence. This empirical exploration provides a more detailed understanding of the space of approximate quantum circuit equivalences. These results will also corroborate prior theoretical findings in [50]. We begin by briefly detailing the experimental setup used. We follow this with a detailed discussion of the results and how they substantiate theoretical findings.

4.1 Experimental Setup

As discussed in the introduction, the essence of local approximate QCO lies in identifying interesting approximate circuit identities or *rewrite rules*. A sensible initial approach is to conduct an exhaustive or random exploration of the space, as there is no clear heuristic for an informed search. This method will enhance our understanding of the space and potentially generate *useful* rewrites.

This raises the question: What constitutes a useful identity? To be useful, a circuit

identity must meet two key criteria:

1. *Applicable*: The identity can be applied locally in larger circuits during optimisation. Practically, this means it should consist of standard gates and be relatively compact in terms of qubit count and depth.

2. *Promote a heuristic*: The identity should result in an overall better circuit or enable further optimisation, leading to an improved circuit.

These largely qualitative properties will guide our approach in the following chapters.

The identification of useful *exact* identities for QCO has been explored previously. Lomont enumerates all exact circuit identities up to three gates for a standard gate set of 25 gates [51]. More recently, Bravyi et al. exhaustively list the optimal circuits for all six-qubit Clifford group elements.

In the approximate regime, exhaustive exploration of the quantum circuit space quickly becomes impractical as the depth and number of qubits increase. The number of possible circuits grows exponentially with respect to depth and qubit number. To address this, we resort to randomly sampling pairs of circuits with a given depth d and calculating their diamond norm.

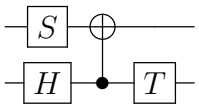
4.1.1 Generating Circuits

We take a straightforward approach to random circuit generation. We pass three parameters to the generator: circuit depth d, number of qubits n, and the gate sets $\mathcal{G}_{\text{single}}$ and $\mathcal{G}_{\text{double}}$. We consider the depth of the circuit to be the number of gates it contains rather than the number of slices. The circuits are generated by choosing a gate, uniformly at random, from the gate set and applying it to one or two of the n qubits randomly.

The generator creates a *standard description* (SD) string which identifies circuits. SD strings are composed of one or more literals of the form GI for $G \in \mathcal{G}_{\text{single}}$ or GI-J

for $\mathcal{G}_{\text{double}}$ with I, J qubit indices.

Setting $\mathcal{G}_{\text{single}}$ and $\mathcal{G}_{\text{double}}$ to the standard Clifford+T gate set, a possible SD string is `S0H1CNOT1-0T1` which encodes the two-qubit circuit



As in [51, 52], we limit the search for circuit identities to discrete gate sets of one or two-qubit gates. This approach is driven by our aim of finding useful identities in the sense discussed above.

4.1.2 Computing Distributions of Norms

After generating random pairs of circuits, we calculate the diamond norm. For each circuit, the corresponding unitary is calculated as described in section 2.2.4 and fed into the hyper-efficient diamond norm algorithm detailed in section 3.3.

We run the random search on two gate sets. The first is the well-known Clifford+T fragment (c.f. section 2.2.4) composed of three single-qubit gates: Hadamard gate H, phase gate S, and T gate T; and one two-qubit gate: controlled-NOT $CNOT$.

We also define a second non-standard gate set for supplementary results. Despite its universality, the Clifford+T fragment comprises gates with large rotation angles ($\geq \frac{\pi}{4}$ radians). Consequently, incorporating a gate with a smaller rotation angle necessitates a notably deep circuit implementation. We call this second gate set the Tiny fragment. It comprises two single-qubit gates, a small z-axis rotation $RzTiny$ and a small x-axis rotation $RxTiny$, and one two-qubit gate: controlled-NOT $CNOT$. $RzTiny$ and $RxTiny$ are rotations of $\frac{\pi}{256}$ radians around the z-axis and x-axis respectively.

We run the computation on a Dell Poweredge R6515 Server with an AMD 7452 2.35 GHz 32 Core (64 Thread) CPU and 256 GB RAM (8 x 32B RDIMM 3200 MT/s).

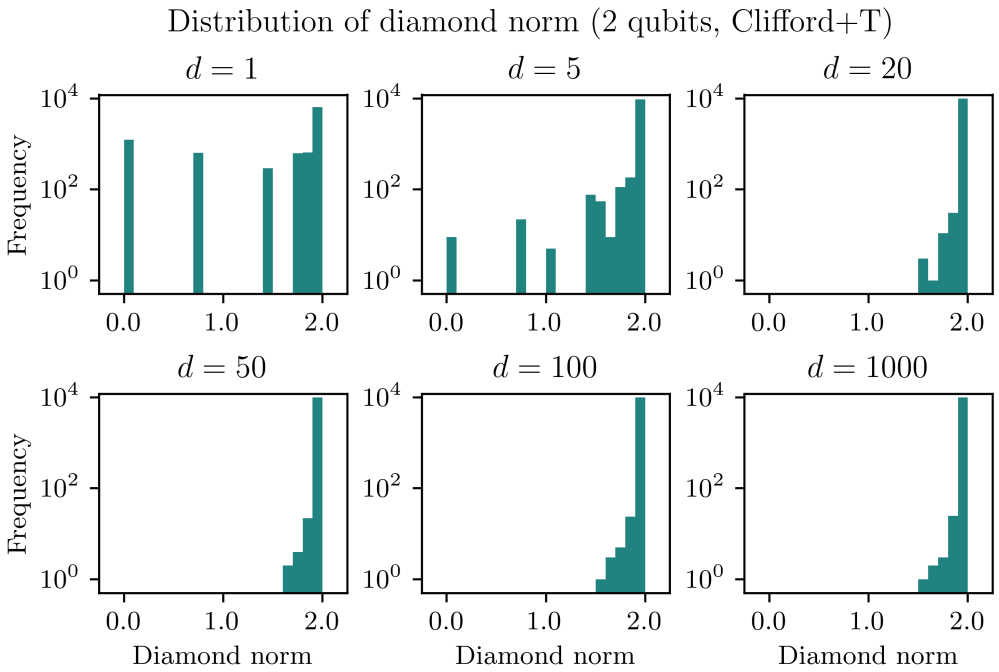


Figure 4.1: Distribution of the diamond Number of qubits distance between random quantum circuits of varying depth

For both gate sets, We run the search on two and three-qubit circuits for depths of 1, 5, 20, 50, 100, and 1000. We run the search for $n = 10000$ iterations and save the diamond distance for each pair.

4.2 Results

After running the random search described in the previous section, we plot the distribution of the diamond norm at each depth. For approximate QCO, we hope to find circuit identities with a small diamond norm that are nevertheless not exact identities.

Figures 4.1 and 4.2 illustrate the distribution of the diamond norm for two and three-qubit Clifford+T circuits. The y-axis, representing frequency, is plotted on a logarithmic scale. Several key observations can be made from these distributions.

First, circuits with relatively shallow depths ($d = 1, 5$) generate many identities with small diamond distances (≤ 0.1). Unfortunately, the histogram is somewhat mis-

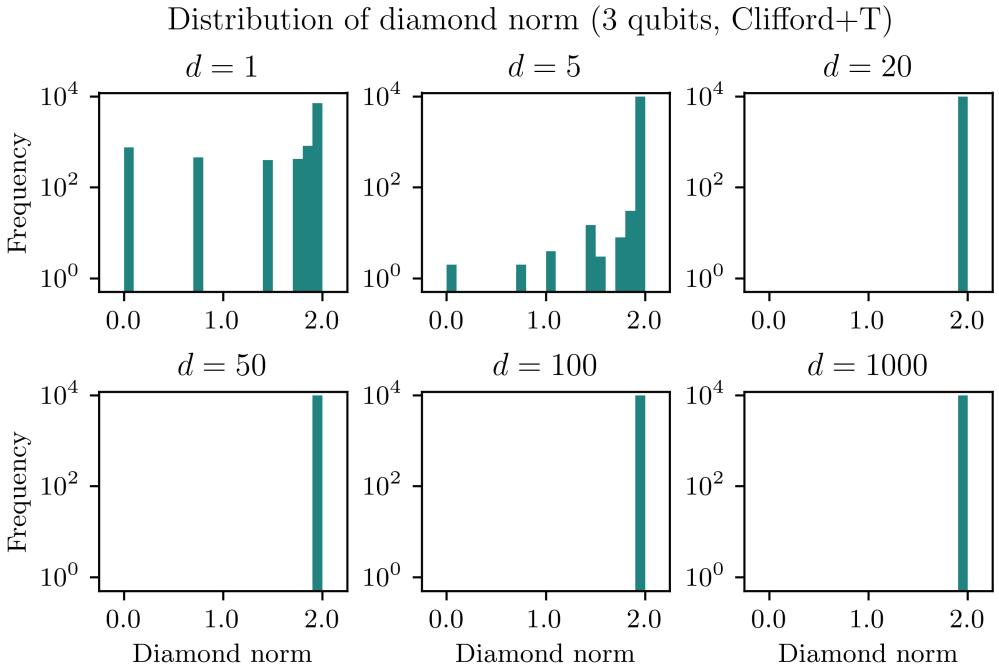


Figure 4.2: Distribution of the diamond Number of qubits distance between random quantum circuits of varying depth

leading. For $d = 1$, additional analysis reveals six distinct diamond norm values of approximately 0.0, 0.765, 1.414, 1.732, 1.848, 1.925, and 2.0. Therefore, the bin between 0 and 0.1 contains only diamond norm values that are precisely 0, i.e., representing exact circuit identities[1].

Second, in the two-qubit case, we observe that a few circuit pairs are not maximally distant at higher depths ($d \geq 20$). This contrasts with the three-qubit case, where all pairs are maximally distant at depths $d \geq 20$.

Lastly, we note the existence of discrete regular bands in the distribution at shallow depths. This regularity highlights specific structured relationships within the circuit identities at these depths, which diminishes as circuit depth increases.

Figure 4.3 plots the diamond distance for two-qubit "Tiny" circuits at various depths. Unlike the Clifford+T case, the distribution here appears more continuous. This can

[1]Note that we do not filter out trivial identities when searching. They thus constitute a majority of the exact identities identified by the search.

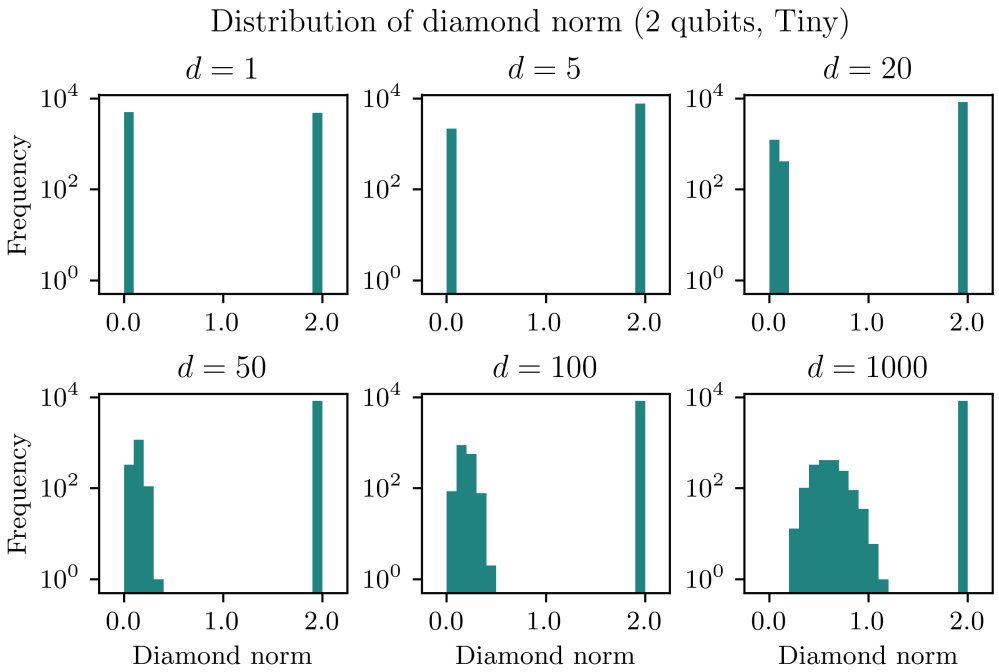


Figure 4.3: Distribution of the diamond Number of qubits distance between random quantum circuits of varying depth

be attributed to the higher granularity in rotation angles, which results in a more continuous-looking distribution.

At all depths, we observe a distinct band in the distribution containing maximally distant circuit pairs. We suspect this results from the addition of the CNOT gate, which induces a rotation by a large angle (x-axis rotation of π radians). The presence of such a significant rotation contributes to the formation of these maximally distant pairs, as it introduces considerable separation in the state space between circuits.

4.3 Discussion

This initial empirical exploration of the space of approximate circuit equivalences provides several insights. Ultimately, the results are rather discouraging for local approximate QCO. Unsurprisingly, there doesn't seem to be an abundance of useful approximate circuit identities within standard gate set.

Random exploration of the Clifford+T space of shallow circuits typically yields either exact identities or identities too distant to be of practical use, with an approximation error of $|\cdot|_\diamond \approx 0.765$. Exact equalities are of little interest to us, as they have already been extensively studied in the context of exact QCO.

Section 4.2 reported on the presence of discrete bands in the distribution for shallow circuits. Using our analysis in section 3.4.3, we can show these are artefacts of phase-squashing. We know the T gate is essentially a phase rotation by $\frac{\pi}{4}$. Plugging this into Equation 3.9 gives us:

$$\|\mathbb{1}_{L(\mathcal{X})} - T \cdot T^\dagger|_\diamond = 2\sqrt{1 - \cos^2\left(\frac{\pi}{8}\right)} \approx 0.765 \tag{4.1}$$

Doing the same analysis on the S gate yields a diamond distance of ≈ 1.414, both of which are bands in our distribution. This suggests that phase squashing may play a larger role in approximate QCO than initially assumed. This hypothesis will be discussed further in chapter 5

As circuit depth increases, we might hope to find more close identities. Indeed, the Solovay–Kitaev theorem[2] [43] tells us that the Clifford+T gate set can efficiently implement small rotation gates, resulting in circuits with small diamond distances (c.f. section 3.4.3). However, implementing these small rotations still necessitates a certain circuit depth.

At these higher depths, however, the vastness of the circuit space means that random search predominantly yields maximally distant circuits, significantly reducing the practical utility of random exploration for identifying useful approximate circuit identities.

To mitigate this, we also conduct a search over the "Tiny" gate set. This has two effects. On the one hand, small rotations can be implemented using only one gate,

[2]Solovay and Kitaev's statement of the theorem uses the operator norm rather than the diamond norm, but we ignore this detail in our discussion

meaning we find close approximate circuit identities at shallower depths ($d \leq 20$). On the other hand, because the gate set is already composed of small rotation gates, the search becomes more directed, obviating the need to randomly search the circuit space in hopes of finding close pairs. The added granularity from small rotations constrains the search, inherently generating circuits that are close (modulo the effects of controlled-NOT).

These empirical results largely corroborate a theoretical result in [50]:

Proposition 4.3.1 (Nechita 19)**.** *Let $U, V \in U(\mathcal{X})$ be Harr random unitary operators for $\dim(\mathcal{X}) = d$. Then:*

$$\mathbb{P}\left(\|U \cdot U^\dagger - V \cdot V^\dagger\|_\diamond = 2\right) \geq 1 - \exp\left(-\frac{\log 2}{2} d^2\right)$$

Note that Proposition 4.3.1 requires the unitaries to be sampled from a Harr random distribution. Further work is necessary to show that our sampling strategies meet this requirement. To do this, we use the notion of expressibility as it is defined in [32, 53].

For our purposes, it is sufficient to show that the ensemble we sample from forms a 1-design, meaning it replicates the properties of the Haar random distribution for polynomials of degree one or less. Consider an ansatz that generates a unitary ensemble. The extent to which this ensemble approximates a t-design is referred to as the expressibility of that ansatz. This measure is generally defined in terms of the following superoperator [32, 53]:

$$\mathcal{A}_{\mathbb{U}}^{(t)}(\cdot) = \int_{\mathcal{U}(d)} d\mu(V) V^{\otimes t}(\cdot)(V^\dagger)^{\otimes t} V^{\otimes t}(\cdot)(V^\dagger)^{\otimes t} - \int_{\mathbb{U}} dU U^{\otimes t}(\cdot)(U^\dagger)^{\otimes t} \tag{4.2}$$

where $\mathcal{U}(d)$ is the uniform Harr distribution of dimension d, $d\mu(V)$ is its associated volume element, and dU is the volume element corresponding to the uniform distribution of $\mathbb{U}$. When we have $\mathcal{A}_{\mathbb{U}}^{(t)}(X) = 0$ for all operators $X \in L(\mathcal{X})$, $\mathbb{U}$ forms a t-design.

Using this superoperator, we can define an operator-independent measure of expressibility using the diamond norm. From now on, when we speak of the expressibility of an ansatz, we refer to:

$$\epsilon_{\mathbb{U}}^{t} = \|\mathcal{A}_{\mathbb{U}}^{(t)}\|_{\diamond} \tag{4.3}$$

If $\epsilon_{\mathbb{U}}^{t} = 0$ the ensemble $\mathbb{U}$ forms a t-design.

Holmes et al. [32] demonstrate that ansatz expressibility is directly correlated to its depth. As circuit depth increases, so does the expressibility. Higher expressibility, in turn, is correlated with a higher probability of sampling maximally distant unitary channels.

We observe this phenomenon in our results, most obviously in the Clifford+T case. Our sampling strategies, particularly in the Tiny case, significantly deviate from Haar random sampling and likely do not even qualify as a 1-design. Nevertheless, as we increase the depth of the circuits and approach a 1-design, Proposition 4.3.1 becomes more applicable. This effect is most pronounced in Figure 4.2 for circuits with $d = 1000$. Our results indicate that as the number of qubits and circuit depth increase, the distribution skews more towards maximally distant identities.

In conclusion, these results are rather pessimistic for approximate QCO. The search within the Clifford+T gate set fails to yield identities that are applicable and promote a heuristic. Although it is likely that close identities exist within this gate set, our empirical exploration suggest these will be deep circuits with limited applications in QCO.

Chapter 5

Informed Unit Search

The random exploration of approximate circuit equivalences in Chapter 4 did not yield any useful identities for quantum circuit optimisation (QCO). It appears that identifying interesting circuit identities requires moving beyond the Clifford+T gate set and utilising more general parameterised rotation gates.

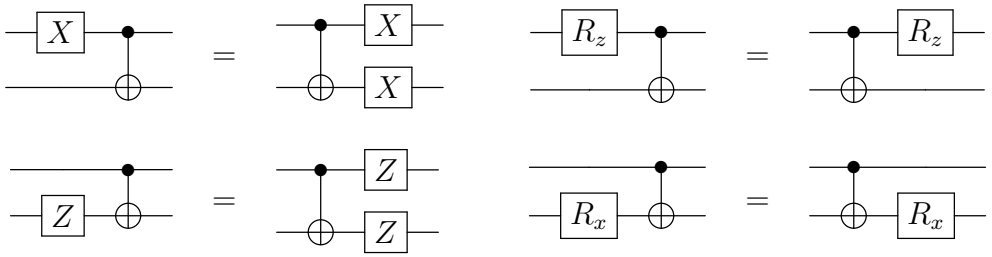
As we are no longer working with a small discrete gate set, an exhaustive approach, even for very short circuits ($d = 5$), becomes infeasible. Therefore, we propose a more informed approach.

The key idea is to construct a parameterised circuit identity, where parameters are the rotation angles for single-qubit rotation gates, which is likely to result in a useful identity. Specifically, we focus our efforts on an identity that would allow us to approximately commute gates past each other. We call this parameterised identity a *unit*.

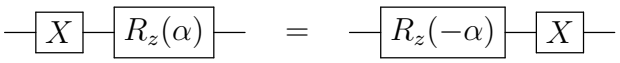
Using this unit, we can run a search on the parameter space at different resolutions to find an identity that minimises the diamond approximation error.

5.1 Experimental Setup

An essential component of our approach is selecting a suitable unit for minimisation. In exact QCO, a particularly useful set of circuit identities is called *commutation rules* [54]. These rules enable a compiler to push commuting gates past one another. For instance, we can push single-qubit z- and x-axis rotation gates through controlled-NOT gates using the following rules:



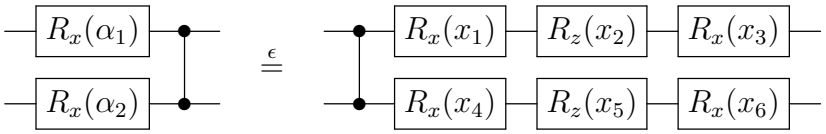
Likewise, we can push an X gate through an arbitrary z-axis rotation with the following identity:



(5.1)

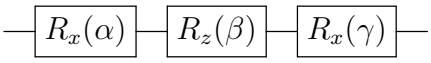
Based on these examples, a potential strategy for approximate QCO is to identify approximate analogues to these *commutation rules*. Ideally, we aim to find general parameterised rules applicable in QCO, but even *ad hoc* circuit identities can serve as proofs of concept. The rules above apply when the commuting gate has a rotation angle of π or any rotation angle. We anticipate that commuting gates with different angles where no exact commutation rule exists will yield interesting approximation rules.

The unit we use for the informed search attempts to commute single-qubit x-axis rotation gates past the controlled-phase gate (denoted as ). For fixed parameters α_1 and α_2 the search attempts to minimise ϵ in:

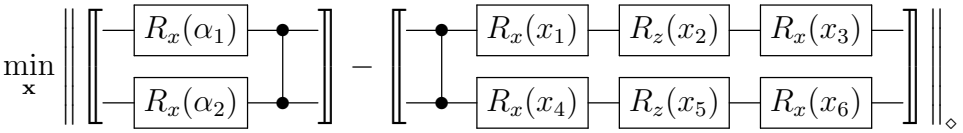


(5.2)

Where $x_i \in [0, 2\pi)$. Note that by the Euler decomposition for single-qubit quantum gates, it is possible to express any single qubit rotation by the following circuit:

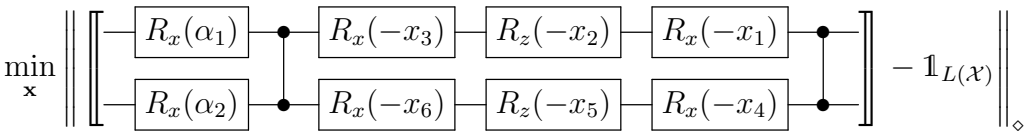


We use this general formula in our unit following the controlled-phase. Formulated as a minimisation problem, we solve:



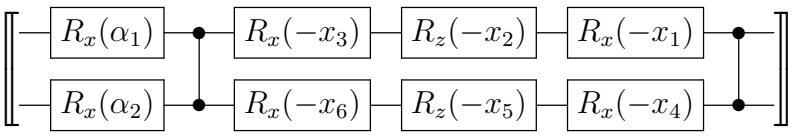
(5.3)

Where $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$ and $[\![C]\!]$ is the unitary channel represented by the an n-qubit circuit C. We can simplify this search using the unitary invariance of the diamond norm. Multiplying the left circuit by the adjoint of right one gives:



(5.4)

To calculate the approximation error for parameters $\mathbf{x}$ it suffices to diagonalise:



To run the minimisation, we discretise the parameter space and run an exhaustive search. We subdivide the continuous range of rotation parameters x_i into a finite set of discrete values, enabling an evaluation of the diamond norm approximation error at each defined point.

Concretely, for each rotation parameter x_i we partition the the interval $[0, 2\pi)$ into 2^r angles, where r is the resolution. This results in 2^r values, for each x_i, expressed as

$$x_i = \frac{k\pi}{2^r}$$

where r ranges from 0 to $2^r - 1$.

Unit	α_1	α_2
1	5.27	2.47
2	4.9	0.9
3	0.8	0.9
4	4.8	4.9
5	4.8	3.9
6	1.5707	4.7123
7	0.0	4.7123
8	5.076438	3.98654
9	0.0076	0.003

Table 5.1: Table of fixed parameters for each candidate unit.

Alternative optimisation strategies were trialled, but fail. We hypothesise this is due to the barren plateau phenomenon discussed in chapter 1. As the value of the diamond approximation error evaluates to 2 for a vast majority of the parameter space, agnostic optimisation strategies relying on function gradients fail.

We conduct the minimisation for 10 units at a resolution of 5. The fixed parameters for these units are listed in Table 5.1. The computations are performed on a Dell PowerEdge R6515 Server equipped with an AMD 7452 2.35 GHz 32 Core (64 Thread) CPU and 256 GB RAM (8 x 32B RDIMM 3200 MT/s). The computation is parallelised to take advantage of all 32 cores.

In addition to the exhaustive search at resolution 5, we run a local search for unit number 9. As the fixed parameters for this unit are very small, we anticipate the minimising parameters to be situated around 0. We therefore run the search for $x_i \in [-0.1, 0.1)$ at a resolution of 10.

5.2 Results

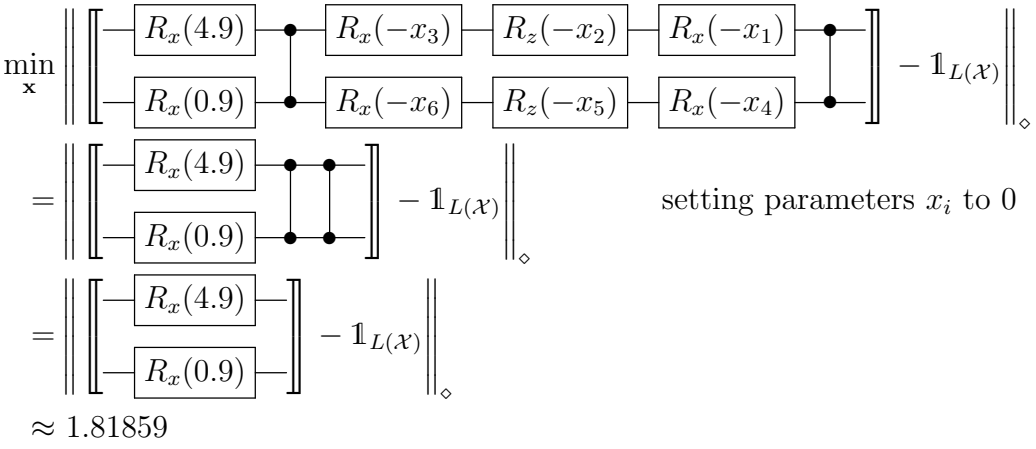
We present the findings of our exhaustive search conducted at a resolution of 5, tabulated in Table 5.2. The optimal solution for each unit $\mathbf{x}_{\min}$ and the corresponding diamond approximation error are provided.

Unit	Diamond Norm	x_1	x_2	x_3	x_4	x_5	x_6
1	≈ 1.49247	0	π	0	0	0	π
2	≈ 1.81859	0	0	0	0	0	0
3	≈ 1.50256	0	0	0	0	0	0
4	≈ 1.98109	0	0	0	0	0	0
5	≈ 1.80089	0	π	0	0	0	π
6	≈ 1.99999	0	π	0	0	0	π
7	≈ 1.414150	0	π	0	0	0	π
8	≈ 1.710307	0	π	0	0	0	π
9	≈ 0.010599	0	0	0	0	0	0

Table 5.2: Table of search results.

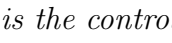
These results provide us with a number of insights. Notably, we observe that optimal parameters for all units are restricted to either 0 or π. Remarkably, for several of the units, the optimal commutation rule consists in deleting the x-axis rotation gate rather than commuting anything through. This observation lends further support to the hypothesis, discussed in section 4.3, that phase-squashing plays a pivotal role in approximate QCO.

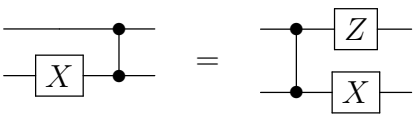
Note that we can extend the notion of phase-squashing to any single-qubit rotation gate. In our case, the optimal approximation error is reached when the single-qubit R_x rotations are squashed. We break down the approach, taking unit 2 as an example:



However, not all optimal solutions are $\mathbf{x}_{\min} = \mathbf{0}$. Although this initially suggests a different approach from phase-squashing, we shown that these cases can also be re-

duced to phase-squashing. We use the following lemma, which can be proved trivially in the ZX-calculus [41, 55]:

Lemma 5.2.1. *Where*  *is the controlled phase, and*  *and* *are the Pauli X and Z gates respectively, we have the following equality:*



Taking unit 1 as example, we show that, the commutation rule yielded by the search reduces to phase-squashing.

$$
\begin{aligned}
\min_{\mathbf{x}} & \left\| \left[\!\!\left[\begin{array}{cccccccc} R_x(5.27) & \bullet & R_x(-x_3) & R_z(-x_2) & R_x(-x_1) & \bullet \\ R_x(2.47) & \bullet & R_x(-x_6) & R_z(-x_5) & R_x(-x_4) & \bullet \end{array} \right]\!\!\right] - \mathbb{1}_{L(\mathcal{X})} \right\|_{\diamond} \\
= & \left\| \left[\!\!\left[\begin{array}{ccccc} R_x(5.27) & \bullet & R_z(\pi) & \bullet \\ R_x(2.47) & \bullet & R_x(\pi) & \bullet \end{array} \right]\!\!\right] - \mathbb{1}_{L(\mathcal{X})} \right\|_{\diamond} \\
= & \left\| \left[\!\!\left[\begin{array}{ccccc} R_x(5.27) & \bullet & Z & \bullet \\ R_x(2.47) & \bullet & X & \bullet \end{array} \right]\!\!\right] - \mathbb{1}_{L(\mathcal{X})} \right\|_{\diamond} \\
= & \left\| \left[\!\!\left[\begin{array}{ccccc} R_x(5.27) & & \bullet & \bullet \\ R_x(2.47) & X & \bullet & \bullet \end{array} \right]\!\!\right] - \mathbb{1}_{L(\mathcal{X})} \right\|_{\diamond} \qquad \text{by Lemma 5.2.1} \\
\approx & \left\| \left[\!\!\left[\begin{array}{c} R_x(5.27) \\ R_x(-0.67) \end{array} \right]\!\!\right] - \mathbb{1}_{L(\mathcal{X})} \right\|_{\diamond} \\
\approx & \ 1.49247
\end{aligned}
$$

These findings corroborate the hypothesis that phase-squashing up to π emerges as the predominant strategy for local approximate circuit compilation.

Further exploration using local search with resolution 10 on unit 9 reveals no additional insights, as the optimal parameters remain at zeros. This indicates that the absence of a novel approach cannot be attributed to a coarse discretisation of the parameter space.

Finally, Figure 5.1 illustrates a simplified optimisation landscape with respect to two parameter pairs, (x_1, x_2) and (x_1, x_3). Although primarily qualitative, it offers a useful albeit incomplete depiction of the optimisation landscape. As anticipated optimal values are situated at 0. The canyon visible in the (x_1, x_3) plot rises beacuse to the opposing angles of the two rotations, resulting in their mutual cancellation.

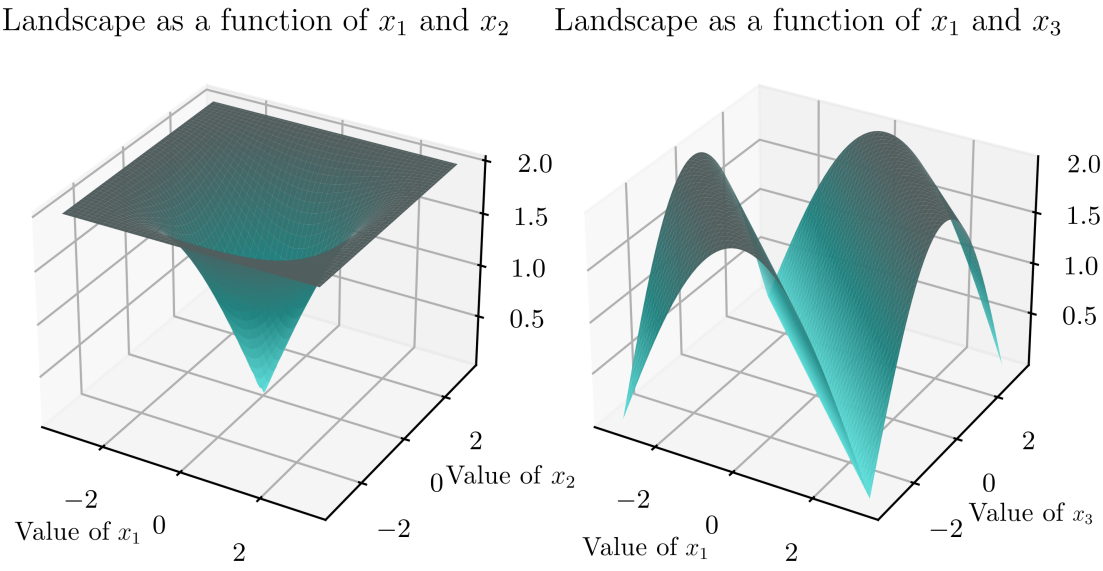


Figure 5.1: 3D plots depicting the diamond error for unit 9 with varying parameters (x_1, x_2) and (x_1, x_3). All other x_i parameters remain constant at 0.

Chapter 6

Conclusion

In this thesis, we undertake a comprehensive study on the feasibility of *local approximate QCO*. We began by elaborating a rigorous mathematical framework to quantify the "distance" or closeness of two quantum channels in chapters 2 and 3.

This prompted a detailed analysis of the diamond norm $\|\cdot\|_\diamond$, an operationally meaningful metric discovered by Kitaev [43]. After reviewing its properties, we provided a novel proof of an unproved result in [38] in section 3.2.

This proof enabled the design of an innovative algorithm for evaluating the diamond norm of a difference of unitary channels. We achieved a speedup of six orders of magnitude on 4-qubit channels compared to existing implementations of the norm.

Using the novel interpretation of the diamond norm, we illustrated its operation with respect to phase-squashing and derived an analytical formula for the approximation error.

In chapters 4 and 5, we apply these theoretical results in an extensive empirical exploration of the space of approximate circuit identities. This exploration served two primary purposes: corroborating existing theoretical results, providing a quantitative understanding of the space, and deriving practical insights on the feasibility of local approximate QCO.

The primary takeaway from this exploration was that contrary to initial assumptions, there are few shallow circuit identities which have practical use in QCO. Despite considerable effort, all the identities we identified reduced to phase-squashing up to π. This did not imply the absence of alternative optimisation strategies; however, it is unlikely that such strategies would be found in shallow circuits using standard gate sets.

Although somewhat discouraging for approximate QCO, these results still have practical relevance. As discussed in chapter 1, outside of the AQFT, approximate QCO has rarely been used and has yet to be implemented in quantum compilation libraries. Implementing phase-squashing in standard compilers, using the analytical formula of section 3.4.3, would be trivial and might significantly improve the performance of these methods.

A natural step to improve this work is further exploration of the space of approximate identities. Moreover, while our empirical investigation provides valuable insights, it is primarily qualitative. Providing theoretical results supporting the phase squash to π hypothesis would be a significant improvement.

More pragmatically, integrating the phase-squash to π strategy into mainstream QCO libraries holds promise for enhancing their efficacy, particularly when coupled with exact optimisation methodologies. This integration has the potential to yield substantial enhancements in performance, resulting in tangible gains in practical quantum computing applications.

References

[1] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge: Cambridge University Press, 10th anniversary ed. ed., 2010.

[2] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, no. 6, pp. 467–488, 1982.

[3] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry," *Reviews of Modern Physics*, vol. 92, no. 1, p. 015003, 2020.

[4] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.

[5] L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack," *Physical review letters*, vol. 79, no. 2, p. 325, 1997.

[6] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, "Quantum computation by adiabatic evolution," *arXiv preprint quant-ph/0001106*, 2000.

[7] Y. Alexeev, D. Bacon, K. R. Brown, R. Calderbank, L. D. Carr, F. T. Chong, B. DeMarco, D. Englund, E. Farhi, B. Fefferman, *et al.*, "Quantum computer systems for scientific discovery," *PRX quantum*, vol. 2, no. 1, p. 017001, 2021.

[8] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[9] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.

[10] J. Preskill, "Quantum computing in the nisq era and beyond," *Quantum*, vol. 2, p. 79, 2018.

[11] J. Preskill, "Quantum computing and the entanglement frontier," *arXiv preprint arXiv:1203.5813*, 2012.

[12] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," *arXiv preprint arXiv:1411.4028*, 2014.

[13] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New Journal of Physics*, vol. 18, no. 2, p. 023023, 2016.

[14] F. Fürrutter, G. Muñoz-Gil, and H. J. Briegel, "Quantum circuit synthesis with diffusion models," *arXiv preprint arXiv:2311.02041*, 2023.

[15] P. Rakyta and Z. Zimborás, "Approaching the theoretical limit in quantum gate decomposition," *Quantum*, vol. 6, p. 710, 2022.

[16] Y.-H. Zhang, P.-L. Zheng, Y. Zhang, and D.-L. Deng, "Topological quantum compiling with reinforcement learning," *Physical Review Letters*, vol. 125, no. 17, p. 170501, 2020.

[17] G. Meuli, F. Mozafari, B. Schmitt, H. Riener, M. Soeken, and G. De Micheli, "Quantum compilation." `https://si2.epfl.ch/~demichel/research/quantum.html`. Accessed: 07/05/2024.

[18] T. Fösel, M. Y. Niu, F. Marquardt, and L. Li, "Quantum circuit optimization with deep reinforcement learning," *arXiv preprint arXiv:2103.07585*, 2021.

[19] M. Amy, D. Maslov, and M. Mosca, "Polynomial-time t-depth optimization of clifford+t circuits via matroid partitioning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 10, pp. 1476–1489, 2014.

[20] A. Barenco, A. Ekert, K.-A. Suominen, and P. Törmä, "Approximate quantum fourier transform and decoherence," *Physical Review A*, vol. 54, no. 1, p. 139, 1996.

[21] M. Amy and M. Mosca, "T-count optimization and reed–muller codes," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4771–4784, 2019.

[22] L. E. Heyfron and E. T. Campbell, "An efficient quantum compiler that reduces t count," *Quantum Science and Technology*, vol. 4, no. 1, p. 015004, 2018.

[23] F. Zhang and J. Chen, "Optimizing t gates in clifford+t circuit as $\pi/4$ rotations around paulis," *arXiv preprint arXiv:1903.12456*, 2019.

[24] A. Kissinger and J. van de Wetering, "Pyzx: Large scale automated diagrammatic reasoning," *arXiv preprint arXiv:1904.04735*, 2019.

[25] A. Kissinger and J. van de Wetering, "Reducing the number of non-clifford gates in quantum circuits," *Physical Review A*, vol. 102, no. 2, p. 022406, 2020.

[26] M. Ostaszewski, L. M. Trenkwalder, W. Masarczyk, E. Scerri, and V. Dunjko, "Reinforcement learning for optimization of variational quantum circuit architectures," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18182–18194, 2021.

[27] L. Madden and A. Simonetto, "Best approximate quantum compiling problems," *ACM Transactions on Quantum Computing*, vol. 3, no. 2, pp. 1–29, 2022.

[28] N. F. Robertson, A. Akhriev, J. Vala, and S. Zhuk, "Escaping barren plateaus in approximate quantum compiling," *arXiv preprint arXiv:2210.09191*, 2022.

[29] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, "Quantum-assisted quantum compiling," *Quantum*, vol. 3, p. 140, 2019.

[30] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," *Nature communications*, vol. 12, no. 1, p. 6961, 2021.

[31] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, "Noise resilience of variational quantum compiling," *New Journal of Physics*, vol. 22, no. 4, p. 043006, 2020.

[32] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," *PRX Quantum*, vol. 3, no. 1, p. 010313, 2022.

[33] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature communications*, vol. 9, no. 1, p. 4812, 2018.

[34] Y. Nam, Y. Su, and D. Maslov, "Approximate quantum fourier transform with $O(n \log(n))$ t gates," *NPJ Quantum Information*, vol. 6, no. 1, p. 26, 2020.

[35] D. Deutsch, "Unpublished report on aqft." Referenced in [20].

[36] D. Coppersmith, "An approximate fourier transform useful in quantum factoring," Tech. Rep. RC 19642, IBM, December 1994.

[37] Y. Nam and R. Blümel, "Analytical formulas for the performance scaling of quantum processors with a large number of defective gates," *Physical Review A*, vol. 92, no. 4, p. 042301, 2015.

[38] D. Aharonov, A. Kitaev, and N. Nisan, "Quantum circuits with mixed states," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 20–30, 1998.

[39] J. R. Johansson, P. D. Nation, and F. Nori, "Qutip: An open-source python framework for the dynamics of open quantum systems," *Computer Physics Communications*, vol. 183, no. 8, pp. 1760–1772, 2012.

[40] Qiskit contributors, "Qiskit: An open-source framework for quantum computing," 2023.

[41] B. Coecke and A. Kissinger, *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2017.

[42] J. Watrous, *The theory of quantum information*. Cambridge university press, 2018.

[43] A. Y. Kitaev, "Quantum computations: algorithms and error correction," *Russian Mathematical Surveys*, vol. 52, no. 6, p. 1191, 1997.

[44] A. Gilchrist, N. K. Langford, and M. A. Nielsen, "Distance measures to compare real and ideal quantum processes," *Physical Review A*, vol. 71, no. 6, p. 062310, 2005.

[45] J. Watrous, "Semidefinite programs for completely bounded norms," *arXiv preprint arXiv:0901.4709*, 2009.

[46] J. Watrous, "Simpler semidefinite programs for completely bounded norms," *arXiv preprint arXiv:1207.5726*, 2012.

[47] N. Johnston, D. W. Kribs, and V. I. Paulsen, "Computing stabilized norms for quantum operations via the theory of completely bounded maps," *arXiv preprint arXiv:0711.3636*, 2007.

[48] A. Ben-Aroya and A. Ta-Shma, "On the complexity of approximating the diamond norm," *arXiv preprint arXiv:0902.3397*, 2009.

[49] A. Cowtan, S. Dilkes, R. Duncan, W. Simmons, and S. Sivarajah, "Phase gadget synthesis for shallow circuits," *arXiv preprint arXiv:1906.01734*, 2019.

[50] I. Nechita, Z. Puchała, Ł. Pawela, and K. Życzkowski, "Almost all quantum channels are equidistant," *Journal of Mathematical Physics*, vol. 59, no. 5, 2018.

[51] C. Lomont, "Quantum circuit identities," *arXiv preprint quant-ph/0307111*, 2003.

[52] S. Bravyi, J. A. Latone, and D. Maslov, "6-qubit optimal clifford circuits," *npj Quantum Information*, vol. 8, no. 1, p. 79, 2022.

[53] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019.

[54] T. Itoko, R. Raymond, T. Imamichi, A. Matsuo, and A. W. Cross, "Quantum circuit compilers using gate commutation rules," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*, pp. 191–196, 2019.

[55] J. van de Wetering, "Zx-calculus for the working quantum computer scientist," *arXiv preprint arXiv:2012.13966*, 2020.