Look at Chapter 4 for the upper bounds stuff
in particular, page 53

Classical Simulations of Quantum Systems Using Stabilizer Decompositions

by

Hammam Qassim

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Barry Sanders
Professor, Dept. of Physics and Astronomy, University of Calgary

Supervisor(s): Joseph Emerson
Professor, Dept. of Applied Mathematics, University of Waterloo

Joel Wallman
Professor, Dept. of Applied Mathematics, University of Waterloo

Internal-External Member: Pierre-Nicholas Roy
Professor, Dept. of Chemistry, University of Waterloo

Other Member(s): Raymond Laflamme
Professor, Dept. of Physics and Astronomy, University of Waterloo

Richard Cleve
Professor, School of Computer Science, University of Waterloo

David Cory
Professor, Dept. of Chemistry, University of Waterloo

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

One of the state of the art techniques for classically simulating quantum circuits relies on approximating the output state of the circuit by a superposition of stabilizer states. If the number of non-Clifford gates in the circuit is small, such simulations can be very effective. This thesis provides various improvements in this framework. First, we describe an improved method of computing approximate stabilizer decompositions, which reduces the time cost of computing a single term in the decomposition from $O(\ell n^2)$ to $O(mn^2)$, where ℓ is the total number of gates in the circuit, and m is the number of non-Clifford gates. Since this subroutine has to be repeated exponentially many times, this improvement can be significant in practice whenever $\ell \gg m$. Our method uses a certain re-writing of the circuit, which in some cases allows for a significant amelioration of the exponential scaling of the required classical resources.

Furthermore, we describe a method of constructing exact, low-rank stabilizer decompositions of $|\psi\rangle^{\otimes m}$, where $|\psi\rangle$ is either a magic state or an equatorial state. For any single qubit magic state $|\psi\rangle$, we find stabilizer decompositions of $|\psi\rangle^{\otimes m}$ with $2^{m \log_2(3)/4}$ terms. This improves on the best known bound of $2^{m \log_2(7)/6}$. Similarly, for any single qubit equatorial state $|\psi\rangle$, we give a stabilizer decomposition of $|\psi\rangle^{\otimes m}$ with $2^{m/2}$ terms. To our knowledge no such decompositions were previously known. These results translate to milder exponential scaling of the classical resources required for estimating probabilities of quantum circuits up to a polynomially small multiplicative error, as well as allowing more types of circuits to be simulated in this way.

We also consider certain obstructions to classical simulations. It has been argued in various contexts that contextuality and non-locality hamper classical simulations of quantum circuits. Linear constraint systems (LCSs) are a generalization of the well-known Peres-Mermin magic square, which has been recently used to prove a separation between the power of constant depth classical and quantum circuits.

While binary LCSs have been studied in detail, d-ary LCSs are less well-understood. In this thesis we consider linear constraint systems modulo $d > 2$. We give a simple proof, of the previously known fact, that any linear constraint system which admits a quantum solution consisting of generalized Pauli observables in odd dimension must be classically satisfiable. We further prove that, for odd d, if a Pauli-like commutation relation between two variables in the LCS arises, then it has no quantum solutions in any dimensions, in stark contrast to the even d case. We apply this result to various examples, for instance showing that many generalizations of the Peres-Mermin magic square do not give rise to a quantum vs. classical satisfiability gap.

Acknowledgements

I would like to thank my supervisor Joseph Emerson and co-supervisor Joel Wallman for their continuous help, patience, and support during my graduate program. I also thank IQC faculty members Richard Cleve, David Gosset, John Watrous, and William Slofstra for being generous with their time and advice. I would also like to thank all the friends who made my time in Waterloo an unforgettable experience.

Dedication

This thesis is dedicated to my parents.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

There is overwhelming evidence that quantum computers are more powerful than classical computers. This evidence arises from strong complexity-theoretic conjectures, such as the non-collapse of the polynomial hierarchy, as well as independently of such conjectures in relativized models of computation. Despite this, seemingly innocuous restrictions on the quantum computation can make it easy to simulate classically. One of the most well-known examples is polynomial-sized quantum circuits consisting of Clifford gates and Pauli measurements. Such circuits are efficiently simulable by the celebrated Gottesman-Knill theorem [1]. Augmenting this model with the ability to implement any non-Clifford gates, for example a $\pi/8$ Z-rotation on a fixed qubit, allows for universal quantum computation. Another example is nearest-neighbour matchgate circuits, which encode the physics of non-interacting fermions [2, 3]. Such circuits can be simulated efficiently on a classical computer, although adding the ability to coherently swap nearest neighbour qubits allows for universal quantum computation [4].

For certain restricted quantum computations that can be simulated classically, the associated classical simulation algorithms become inefficient as we relax the restriction. For slight deviations the required classical resources may scale exponentially in a different parameter than the number of qubits. For instance, one can classically simulate an n-qubit Clifford circuit augmented with m non-Clifford gates using classical resources that scale as $4^m\text{poly}(n)$, as shown in Aaranson & Gottesman [5]. Since such circuits are universal for quantum computation[1], this allows for classically simulating universal quantum circuits

[1] provided we take $m = \Omega(n)$.

with an amount of resources that scales exponentially only in the number of non-Clifford gates. The works of Bravyi, Smith & Smolin [6], Bravyi & Gosset [7], and Bravyi et al [8], show that this scaling can be significantly ameliorated. For instance, if all non-Clifford gates are $\pi/8$ Z-rotations, the classical cost scales like $2^{\gamma m}$, where $\gamma \approx 0.228$. In practice this allows for cheap classical simulations of useful quantum algorithms on as many as 50 qubits [8].

This exponential scaling is in fact much milder than recently reported similar Monte Carlo methods. These alternative methods are based on importance-sampling the entries of the gates in the standard basis [9], quasi-probability distributions [10], robustness of magic [11], or Clifford-channel decompositions [12]. The advantage holds even though the above mentioned alternatives can only estimate quantum circuit probabilities up to inverse-polynomial additive error. In comparison the methods in [7, 8] can do something much stronger; estimate probabilities up to inverse-polynomial multiplicative error, and approximately sample from the quantum circuit distribution with inverse-polynomial error in total variation distance.

Prior works [6, 7, 8] introduce and study the notion of the *stabilizer rank*. For a multi-qubit quantum state $|\psi\rangle$, the stabilizer rank, denoted $\chi(\psi)$ is the minimum integer k such that a decomposition of the form

$$|\psi\rangle = \sum_{i=1}^{k} c_i |\phi_i\rangle \tag{1.1}$$

exists, where $|\phi_i\rangle$ are multi-qubit stabilizer states, and c_i are complex coefficients. Given any decomposition of the form in eq. (1.1), the classical cost of simulating measurements of $|\psi\rangle$ in the computational basis scales linearly in k, and polynomially in the number of qubits, provided we can compute each term efficiently. In many interesting cases one can find decompositions of the form in eq. (1.1) for which $k \ll 2^n$. A decomposition with small k is referred to as a *low-rank* stabilizer decomposition. Occasionally we will also refer to it as a *sparse* stabilizer decomposition.

The fact that the number of stabilizer states of n qubits grows like 2^{n^2} makes it extremely challenging to compute low-rank stabilizer decompositions, or to even determine whether there is a decomposition of some rank k, even for small k and n. However, when the state $|\psi\rangle$ is the output state of a quantum circuit, we can use the structure of the circuit to find such decompositions.

One way to do this relies on finding stabilizer decompositions of tensor product states of the form $|A\rangle^{\otimes m}$, where $|A\rangle$ is some single qubit state. These can typically be decomposed using only $2^{\beta m}$ stabilizer states, for some exponent $0 < \beta < 1$. The link between such decompositions and quantum circuits comes from considering a gate A such that $|A\rangle = A|+\rangle$. Any quantum computation that can be implemented by a circuit of Clifford gates and m instances of the gate A can be simulated classically using an amount of resources that scales as $2^{\beta m}$ [7, 8]. Reducing the exponent β can be done by finding a stabilizer decomposition of $|A\rangle^{\otimes r}$ for some small $r < m$ using some number k_r of stabilizer terms, and then taking an (m/r)-fold tensor product of the decomposition, which gives $\beta = \log_2(k_r)/r$. This method is throttled by the difficulty of numerically finding low rank stabilizer decompositions even for small r; typically r is at most 6 or 7.

Another way to find sparse stabilizer decompositions relies on an approximate version of the stabilizer rank. In this version we allow an error ϵ, and ask for the minimum integer k such that

$$\||\psi\rangle - \sum_{i=1}^{k} c_i|\phi_i\rangle\| \leq \epsilon, \tag{1.2}$$

for some stabilizer states $|\phi_i\rangle$ and coefficients c_i. Obviously this can only reduce the required number of stabilizer terms below what is needed for the exact case.

Furthermore, starting from an exact decomposition as in eq. (1.1), which may not be sparse, Ref. [8] shows that it can be made sparse by picking random terms in an importance sampling scheme in a way that produces a low-rank approximation of $|\psi\rangle$. This method is known as randomized sparsification, and combining it with stabilizer rank classical simulations allows for significant improvement in the required resources as well as in the scope of the type of circuits that can be simulated [8]. Using these methods the amount of classical resources scales as $\|c\|_1^2$, rather than k, provided we can efficiently sample the distribution $p(i) = |c_i|/\|c\|_1$ and compute $|\phi_i\rangle$ for any given i.

Most of the first part of this thesis discusses various improvements in this framework. These improvements translate to reducing the cost of simulating quantum circuits, in some situations leading to runtimes that have a milder exponential scaling than previously reported.

In the second part of the thesis, we consider obstructions to efficient classical simulations. It has been argued in various contexts that quantum computational advantage is deeply related to phenomena such as non-locality and contextuality. For instance, it is known that the absence of contextuality makes any quantum computation classically simulable efficiently in a certain restricted setting [13, 14]. Furthermore, it was recently shown that the presence of non-local correlations between the qubits in a state generated by a constant-depth quantum circuit can thwart constant-depth classical simulations of the circuit [15, 16].

Linear constraint systems (LCS), and their associated non-local games [17, 18], are two closely related manifestations of contextuality and non-locality. The idea behind these constructions is to start with a system of linear constraints $Ax = b$ over a mathematical field, and promote the scalar-valued variables to quantum observables while requiring that every constraint can be checked by a simultaneous measurement of the observables it depends on. A quantum solution is a set of observables over some Hilbert space that satisfies every constraint. If the system $Ax = b$ has no classical solutions and yet admits a quantum solution, we call that a satisfiability gap. In terms of hidden variable theories a satisfiability gap is a state-independent proof of contextuality [19, 20]. When a satisfiability gap exists in a LCS, one can construct a bipartite non-local game which can be won with unit probability by quantum players but not by classical players, a situation sometimes referred to as quantum pseudo-telepathy [21].

Non-local games constructed from LCSs are behind recent breakthroughs in proving a separation between fixed-depth classical and quantum computation that is independent of complexity-theoretic conjectures [15, 16]. Certain non-local correlations can arise between the registers of a fixed-depth quantum circuit that are impossible to produce by any fixed-depth classical circuit, even if the former is restricted to short-range gates and the latter is allowed arbitrary long-range gates, and even if noise is present [16]. The proof of this statement is based on considering correlations that win certain variants of the Peres-Mermin magic square game, which is in fact the canonical example of a LCS with a satisfiability gap.

Non-local games constructed from LCSs have also been instrumental in significant recent progress in understanding quantum non-locality. An important problem in non-locality is to distinguish sets of quantum correlations defined using different models of entanglement, for instance the tensor-product and the commuting-operator models [22]. In this framework the use of LCSs led to ground-breaking results [23, 24, 25] [2].

The Pauli group plays an interesting role in the framework of LCSs. The commutation relation $UV = -VU$ encodes the simplest classically inconsistent linear constraint on two binary variables. Whenever such relations appear in the so-called *solution group* of a quantum-feasible LCS, it implies a satisfiability gap. The results in [24, 25] rely in an important way on embedding this commutation relation into certain families of binary

[2] A variant of such questions is equivalent to one of the most famous and long-standing conjectures in operator theory, namely Connes' embedding conjecture. This conjecture was very recently refuted using a construction involving certain bipartite quantum nonlocal games [26] (although not of the LCS variety).

LCSs.

This thesis considers a generalization of the Pauli group for higher dimensional systems, known as the Weyl-Heisenberg group. While other generalizations of the qubit Pauli group exist, the Weyl-Heisenberg group is arguably the most natural one, since it preserves the commutativity structure and error-correction properties of the qubit Pauli group [27]. We consider how the Weyl-Heisenberg group relates to d-ary LCSs, i.e. those defined modulo some integer d. When the dimension of the quantum system is odd, it is known that one cannot have a state-independent proof of contextuality using these generalized Pauli observables [13, 28].

Somewhat more surprisingly, when the *modulus* d is odd, and regardless of the quantum dimension, we find in this thesis that any commutation relation of the generalized Pauli group, $UV = \omega^j VU$, where ω is a dth root of unity and $j \neq 0$, cannot be embedded in the solution group of a LCS if a quantum solution exists in any (infinite or finite) dimension. In other words, no satisfiability gap can ever arise from classically inconsistent relations of that form, in stark contrast to the binary case.

Another piece of the puzzle is revealed by the so-called discrete Wigner function. For qudits of odd-prime dimension, it is known that any quantum circuit consisting of measurement of generalized Pauli observables and unitary transformations that preserve them (i.e. the qudit Clifford group) can be simulated efficiently using a non-negative discrete Wigner function. For such systems the discrete Wigner function provides a local hidden variable theory [19, 20]. In other words, correlations produced from such circuits cannot produce quantum correlations that win any nonlocal game. Remarkably, the simulation algorithm based on the discrete Wigner function can be executed in the same depth as the quantum circuit, which precludes any advantage for fixed depth quantum circuits of this type, i.e. it precludes any construction such as in Refs. [15, 16].

1.2 Organization and contributions of this thesis

We begin in Chapter 2 by reviewing preliminary material related to stabilizer rank methods, and summarizing important tools introduced in [7, 8].

In Chapter 3, we consider the problem of computing stabilizer decompositions of the output state of a quantum circuit consisting of ℓ gates in total, m of which are non-Clifford gates. We describe an improved method of performing randomized sparsification, which reduces the time of cost of computing a single term of the stabilizer decomposition from $O(\ell n^2)$ to $O(mn^2)$. Since this operation has to be repeated for each term in the decomposition, the reduction in runtime can be significant if $\ell \gg m$. This method uses a certain re-writing of the circuit, which also allows for exploiting more sparse Clifford decompositions of products of non-Clifford gates. In some cases this gives a significant amelioration of the exponential scaling of the required classical resources. We suggest a straight-forward simulation using importance-sampling Monte Carlo to estimate the circuit amplitudes up to additive error. Such estimates can be used with the Markov Chain Monte Carlo variant of stabilizer rank methods, as shown in [8]. We briefly consider generalizing this method of estimating amplitudes to other gate sets. The results in Chapter 3 are published in [29].

In Chapter 4, we describe a method of constructing exact, low-rank stabilizer decompositions of $|\psi\rangle^{\otimes m}$, where $|\psi\rangle$ is a single qubit state. We focus on the case where $|\psi\rangle$ is either a magic state or an equatorial state.

Magic states are the non-stabilizer eigenstates of single qubit Clifford unitaries [30]. They come in two flavors, depending on whether they are Clifford-equivalent to the T state $|T\rangle = 2^{-1/2}(|0\rangle + e^{i\pi/4}|1\rangle)$ or the face state $|F\rangle = \cos(\beta)|0\rangle + e^{i\pi/4}\sin(\beta)|1\rangle$, where β is an angle such that $\cos(2\beta) = 1/\sqrt{3}$. Equatorial states are those of the form $|\theta\rangle \equiv 2^{-1/2}(|0\rangle + e^{i\theta}|1\rangle)$.

We find stabilizer decompositions of $|T\rangle^{\otimes m}$ and $|F\rangle^{\otimes m}$ with $2^{m\log_2(3)/4}$ terms. This improves on the best known bound of $2^{m\log_2(7)/6}$, and is based on a stabilizer decomposition of $|T\rangle^{\otimes 6}$ with six terms. Previously [6] the stabilizer rank of $|T\rangle^{\otimes 6}$ was conjectured to be seven (which this result refutes), and the stabilizer rank of the two types of magic states was conjectured to be the same (which this result provides further evidence for). Also, for any equatorial state $|\theta\rangle$ we give a stabilizer decomposition of $|\theta\rangle^{\otimes m}$ with $2^{m/2}$ terms. To our knowledge no such decompositions were previously known. Exact stabilizer decompositions of this form allow us to estimate probabilities of quantum circuits up to polynomially small multiplicative error, which we cannot do with approximate stabilizer decompositions. Our results translate to milder exponential scaling of the classical resources required for this task, as well as allowing more types of circuits to be simulated in this way.

The results in Chapter 4 are based on unpublished joint work with David Gosset [3] and Hakop Pashayan [4].

In Chapter 5, we consider using stabilizer decompositions to simulate the time evolution under a time-independent spin Hamiltonian, that is, to simulate the state $e^{iHt}|0\rangle^{\otimes n}$. If the

[3] Institute for Quantum Computing, Waterloo, Ontario, and the Department of Combinatorics and Optimization, University of Waterloo, Ontario.

[4] Perimeter Institute, Waterloo, Ontario, and the Institute for Quantum Computing, Waterloo, Ontario.

Hamiltonian is given as a sum of operators that are easy to exponentiate,

$$H = \sum_{j=1}^{K} a_j A_j, \tag{1.3}$$

then this method allows for approximate-sampling from the probability distribution associated with measuring any subset of the qubits in the computational basis in time which scales as $\exp(\gamma t \|a\|_1)$, where γ is a constant that depends on which Lie product formula is used to approximate the time evolution, and on the sparsity of the Clifford decomposition of the associated circuit. For example, using a decomposition of the Hamiltonian into the Pauli basis together with the first-order Suzuki formula gives $\gamma = 2\sqrt{2} - 2 \approx 0.83$. We describe various ways to reduce γ. This kind of simulation may be useful for short time evolution when $\exp(\gamma t \|a\|_1) \ll 2^n$, and reducing γ allows longer times for the same computational budget.

In the last two chapters we consider linear constraint systems (LCS). We start by reviewing background material in Chapter 6, then in Chapter 7 we consider linear constraint systems modulo an integer $d > 2$. First we give a simple proof, of the previously known fact [13, 28], that any linear constraint system which has a generalized Pauli solution of odd dimension must be classically satisfiable. We further prove that, for odd d, embedding a Pauli-like commutation relation between any two elements of the solution group implies that the LCS has no quantum solutions in any dimensions, and therefore cannot display a satisfiability gap. We apply this result to various examples, for instance showing that many generalizations of the Peres-Mermin magic square and pentagram cannot have a satisfiability gap. The results in Chapter 7 are published in [31].

Chapter 2

The stabilizer rank and classical simulation of quantum circuits

2.1 Preliminaries

2.1.1 Notation

This thesis deals with finite-dimensional quantum systems. We use the bra-ket notation to denote elements of the d-dimensional complex vector space $\mathbb{C}^d$. For n-qubits the dimension is 2^n, and we identify the standard $0,1$ basis with elements of the binary vector space $\mathbb{F}_2^n = \{0,1\}^n$. We use small letters $x, y, ...$ to denote bit-strings in $\mathbb{F}_2^n$, but reserve capital letters $A, B, \ldots$ to denote operators on $\mathbb{C}^d$.

2.1.2 Stabilizer states and the Clifford group

Stabilizer states and the Clifford group are ubiquitous in the theory of quantum information. In a nutshell, the Clifford group of n qubits is a finite subgroup of the unitary group $\mathcal{U}(2^n)$, and the set of stabilizer states is the orbit of the state $|0\rangle^{\otimes n}$ under this group.

More specifically, we start by defining the usual single qubit Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.1}$$

The n-qubit Pauli group $\mathcal{P}(n)$ consists of all n-fold tensor products of Pauli matrices with an arbitrary global phase $\mathcal{U}(1)$. The Clifford group of n qubits $\mathcal{C}(n)$ is defined as the subgroup of unitaries which preserve $\mathcal{P}(n)$, i.e.

$$\mathcal{C}(n) := \{U \in \mathcal{U}(2^n) \ : \ UPU^\dagger \in \mathcal{P}(n) \text{ for all } P \in \mathcal{P}(n)\}. \tag{2.2}$$

The set of n-qubit stabilizer states $\mathcal{S}(n)$ is the orbit

$$\mathcal{S}(n) := \{U|0\rangle^{\otimes n} \ : \ U \in \mathcal{C}(n)\}. \tag{2.3}$$

This definition coincides with the usual definition of a stabilizer state as a simultaneous $+1$ eigenstates of a maximal stabilizer group, that is, a subgroup generated by n independent, Hermitian, and mutually commuting Pauli operators.

Another definition of stabilizer states, which we will find useful in proving some results later, is based on characterizing both the support of these states in the computational basis and the set of allowed relative phases.

Namely, it is known that the support of a stabilizer state in the computational basis is always an affine subspace $A \subseteq \mathbb{F}_2^n$ [1]. The allowed relative phases between the states in the support of a stabilizer state take the form of *linear* and *quadratic phases* [32]. Specifically, the phase on the component $x \in A$ is given by $i^{\ell(x)}(-1)^{q(x)}$, where q is a quadratic polynomial on $\mathbb{F}_2^n$, and ℓ is a linear function on $\mathbb{F}_2^n$. A stabilizer state is then determined by the triplet (A, q, ℓ);

$$|A, q, \ell\rangle := \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{\ell(x)}(-1)^{q(x)}|x\rangle. \tag{2.4}$$

One can easily check that Clifford unitaries preserve this form up to a global phase. For instance, the set of gates

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad CZ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{2.5}$$

on all qubits (for H and S) and pairs of qubits (for CZ), generate the Clifford group, and one can quickly check that they preserve the form (2.4).

[1] Recall that an affine subspace of $\mathbb{F}_2^n$ is a translated linear subspace, i.e. a set of the form $A = V \oplus a$ where V is a linear code and a is a translation vector.

Yet another equivalent classical description of stabilizer states is given by the so-called CH-form. This description is suitable for certain tasks, such as computing phase-sensitive Clifford circuit amplitudes, which is a key component in the classical simulation schemes in [8] and in this thesis.

The CH-form description of a stabilizer state is given by a tuple (w, s, h, U), where $w \in \mathbb{C}$ is a complex number, $s, h \in \mathbb{F}_2^n$ are binary vectors, and U is a circuit consisting of gates from the gate set {CNOT, CZ, S}, where

$$CNOT_{1,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2.6}$$

Note that any such U satisfies

$$U|0\rangle^{\otimes n} = |0\rangle^{\otimes n}, \tag{2.7}$$

which fixes the global phase convention. The vector representation of the stabilizer state associated with the tuple (w, s, h, U) is given by

$$|w, s, h, U\rangle := wUH(h)|s\rangle, \tag{2.8}$$

where $H(h)$ acts as the Hadamard gate on qubit i if $h_i = 1$ and as the identity otherwise. Note that the circuit U can be stored using $O(n^2)$ bits in the form of a stabilizer tableau specifying its action on Paulis $X_1, \ldots, X_n, Z_1, \ldots, Z_n$, see Ref. [5]. Hence storing the CH-description of an n-qubit stabilizer state requires $O(n^2)$ bits of classical memory. This is the same cost as the more standard approach in [5], but it allows us to keep track of the global phase in the simulation by updating w.

2.1.3 The stabilizer rank, the approximate stabilizer rank, and the stabilizer extent

In this subsection we define the stabilizer rank and related quantities, and discuss some of their properties and relations.

Definition 1. The *stabilizer rank* of a quantum state $|\psi\rangle$, denoted by $\chi(\psi)$, is the minimum number of terms in a stabilizer decomposition of the state.

In other words, $\chi(\psi)$ is the minimum integer k such that there exists a superposition

$$|\psi\rangle = \sum_{i=1}^{k} c_i |\phi_i\rangle, \tag{2.9}$$

where $|\phi_i\rangle$ are stabilizer states and c_i are complex coefficients. The stabilizer rank is sub-multiplicative under the tensor product

$$\chi(\psi \otimes \phi) \leq \chi(\psi)\chi(\phi), \tag{2.10}$$

since we can decompose each state separately and take the tensor product of the decompositions. It is also invariant under Clifford operations and sub-multiplicative under projections onto stabilizer codes, that is

$$\chi(C\psi) = \chi(\psi), \quad \chi(\Pi\psi) \leq \chi(\psi), \tag{2.11}$$

for any Clifford C and projector Π onto a stabilizer code.

The stabilizer rank of an n-qubit state is extremely difficult to compute even for product states and modest n. The difficulty comes from having to check all k-tuples of stabilizer states to determine whether a decomposition with k terms exists. The number of n-qubit stabilizer states grows like 2^{n^2}, making such brute force technique impractical. Numerical methods for computing $\chi(\psi)$ (and finding low rank decompositions) have been proposed in [6] and [8], based on a heuristic simulated annealing algorithm, but these methods miss the mark for product states on as low as six qubits, as we will elaborate on in Chapter 4.

For m copies of a state ψ, sub-multiplicativity implies that

$$\chi(\psi^{\otimes m}) \leq \chi(\psi)^m.$$

In general, we can do much better than a naive product decomposition. An important example is the quantum state [2]

$$|T\rangle = 2^{-1/2}(|0\rangle + e^{i\pi/4}|1\rangle). \tag{2.13}$$

[2] This state is related to the T gate,

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{pmatrix}, \tag{2.12}$$

via $|T\rangle = T|+\rangle$. The state $|T\rangle$ is Clifford-equivalent to a state known as the H-type magic state (the $+1$ eigenstate of Hadamard), and in the past has been referred to as $|H\rangle$ (incidentally the second type of magic state, the so-called face state, has been denoted by $|T\rangle$). In this thesis we use $|T\rangle$ for the first type, and $|F\rangle$ for the second type.

Two copies of $|T\rangle$ can be written as a sum of two stabilizer states:

$$|T\rangle^{\otimes 2} = \frac{1}{2}\Big(|00\rangle + i|11\rangle\Big) + \frac{e^{i\pi/4}}{2}\Big(|01\rangle + |10\rangle\Big). \tag{2.14}$$

This makes it possible to decompose $|T\rangle^{\otimes m}$ using only $k = 2^{m/2}$ stabilizer terms:

$$|T\rangle^{\otimes m} = \left(|T\rangle^{\otimes 2}\right)^{\otimes m/2}. \tag{2.15}$$

Better decomposition strategies allow us to achieve $k = 2^{\alpha m}$ where $\alpha = \log_2(3)/4 \approx 0.3962$, as we shall prove in Chapter 4 of this thesis. An optimal decomposition achieves $k = \chi(T^{\otimes m})$; the stabilizer rank of m copies of $|T\rangle$. Conjectures aside, the value of $\chi(T^{\otimes m})$ is currently unknown for $m > 3$, and determining its asymptotic scaling is one of the biggest open problems in this field.

In the approximate version of the stabilizer rank, our goal is to approximate a state ψ up to error ϵ by a superposition of stabilizer states. That is, to find $\tilde{\psi}$ such that

$$\|\psi - \tilde{\psi}\| \leq \epsilon \qquad |\tilde{\psi}\rangle = \sum_{i=1}^{k} c_i|\phi_i\rangle, \tag{2.16}$$

where ϕ_i are stabilizer states.

Definition 2. The *approximate stabilizer rank*, $\chi_\epsilon(\psi)$, is defined as the minimum k for which (2.16) can hold.

Clearly, $\chi_\epsilon(\psi) \leq \chi(\psi)$ for any state ψ and $\epsilon > 0$. Interestingly, it turns out that the approximate stabilizer rank is upper bounded by a quantity which is easier to handle, namely the *stabilizer extent* [8].

Definition 3. The *stabilizer extent* of ψ, denoted by $\xi(\psi)$, is the minimum 1-norm squared of a vector of coefficients $c \in \mathbb{C}^k$ in a stabilizer decomposition of ψ.

Here the 1-norm of a vector $c \in \mathbb{C}^k$ is $\|c\|_1 = |c_1| + \cdots + |c_k|$. In other words, the stabilizer extent is the solution to the (convex) optimization problem:

$$\begin{aligned} &\text{minimize} && \|c\|_1^2 \\ &\text{subject to} && |\psi\rangle = \sum_i c_i|\phi_i\rangle \\ & && \phi_i \text{ are stabilizer states} \end{aligned} \tag{2.17}$$

This optimization is slightly less computationally demanding than computing the stabilizer rank. It has also led to very interesting insights. For instance, by examining its dual program it was shown in [8] that, for a large class of quantum states, the stabilizer extent is multiplicative under the tensor product, i.e.

$$\xi(\psi^{\otimes m}) = \xi(\psi)^m. \tag{2.18}$$

The class of states for which this is known to hold is called *stabilizer-aligned* states, and includes all states ψ of three or less qubits. Interestingly, for $n > 3$ not all n qubit states are stabilizer-aligned.

The main result relating the approximate stabilizer rank of a state to its stabilizer extent is known as the *randomized sparsification lemma.*

Lemma 4 (Randomized Sparsification, Ref. [8])**.** *For any quantum state ψ and $\epsilon > 0$, it holds that*

$$\chi_\epsilon(\psi) \leq \epsilon^{-2}\xi(\psi). \tag{2.19}$$

Proof sketch. Suppose we're given a stabilizer decomposition

$$|\psi\rangle = \sum_i c_i|\phi_i\rangle. \tag{2.20}$$

Without loss of generality, we can assume that the c_i are real and positive by absorbing all phases into the stabilizer states ϕ_i. Consider the random vector $|w_N\rangle$ obtained by averaging N random vectors of the form $\|c\|_1|\phi\rangle$, where $|\phi\rangle$ is chosen independently from the set of stabilizer states $\{\phi_i\}$ according to the probability distribution $c_i/\|c\|_1$. The average of this random vector is $|\psi\rangle$. Using standard probability tail bounds, one can show that, with high probability,

$$\||\psi\rangle - |w_N\rangle\| \leq \epsilon$$

provided that $N \geq \|c\|_1^2\epsilon^{-2}$. This implies that there exists a quantum state, $|\tilde{\psi}\rangle$, of the form

$$|\tilde{\psi}\rangle = \sum_{j=1}^{\lceil\|c\|_1^2\epsilon^{-2}\rceil} \alpha_j|\theta_j\rangle, \tag{2.21}$$

where θ_j are stabilizer states, such that $\|\tilde{\psi} - \psi\| \leq \epsilon$. □

A more detailed version of the above proof can be found in Ref. [8]. This construction plays a key role in the classical simulation schemes we describe in this thesis.

The three quantities we have defined in this section, the stabilizer rank, approximate stabilizer rank, and stabilizer extent, can all be defined for unitary operators as well. For example, we may define the stabilizer rank of a unitary U as the minimum number of Clifford unitaries in a linear decomposition of U. However, it turns out that this is not the most general thing we can do. For instance, we can also allow the terms of the decomposition to be projectors onto stabilizer codes, or even products of such projectors. The point is that such operators preserve stabilizer states, and can be used just as well as Cliffords for classical simulation purposes.

Such generalized notions of the stabilizer rank will not be discussed in this thesis, with one exception, namely the stabilizer extent. For a unitary U we define $\xi(U)$ as the minimum $\|c\|_1^2$ for any vector of coefficients in a linear Clifford decomposition of U. For the cases we consider, we find that nothing is gained by allowing stabilizer projectors or their products, i.e. that the minimum 1-norm is always achieved with a decomposition in which every term is a Clifford unitary. For this reason we refrain from making that generalization in this thesis.

Simulating a quantum circuit consisting of Clifford gates augmented with m T gates can be done classically in time that scales as $\xi(T)^m \approx 2^{0.228m}$ [8]. Here the T gate is defined as

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}. \tag{2.22}$$

Note that for the closely related T state $|T\rangle = T|+\rangle$, we have $\xi(T) = \xi(|T\rangle)$. More generally this holds for diagonal gates from the third level of the Clifford hierarchy [8].

For unitary gates, the stabilizer extent is sub-multiplicative under the tensor product:

$$\xi(A \otimes B) \leq \xi(A)\xi(B), \tag{2.23}$$

as well as under matrix multiplication;

$$\xi(AB) \leq \xi(A)\xi(B). \tag{2.24}$$

In Chapter 3, we will show how a certain re-writing of the quantum circuit allows us to take advantage of this fact, resulting in a milder exponential scaling of the time cost of classically simulating certain quantum circuits.

Operation	**Time cost**
PHASE, CNOT, or CPHASE gates	$O(n)$
Hadamard gates	$O(n^2)$
Multi-qubit rotations $\exp(i\pi P/4)$	$O(n^2)$
Post-measurement state after measuring b qubits	$O(bn^2)$
Compute an amplitude $\langle x\|\phi\rangle$	$O(n^2)$

Table 2.1: The time cost of performing various phase-sensitive operations on a stabilizer state specified by its CH-form using the methods of [8], described in this section.

2.1.4 Efficient simulation of Clifford circuits: the phase-sensitive version

The Gottesman-Knill theorem gives an efficient classical simulation of quantum circuits consisting of Clifford gates acting on stabilizer state inputs [1]. The classical algorithm is based on storing the quantum state in terms of generators of its stabilizer group, and updating the generators as Clifford gates are applied. Such a description of the state is correct up to an irrelevant global phase. In this section we describe a recent variant of these simulations which also keeps track of the global phase on the quantum state.

Keeping track of the global phase is necessary to maintain coherence between the terms in a stabilizer decomposition of a non-stabilizer state, which is a prerequisite for the simulation schemes in this thesis. Peculiarly, tracking the global phase adds a considerable time requirement. For instance, it increases the number of basic operations required to update the description of a stabilizer state under a single qubit Clifford gate from $O(n)$ to $O(n^2)$.

In the rest of this subsection we give an overview of how to perform Clifford operations and measurement updates while keeping track of the global phase. We describe these update rules in quite some detail, although we leave out the proof of one key lemma (Lemma 5). The interested reader can go to [8] for further reference. The impatient reader can skip to the next section. For convenience the time cost of the update rules described in this section is summarized in Table 2.1.

Unitary updates

Recall that the CH description of a stabilizer state is a tuple (w, s, h, U), where w is a complex scalar, s and h are n-bit strings, and U is a Clifford circuit consisting of gates from the set $J = \{CZ, CX, S\}$. The correspondence is given by

$$|w, s, h, U\rangle = wUH(h)|s\rangle. \tag{2.25}$$

Also recall that U is stored in classical memory in the form of a stabilizer tableau, i.e. a binary matrix of $2n$ rows where each row specifies the Pauli operator $UPU^\dagger$ with P ranging over $X_1, \ldots, X_n, Z_1, \ldots, Z_n$.

Given a stabilizer state specified by its CH-form (w, s, h, U), and a Clifford gate C from the gate set $\{CZ, CX, S, H\}$, our goal is to compute (w', s', h', U') such that

$$C|w, s, h, U\rangle = |w', s', h', U'\rangle. \tag{2.26}$$

If the gate C itself belongs to J, only U needs to be updated. We simply compute the tableau of the circuit $U' = CU$ by computing the action of C on at most $2n$ Pauli operators, using standard binary linear algebra. Since C acts on $O(1)$ qubits, the update $(w, s, h, U) \to (w, s, h, U')$ can be computed in time $O(n)$.

If C is a Hadamard gate on one of the qubits, for example $C = H_j$, the situation is more complicated. The identity

$$H_j = \frac{1}{\sqrt{2}}(X_j + Z_j) \tag{2.27}$$

implies

$$H_j|w, s, h, U\rangle = \frac{1}{\sqrt{2}}\left(X_j|w, s, h, U\rangle + Z_j|w, s, h, U\rangle\right). \tag{2.28}$$

By definition, $|w, s, h, U\rangle = wUH(h)|s\rangle$, and therefore

$$\begin{aligned} X_j|w, s, h, U\rangle &= wUH(h)P|s\rangle \\ Z_j|w, s, h, U\rangle &= wUH(h)Q|s\rangle \end{aligned}$$

where $P = (UH(h))^\dagger X_j(UH(h))$ and $Q = (UH(h))^\dagger Z_j(UH(h))$ are Pauli operators. Note that we can compute P and Q in time $O(n^2)$ using the tableau associated with U and the fact that H permutes X and Z. We can then compute strings $x, z \in \mathbb{F}_2^n$ and integers $p, q \in \mathbb{Z}_4$, in time $O(n)$, such that $P|s\rangle = i^p|x\rangle$ and $Q|s\rangle = i^q|z\rangle$. We now have

$$H_j|w, s, h, U\rangle = i^p wUH(h)\left(\frac{|x\rangle + i^{q-p}|z\rangle}{\sqrt{2}}\right). \tag{2.29}$$

To proceed we need to use a certain decomposition of Clifford circuits.

Lemma 5 (Proposition 4 in [8])**.** *Let $x, z \in \mathbb{F}_2^n$, $x \neq z$. For any $h \in \mathbb{F}_2^n$ and $m \in \mathbb{Z}_4$, there exists a circuit V consisting of $O(n)$ gates from the set $\{CZ, CX, S\}$, a vector $h' \in \mathbb{F}_2^n$, and a complex number $\alpha \in \mathbb{C}$, such that*

$$H(h)\left(\frac{|x\rangle + i^m|z\rangle}{\sqrt{2}}\right) = \alpha V H(h')|s'\rangle. \tag{2.30}$$

Moreover, V, h', and α, can be computed in time $O(n)$.

Applying Lemma 5 to the right-hand side of eq. (2.29) yields

$$H_j|w, s, h, U\rangle = \begin{cases} (i^p e^{\pm i\pi/4} w) U H(h)|x\rangle & \text{if } x = z \\ (\alpha i^p w) U V H(h \oplus h')|s'\rangle & \text{if } x \neq z, \end{cases} \tag{2.31}$$

where V, α, and h' are the items referred to in Lemma 5. Finally, computing the stabilizer tableau of $U' := UV$ given that of U can be done in time $O(n^2)$, since there are $O(n)$ gates in V each of which take time $O(n)$ to absorb into U.

To conclude, we can update the CH description of a stabilizer state under the action of a Hadamard gate, i.e. compute (w', s', h', U') such that $H_j|w, s, h, U\rangle = |w', s', h', U'\rangle$, in time $O(n^2)$.

Another set of Clifford gates which we make use of in this chapter are Pauli rotations of the type

$$\{\exp(i(\pi/4)P) : P \in \mathcal{P}(n), P = P^\dagger\}. \tag{2.32}$$

These unitaries typically entangle many qubits, but it turns out that one can update the CH-description of a stabilizer state under such gates in time $O(n^2)$, the same time requirement for a single qubit Hadamard gate.

The above assertion follows immediately from Lemma 5: given a stabilizer state (w, s, h, U) and a Pauli P we can compute another Pauli P' in time $O(n^2)$ such that

$$e^{i\frac{\pi}{4}P}|w, s, h, U\rangle = wUH(h)e^{i\frac{\pi}{4}P'}|s\rangle, \tag{2.33}$$

namely, $P' = (UH(h))^\dagger P(UH(h))$. We can then compute a string z and an integer p such that $e^{i\frac{\pi}{4}P'}|s\rangle = 2^{-1/2}(|s\rangle + i^p|z\rangle)$, and proceed similarly to the Hadamard case, using Lemma 5 if necessary.

Measurement updates

Next we summarize how to compute the CH-form of the post-measurement state after measuring a subset of the qubits and obtaining some outcome. For any Hermitian Pauli R, projecting onto the $+$ eigenspace of R corresponds to an update rule $(w, s, h, U) \to (w', s', h', U')$ such that

$$\frac{1}{2}(I + R)|w, s, h, U\rangle = |w', s', h', U'\rangle. \tag{2.34}$$

As before, we have

$$R|w, s, h, U\rangle = wRUH(h)|s\rangle = wUH(h)R'|s\rangle \tag{2.35}$$

where $R' = (UH(h))^\dagger R(UH(h))$ is a Pauli operator that can be computed in time $O(n^2)$. We now have

$$\frac{1}{2}(I + R)|w, s, h, U\rangle = wUH(h)(|s\rangle + i^r|s'\rangle) \tag{2.36}$$

where $r \in \mathbb{Z}_4$ and $s' \in \mathbb{F}_2^n$ satisfy $R|s\rangle = i^r|s'\rangle$. If $s = s'$ the update is trivial, and if $s \neq s'$ it reduces to Lemma 5. Therefore computing (w', s', h', U') takes time $O(n^2)$.

Let Π_x denote the projector onto the subspace corresponding to some computational basis outcome x. In other words Π_x acts as $|x\rangle\langle x|$ on the measured qubits and as identity on the rest of the qubits. We have

$$\Pi_x = 2^{-n'} \prod_{j=1}^{n'} (I + (-1)^{x_j} Z_j), \tag{2.37}$$

where n' is the number of qubits being measured. Computing the CH-form of the post-measurement state therefore takes time $O(n'n^2)$.

Computing amplitudes

Finally, let's describe how to compute an amplitude of a stabilizer state given its CH-form. Namely, we want to compute $\langle x|w, s, h, U\rangle$, where $x \in \mathbb{F}_2^n$. This can be done in time only $O(n^2)$, compared to the $O(n^3)$ required to update the CH-description using the post-measurement update rules.

We have

$$\begin{aligned}\langle x|w,s,h,U\rangle &= w\langle x|UH(h)|s\rangle \\ &= w\langle 0^n|(U^\dagger X(x)U)H(h)|s\rangle \\ &= w\langle 0^n|QH(h)|s\rangle \end{aligned} \tag{2.38}$$

for some Pauli Q which can be computed in time $O(n^2)$. Here we crucially made use of the fact that $U|0^n\rangle = |0^n\rangle$. Writing $Q = i^p Z(t)X(u)$ we obtain

$$\begin{aligned}\langle x|w,s,h,U\rangle &= wi^p\langle 0^n|X(u)H(h)|s\rangle \\ &= 2^{-|h|/2}wi^p \prod_{j:h_j=0}\langle u_j|s_j\rangle \prod_{j:h_j=1}(-1)^{h_js_j}\end{aligned} \tag{2.39}$$

Once Q is computed, the above expression can be evaluated in time $O(n)$. Thus the total runtime for computing an amplitude of the state $|w,s,h,U\rangle$ is $O(n^2)$.

2.2 Classical simulation using stabilizer decompositions

In this section, we give an overview of a family of techniques for classically simulating quantum circuits based on stabilizer decompositions. These techniques are useful when the number of non-Clifford gates in the circuit is roughly the same as the number of qubits. The discussion presented in this section is based on Refs. [7, 8].

Let U be a circuit acting on n qubits initialized in the state $|0\rangle^{\otimes n}$. The methods described in this section pertain to two different simulation tasks;

1. Estimating the probability $P(x)$ of obtaining outcome $x \in \mathbb{F}_2^s$ when measuring a subset of $s \leq n$ qubits of the state $U|0\rangle^{\otimes n}$, and
2. Sampling an outcome $x \in \mathbb{F}_2^s$ according to a distribution $\tilde{P}(x)$ that approximates $P(x)$ in total variation distance.

The exact versions of these tasks are colloquially known as *strong* and *weak* simulations, respectively. We caution that these terms can be somewhat misleading. While it is true that the ability to compute the quantum circuit probabilities exactly can be used to sample the circuit distribution, this implication hinges on the ability to compute marginal probabilities. For example, for a given quantum circuit U, a classical simulator that can only compute $P(x) = |\langle x|U|0^n\rangle|^2$ for $x \in \mathbb{F}_2^n$ needs to be run exponentially many times to produce a sample from the circuit distribution.

A key component in the simulation schemes based on stabilizer decompositions is a randomized algorithm for approximating the norm of a linear combination of stabilizer states.

Lemma 6 (Refs. [7, 8]). *Let $|\psi\rangle \in (\mathbb{C}^2)^{\otimes n}$ be given as a linear combination of stabilizer states;*

$$|\psi\rangle = \sum_{i=1}^{k} c_i|\phi_i\rangle. \tag{2.40}$$

There exists a randomized classical algorithm which queries the terms $\{(c_i, \phi_i)\}$ one at a time and with probability at least $1-\delta$ outputs an estimate b which approximates $\|\psi\|^2$ up to multiplicative error ϵ, i.e.

$$(1-\epsilon)\|\psi\|^2 \leq b \leq (1+\epsilon)\|\psi\|^2. \tag{2.41}$$

The runtime of the algorithm is

$$O(kn^3\epsilon^{-2}\log\delta^{-1}). \tag{2.42}$$

The randomized algorithm in Lemma 6 is based on computing inner products of $|\psi\rangle$ with uniformly random stabilizer states. It relies on two key properties of stabilizer states. First, computing an inner product $\langle\phi_i|\phi_j\rangle$ between two stabilizer states can be performed in time $O(n^3)$. Second, for a uniformly random stabilizer state $|\theta\rangle$ and any fixed vector $|v\rangle$, the random variable $2^n|\langle\theta|v\rangle|^2$ has mean and standard deviation both equal to $\|v\|^2$. This follows from the fact that stabilizer states form a 2-design, but being a 2-design is not necessary for this to hold [3]. Thus with high probability the average of a large number of evaluations of this random variable approximates $\|v\|^2$ with a small multiplicative error (for instance by a standard application of Chebyshev's inequality).

2.2.1 Overview of known simulation algorithms

Estimating probabilities with bounded multiplicative error

Lemma 6 can be used to estimate outcome probabilities of a quantum circuit U given a stabilizer decomposition of the output state. To see this, suppose we're given a stabilizer

[3] It was shown in [8] that in fact one can take a much smaller subset of stabilizer states, namely the so-called equatorial stabilizer states.

decomposition

$$U|0\rangle^{\otimes n} = \sum_{i=1}^{k} c_i|\phi_i\rangle. \tag{2.43}$$

Suppose that a subset of $s \le n$ qubits is measured in the computational basis, and that we want to estimate the probability $P(x)$ of some outcome $x \in \mathbb{F}_2^s$. The probability $P(x)$ is equal to the square of the norm of the vector $\Pi_x U|0\rangle^{\otimes n}$, where Π_x is the projector that acts as $|x\rangle\langle x|$ on the measured qubits and as identity operator on the rest. Using the phase-sensitive measurement updates of section 2.1.4, we can compute the CH-form of the vector $\Pi_x|\phi_i\rangle$ with its global phase in time $O(sn^2)$, and therefore we can compute a stabilizer decomposition

$$\Pi_x U|0\rangle^{\otimes n} = \sum_{i=1}^{k} c_i \Pi_x|\phi_i\rangle \tag{2.44}$$

in time $O(ksn^2)$. Combining this with Lemma 6, we obtain an estimate of $P(x)$, correct up to multiplicative error ϵ, in time $O(ksn^2) + O(kn^3\epsilon^{-2}\log\delta^{-1})$.

Approximate sampling from the circuit distribution

Now we briefly describe how the above method for estimating outcome probabilities $P(x)$ with bounded multiplicative error can be used to sample bit strings $x \in \mathbb{F}_2^s$ from a probability distribution which approximates P. Here s is the number of qubits measured. More specifically, the goal is to sample $x \in \mathbb{F}_2^s$ according to a distribution $\tilde{P}$ such that

$$\|P - \tilde{P}\|_1 := \frac{1}{2}\sum_{x\in\mathbb{F}_2^s} |P(x) - \tilde{P}(x)| < \epsilon. \tag{2.45}$$

Suppose we are given a stabilizer decomposition of the output state of the circuit

$$U|0^n\rangle = \sum_{i=1}^{k} c_i|\phi_i\rangle. \tag{2.46}$$

Using the chain rule for probabilities we have, for $x = (x_1, \ldots, x_s) \in \mathbb{F}_2^s$,

$$P(x) = P(x_1)P(x_2|x_1)P(x_3|x_1,x_2)\ldots P(x_s|x_1,\ldots,x_{s-1}). \tag{2.47}$$

The idea is to sample x_1 according to the distribution $\tilde{P}(x_1)$ obtained by estimating $P(x_1)$ using eq. (2.44) and Lemma 6, then sample x_2 according to the distribution $\tilde{P}(x_2|x_1)$ obtained by estimating $P(x_2|x_1)$, and so forth. This produces a sample $x \in \mathbb{F}_2^s$ according to a certain distribution $\tilde{P}(x)$.

To ensure that $\tilde{P}$ satisfies eq. (2.45), the multiplicative error on the estimates of each conditional probability must be small enough. Each conditional probability can be written as

$$P(x_i|x_1,\ldots,x_{i-1}) = \frac{P(x_1,\ldots,x_{i-1},x_i)}{P(x_1,\ldots,x_{i-1})}. \tag{2.48}$$

Let ν be the multiplicative error on the approximations of each of the numerator and denominator. Then the multiplicative error on the ratio is at most 2ν, and the multiplicative error on approximating $P(x)$ by

$$\tilde{P}(x) := \tilde{P}(x_1)\tilde{P}(x_2|x_1)\tilde{P}(x_3|x_1,x_2)\ldots\tilde{P}(x_s|x_1,\ldots,x_{s-1}) \tag{2.49}$$

is at most $O(s\nu)$. To ensure that eq. (2.45) holds, it's enough to pick $\nu = O(\epsilon/s)$.

The scaling of the total runtime can be determined as follows. To produce a sample, we need to make $2s$ calls to the multiplicative error estimation algorithm of the previous section, each with error tolerance $O(\epsilon/s)$. Thus the runtime for sampling a bitstring $x \in \mathbb{F}_2^s$ from a distribution that approximates P up to error ϵ in total variation distance is

$$O(kn^3s^3\epsilon^{-2}\log(\delta^{-1})).$$

Suppose now that, instead of eq. (2.46), we only have an approximate stabilizer decomposition of $U|0^n\rangle$. That is, we have a sum of stabilizer states

$$|\psi\rangle = \sum_{i=1}^{k} c_i|\phi_i\rangle \tag{2.50}$$

such that

$$\||\psi\rangle - U|0^n\rangle\| \leq \epsilon. \tag{2.51}$$

Define a probability distribution Q over s-bit strings by

$$Q(x) := \frac{\langle\psi|\Pi_x|\psi\rangle}{\|\psi\|^2}. \tag{2.52}$$

Since $|\psi\rangle$ is ϵ-close to a normalized state, we have $1-\epsilon \leq \|\psi\| \leq 1+\epsilon$, which implies $(1+\epsilon)^{-2} \leq \|\psi\|^{-2} \leq (1-\epsilon)^{-2}$. Using $(1 \pm \epsilon)^{-2} = 1 \mp O(\epsilon)$ we get

$$\| \|\psi\|^{-2} |\psi\rangle - U|0^n\rangle \| \leq c\epsilon, \tag{2.53}$$

for a small constant $c > 0$ that is not hard to determine from the Taylor expansion of $(1 \pm \epsilon)^{-2}$ (for example for $\epsilon \leq 1/2$ we may take $c = 6$). This implies

$$\|P - Q\|_1 \leq c\epsilon. \tag{2.54}$$

By appropriately re-scaling ϵ, we can ensure that approximate sampling from Q produces a sample from a distribution which approximates P to within any desired error in total variation distance. We can approximately sample Q by using the same method as in the exact case, replacing $U|0^n\rangle$ by the state

$$|\psi\rangle/\|\psi\|^2 = \sum_{i=1}^{k} (c_i/\|\psi\|^2)|\phi_i\rangle.$$

Namely, every marginal distribution of Q, obtained by measuring a subset of r qubits, can be written as

$$Q(x_1, \ldots, x_r) = \frac{\langle\psi|\Pi|\psi\rangle}{\|\psi\|}$$

for some stabilizer projector Π, and both the numerator and denominator can be estimated up to multiplicative error using Lemma 6.

2.2.2 Finding exact and approximate low-rank stabilizer decompositions

Next we discuss how to obtain stabilizer decompositions of the output state of the circuit; i.e. a decomposition of the form

$$U|0^n\rangle = \sum_{i=1}^{k} c_i|\phi_i\rangle, \tag{2.55}$$

where ϕ_i are stabilizer states. Here we seek decompositions which are sparse, that is, the number of stabilizer terms in the decomposition needs to be $k \ll 2^m$. Note that we can typically write a non-Clifford gate as a linear combination of two Clifford gates, which gives a stabilizer decomposition of the form in eq. (2.55) with $k = 2^m$ terms. We seek to find more compressed decompositions than this baseline.

Two methods are known for producing such decompositions. The first method relies on constructing an equivalent circuit to U in which every non-Clifford gate is replaced by a Clifford gate and an ancilla qubit initialized in a magic state. This is the same state-injection gadget from fault-tolerant quantum computing, where non-Clifford gates are applied by consuming ancilla qubits prepared using some magic state distillation routine [30]. The upshot is that this method allows us to convert compressed stabilizer decompositions of copies of a magic state into stabilizer decompositions of the output state of the circuit.

The second method, known as the *sum-over-Cliffords* method, is more direct, but produces only approximate stabilizer decompositions. It relies on decomposing each non-Clifford gate in the circuit as a linear combination of Cliffords and then using randomized sparsification to find a sparse approximation. Below we give a more detailed description of the two methods.

Decompositions using state injection gadgets

In this construction, a non-Clifford gate A can be applied by consuming an ancilla qubit initialized in the state $A|+\rangle$. A CNOT gate is applied between the qubit we want to apply A to and the ancilla qubit, followed by measuring the ancilla qubit, see Figure 2.1. Conditioned on obtaining the outcome 0, the correct gate is applied to the data qubit. If outcome 1 is obtained, a correction is needed. The correction gate is guaranteed to be a Clifford gate provided that A belongs to the third level of the Clifford hierarchy, otherwise it is some arbitrary non-Clifford gate (in which case the gadget is not very useful for purposes of fault-tolerant quantum computing, but still useful for classical simulations).

The ancilla measurement returns either 0 or 1 with equal probability. Suppose that we post-select on the ancilla measurement returning the outcome 0. In this case no correction is required. By replacing every non-Clifford gate in the circuit U by its corresponding state-injection gadget and post-selecting, we obtain an equivalent Clifford circuit V, acting on the original n qubits and m additional ancilla qubits, that is equivalent to U under post-selection.

Suppose for example that each of the m non-Clifford gates in the circuit is a T gate, where $T = diag(1, e^{i\pi/4})$, and let $|T\rangle := T|+\rangle$. Using state-injection we obtain a Clifford circuit V on $n + m$ qubits which satisfies

$$U|0\rangle^{\otimes n} = 2^{m/2} \left(I_{1,\ldots,n} \otimes \langle 0^{\otimes m}|\right) V(|0\rangle^{\otimes n}|T\rangle^{\otimes m}). \tag{2.56}$$

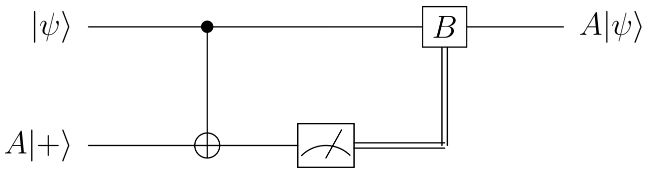

Figure 2.1: The state-injection gadget. A non-Clifford gate A can be applied by consuming an ancilla qubit initialized in the state $A|+\rangle$. If the measurement outcome is 0, the gate A is successfully applied to the data qubit. Otherwise a correction gate B needs to be applied. If A is in the third level of the Clifford hierarchy (such as the T gate), the correction gate B is a Clifford gate.

The factor $2^{m/2}$ corrects the norm of the state to account for the post-selection probability.

For every $(n+m)$-qubit stabilizer state $|\phi\rangle$, the state $\langle 0^{\otimes m}|\phi\rangle$ is an n qubit stabilizer state which can be computed efficiently. The idea now is to start with a stabilizer decomposition of $|T\rangle^{\otimes m}$ and use it to compute a stabilizer decomposition of

$$\left(I_{1,\ldots,n} \otimes \langle 0^{\otimes m}|\right) V(|0\rangle^{\otimes n}|T\rangle^{\otimes m}).$$

This gives a stabilizer decomposition of $U|0^n\rangle$ using the same number of terms as the decomposition of $|T\rangle^{\otimes m}$.

Finding a low-rank decomposition of $U|0^n\rangle$ is now reduced to the problem of finding low-rank decompositions of $|T\rangle^{\otimes m}$. This problem is still hard, but we can use decompositions of a small number of copies by breaking the m tensor factors into blocks. For example, it is known that we can decompose 6 copies of $|T\rangle$ using 7 stabilizer states. This gives a decomposition of m copies in terms of $7^{m/6} \approx 2^{0.468m}$ stabilizer states. In Chapter 4 in this thesis, we describe how that can be improved to reduced $2^{0.3962m}$.

The method outlined above requires an *exact* stabilizer decomposition of $|T\rangle^{\otimes m}$. This is easy to see from eq. (2.56); an additive error ϵ on $|T\rangle^{\otimes m}$ will be exponentially amplified by the factor $2^{m/2}$.

Decompositions by sparsifying a sum-over-Cliffords

The randomized sparsification idea of Lemma 4 can be used to produce approximate stabilizer decompositions of the state $U|0\rangle^{\otimes n}$ in a way that bypasses the need for introducing state-injection gadgets.

Suppose the circuit is given in the form

$$U = C_m U_m C_{m-1} U_{m-1} \dots C_1 U_1 C_0, \tag{2.57}$$

where C_i are Clifford blocks and each U_i is a non-Clifford gate. By expanding each U_i as a sum of Cliffords,

$$U_j = \sum_k b_{jk} V_{jk}, \tag{2.58}$$

we can write the entire circuit U as

$$U = \sum_i a_i W_i, \tag{2.59}$$

where each W_i is a Clifford circuit, and $\|a\|_1 = \prod_j \|b_j\|_1$. Since each U_j requires at least 2 Clifford gates to decompose, the number of terms in the above Clifford expansion of U is at lease 2^m. Thus $k = 2^m$ is the baseline for the number of stabilizer terms. We will improve on it by making use of randomized sparsification.

To implement randomized sparsification, we pick a random subset of N of the terms in eq. (2.59), where each element is independently randomly selected from the set $\{W_i\}$ with a frequency determined by $|a_i|/\|a\|_1$. This can be done by selecting V_{jk} with probability $|b_{jk}|/\|b_j\|_1$ and replacing each U_j with V_{jk}. This produces a random Clifford circuit $\mathcal{C}$ according to some distribution.

Let $\mathcal{C}_1, \dots, \mathcal{C}_N$ be independent instances produced in this way. By a complex-valued version of Hoeffding's inequality, we have

$$\|U|0\rangle^{\otimes n} - \frac{\|a\|_1}{N} \sum_{i=1}^{N} \mathcal{C}_i |0\rangle^{\otimes n}\| < \epsilon \tag{2.60}$$

with probability at least $1 - \delta$ provided that

$$N = c\|a\|_1^2 \epsilon^{-2} \log \delta^{-1}. \tag{2.61}$$

where c is a small constant. For such N the state

$$\frac{\|a\|_1}{N} \sum_{i=1}^{N} \mathcal{C}_i |0\rangle^{\otimes n} \tag{2.62}$$

approximates the output state of the circuit. Since each $\mathcal{C}_i$ is a Clifford circuit, each term in eq. (2.62) is a stabilizer state. Hence we found an approximation of $U|0^n\rangle$ that is a sum of roughly $\|a\|_1^2 \epsilon^{-2} \log \delta^{-1}$ stabilizer states.

In order to reduce the number of terms N, we should pick decompositions in eq. (2.58) with small 1-norm, i.e. small $\|b_j\|_1 \equiv \sum_k |b_{jk}|$. If each U_j is a single-qubit rotation of the form $\exp(i\theta_j Z)$ for $0 \leq \theta_j \leq \pi/2$, then the optimal Clifford decomposition achieves $\|b_j\|_1 = \cos\theta_j + \tan(\pi/8)\sin(\theta_j)$, and in this case

$$\|a\|_1 = \prod_{j=1}^{m} \|b_j\|_1 = \prod_{j=1}^{m} \left(\cos(\theta_j) + \tan(\pi/8)\sin(\theta_j)\right). \tag{2.63}$$

For example, if each U_j is a T gate, i.e. $\theta_j = \pi/4$ for all j, then the number of stabilizer terms in the approximate state in eq. (2.62) is $k \approx 2^{0.228m} \epsilon^{-2} \log \delta^{-1}$. The time cost of simulating the circuit using these approximate decompositions is determined by this scaling.

More generally, the optimal product decomposition of this type has 1-norm

$$\|a\|_1^2 = \prod_{j=1}^{m} \xi(U_j). \tag{2.64}$$

where ξ is the stabilizer extent. In section 3.3, we describe how we can go beyond such product decompositions by re-writing the circuit in such a way that allows us to use Clifford decompositions of *matrix products* of gates.

Chapter 3

Improved simulation by Clifford recompilation

3.1 Introduction

When a quantum circuit U contains m non-Clifford single qubit gates, the output state $U|0\rangle^{\otimes n}$ can be written as sum of $k = 2^m$ stabilizer states. Randomized sparsification allows us to reduce k significantly at the expense of introducing an approximation error. Each term of this approximate stabilizer decomposition is computed by replacing each non-Clifford gate in the circuit with a Clifford gate chosen randomly according to its relative weight in the Clifford decomposition of the non-Clifford gate, and then classically simulating the circuit using the phase-sensitive Clifford simulations of section 2.1.4.

The time cost of producing the approximate state in eq. (2.62) is roughly the number of required stabilizer states k to approximate $U|0^n\rangle$ with high probability multiplied by the cost of computing each stabilizer term. Denoting by ℓ the total number of gates in the circuit, the update rules in section 2.1.4 show that we can compute the CH-form of each stabilizer term in time $O(\ell n^2)$, and therefore the total time cost for producing a sparse approximation of $U|0\rangle^{\otimes n}$ is $O(k\ell n^2)$.

In Section 3.2, we provide an alternative method of computing the terms which requires only $O(mn^2)$ basic operations and a modest increase in memory. As k is typically exponentially large, such an improvement can result in a useful speed up whenever $\ell \gg m$. This is done by absorbing the action of the Clifford part of the circuit into the input state $|0^n\rangle$, while simultaneously keeping track of an efficient description of certain many-qubit gates that is well-suited for phase-sensitive Clifford simulation.

In Section 3.3, we show that the improved method allows for reducing the exponential scaling of the runtime by utilizing low rank approximate decompositions of *matrix products* of non-Clifford gates.

In section 3.4 we suggest a straight-forward method of estimating amplitudes with inverse polynomial additive error using this canonical form of the circuit. Such estimates are useful when the probability we want to estimate is not exponentially small, as well as in approximate-sampling from the circuit distribution using the heuristic Metropolis algorithm in [8].

3.2 Clifford recompilation

We now describe an improved method for generating low stabilizer rank approximations of the output state $U|0\rangle^{\otimes n}$ of a quantum circuit U based on a re-writing procedure, which we call *Clifford recompilation*. Here the goal is to simulate a quantum circuit U given as

$$U = C_m U_m C_{m-1} U_{m-1} \dots C_1 U_1 C_0, \tag{3.1}$$

where C_i are multi-qubit Clifford sub-circuits, and U_i are non-Clifford gates. The first step is to commute all non-Clifford gates to the end of the circuit

$$U = U'_m \dots U'_1 C_m C_{m-1} \dots C_1 C_0, \tag{3.2}$$

where $U'_i = C_{m:i} U_i C^{\dagger}_{m:i}$, with the concise notation $C_{j:k} := C_j C_{j-1} \dots C_k$. Of course the U'_i no longer act on a fixed number of qubits, and one might expect them to be too complicated to store and manipulate. However, as we will explain below, we can efficiently keep track of a compact description of the U'_i that is well-suited for randomized sparsification.

We can absorb the action of the Clifford gates into the input state, so that the output state is given by

$$U|0^n\rangle = U'_m \dots U'_1 |\phi\rangle \tag{3.3}$$

where $|\phi\rangle \equiv C_{m:0}|0^n\rangle$ is a stabilizer state. Note that the CH-form of $|\phi\rangle$ can be computed in time $O(cn^2)$, where $c \equiv \ell - m$ is the number of Clifford gates in U. In any case, it is a one-time computation that does not need to be iterated over in the simulation. Storing the CH-form of $|\phi\rangle$ in memory requires $O(n^2)$ bits.

Algorithm 1: Monte Carlo algorithm for improved randomized sparsification

```
Input  : The CH-form $(\omega_0, U_C^0, U_H^0)$ of the stabilizer state $|\phi\rangle$ in eq. (3.3). A
         decomposition of the form in eq. (3.7) for $i = 1, \ldots, m$.
Output: A random vector $|z\rangle$ such that $\mathbb{E}(|z\rangle) = U|0\rangle^{\otimes n}$.
begin
    Set $(\omega, U_C, U_H) \leftarrow (\omega_0, U_C^0, U_H^0)$
    for $i$ =1, ..., $m$ do
        Choose $V'_{ij}$ with probability $p_i(j) \equiv |a_{ij}|/\|a_i\|_1$;
        Update $(\omega, U_C, U_H)$ under $V'_{ij}$;
        Set $\omega \leftarrow (a_{ij}/|a_{ij}|)\omega$;
    Output $|z\rangle = \|a\|_1 w U_C U_H |s\rangle$.
end
```

Next we discuss how to compute and store the U'_i. Suppose we are given a Clifford decomposition of each non-Clifford gate in U;

$$U_i = \sum_j a_{ij} V_{ij}. \tag{3.4}$$

We will use exponentiated Pauli matrices to compute and store U'_i efficiently. First note that each V_{ij} can be written as a product of Clifford gates of the form $\exp(i\theta P)$ where P is a Pauli and $\theta \in (\pi/4)\mathbb{Z}$ [1]. We can therefore write

$$U_i = \sum_j a_{ij} \prod_{k=1}^{O(1)} \exp(i\theta_{ijk} P_{ijk}), \tag{3.6}$$

where a_{ij} absorbs any phases from the elementary gates, and $\{P_{ijk}\}$ is some collection of

[1] For example, the Hadamard, PHASE, and CZ gates can be written as

$$\begin{aligned} H_j &= e^{i\pi/2} e^{-i\frac{\pi}{2} Z_j} e^{i\frac{\pi}{4} Y_j}, \\ S_j &= e^{i\pi/4} e^{-i\frac{\pi}{4} Z_j}, \\ CZ_{jk} &= e^{i\pi/4} e^{i\frac{\pi}{4} Z_j Z_k} e^{-i\frac{\pi}{4} Z_j} e^{-i\frac{\pi}{4} Z_k}. \end{aligned} \tag{3.5}$$

Similar identities can be derived for any Clifford gate.

Pauli matrices. Then

$$\begin{aligned} U'_i &= \sum_j a_{ij} C_{m:i} \left[\prod_k \exp(i\theta_{ijk} P_{ijk}) \right] C^\dagger_{m:i} \\ &= \sum_j a_{ij} \prod_k \exp(i\theta_{ijk} C_{m:i} P_{ijk} C^\dagger_{m:i}) \\ &= \sum_j a_{ij} \prod_k \exp(i\theta'_{ijk} P'_{ijk}) \\ &= \sum_j a_{ij} V'_{ij}. \end{aligned} \tag{3.7}$$

where

$$V'_{ij} \equiv \prod_k \exp(i\theta'_{ijk} P'_{ijk}), \tag{3.8}$$

$\theta'_{ijk} \in \{\pm\theta_{ijk}\}$, and the P'_{ijk} are Pauli matrices each of which can be computed efficiently in time $O(c)$. Storing this description of U' requires $O(nm)$ bits of memory, since each Pauli matrix requires $O(n)$ bits to specify.

Algorithm 1 describes the improved randomized sparsification Monte Carlo. By Hoeffding's inequality, the required number of iterations of the Monte Carlo in order to bound the error by ϵ with probability at least $1-\delta$ is $O(\|b\|_1^2 \epsilon^{-2} \log(1/\delta))$. This is the same number of iterations required of the vanilla randomized sparsification method in [8] (see eq. (2.61)), but the point is that each iteration in the improved algorithm requires only $O(mn^2)$ basic operation.

In short, our improved randomized sparsification removes the dependence of the runtime on how many Clifford gates there are. This is achieved by recompiling the circuit and absorbing the action of all Clifford gates in the circuit into the input state, which only needs to be computed once.

3.3 Clifford decomposition of products of multi-qubit gates

As we mentioned earlier in this chapter, finding a Clifford decomposition of a non-Clifford gate U_i with minimal 1-norm (i.e. minimum $\|a_i\|_1$ in eq. (3.4)) can be done using a convex optimization software [11, 8]. These brute-force optimizations can only be performed for one- or two-qubit gates, as it is too computationally intensive for more qubits. In this section we describe how recompiling the circuit as in the previous section allows us to search for more sparse approximate stabilizer decompositions of the output state.

The stabilizer extent for unitaries is sub-multiplicative under composition, i.e.

$$\xi(AB) \leq \xi(A)\xi(B). \tag{3.9}$$

For two operators A and B, let us call a Clifford decomposition of the product AB *contractive* if its 1-norm is strictly less than $\xi(A)\xi(B)$. Finding a contractive Clifford decomposition of AB is generally hard unless A and B are at most 2-qubit gates acting on the same pair of qubits, so that we can use convex optimization.

An exception is when the two operators can be mapped to the same pair of qubits via a Clifford C. In that case, one can find a contractive decomposition of $CABC^\dagger$ using convex-optimization, and then apply the inverse $C^\dagger$, using the idea of eq. (3.7) to keep the terms in a form enabling fast phase-sensitive Clifford updates. This method is well-suited to the case of Clifford+$e^{i\theta P}$ circuits, which we discuss next. Furthermore, if better numerical methods can be found for minimizing the 1-norm, so that one can find the optimal decomposition for 3 qubit gates, this Clifford-equivalence method can be used with a much greater effect.

Example: Clifford+$e^{i\theta P}$ circuits. As an example of how these ideas can be used in practice, we consider circuits where each non-Clifford gate has the form $e^{i\theta P}$, where P is a Pauli operator, and $\theta \in [-\pi, \pi]$. This class of circuits includes important cases such as Clifford+T circuits, and more generally Clifford+Z rotation circuits. For a circuit U of this type containing m non-Clifford gates, eq. (3.3) reduces to

$$U|0^n\rangle = e^{i\theta_m Q_m} \dots e^{i\theta_1 Q_1}|\phi\rangle, \tag{3.10}$$

for a list of Hermitian n-qubit Pauli operators $Q_1, \dots, Q_m$, a list of angles $\theta_1, \dots, \theta_m$, and a stabilizer state $|\phi\rangle$.

Note that the unitary $e^{i\theta P}$ is Clifford if and only if $\theta \in (\pi/4)\mathbb{Z}$. Thus we may, without loss of generality, assume that $\theta_i \in (0, \frac{\pi}{4})$, as otherwise we can "factor out" Clifford gates $e^{\pm i\frac{\pi}{4}Q_i}$ and absorb them into the state $|\phi\rangle$, at the cost of applying a Clifford mapping to the list of Pauli operators Q_j with $j < i$. The Clifford mapping can be quickly computed using the relation

$$e^{i\frac{\pi}{4}P} Q e^{-i\frac{\pi}{4}P} = \begin{cases} Q & \text{if } [P, Q] = 0 \\ iPQ & \text{if } \{P, Q\} = 0. \end{cases} \tag{3.11}$$

We refer to eq. (3.10) as the *canonical form* of the output state of a Clifford+$e^{i\theta P}$ circuit.

What is the optimal Clifford decomposition of the product

$$e^{i\theta_m Q_m} \dots e^{i\theta_1 Q_1},$$

and how can we find it? Here optimal is in the sense of having a vector of coefficients with minimum 1-norm. Not only is this question too hard to settle using currently known methods, it is unlikely that there will ever be tractable numerical or analytical tools that can answer it in general.

We therefore restrict our attention to the task of *reducing* the 1-norm below that which is achieved by optimally decomposing each exponential separately. Recall that, for a single qubit Pauli operator P, we can *numerically* find the optimal decomposition of $e^{i\theta P}$ into Clifford gates, and it is given by

$$e^{i\theta P} = (\cos\theta - \sin\theta) I + (\sqrt{2}\sin\theta) e^{i\frac{\pi}{4}P}. \tag{3.12}$$

The above decomposition achieves the stabilizer extent

$$\xi(e^{i\theta P}) = \Big[\cos\theta + (\sqrt{2}-1)\sin\theta\Big]^2 \qquad (0 \leq \theta \leq \frac{\pi}{4}). \tag{3.13}$$

This numerically obtained fact for single qubit Pauli operators can be generalized to n-qubit Pauli operators, since we can map an n-qubit Pauli to a Pauli acting on a single qubit via Clifford conjugation. This assertion follows from i) invariance of ξ under multiplication by Clifford unitaries, ii) invariance of ξ under tensoring by the identity, i.e. $\xi(U \otimes I) = \xi(U)$.

Let's now consider a product of a small number of exponentials. If a sub-sequence of the exponentials in eq. (3.10) multiplies to a Clifford gate, it can be commuted past the gates prior to it and absorbed into the input stabilizer state $|\phi\rangle$, effectively reducing the number of non-Clifford gates in the circuit. As we are dealing with a universal set of gates, it is generally intractable to determine whether any such sub-sequence belongs to the Clifford group.

For a sub-sequence of length two; $e^{i\theta P} e^{i\phi Q}$, with $\theta, \phi \in (0, \pi/4)$, $PQ \neq \pm I$, there exists no better decomposition of $e^{i\theta P} e^{i\phi Q}$ than the product one, obtained by decomposing each gate separately and multiplying. In other words,

Lemma 7. *For any two n-qubit Paulis P and Q, $PQ \neq \pm I$, and $\theta, \phi \in (0, \pi/4)$, we have*

$$\xi\left(e^{i\theta_1 P} e^{i\theta_2 Q}\right) = \xi\left(e^{i\theta_1 P}\right) \xi\left(e^{i\theta_2 Q}\right). \tag{3.14}$$

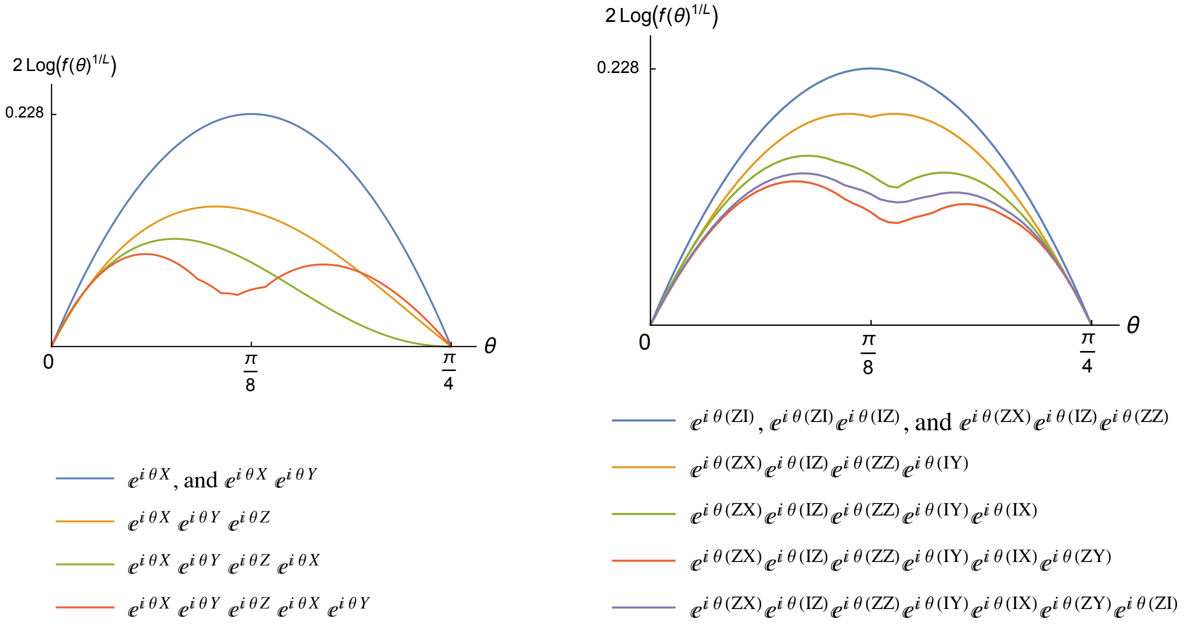

Figure 3.1: Theoretical time cost of simulating various one- and two-qubit exponential sequences. The angles are chosen to be equal for simplicity. For a sequence of length L, the function $f(\theta)$ is the 1-norm of the decomposition of the sequence into Clifford gates, and $f(\theta)^{1/L}$ is the effective 1-norm per gate. The function $2\log(f(\theta)^{1/L})$ is the contribution of each gate to α, where 2^α is the exponential scaling factor in the runtime of the simulation. The data is obtained using a convex optimization software (CVX with the SDPT3 solver).

Proof. This can be verified numerically for two-qubit Paulis, and the general case follows from the fact that any pair of n-qubit Paulis P and Q can be mapped by a Clifford circuit to a pair of Paulis with support on the first two qubits only. □

For products of three or more exponentials, we can find decompositions with significantly reduced 1-norm in certain cases. Numerically, we observe that the possibility of 1-norm reduction is correlated with the rank of the binary representation of the Paulis. Specifically, for $x, z \in \mathbb{Z}_2^n$ write $X[x] \equiv X^{x_1} \otimes \cdots \otimes X^{x_n}$, $Z[z] \equiv Z^{z_1} \otimes \cdots \otimes Z^{z_n}$, and

$P_{(x,z)} \equiv i^{x\cdot z} X[x]Z[z]$. Then numerics in the two qubit case suggest that,

$$\xi\left(\prod_j e^{i\theta_j P_{v_j}}\right) < \prod_j \xi(e^{i\theta_j P_{v_j}}) \tag{3.15}$$

whenever $\mathrm{rank}(v_1, v_2, \cdots, v_k) < k$ with $\mathrm{rank}(\cdot)$ computed mod 2.

For example, it is possible to numerically verify that

$$\xi\left(e^{i\theta_1 X} e^{i\theta_2 Y} e^{i\theta_3 Z}\right) < \xi\left(e^{i\theta_1 X}\right) \xi\left(e^{i\theta_2 Y}\right) \xi\left(e^{i\theta_3 Z}\right), \tag{3.16}$$

for any choice of angles $(\theta_1, \theta_2, \theta_3)$. We can also numerically find the optimal Clifford expansion

$$e^{i\theta_1 X} e^{i\theta_2 Y} e^{i\theta_3 Z} = \sum_i \alpha_i C_i, \tag{3.17}$$

for some collection of single qubit Cliffords C_i. For example, for any $0 \leq \theta \leq \pi/8$, the optimal expansion is given by

$$e^{i\theta X} e^{i\theta Y} e^{i\theta Z} = c_0(\theta) I + c_1(\theta) SH + c_2(\theta) HX + c_3(\theta) SHS \tag{3.18}$$

where

$$\begin{aligned} c_0(\theta) &= \cos 3\theta - \sin\theta \\ c_1(\theta) &= e^{i\pi/4}[\sin 2\theta(\cos\theta - \sin\theta)] \\ c_2(\theta) &= c_3(\theta) = (\sqrt{2}\sin 2\theta \sin\theta). \end{aligned} \tag{3.19}$$

More generally, the Clifford terms in the optimal expansion eq. (3.17) vary depending on $(\theta_1, \theta_2, \theta_3)$, and a general expression for the optimal decomposition is too tedious to write down, although finding it numerically is easy on a case-by-case basis.

Similarly to what we did in the case of two exponentials, we can now conclude that, for any triple of n-qubit Paulis (P, Q, W) which are equivalent to (X_1, Y_1, Z_1) via conjugation by a Clifford V, the optimal decomposition is given by

$$e^{i\theta_1 P} e^{i\theta_2 Q} e^{i\theta_3 W} = \sum_i \alpha_i V^\dagger (C_i \otimes I^{\otimes n-1}) V. \tag{3.20}$$

As before, each Clifford term $V^\dagger(C_i \otimes I^{\otimes n-1})V$ is easy to write as a product of a small number of Cliffords of the form $e^{i\phi R}$ where R is a Pauli and $\phi \in (\pi/4)\mathbb{Z}$, using the same idea in eq. (3.7). This way we obtain a decomposition of the many-body operator $e^{i\theta_1 P}e^{i\theta_2 Q}e^{i\theta_3 W}$ which is both optimal and in a form enabling fast phase-sensitive Clifford updates.

The reduction in the 1-norm by decomposing products of gates can be quite large, see for example Figure 3.1, where the log of the 1-norm is plotted for different one and two qubit gate sequences of varying lengths. This directly translates to classical simulation schemes with milder exponential scaling of the runtime.

3.4 Additive-error approximation of amplitudes

In section 2.2 we described how a marginal circuit probability [2] can be estimated using stabilizer decompositions of the output state. To bound the *multiplicative* error on the estimate by ϵ, the required runtime of this algorithm depends linearly on the number of terms in an *exact* stabilizer decomposition of the output state, and polynomially on the number of qubits and $1/\epsilon$.

For estimating an amplitude, (i.e. a probability of a fine-grained event, where we measure all of the qubits), the time cost of the above mentioned algorithm is $O(kn^5\epsilon^{-2})$, where k is the number of terms in the stabilizer decomposition. The n^5 factor is significant in practice, as was noted in [8].

In this section we instead consider an approximation correct up to *additive* error ϵ whose runtime scales like $O(kmn^2)$, where m is the number of non-Clifford gates in the circuit and k is the number of stabilizer terms in an approximate decomposition of the output state. Such estimates are useful when the probability we want to estimate is not exponentially small. Furthermore estimates obtained in this way can be used with the Markov Chain Monte Carlo variant of stabilizer rank methods, as shown in [8], which was shown to outperform other methods, despite having no rigorous error guarantees.

We describe how such an algorithm may be extended to other gate sets, such as nearest-neighbor matchgate circuits, and highlight how sampling Clifford-like unitary trajectories requires a significantly smaller sample size to guarantee convergence up to a required precision than other, similar approaches.

[2] i.e. the probability $P(x)$ of obtaining outcome x when measuring a subset of the qubits after applying a quantum circuit.

3.4.1 A simplified algorithm for estimating amplitudes

A simple Monte Carlo algorithm based on randomized sparsification can be used to estimate the amplitudes $\langle x|U|0^n\rangle$ of the circuit U. Indeed the output of Algorithm 1 is an unnormalized random vector $|\eta\rangle$, proportional to a stabilizer state, whose mean is $\mathbb{E}(|\eta\rangle) = U|0^n\rangle$. Thus for any $x \in \mathbb{F}_2^n$

$$\mathbb{E}(\langle x|\eta\rangle) = \langle x|U|0^n\rangle. \tag{3.21}$$

Given an instant $|\eta\rangle$, we can efficiently compute $\langle x|\eta\rangle$ using the methods in section 2.1.4. This is because $|\eta\rangle$ is proportional to a stabilizer state. Since $\langle x|\eta\rangle$ is an unbiased estimator of $\langle x|U|0^n\rangle$, we can use an empirical mean to estimate it, invoking a complex-valued version of Hoeffding's inequality [33, 9] to bound the probability of the additive error exceeding ϵ. An empirical mean R obtained by averaging k instances satisfies

$$\Pr\left[|R - \langle x|U|0^n\rangle| \geq \epsilon\right] \leq \delta, \tag{3.22}$$

provided we take $k \approx \|a\|_1^2\epsilon^{-2}\log(\delta^{-1})$. Thus by averaging enough evaluations of $\langle x|z\rangle$, one obtains a good estimate of $\langle x|U|0^n\rangle$ with high probability.

Given the CH-form of an instance of $|\eta\rangle$, it takes time $O(n^2)$ to compute the amplitude $\langle x|\eta\rangle$ using the methods in section 2.1.4. This gives a runtime scaling of $O(\|a\|_1^2 mn^2\epsilon^{-2}\log(\delta^{-1}))$ for computing an additive error estimate of the amplitude, using the improved sparsification Monte Carlo Algorithm 1.

Typically we are interested in estimating the probability $\Pr(x) = |\langle x|U|0^n\rangle|^2$, rather than the amplitude. It is straightforward to check that if $|R - \langle x|U|0^n\rangle| \leq c\epsilon$, where $c = \sqrt{2} - 1$, then $||R|^2 - P(x)| \leq \epsilon$ (assuming $\epsilon < 1$). In other words, estimating the probability instead of the amplitude only requires multiplying k by the constant $c^{-2} \approx 6$.

3.4.2 A digression on sampling computational paths

The algorithm just described for estimating $\langle x|U|0^n\rangle$ relies on sampling a large number of Clifford-like computational paths and averaging their contributions to the amplitude. This is achieved by writing the circuit U as sum of Clifford circuits,

$$U = \sum_i a_i W_i = \|a\|_1 \sum_i \frac{|a_i|}{\|a\|_1} W_i, \tag{3.23}$$

and then averaging many contributions $\|a\|_1 \langle x|W_i|0^n\rangle$, each chosen independently with probability $p(i) = \frac{|a_i|}{\|a\|_1}$ (here we assume that the a_i are positive by absorbing phases into the W_i). Each Clifford circuit W_i is associated with a computational trajectory. These trajectories branch at every non-Clifford gate in the circuit. The distribution p over trajectories can be sampled by implementing a suitable probabilistic update rule at every non-Clifford time step. This is an intuitive description of what algorithm 1 does.

Estimating the amplitude in this way clearly suffers from the problem of large cancellations. This is because $\|a\|_1 \gg 1$, and the contributions from different computational trajectories, namely $\|a\|_1 \langle x|W_i|0^n\rangle$, can have different phases. Catastrophic destructive interference can make the estimate of the amplitude unstable.

Despite this, we can still obtain rigorous error guarantees by averaging a large number of trajectories, roughly $\|a\|_1^2/\epsilon^2$. It is possible to obtain error guarantees with less trajectories if we had more information about the circuit, in particular the variance of the associated random variable. In general we do not have access to the variance and so we choose the most pessimistic tail-bound, i.e. the one specified by Hoeffding's inequality.

The main advantage of using Clifford decompositions of the circuit is that the required number of trajectories to obtain error guarantees has a mild exponential scaling with the number of non-Clifford gates. For instance, a circuit consisting of Clifford gates and m T gates requires averaging roughly $\|a\|_1^2/\epsilon^2 \approx 2^{0.228m}/\epsilon^2$ trajectories, a dramatic improvement over the total number of trajectories 2^m. This exponential scaling is in fact much milder than similar Monte Carlo methods for simulating Clifford+T circuits that have been recently reported. These alternative methods are based on Monte Carlo using entries of the gates in the computational basis [9], quasi-probability distributions [10], robustness of magic [11], and channel decompositions [12].

What is it about unitary Clifford trajectories that gives rise to this phenomenon? The obvious explanation is not very satisfactory; it is simply the geometric fact that there exists a decomposition

$$T = c_1 C_1 + c_2 C_2, \tag{3.24}$$

where C_i are Clifford gates, such that $\|c\|_1^{2m} = 2^{0.228m}$, where $\|c\|_1 = |c_1| + |c_2|$ is the 1-norm. Can we instead decompose the circuit as a linear combination over some other set of circuits and obtain a similarly mild exponential scaling? In the next subsection we explore such possibilities.

3.4.3 Generalizations to other gate sets

Suppose we want to estimate the amplitude $\langle x|U|0^n\rangle$ by decomposing U as a linear combination of some set of unitaries W_i. A necessary requirement is that we are able to efficiently compute the contribution from each term $\langle x|W_i|0^n\rangle$ on a classical computer.

As an example, given a circuit U in the gate set {HADAMARD, CNOT, NOT, T}, we may write U as a sum of circuits that preserve the computational basis. This can be done by decomposing each Hadamard gate in the circuit as a sum of Paulis;

$$H = \frac{1}{\sqrt{2}}(X + Z). \tag{3.25}$$

Denoting by h the total number of Hadamard gates in the circuit, the resulting expansion

$$U = \sum_i a_i W_i$$

has $\|a\|_1 = (2/\sqrt{2})^h = 2^{h/2}$. Therefore the required number of computational trajectories for constant additive error ϵ is roughly $\|a\|_1^2 = 2^h$. This offers no advantage in terms of the exponential scaling of the runtime over simply computing the sum of all 2^h branches.

Perhaps the most natural candidate for generalizing stabilizer rank methods is fermionic Gaussian states and nearest-neighbor matchgates. Nearest-neighbor matchgates have been known to be classically simulable since a seminal work of Valiant [2], who gave efficient algorithms for computing circuit probabilities and sampling the circuit distribution. An extension of these method and an interpretation of them in terms of non-interacting fermions was later identified by Terhal & Divincenzo [3]. Furthermore, a variant of the norm estimation algorithm, Lemma 6, specific to fermionic Gaussian states was derived by Bravyi & Gosset in the context of simulating quantum impurity models [34].

Importantly for our purpose, nearest-neighbor matchgate circuit amplitudes can be computed (i.e. with the phase) in classical polynomial time, see for instance the appendix in [35] for a quick recipe. Augmenting nearest-neighbor matchgates with nearest-neighbor SWAP gates is sufficient for universal quantum computation. This motivates searching for matchgate decompositions of the SWAP gate.

We can use a straightforward numerical optimization to minimize the 1-norm over matchgate decompositions of the SWAP gate. Namely, every two-qubit matchgate has the form

$$G(U,V) = \begin{pmatrix} u_{11} & 0 & 0 & u_{12} \\ 0 & v_{11} & v_{12} & 0 \\ 0 & v_{21} & v_{22} & 0 \\ u_{21} & 0 & 0 & u_{22} \end{pmatrix} \tag{3.26}$$

where $U, V \in U(2)$ satisfy $\det(U) = \det(V)$. Discarding a global phase we may take $U, V \in SU(2)$. Note that the SWAP gate cannot be written in this form since $\det(X) = -\det(I)$.

Our numerical search suggests that the decomposition

$$\text{SWAP} = \frac{e^{i\pi/4}}{\sqrt{2}} G(I, -iX) + \frac{e^{-i\pi/4}}{\sqrt{2}} G(I, iX) \tag{3.27}$$

has minimal 1-norm over all matchgate-decompositions of SWAP, where $G(I, \pm iX)$ is the $\pm i$SWAP gate

$$G(I, \pm iX) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \pm i & 0 \\ 0 & \pm i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.28}$$

This decomposition has 1-norm equal to $\sqrt{2}$, which implies that amplitude estimation using this decomposition has a runtime scaling like 2^m, where m is the number of SWAP gates. Since it is possible to exactly compute the amplitude in time $2^m\text{poly}(n)$ by computing the 2^m matchgate contributions, this decomposition is not useful for this purpose.

Our numerics are not exhaustive, and furthermore do not exclude the possibility of reducing the 1-norm by decomposing parallel SWAP gates into nearest-neighbor matchgate circuits involving more than two qubits.

Another possible generalization involves product vs. entangling gates. Considering the universal gate set $\{\text{HADAMARD}, \text{PHASE}, \text{T}, \text{CPHASE}\}$, we seek to find decompositions of the CPHASE gate into a sum of tensor product unitaries. We find numerical evidence that the decomposition

$$\text{CPHASE} = \frac{e^{i\pi 4}}{\sqrt{2}} (S_1^\dagger S_2^\dagger - iS_1 S_2) \tag{3.29}$$

achieves the minimum 1-norm over all decompositions of CPHASE as a sum of tensor product gates (not necessarily Cliffords), where S_i is the PHASE gate on qubit i. Similar to the matchgate case, the 1-norm of this decomposition is $\sqrt{2}$, too large to provide an improvement over brute force methods. Nevertheless, our numerical search is not exhaustive and a better decomposition may exist. It may also be the case that a decomposition of the form

$$\mathrm{CPHASE}^{\otimes k} = \sum_j a_j V_j, \tag{3.30}$$

exists such that each V_j is a product gate on $2k$ qubits and $\|a\|_1 < 2^{k/2}$. Our numerical search did not detect such decompositions for $k = 2$ allowing for up to 4 terms in the sum.

The reason it is not easy to perform conclusive numerical search, both for matchgate and product decompositions, is that the number of terms in the optimal decomposition may be large, and each term requires 6 extra search parameters (three for each gate in $SU(2)$). The decompositions given in eqs. (3.27) and (3.29) are optimal allowing up to 8 terms in the sum.

It is possible that much better decompositions can be achieved in the continuum limit, for example, via a complex-valued function f such that

$$\mathrm{CPHASE} = \int dU dV f(U,V) U \otimes V, \qquad \|f\|_1 = \int dU dV |f(U,V)| < \sqrt{2},$$

where integration is with respect to the Haar measure over SU(2). This would reduce the runtime of the amplitude estimation algorithm to $\|f\|_1^{2m} < 2^m$, where m is the number of CPHASE gates, provided we can efficiently sample the distribution $p(U,V) = |f(U,V)|/\|f\|_1$. A similar question arises for decompositions of SWAP in terms of matchgates, i.e. whether there exists a function g such that

$$\mathrm{SWAP} = \int dU dV g(U,V) G(U,V), \qquad \|g\|_1 = \int dU dV |g(U,V)| < \sqrt{2}.$$

Determining whether such f and g exist (whose associated distributions can be sampled efficiently) is an interesting question, but outside the scope of this discussion.

Finally, let us return to the example of decomposing Hadamard gates as a linear combination of computational-basis-preserving gates. It is easy to check that the decomposition

$$H = \frac{1}{\sqrt{2}}(X + Z), \tag{3.31}$$

is in fact optimal over all such single-qubit gates. It is easy to prove a somewhat stronger statement that lower-bounds the 1-norm in any decomposition of parallel Hadamard gates $H^{\otimes k}$ over the set $\mathcal{B}_k$ of computational-basis-preserving k-qubit unitaries. Equivalently $\mathcal{B}_k$ is the group of permutation matrices with a phase-mask: every element in $\mathcal{B}_k$ is obtained by taking a 2^k by 2^k permutation matrix and replacing its non-zero entries with phases.

In what follows we treat $\mathcal{B}_k$ as if it were a discrete group (it is in fact a continuous group). This does not lose any generality (just replace the sum with integral with respect to Haar measure over the group $\mathcal{B}_k$).

Proposition 8. *Suppose*

$$H^{\otimes k} = \sum_{W \in \mathcal{B}_k} a_W W. \tag{3.32}$$

Then $\|a\|_1 \geq 2^{k/2}$.

Proof. Every entry in $H^{\otimes k}$ is non-zero with modulus $2^{-k/2}$. Let $\mathcal{B}_k(i,j) \subset \mathcal{B}_k$ be the subset in which the (i,j)th entry is nonzero. We must have, for every (i,j),

$$2^{-k/2} = |\sum_{W \in \mathcal{B}_k(i,j)} a_W| \leq \sum_{W \in \mathcal{B}_k(i,j)} |a_W|.$$

As every element of $\mathcal{B}_k$ contains exactly one non-zero entry in every row, we have $\mathcal{B}_k(i,j) \cap \mathcal{B}_k(i,j') = \emptyset$ for $j \neq j'$, which implies

$$2^{k/2} = \sum_j |\sum_{W \in \mathcal{B}_k(i,j)} a_W| \leq \sum_j \sum_{W \in \mathcal{B}_k(i,j)} |a_W| \leq \|a\|_1.$$

□

Proposition 8 shows that no such decomposition can speed up the amplitude estimation Monte Carlo beyond the baseline scaling of 2^h, where h is the number of Hadamard gates in the circuit.

To conclude, up to validity of the numerical evidence we gathered, natural generalizations of the amplitude estimation and randomized sparsification to other gate sets do not offer any advantage in terms of the scaling of the simulation runtime. Only Clifford unitaries and stabilizer states seem to have that feature.

Chapter 4

Upper bounds on the exact stabilizer rank

4.1 Introduction

Finding low-rank stabilizer decompositions of the output state of a quantum circuit can significantly reduce the runtime of simulating the circuit on a classical computer. The sparsification lemma offers one way to construct low-rank approximate decompositions. One drawback of approximate stabilizer decompositions is that we do not know how to use them to estimate circuit probabilities with multiplicative error ϵ in time $k \cdot poly(1/\epsilon)$, where k is the number of stabilizer terms and we are ignoring the dependence on other variables. In contrast, exact stabilizer decompositions can be used for this purpose.

In this chapter we describe a method of constructing exact, low-rank stabilizer decompositions of $|\psi\rangle^{\otimes m}$, where $|\psi\rangle$ is a single qubit state. Using the gadget-based methods of [7, 8] (see section 2.2), such decompositions allow us to estimate marginal probabilities of quantum circuits consisting of Cliffords gates and m instances of a certain non-Clifford gate associated with $|\psi\rangle$, with a resource cost scaling as $k \cdot poly(1/\epsilon)$, where k is the number of terms in the decomposition and ϵ is the multiplicative error.

We focus on the case where $|\psi\rangle$ is either a magic state or an equatorial state. Recall that magic states are the non-stabilizer eigenstates of single qubit Clifford unitaries. The set of magic states decomposes into two Clifford orbits, of either the T state $|T\rangle = 2^{-1/2}(|0\rangle + e^{i\pi/4}|1\rangle)$ or the face state $|F\rangle = \cos(\beta)|0\rangle + e^{i\pi/4}\sin(\beta)$, where β is an angle such that $\cos(2\beta) = 1/\sqrt{3}$. Equatorial states are those of the form $|\theta\rangle = 2^{-1/2}(|0\rangle + e^{i\theta}|1\rangle)$.

We demonstrate a family of decompositions with improved asymptotic scaling for such states, namely a stabilizer decomposition of $|T\rangle^{\otimes m}$ with $2^{m\log_2(3)/4}$ terms. This improves on the best known bound of $2^{m\log_2(7)/6}$, and is based on a stabilizer decomposition of $|T\rangle^{\otimes 6}$ with 6 terms. The stabilizer rank of $|T\rangle^{\otimes 6}$ was conjectured to be 7 in [6]. We also find a decomposition of $|F\rangle^{\otimes m}$ with the same number of terms. The stabilizer rank of $|F\rangle^{\otimes m}$ was conjectured in [6] to be equal to that of $|T\rangle^{\otimes m}$, and the fact that our method applies to both types of magic state may be seen as evidence supporting this conjecture. Finally, for any equatorial state $|\theta\rangle$ we give a stabilizer decomposition of $|\theta\rangle^{\otimes m}$ with $2^{m/2}$ terms. To our knowledge no such decompositions were previously known.

The results in this chapter are based on joint work with David Gosset [1] and Hakop Pashayan [2].

4.2 The stabilizer rank of cat states

In this chapter we will make use of relations involving the stabilizer rank of certain entangled states. In this section we summarize these relations. For a single qubit state $|\psi\rangle$, define the cat state

$$|\mathrm{cat}_n(\psi)\rangle := 2^{-1/2}(|\psi\rangle^{\otimes n} + |\psi^{\perp}\rangle^{\otimes n}), \tag{4.1}$$

where $|\psi^{\perp}\rangle$ is an orthogonal state to $|\psi\rangle$. The phase on $|\psi^{\perp}\rangle$ is fixed by assuming that $\langle 0|\psi^{\perp}\rangle$ is real and positive (the case $|\psi\rangle = |0\rangle$ is unimportant and we can discard it from this definition).

The stabilizer ranks of $|\psi\rangle^{\otimes n}$ and $|\mathrm{cat}_n(\psi)\rangle$ are related as follows.

Proposition 9. *For any single qubit quantum state $|\psi\rangle$, we have*

$$\chi\left(\psi^{\otimes n}\right) \leq 4\chi\left(cat_n(\psi)\right).$$

Moreover, if $|\psi\rangle$ is a magic state then

$$\chi\left(\psi^{\otimes n}\right) \leq 2\chi\left(cat_n(\psi)\right).$$

[1] Institute for Quantum Computing, Waterloo, Ontario, and the Department of Combinatorics and Optimization, University of Waterloo, Ontario.

[2] Perimeter Institute, Waterloo, Ontario, and the Institute for Quantum Computing, Waterloo, Ontario.

Proof. We have

$$|\psi\rangle^{\otimes n} \propto (|\psi\rangle\langle\psi| \otimes I)|\mathrm{cat}_n(\psi)\rangle.$$

Decomposing $|\psi\rangle\langle\psi|$ into the four Pauli matrices implies

$$\chi(\psi^{\otimes n}) \leq 4\chi\left(\mathrm{cat}_n(\psi)\right). \tag{4.2}$$

If $|\psi\rangle$ is a magic state then it is a $+1$ eigenstate of a Hermitian Clifford C, and therefore $|\psi\rangle\langle\psi| = \frac{1}{2}(I + C)$ and $\chi(\psi^{\otimes n}) \leq 2\chi\left(\mathrm{cat}_n(\psi)\right)$. □

Next we describe sparse stabilizer decompositions of cat states of up to six qubits. In particular, we focus on cat states constructed from single qubit equatorial states and from magic states.

Cat states constructed from equatorial states

Let $|\theta\rangle = 2^{-1/2}(|0\rangle + e^{i\theta}|1\rangle)$. For $n \leq 6$, the states $\mathrm{cat}_n(|\theta\rangle)$ can be decomposed in terms of at most 4 permutation-invariant stabilizer states, namely

$$|0\rangle^{\otimes n}, \quad |1\rangle^{\otimes n}, \quad |E_n\rangle \equiv 2^{-(n-1)/2} \sum_{x \text{ even}} |x\rangle, \quad |K_n\rangle \equiv 2^{-(n-1)/2} \sum_{x \text{ even}} (-1)^{S(x)}|x\rangle, \tag{4.3}$$

where $S(x) = \sum_{j<k} x_j x_k$. The exact decompositions are given by

$$\begin{aligned}
|\mathrm{cat}_2(\theta)\rangle &= (|0\rangle^{\otimes 2} + e^{2i\theta}|1\rangle^{\otimes 2})/\sqrt{2}, \\
|\mathrm{cat}_3(\theta)\rangle &= \frac{1}{2}|E_3\rangle + (\frac{1}{2} - e^{2i\theta})|K_3\rangle \\
|\mathrm{cat}_4(\theta)\rangle &= \frac{-ie^{i\theta}\sin\theta}{\sqrt{2}}|0\rangle^{\otimes 4} + \frac{ie^{3i\theta}\sin\theta}{\sqrt{2}}|1\rangle^{\otimes 4} + e^{2i\theta}|E_4\rangle, \\
|\mathrm{cat}_5(\theta)\rangle &= (1/4)(1 - e^{4i\theta})|0\rangle^{\otimes 5} + e^{3i\theta}\cos\theta|E_5\rangle + ie^{3i\theta}\sin\theta|K_5\rangle \\
|\mathrm{cat}_6(\theta)\rangle &= (1/4\sqrt{2})(1 - e^{5i\theta})|0\rangle^{\otimes 6} + (1/4\sqrt{2})(e^{6i\theta} - e^{2i\theta})|1\rangle^{\otimes 6} \\
&\quad + e^{3i\theta}\cos\theta|E_6\rangle + ie^{3i\theta}\sin\theta|K_6\rangle.
\end{aligned}$$

Therefore for any θ we have the upper bounds

$$\chi\left(\mathrm{cat}_2(\theta)\right) = \chi\left(\mathrm{cat}_3(\theta)\right) = 2, \quad \chi\left(\mathrm{cat}_4(\theta)\right) \leq 3, \quad \chi\left(\mathrm{cat}_5(\theta)\right) \leq 3, \quad \chi\left(\mathrm{cat}_6(\theta)\right) \leq 4.$$

Cat states constructed from the T state

For some values of θ these upper bounds are not tight. For example, the states $\mathrm{cat}_n(T)$ constructed from $|T\rangle = |0\rangle + e^{i\pi/4}|1\rangle$ satisfy

$$\chi\left(\mathrm{cat}_2(T)\right) = 1, \quad \chi\left(\mathrm{cat}_4(T)\right) = 2, \quad \chi\left(\mathrm{cat}_6(T)\right) \leq 3.$$

The above can be verified by noting that for $\theta = \pi/4$ certain terms in the above decompositions combine into stabilizer states, reducing the overall number of stabilizer terms.

Cat states constructed from the F state

The face state is given by

$$|F\rangle = \cos(\beta)|0\rangle + e^{i\pi/4}\sin(\beta)|1\rangle \tag{4.4}$$

where $\beta \in [0, \pi/2]$ is an angle such that $\cos(2\beta) = 1/\sqrt{3}$. For $n \leq 6$, the states $\mathrm{cat}_n(F)$ admit low rank stabilizer decompositions similar to those of the T state. The most important two are $\mathrm{cat}_2(F)$ and $\mathrm{cat}_6(F)$:

$$\begin{aligned}
|\mathrm{cat}_2(F)\rangle &= (|0\rangle^{\otimes 2} + i|1\rangle^{\otimes 2})/\sqrt{2}, \\
|\mathrm{cat}_6(F)\rangle &= (2/3)|\psi_1\rangle + e^{3i\pi/4}(2/3)|\psi_2\rangle - e^{i\pi/4}(2/3)|\psi_3\rangle
\end{aligned}$$

where

$$\begin{aligned}
|\psi_1\rangle &= 2^{-1/2}(|0\rangle^{\otimes 6} - i|1\rangle^{\otimes 6}) \\
|\psi_2\rangle &= 2^{-3} \sum_{x \in \{0,1\}^6} (-i)^{|x| \bmod 2}|x\rangle \\
|\psi_3\rangle &= 2^{-3} \sum_{x \in \{0,1\}^6} (-1)^{S(x)+|x|}|x\rangle.
\end{aligned} \tag{4.5}$$

These decompositions imply

$$\chi(\mathrm{cat}_2(F)) = 1, \qquad \chi(\mathrm{cat}_6(F)) \leq 3. \tag{4.6}$$

Contracting cat states

The stabilizer rank of $|\mathrm{cat}_n\rangle$, constructed from any single qubit state $|\psi\rangle$ (which we omit to ease the notation), can be upper bounded via certain tensor contractions. Here, we explain the idea by considering an example. We can prepare cat states by contracting other cat states together, for instance

$$\left(\langle T|^{\otimes 2} + \langle T^{\perp}|^{\otimes 2}\right)_{\text{on qubits 6 \& 7}}\left(\left(|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}\right)\left(|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}\right)\right) = |T\rangle^{\otimes 10} + |T^{\perp}\rangle^{\otimes 10} \tag{4.7}$$

Proposition 10. *Let $|\phi\rangle$ be an n qubit state and $|\phi'\rangle$ be a state on $s < n$ qubits. Then the stabilizer rank of the state $\langle\phi'|\phi\rangle$, where the contraction is on some subset of s qubits, satisfies*

$$\chi(\langle\phi'|\phi\rangle) \leq \chi(\phi)\chi(\phi'). \tag{4.8}$$

Proof. For every stabilizer decomposition of $|\phi\rangle$ and $|\phi'\rangle$ using stabilizer states $\{|q\rangle\}$ and $\{|q'\rangle\}$ respectively, there is a decomposition of $\langle\phi'|\phi\rangle$ using stabilizer states $\{\langle q'|q\rangle\}$. □

Combining eq. (4.7) and Proposition 10, we obtain the upper bound

$$\chi(\mathrm{cat}_{10}(T)) \leq \chi(\mathrm{cat}_2(T))\chi(\mathrm{cat}_6(T))^2 \leq 9, \tag{4.9}$$

where we used $\chi(\mathrm{cat}_2(T)) = 1$ and $\chi(\mathrm{cat}_6(T)) \leq 3$.

More generally, we can place k copies of $|\mathrm{cat}_n\rangle$ in a 1D chain, and then connect each pair of nearest neighbours by acting with the contraction $\langle\mathrm{cat}_2|$ on one qubit from each. This prepares the state $|\mathrm{cat}_m\rangle$, where $m = nk - 2(k-1)$, implying that

$$\chi(\mathrm{cat}_m) \leq \chi(\mathrm{cat}_2)^{2(k-1)}\chi(\mathrm{cat}_n)^k. \tag{4.10}$$

4.3 Upper bounds on the stabilizer rank of n copies of single qubit states

We now consider the stabilizer rank of $|\psi\rangle^{\otimes n}$, where $|\psi\rangle$ is a single qubit state. We start by giving a quick proof of two improved upper bounds, and then, in Theorems 11 and 12, show how to improve them further.

Proposition 9 implies that

$$\chi(T^{\otimes 6}) \leq 6, \tag{4.11}$$

which follows since $\chi(\mathrm{cat}_6(T)) \leq 3$ and $|T\rangle\langle T| = \frac{1}{2}(I + C)$, where C is the Clifford unitary whose $+1$ eigenstate is $|T\rangle$ [3]. Equation (4.11) implies an improved upper bound

$$\chi(T^{\otimes m}) \leq \chi(T^{\otimes 6})^{m/6} \leq 2^{\tau m} \tag{4.12}$$

where $\tau = \log_2(6)/6 \approx 0.431$. An identical argument holds for the face state $|F\rangle$.

Similarly, for any equatorial single qubit state $|\theta\rangle = |0\rangle + e^{i\theta}|1\rangle$, the stabilizer rank of $|\theta\rangle^{\otimes m}$ is at most $2^{2m/3}$. This is true since

$$\chi(\theta^{\otimes 6}) \leq 16, \tag{4.13}$$

where we used Proposition 9 and noted that the stabilizer rank of $|\mathrm{cat}_6(\theta)\rangle$ is at most four. Therefore

$$\chi(\theta^{\otimes m}) \leq \chi(\theta^{\otimes 6})^{m/6} \leq 2^{2m/3}. \tag{4.14}$$

We can improve the two upper bounds in eq. (4.12) and eq. (4.14) by contracting cat states, which is shown in the next two theorems. Namely, the improved upper bounds for the stabilizer rank of m copies of magic/equatorial states are respectively given by $2^{\alpha m}$ and $2^{m/2}$, where $\alpha = \log_2(3)/4 \approx 0.3962$. Moreover, the proofs are constructive, meaning that we obtain decompositions which achieve these bounds.

Theorem 11. *Let $|\psi\rangle$ be a single qubit magic state. Let $\alpha > 0$ be the smallest number such that $\chi(\psi^{\otimes m}) \leq 2^{\alpha m}$ for all m. Then*

$$\alpha \leq \frac{\log_2(3)}{4} \approx 0.3962. \tag{4.15}$$

Proof. Since $|\psi\rangle$ is a magic state it is Clifford equivalent to either $|T\rangle$ or $|F\rangle$. Without loss of generality let it be equivalent to $|T\rangle$. For ease of notation set

$$|\Psi_n\rangle = |\mathrm{cat}_n(T)\rangle.$$

[3] Namely $C = e^{-i\pi/4}SX$ where S is phase gate

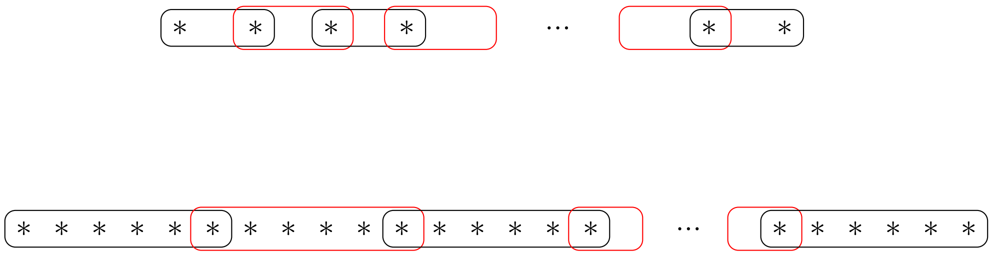

Figure 4.1: **Top**: the chain used for proving Theorem 11. There are k vertices and $k-1$ edges. Every vertex is a $|\Psi_6\rangle$ state and every edge is a $\langle\Psi_2|$ contraction on one qubit from each of the vertices it contains. **Bottom**: the chain used for proving Theorem 12. There are $5k+1$ vertices and k hyper-edges. Every vertex is a $|\Phi_6\rangle$ state and every hyper-edge is a $\langle\Phi_6|$ contraction on one qubit from each of the vertices it contains.

Consider the 1D chain in Figure 4.1 Top. There are k vertices and $k-1$ edges. Every vertex is a $|\Psi_6\rangle$ state and every edge is a $\langle\Psi_2|$ contraction on one qubit from each of the vertices it contains. The resulting state after the contraction is

$$\left(\langle\Psi_2|^{\otimes k-1}\right)_{\text{edges}}\left(|\Psi_6\rangle^{\otimes k}\right)_{\text{vertices}} \propto |\Psi_{4k+1}\rangle. \tag{4.16}$$

Since $\chi(\Psi_2)=1$ and $\chi(\Psi_6)\leq 3$, we have

$$\chi\left(\Psi_{4k+1}\right)\leq 3^k, \tag{4.17}$$

which implies

$$\chi(T^{\otimes 4k+1})\leq 2\times 3^k, \tag{4.18}$$

where we used Proposition 9 and noted that $|T\rangle\langle T|=\frac{1}{2}\chi(I+C)$ for some Clifford unitary C. Using Clifford invariance of the stabilizer rank, sub-multiplicativity of the stabilizer rank, and the bound eq. (4.18), we obtain

$$\chi(\psi^{\otimes m})=\chi(T^{\otimes m})\leq\chi(T^{\otimes 4k+1})^{m/(4k+1)}\leq(2\times 3^k)^{m/(4k+1)}\to 2^{\alpha m}, \tag{4.19}$$

as $k\to\infty$, where $\alpha=\frac{\log_2(3)}{4}$. □

Theorem 12. *For any equatorial single qubit state $|\theta\rangle=|0\rangle+e^{i\theta}|1\rangle$, the smallest number $\beta>0$ such that $\chi(\theta^{\otimes m})\leq 2^{\beta m}$ for all m satisfies*

$$\beta\leq\frac{1}{2}. \tag{4.20}$$

Proof. For ease of notation set

$$|\Phi_n\rangle = |\mathrm{cat}_n(\theta)\rangle.$$

Consider the chain in Figure 4.1 Bottom. There are $5k+1$ vertices and k hyper-edges. Every vertex is a $|\Phi_6\rangle$ state and every hyper-edge is a $\langle\Phi_6|$ contraction acting on one qubit from each of the vertices it contains. The resulting state after the contraction is given by

$$\left(\langle\Phi_6|^{\otimes k}\right)_{\text{hyper-edges}}\left(|\Phi_6\rangle^{\otimes 5k+1}\right)_{\text{vertices}} \propto |\Phi_{24k+6}\rangle. \tag{4.21}$$

Since $\chi(\Phi_6) \le 4$ for any equatorial state, we have

$$\chi\left(\Phi_{24k+6}\right) \le 4^{6k+1}, \tag{4.22}$$

which implies

$$\chi(\theta^{\otimes 24k+6}) \le 4^{6k+2}, \tag{4.23}$$

where we used eq. (4.2) and noted that $|\theta\rangle\langle\theta|$ can be written as a sum of four Cliffords. Using sub-multiplicativity of the stabilizer rank and the bound eq. (4.23), we obtain

$$\chi(\theta^{\otimes m}) \le \chi(\theta^{\otimes 24k+6})^{m/(24k+6)} \le (4^{6k+2})^{m/(24k+6)} \to 2^{m/2}, \tag{4.24}$$

as $k \to \infty$. □

4.4 Discussion, conjectures, and open problems

A key idea in this chapter is that the stabilizer rank of cat states grows more slowly than the stabilizer rank of the state they are constructed from. To get a feel of why this is the case, let us consider $|T\rangle = |0\rangle + e^{i\pi/4}|1\rangle$ and expand (ignoring normalization)

$$\begin{aligned}|T\rangle^{\otimes m} &= \sum_{x\in\mathbb{F}_2^m} \omega^{|x|}|x\rangle,\\ |T\rangle^{\otimes m} + |T^{\perp}\rangle^{\otimes m} &= \sum_{x\in\mathbb{F}_2^m\,:\,x\text{ even}} \omega^{|x|}|x\rangle,\end{aligned} \tag{4.25}$$

where $\omega = \exp(i\pi/4)$ and $|x|$ is the Hamming weight of the bit string x. The cat state looks the same as the original state, except its support is restricted to the even parity subspace of $\mathbb{F}_2^m$. In both cases in eq. (4.25), the problem is to find a sparse decomposition of the hamming weight phase $\omega^{|x|}$ in terms of phases of the form $i^{\ell(x)}(-1)^{q(x)}$ where ℓ and q are linear and quadratic functions[4]. On the even parity subspace we have $\omega^{|x|} = i^{|x|/2} \in \{\pm 1, \pm i\}$, and moreover, on this subspace the function $|x|/2$ is congruent modulo 4 to the totally symmetric quadratic polynomial $S(x) = \sum_{i<j} x_i x_j$, and therefore $i^{|x|/2} = i^{S(x)}$. This almost has the right form of a stabilizer state. It is then not too surprising that the stabilizer rank of the highly entangled cat states can be lower than the stabilizer rank of the product states they are constructed from.

For concreteness let us discuss two examples; the two-qubit and six-qubit cat states. The fact that these states have low stabilizer rank is a key ingredient in the proof of Theorem 11. In both cases, special cancellations/combinations allow for low rank stabilizer decompositions. The two-qubit cat state is a stabilizer state

$$|T\rangle^{\otimes 2} + |T^{\perp}\rangle^{\otimes 2} = |00\rangle + i|11\rangle. \tag{4.27}$$

For $|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}$, the even parity subspace of 6-bit strings contains one string of weight 0, one string of weight 6, and an equal number of strings of weights 2 and 4. Let's consider two quadratic phase functions: the first is the trivial phase $f_1(x) = 1$ and the second is the totally-symmetric quadratic phase $f_2(x) = (-1)^{\sum_{i<j} x_i x_j}$. The function f_1 is of course constant on all bit strings, and the function f_2 distinguishes between strings of weights 2 and 4. Therefore we can get the correct phase $\omega^{|x|}$ on all bit strings of weights 2 and 4 by a linear combination of the two stabilizer states

$$\sum_{x \text{ even}} f_1(x)|x\rangle, \quad \sum_{x \text{ even}} f_2(x)|x\rangle.$$

However, the phases on $|0\rangle^{\otimes 6}$ and $|1\rangle^{\otimes 6}$ will be incorrect. By a stroke of luck, the required corrections on $|0\rangle^{\otimes 6}$ and $|1\rangle^{\otimes 6}$ differ only by a phase multiple $-i$. Therefore both corrections can be applied with a single additional stabilizer state, namely

$$|0\rangle^{\otimes 6} - i|1\rangle^{\otimes 6}.$$

This gives us a decomposition of $|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}$ in terms of only 3 stabilizer states.

[4] Recall that every stabilizer state on m qubits has the form

$$\sum_{x \in A} i^{\ell(x)}(-1)^{q(x)}|x\rangle. \tag{4.26}$$

where A is an affine subspace of $\mathbb{F}_2^m$, q is a quadratic function on A, and ℓ is a linear function on A.

Unfortunately, for $m > 6$ a similar construction fails. A simple explanation is that there are too many m-bit strings of weight 6 in this case, and the correction we need to apply requires too many additional stabilizer terms. The best decomposition we found for $|T\rangle^{\otimes 7} + |T^{\perp}\rangle^{\otimes 7}$ requires six stabilizer states, which is too many for our purposes. It is an open question whether more sparse decompositions exist for $m = 7$.

In order for a decomposition of $|T\rangle^{\otimes 7} + |T^{\perp}\rangle^{\otimes 7}$ to improve the upper bound in Theorem 11, it needs to consist of no more than three stabilizer states, which in our opinion is unlikely to exist. For $|T\rangle^{\otimes 8} + |T^{\perp}\rangle^{\otimes 8}$, we need a decomposition consisting of no more than five stabilizer states. We leave it as an open question whether such a decomposition exists. Note that it is not hard to construct a decomposition of $|T\rangle^{\otimes 8} + |T^{\perp}\rangle^{\otimes 8}$ consisting of six stabilizer terms, for instance via the contraction

$$\left(\langle T|^{\otimes 2} + \langle T^{\perp}|^{\otimes 2}\right)_{\text{on qubits 4 \& 5}} \left(\left(|T\rangle^{\otimes 4} + |T^{\perp}\rangle^{\otimes 4}\right)\left(|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}\right)\right) = |T\rangle^{\otimes 8} + |T^{\perp}\rangle^{\otimes 8} \tag{4.28}$$

and the fact that $|T\rangle^{\otimes 2} + |T^{\perp}\rangle^{\otimes 2}$, $|T\rangle^{\otimes 4} + |T^{\perp}\rangle^{\otimes 4}$, and $|T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}$ can be decomposed in terms of one, two, and three stabilizer states, respectively.

The second important idea is contracting cat states to produce other cat states, as in eq. (4.28) above. Intuitively, this provides a systematic way of constructing decompositions of $|T\rangle^{\otimes m} + |T^{\perp}\rangle^{\otimes m}$ that consist of highly entangled stabilizer states. These decompositions can then be converted to stabilizer decompositions of $|T\rangle^{\otimes m}$ itself, for instance by projecting the first qubit onto $|T\rangle$. Thus cat states are useful not only because of their low stabilizer rank, but also because they allow us to include in the decomposition of $|T\rangle^{\otimes m}$ stabilizer states which have long range entanglement, which can only improve the sparsity compared to decompositions which are local in this sense.

Can the bound in Theorem 11 be improved? We can consider a generalization of Figure 4.1 Top to an arbitrary graph G with vertex set V and edge set E. Proceeding similarly to the proof of Theorem 11, we place a cat state of 6 qubits at each vertex, and contract pairs in the edge set E with cat states of two qubits. If the graph is connected, then the resulting state after the contraction is a cat state on $6|V| - 2|E|$ qubits. The number of stabilizer terms in the decomposition will be $3^{|V|}$. Proceeding as in the proof of Theorem 11, the upper bound on α (defined in the statement of Theorem 11) obtained in this way is

$$\alpha \leq \frac{\log_2(2 \times 3^{|V|})}{6|V| - 2|E|}. \tag{4.29}$$

To minimize this upper bound for fixed $|V|$ we need to have the minimum number of edges, subject to the graph being connected. This is equivalent to picking G to be a tree, so that the number of edges is $|E| = |V| - 1$. Therefore the graph used to prove Theorem 11 is optimal amongst such strategies. We can also show that we cannot do any better by a combination of vertices and edges consisting of cat states on $m \leq 6$ qubits. In other words, improving this upper bound will require something beyond simply considering other contraction graphs and other known decompositions of cat states.

On the other hand, we conjecture that the upper bound in Theorem 11, $\alpha \leq \log_2(3)/4 \approx 0.3962$, is not tight. The reason is the following. By sub-multiplicativity of the stabilizer rank, we know that the stabilizer rank of two copies of $|\mathrm{cat}_6\rangle = |T\rangle^{\otimes 6} + |T^{\perp}\rangle^{\otimes 6}$ is at most 9, with equality iff $\chi(|\mathrm{cat}_6\rangle|\mathrm{cat}_6\rangle) = \chi(|\mathrm{cat}_6\rangle)\chi(|\mathrm{cat}_6\rangle)$. There are no known examples of quantum states whose stabilizer rank is multiplicative for two copies, and this requirement seems quite stringent. Therefore it is quite likely that

$$\chi(|\mathrm{cat}_6\rangle|\mathrm{cat}_6\rangle) \underset{\text{conjecture}}{\leq} 8. \tag{4.30}$$

Plugging this in the proof of Theorem 11 gives the upper bound

$$\alpha \underset{\text{conjecture}}{\leq} \frac{3}{8} \approx 0.375. \tag{4.31}$$

Chapter 5

The stabilizer rank and continuous-time evolution

5.1 Introduction

Stabilizer-rank methods are useful for simulating quantum circuits, and a natural question is whether these ideas can be applied to simulate other types of quantum dynamics, such as continuous-time evolution. In this chapter we consider such applications.

The question we consider in this chapter is how to find low-rank stabilizer decompositions of the state $e^{iHt}|0\rangle^{\otimes n}$, where H is an n-qubit Hamiltonian. As we have previously seen, the cost of classically simulating this state scales linearly in the number of stabilizer terms in such a decomposition, provided we can compute the terms efficiently. Here by classical simulation we mean sampling bit strings from a distribution close to the distribution obtained by measuring a subset S of the qubits in the computational basis, in other words, approximately sampling from the probability distribution

$$P(x) := \langle 0^n|e^{-iHt}\Pi_x e^{iHt}|0^n\rangle, \quad x \in \mathbb{F}_2^{|S|}, \tag{5.1}$$

where Π_x acts as $|x\rangle\langle x|$ on S and as identity on all other qubits.

One straightforward way to construct low rank stabilizer decompositions of $e^{iHt}|0\rangle^{\otimes n}$ is to find an approximate Clifford decomposition of the operator e^{iHt} with small 1-norm $\|c\|_1$, and then use randomized sparsification to find an approximate stabilizer decomposition of $e^{iHt}|0\rangle^{\otimes n}$ within error ϵ using $\|c\|_1^2/\epsilon^2$ stabilizer states. As before, this is useful as long as $\|c\|_1^2/\epsilon^2 \ll 2^n$.

There are a number of different ways to find approximate decompositions of the operator e^{iHt} as a sum of Cliffords. For example, we can start with either a Lie product formula or a truncated series expansion, and then apply randomized sparsification to reduce the number of terms. This approach somewhat resembles well known methods such as worldline Monte Carlo and the stochastic series expansion, which are commonly used to compute static properties of the Hamiltonian, such as the ground state energy or the partition function.

The time evolution of a non-eigenstate of a spin Hamiltonian can be much harder to simulate than static properties of the Hamiltonian. For example, the ubiquitous class of stoquastic Hamiltonians, i.e. sign-problem free Hamiltonians, have efficiently-computable static properties [36], and yet their time evolution is universal for quantum computation [37]. As another example, computing the ground state energy of 1D nearest neighbour Hamiltonians can be done efficiently using matrix product states [38], and yet simulating the time evolution of such Hamiltonians is likely to be classically hard [39].

In this chapter the Hamiltonian is given as

$$H = \sum_{j=1}^{K} a_j Q_j, \tag{5.2}$$

where Q_j has an efficient classical description and $\|Q_j\| \leq 1$. For example the operator $\exp(i\theta Q_j)$ may be written as a sum of a small number of Cliffords for all θ. Such Q_j may be non-local (for example, n-qubit Pauli operators). We assume that the number of terms K is polynomial in n, and that the a_j are positive by absorbing signs into the Q_j.

For example, by taking $\{Q_j\}$ to be Pauli operators and using the first-order Suzuki integrator, we find that the required number of stabilizer states is, up to a polynomial factor, $\exp(\gamma\|a\|_1 t)$, where $\gamma = 2\sqrt{2} - 2 \approx 0.83$. If instead we apply randomized sparsification to a truncated Taylor approximation

$$e^{iHt} \approx T_{s,r} = \left(\sum_{j=0}^{r} \frac{(iHt/s)^j}{j!} \right)^s \tag{5.3}$$

then the number of stabilizer states required for the same error guarantee is, up to a polynomial factor, $\exp(2\|a\|_1 t)$.

This reduction in the exponent is interesting because it directly ameliorates the exponential scaling of the required classical resources to simulate short time evolution. For example, for fixed depth $t = 1/\log_2(e)$ and fixed $\|a\|_1/n = 1$, the required classical resources scale as $2^{\gamma n}$. The minimum possible γ is determined by the stabilizer extent of

$e^{iHt}|0\rangle^n$. We give examples of how to reduce γ in section 5.4 by finding more sparse stabilizer decompositions.

5.2 Clifford decompositions of the time evolution operator

There are a number of ways to find approximate decompositions of the operator e^{iHt} as a sum of Cliffords. For example, if we are given H in the form

$$H = \sum_j a_j Q_j, \tag{5.4}$$

where each Q_i is "easy to exponentiate" then we can use a Lie product formula. By easy to exponentiate we mean that, for every angle θ and every j, we can efficiently compute a Clifford decomposition

$$\exp(i\theta Q_j) = \sum_k c_{jk} C_{jk} \tag{5.5}$$

and sample the distribution $|c_{jk}|/\|c_j\|_1$, where $\|c_j\|_1 = \sum_k |c_{jk}|$. If a Lie product formula U consists of products of such exponentials and satisfies

$$\|U - e^{iHt}\| \le \epsilon \tag{5.6}$$

then we can approximate $e^{iHt}|0^n\rangle$ by a sum of stabilizer states obtained by multiplying the Clifford expansion of each exponential. We can then implement randomized sparsification by choosing C_{jk} with probability $|c_{jk}|/\|c_j\|_1$, for each exponential in the product formula. This samples world-line trajectories in which the state at every time step is a stabilizer state. By averaging enough trajectories the state we obtain is close to $e^{iHt}|0^n\rangle$ with high probability. The number of required trajectories, i.e. the number of stabilizer terms in the decomposition, depends on the particular Lie product formula and the decompositions eq. (5.5).

As an example of this approach, we consider the Trotter-Suzuki first order integrator

$$U_s := \left[\left(e^{ita_1Q_1/2s} \ldots e^{ita_KQ_K/2s}\right)\left(e^{ita_KQ_K/2s} \ldots e^{ita_1Q_1/2s}\right)\right]^s. \tag{5.7}$$

To bound the approximation error $\|e^{iHt} - U_s\|$ by δ it is sufficient to take $s = \sqrt{(2K\|a\|_\infty t)^3/\delta}$, where $\|a\|_\infty = \max_i |a_i|$ [40]. More sophisticated product formulas can reduce the number of gates required for a given error tolerance. We will stick to eq. (5.7) to simplify the analysis.

The stabilizer extent of U_s can be upper bounded by decomposing each exponential in terms of Cliffords. For example, if Q is a Pauli operator then we can use the optimal decomposition eq. (3.12)

$$e^{i\theta Q} = (\cos\theta - \sin\theta)I + (\sqrt{2}\sin\theta)e^{i\frac{\pi}{4}Q}. \tag{5.8}$$

Using Lemma 4 (the randomized sparsification lemma) we obtain

$$\xi(U_s) \leq \prod_{j=1}^{K}\Big(\cos(a_j t/2s) + (\sqrt{2}-1)\sin(a_j t/2s)\Big)^{4s}. \tag{5.9}$$

Note that for any s the right hand side is less than $e^{(2\sqrt{2}-2)\|a\|_1 t}$, and for large s we have

$$\lim_{s\to\infty}\prod_{j=1}^{K}\Big(\cos(a_j t/2s) + (\sqrt{2}-1)\sin(a_j t/2s)\Big)^{4s} = e^{(2\sqrt{2}-2)\|a\|_1 t}. \tag{5.10}$$

This implies

$$\xi(U_s) \leq e^{(2\sqrt{2}-2)\|a\|_1 t}. \tag{5.11}$$

To compare this to a more standard approach, let us use the same Pauli decomposition of the Hamiltonian in a truncated series approximation

$$e^{iHt} \approx T_{s,r} := \left(\sum_{j=0}^{r}\frac{(iHt/s)^j}{j!}\right)^s. \tag{5.12}$$

To bound the error by δ one can take (see for example [41])

$$s = \|a\|_1 t, \quad r = \frac{\log\log(s/\delta)}{\log(s/\delta)}. \tag{5.13}$$

To perform randomized sparsification with this series, we start by writing

$$T_{s,r} = \left[N_{s,r}\sum_{j=0}^{r}p(j)\sum_{z\in\mathbb{Z}_K^j}q(z)(i^j Q_{z_1}\ldots Q_{z_j})\right]^s \tag{5.14}$$

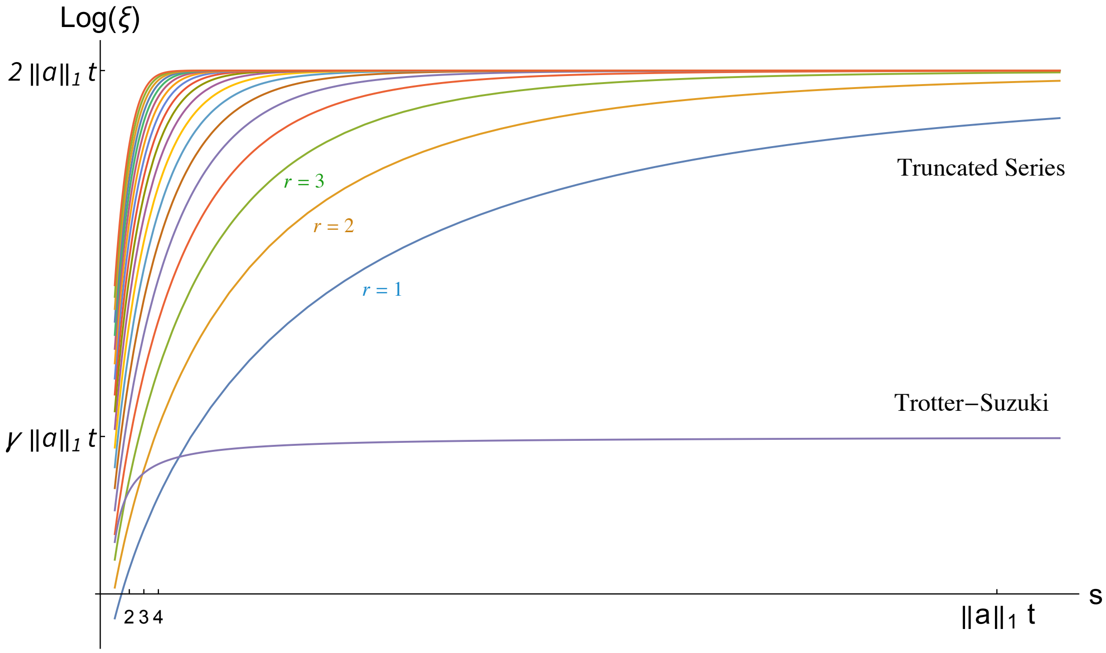

Figure 5.1: The exponent as a function of the slicing parameter s for the Trotter-Suzuki decomposition and the truncated series decomposition for a collection of truncation parameters r. The cost of simulating $\exp(iHt)|0^n\rangle$ scales linearly in ξ. Here $\gamma = 2\sqrt{2} - 2$. For the purpose of illustration we set $t = \pi/4$, $K = 100$, and $\|a\|_1 = K\pi/4$.

where p and q are probability distributions given by

$$p(j) = N_{s,r}^{-1} \frac{(\|a\|_1 t/s)^j}{j!}, \quad q(z) = \frac{a_{z_1} \ldots a_{z_j}}{\|a\|_1^j} \tag{5.15}$$

and $N_{r,s}$ is the normalization constant

$$N_{s,r} = \sum_{j=0}^{r} \frac{(\|a\|_1 t/s)^j}{j!}. \tag{5.16}$$

To randomly sparsify this decomposition, we start with $|0^n\rangle$, and for every time slice we multiply it by a string of Pauli operators and a weight, $(i^j N_{s,r}) Q_{z_1} \ldots Q_{z_j}$, chosen with probability $p(j)q(z)$. We need to average $N_{s,r}^{2s}/\epsilon^2$ such trajectories to guarantee that the resulting state approximates $e^{iHt}|0\rangle^{\otimes n}$ with high probability. Figure 5.1 compares this to the Trotter-Suzuki method.

Interestingly, if instead of eq. (5.8) we use the more usual decomposition

$$e^{i\theta Q} = (\cos\theta)I + (i\sin\theta)Q, \tag{5.17}$$

in the Trotter-Suzuki method, then a similar calculation shows that the number of required trajectories, up to a polynomial factor, is $\exp(2\|a\|_1 t)$. In other words, the amelioration of the exponential scaling is wiped out. An intuitive (but heuristic) explanation is that Pauli operators permute the computational basis up to phases, and so the sampled trajectories for both the truncated series and the Trotter-Suzuki with eq. (5.17) are over a classical state space, whereas by using eq. (5.8) the trajectories are over highly-entangled stabilizer states, which somehow allows for a more sparse representation of $\exp(iHt)|0^n\rangle$.

We should emphasize that there are numerous techniques for improving the worldline and stochastic series approaches, such as using loop and cluster updates. These are typically applied in the context of simulating imaginary time evolution, but some of these ideas can be adapted to the real time case. The above discussion only compares the most crude way to apply these methods. However, there is already a significant improvement at this base level, and more clever ways of constructing low-rank approximate stabilizer decompositions may be possible.

5.3 Simulating continuous-time evolution using stabilizer decompositions

We consider two applications of stabilizer rank methods: the first is approximating an amplitude $\langle x|e^{iHt}|0^n\rangle$, where $x \in \mathbb{F}_2^n$ is a computational basis state (an n-bit string), and the second is sampling bit strings from a distribution which approximates the probability distribution

$$P(y) := \langle 0^n|e^{-iHt}\Pi_y e^{iHt}|0^n\rangle,$$

where Π_y is a projector onto the eigenspace of the outcome $y \in \mathbb{F}_2^{|S|}$ when measuring a subset S of the qubits in the computational basis.

We can apply the approximate sampling algorithm of Ref. [8], and the additive-error amplitude approximation algorithm of section 3.4, to the Trotter circuit U_s. The resulting classical simulations require poly(n) bits of memory and have run times in which the only exponential factor is the number of stabilizer terms in an approximate stabilizer decomposition of $e^{iHt}|0\rangle^{\otimes n}$. For concreteness we will restrict our attention to a Pauli decomposition of the Hamiltonian

$$H = \sum_j a_j Q_j, \tag{5.18}$$

but generalizing the results to any decomposition in which Q_j can be easily exponentiated is straightforward.

Theorem 13. *The time cost of classically approximating $\langle x|e^{iHt}|0^n\rangle$ up to additive error ϵ using the algorithm in section 3.4 is*

$$O\left(e^{(2\sqrt{2}-2)\|a\|_1 t} n^2 (K/\epsilon)^{5/2} (\|a\|_\infty t)^{3/2} \log(\delta^{-1})\right),$$

where $\|a\|_\infty = \max_i a_i$.

Proof. We can produce an estimate η of the amplitude $\langle x|U_s|0^n\rangle$, by using the algorithm in section 3.4, with the Clifford decompositions in eq. (5.8). The number of iterations the algorithm in section 3.4 requires to bound the additive error on η by $\epsilon/2$ with probability at least $1-\delta$ is

$$O\left(e^{(2\sqrt{2}-2)\|a\|_1 t} \epsilon^{-2} \log(\delta^{-1})\right).$$

Each iteration requires performing $2sK$ Clifford updates, each of which takes time $O(n^2)$. If we choose $s = \sqrt{(2K\|a\|_\infty t)^3/\epsilon}$, then $\|e^{iHt} - V_s\| \leq \epsilon/2$. By the triangle inequality our estimate η satisfies $|\eta - \langle x|e^{iHt}|0^n\rangle| \leq \epsilon$ with probability at least $1-\delta$. □

We can also use this stabilizer decomposition to sample from the outcome distribution associated with measuring any subset S of the qubits in the computational basis, i.e. the distribution

$$P(y) \equiv \langle 0^n|e^{-iHt}\Pi(y)e^{iHt}|0^n\rangle, \tag{5.19}$$

where $y \in \mathbb{F}_2^{|S|}$ and $\Pi(y)$ acts as $|y\rangle\langle y|$ on S and identity on the rest of the qubits.

Theorem 14. *The distribution $P(y)$ can be approximately sampled from with error ϵ in total variation distance using the algorithm in section 2.2.1. The time cost scales as*

$$O\left(e^{(2\sqrt{2}-2)\|a\|_1 t} n^2 (K/\epsilon)^{5/2} (t\|a\|_\infty)^{3/2}\right) + O(e^{(2\sqrt{2}-2)\|a\|_1 t} n^3 |S|^3 \epsilon^{-2} \log(\delta^{-1})).$$

where δ is the failure probability of the algorithm.

Proof. As in the proof of Theorem 13, we can choose $s = O\Big(\sqrt{(2K\|a\|_\infty t)^3/\eta}\Big)$ so that

$$\|U_s|0^n\rangle - e^{iHt}|0^n\rangle\| \leq \eta/2.$$

Now we obtain an approximate stabilizer decomposition of $e^{iHt}|0^n\rangle$ by randomly sparsifying $U_s|0^n\rangle$. By eq. (5.11), we can approximate $U_s|0^n\rangle$ up to error $\eta/2$ by a linear combination of $O(e^{(2\sqrt{2}-2)\|a\|_1 t}\eta^{-2})$ stabilizer states. Denote this linear combination by $|\psi\rangle$. Then

$$\||\psi\rangle - e^{iHt}|0^n\rangle\| \leq \eta.$$

Each of the states in the decomposition of $|\psi\rangle$ takes time $O(sKn^2)$ to compute, therefore the time cost of producing this approximation is

$$O\left(e^{(2\sqrt{2}-2)\|a\|_1 t} n^2 (K/\eta)^{5/2} (t\|a\|_\infty)^{3/2}\right). \tag{5.20}$$

Given this approximation, we can proceed as in section 2.2.1. Namely, approximately sampling the distribution $P(y)$ up to error ϵ in total variation distance can be done for a suitable choice of $\eta = O(\epsilon)$ by approximately sampling the distribution

$$\tilde{P}(y) \equiv \frac{\langle\psi|\Pi(y)|\psi\rangle}{\|\psi\|^2}.$$

The time cost of producing a sample in this way scales as $O(e^{(2\sqrt{2}-2)\|a\|_1 t} n^3 |S|^3 \epsilon^{-2}\log(\delta^{-1}))$.

□

5.4 Discussion and open problems

The method discussed in this chapter bears a close resemblance with Quantum Monte Carlo (QMC) methods. One of the most important QMC methods is path integral Monte Carlo [42], which is typically used to simulate static properties of the Hamiltonian, such as its ground state energy and partition function. In its most general form, this method is based on a Markov chain Monte Carlo, in which the state space is defined as the set of trajectories of the form $(v_1, v_2, \ldots, v_L)$, where the v_i are states from a product basis and L is the number of time slices. Update rules, also known as Metropolis moves, are defined in a certain way that makes use of properties of the Hamiltonian. The runtime of such algorithms depends on the mixing time of the chain. In important cases, such as when the

Hamiltonian is 1D and stoquastic, the chain can be shown to mix rapidly under certain update rules [60].

An important variant of stabilizer rank methods relies on a similar Metropolis algorithm in which the probabilities of local bit flip moves are determined by ratios of estimates of the associated amplitudes [8]. The algorithm was found to numerically outperform the method based on the norm estimation algorithm, despite having no rigorous error guarantees if the mixing time of the chain is unknown, as is generally the case.

For the continuous-time case, a similar algorithm can be implemented using additive error estimates of the amplitude $\langle x|e^{iHt}|0^n\rangle$ obtained using Theorem 13. The state space of the Markov chain is the set of computational basis states $\{|y\rangle\}$, and the moves are single bit flips chosen with probability that depends on the ratio of the estimates of the two associated amplitudes. When the chain equilibrates we obtain a final state from a distribution close to $P(y) \equiv \langle 0^n|e^{-iHt}\Pi(y)e^{iHt}|0^n\rangle$, although the mixing time of the chain is generally unknown. The only other super-polynomial scaling comes from the sample complexity required to estimate an amplitude using stabilizer state monte carlo, namely $e^{\gamma t\|a\|_1}$.

To reduce the exponent we need to search for more sparse stabilizer decompositions of the final state. If the Hamiltonian is given as a sum of 2-local terms,

$$H = \sum_j a_j A_j \tag{5.21}$$

then one way to do this is to numerically find the optimal Clifford decomposition of each Trotter gate $\exp(i\theta A_j)$ using convex optimization. For 3-local Hamiltonians this may also be possible with more computational resources.

Sometimes we can use the idea of decomposing matrix products of gates, as in section 3.3, to find more sparse decompositions. One example where this is possible is

$$H = \sum_{<i,j,k>} (X_i X_j X_k + Y_i Y_j Y_k + Z_i Z_j Z_k), \tag{5.22}$$

where the sum is over three-neighbors in a chain. For any neighbours (i, j, k), the set

$$\{X_i X_j X_k, Y_i Y_j Y_k, Z_i Z_j Z_k\}$$

is equivalent to $\{X_i, -Y_i, Z_i\}$ up to conjugation by a Clifford, and therefore we can use the results of section 3.3 to find optimal Clifford decompositions of the Trotter step

$$\exp(i\theta X_i X_j X_k)\exp(i\theta Z_i Z_j Z_k)\exp(i\theta Y_i Y_j Y_k).$$

Using convex optimization, we find the optimal single qubit decomposition

$$\exp\left(i\theta X_i\right)\exp\left(i\theta Z_i\right)\exp\left(-i\theta Y_i\right) = c_1(\theta)I + c_2(\theta)S_i + c_3(\theta)H_iS_i, \tag{5.23}$$

where S_i and H_i are the phase and Hadamard gates on qubit i and

$$\begin{aligned} c_1(\theta) &= \cos(2\theta)[\cos(\theta) - \sin(\theta)] \\ c_2(\theta) &= \frac{e^{i\pi/4}}{\sqrt{2}}[\cos(\theta) - \cos(3\theta)] \\ c_3(\theta) &= e^{i\pi/4}\sin(2\theta)[\cos(\theta) - \sin(\theta)]. \end{aligned} \tag{5.24}$$

We get the optimal decomposition of the Trotter step by conjugating eq. (5.23) with any 3-qubit entangling Clifford which permutes $X_1 \to X_1X_2X_3$, $Z_1 \to Z_1Z_2Z_3$ (and therefore $Y_1 \to -Y_1Y_2Y_3$). This significantly reduces the number of stabilizer terms in the approximate stabilizer decomposition of $e^{iHt}|0^n\rangle$, in a similar way to the decompositions in section 3.3.

More generally, for a decomposition $H = \sum_j A_j$ we try to find Clifford decompositions of the Trotter steps $\exp(i\theta A_j)$ with minimum 1-norm. The framework presented in this chapter allows us to use any decomposition of the Hamiltonian, even if the terms are non-local. Although this involves finding sparse Clifford decompositions for many-qubit gates, which is computationally difficult, it leaves open the possibility of much faster simulations. Combining this with the the fact that there are various ways to directly ameliorate the exponential scaling, suggests that this approach deserves further study. An interesting question is how mild we can make the exponential scaling for a given Hamiltonian. This question is likely to be very difficult to answer. A similar problem, namely proving super-polynomial lower bounds on the approximate stabilizer rank of product states, is remarkably difficult [8].

Chapter 6

Linear constraint systems

6.1 Introduction

Linear constraint systems (LCS) are generalizations of the simple proofs of contextuality of Peres and Mermin [20, 19]. These proofs are based on inconsistent systems of linear equations over the binary field that nevertheless admit solutions if we allow for non-commutative versions of the variables, i.e. binary quantum observables. A gap between classical and quantum satisfiability in this setting is a proof of contextuality [43]. In addition, such a gap can be used to define a non-local game which can be won with certainty by quantum players but not by classical players, i.e. a *pseudo-telepathic* game [21]. This link has turned out to be immensely useful in the study of quantum non-locality [17, 44, 18, 23, 25, 24, 45]. In particular, games constructed from LCSs are suitable for studying the separation between different models of quantum correlations, such as the tensor-product and the commuting-operator models [23, 25, 24]. Further, non-local games constructed in this way can be used for devising robust self-testing protocols [46].

An example of a linear constraint system is the Peres-Mermin square, depicted in fig. 6.1

(Left). It corresponds to a system of linear equations modulo 2

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_4 + x_5 + x_6 &= 0 \\ x_7 + x_8 + x_9 &= 0 \\ x_1 + x_4 + x_7 &= 0 \\ x_2 + x_5 + x_8 &= 0 \\ x_3 + x_6 + x_9 &= 1 \end{aligned} \tag{6.1}$$

It is easy to check that this system has no classical solutions with binary variables x_i. Yet the system of operator equations

$$\begin{aligned} A_1 A_2 A_3 &= I \\ A_4 A_5 A_6 &= I \\ A_7 A_8 A_9 &= I \\ A_1 A_4 A_7 &= I \\ A_2 A_5 A_8 &= I \\ A_3 A_6 A_9 &= -I \end{aligned} \tag{6.2}$$

constrained by the requirement that the operators in each equation pairwise commute and square to the identity, does have a solution, depicted in fig. 6.2 (Left). While it is easy to check the satisfiability of eq. (6.1), determining whether eq. (6.2) has an operator solution is trickier, intuitively because we need to check every possible dimension these operators may act on (including countable infinity).

Quantum satisfiability of LCSs over $\mathbb{F}_2$ has an elegant characterization if every variable appears in exactly two equations. Namely, it was shown in [17] that a LCS of that form is "magic", i.e. has a satisfiability gap for some choice of right-hand side in eq. (6.1), if and only if a certain graph constructed from the LCS is non-planar [1]. Planarity of a graph can be checked efficiently in the size of the graph, and so this offers a complete and efficiently checkable characterization. However, it is unclear how to extend this characterization to $d > 2$, or to the more general case where variables appear in any number of constraints.

Beyond the case where each variable appears in exactly two constraints, whether or not a binary LCS has a quantum solution was addressed in [18, 23], where it was found that

[1] A graph is planar if it can be drawn without the lines representing edges intersecting.

this question has two other equivalent forms. The first form involves quantum strategies for certain non-local games constructed from the LCS, and the second form involves a property of a certain group associated with the LCS, known as the *solution group*.

In this chapter and the next our purpose is to shed light on systems of linear equations modulo d where $d > 2$. This case was considered to varying extents in Refs. [23, 46, 28], but is generally less well-understood.

6.2 Group theory preliminaries

In this section we review elements of group theory. For more details the reader can refer to any standard textbook, such as Ref. [47].

A group is a set G closed under a binary operation, which contains an identity element e and an inverse g^{-1} for each element $g \in G$. The binary operation is called group multiplication. For $h, g \in G$ the product is denoted gh. One way to specify a group is by a multiplication table. A more concise way to define a group is through a *presentation*, which is a list of relations that define the group. For example, we can define a group with two generators s, t which commute $st = ts$ and square to the identity $s^2 = t^2 = e$. This completely specifies the group as $\{e, s, t, st\}$. A finitely presented group is a group defined by a finite number of relations. Such a group may be infinite.

More formally, let $S = \{s, t, u, \dots\}$ be a set of symbols, and define a set of inverses of these symbols $S' = \{s^{-1}, t^{-1}, u^{-1}, \dots\}$. A word in $S \cup S'$ is any string of symbols and their inverses. An example of such a word is

$$sstuu^{-1}t^{-1}st. \tag{6.3}$$

Two words are equivalent if one of them can be reduced to the other by removing collisions of the form $s^{-1}s$ or ss^{-1}. For example, the above word is equivalent to $ssst = s^3t$. The *free group* $\mathcal{F}(S)$ generated by S consists of all words in the alphabet $S \cup S'$, up to this equivalence. Let G be a group. The *normal closure* of a subset $R \subseteq G$ is the subgroup $N(R)$ generated by the set $\{grg^{-1} : g \in G, r \in R\}$. Let $R = \{r_1, \dots, r_k\}$ be a set of words in $\mathcal{F}(S)$. The *finitely presented group* with generators S and relations $\{r_1 = e, \dots, r_k = e\}$ is the quotient $\mathcal{F}(S)/N(R)$. Intuitively this group consists of all words in $\mathcal{F}(S)$ where we identify r_i with the identity element.

A *representation* of a group G is a mapping ϕ from G to invertible matrices over some vector space, which preserves the group operation, i.e. $\phi(gh) = \phi(g)\phi(h)$. A representation is irreducible if it cannot be simultaneously brought into a block diagonal form by a change of basis.

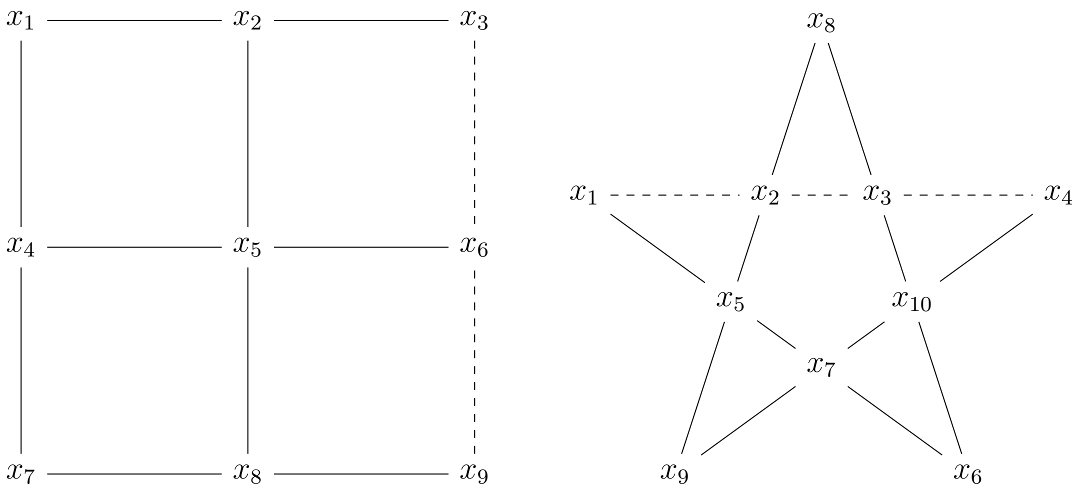

Figure 6.1: Linear constraint systems for the magic square (left) and pentagram (right) [19]. Each variable can take values modulo d. The linear constraints are that the values of the variables along a solid (dashed) line add to zero (one) modulo d. Note that whether the games are "magic" depends upon the constraints, e.g., if all lines are solid then both games are classically satisfiable.

6.3 Linear constraint systems and solution groups

A linear constraint system (LCS) over $\mathbb{Z}_d$ is a pair (M, b), where $M \in \mathbb{M}_{m\times n}(\mathbb{Z}_d)$ and $b \in \mathbb{Z}_d^m$. A classical solution is a vector $a \in \mathbb{Z}_d^n$ such that $Ma = b$ modulo d. Figure 6.1 (a) is a diagrammatic depiction of a LCS that is classically unsatisfiable for all $d \geq 2$, since the sum of the variables along the rows is 0, while the sum of the variables along the columns is 1, which is a contradiction for classical variables.

A quantum solution $\{A_1, \ldots, A_n\}$ is a collection of normal operators [2], acting on some Hilbert space, satisfying:

1. $A_i^d = I$ for all $i \in \{1, \ldots, n\}$;

2. $[A_j, A_k] = 0$ whenever there exists a row i such that $M_{ij} \neq 0$ and $M_{ik} \neq 0$; and

3. $\prod_j A_j^{M_{ij}} = \omega^{b_i}$ for all $i \in \{1, \ldots, m\}$.

[2] An operator A is normal if $AA^\dagger = A^\dagger A$.

By condition (1), each operator A_i is unitary and the eigenvalues are dth roots of unity. By condition (2), whenever x_j and x_k appear in one of the constraints, their operator counterparts A_j and A_k commute. In particular, all operators corresponding to a constraint pairwise commute, and therefore can be simultaneously measured. By condition (3), the expectation values of the operators satisfy the specified constraints. Here and below, we set $\omega \equiv \exp 2\pi i/d$, where d is to be understood from context. Quantum solutions for the magic square and pentagram are shown in Figure 6.2.

To every linear constraint system one can associate a *solution group*, which is a finitely presented group with relations that encode the defining properties of a quantum solution. Namely,

Definition 15. For a LCS (M, b), the *solution group* $\mathcal{G}(M, b)$ is the finitely presented group generated by the symbols $\{J, g_1, \ldots, g_n\}$ and the relations

1. $g_i^d = e$, $J^d = e$,
2. $Jg_i = g_iJ$,
3. $g_jg_k = g_kg_j$ whenever there exists a row i such that $M_{ij} \neq 0$ and $M_{ik} \neq 0$,
4. $\prod_j g_j^{M_{ij}} = J^{b_i}$ for all $i \in \{1, \ldots, m\}$.

Proposition 16. *Let (M, b) be a LCS. Suppose that a quantum solution exists in dimension $D \in \mathbb{N} \cup \{\infty\}$. Then $\mathcal{G}(M, b)$ admits a D-dimensional representation ϕ in which $\phi(J) = \omega I$.*

Proof. Given a quantum solution $\{A_1, \ldots, A_n\}$ in dimension D, it is straight-forward to check that the homomorphic extension of the map $\phi(g_i) = A_i$, $\phi(J) = \omega I$, is a D-dimensional representation of $\mathcal{G}(M, b)$. □

6.4 Non-local games constructed from linear constraint systems

One of the most useful applications of LCSs is the fact that they are related to perfect quantum strategies in certain non-local games, i.e. strategies which win with certainty. For a LCS given by $Mx = b$, we define a bipartite non-local game $\mathcal{L}(M, b)$ as follows. The game involves two co-operating players, Alice and Bob, who receive and reply to questions from

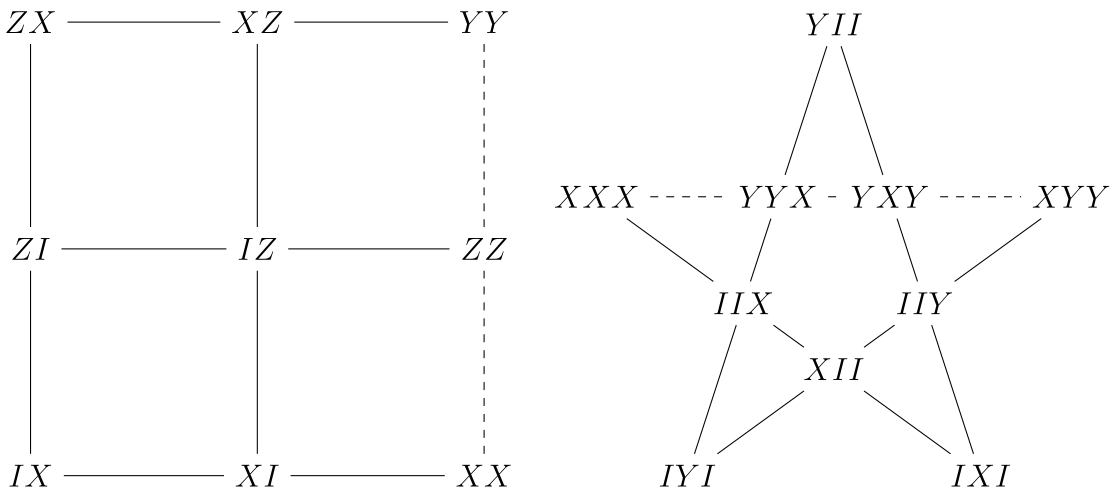

Figure 6.2: Quantum solutions for the magic square (left) and pentagram (right) for $d = 2$ [19]. Each vertex is a unitary operator whose eigenvalues are dth roots of unity. The operators along a solid (dashed) line multiply to I $(-I)$.

a referee. The players are allowed to agree on a strategy beforehand but are not allowed to communicate once the game starts. Alice receives an index of a row of M (corresponding to a constraint), and Bob receives an index of one of the variables in Alice's constraint. Alice is required to assign a value to each variable in the constraint, and Bob is required to assign a value to the single variable he receives. The players win if Alice's assignment satisfies the constraint and Bob's assignment agrees with Alice's.

It is not hard to show that players using classical means cannot win with certainty unless $Mx = b$ has a classical solution. If the players are allowed to use a quantum strategy, they can in some cases win the game $\mathcal{L}(M, b)$ with certainty. Namely, the main result of [23] is the following

Theorem 17 ([23])**.** *Let d be a prime number, and $Mx = b$ be a LCS modulo d. The following statements are equivalent*

1. *$Mx = b$ has a quantum solution (possibly infinite-dimensional),*
2. *The non-local game $\mathcal{L}(M, b)$ has a perfect quantum strategy (possibly of the commuting-operator type),*
3. *The solution group $\mathcal{G}(M, b)$ has the property that $J \neq e$.*

Note that in (2) the quantum strategy may be of the commuting-operator type. Such a strategy is defined by a quantum state in some Hilbert space $|\psi\rangle \in \mathcal{H}$ and observables $\{A_i^{(\ell)}\}$, $\{B_i\}$ respectively for Alice and Bob. Here ℓ is an index for the constraint, and i is an index for the variable. Upon receiving the constraint ℓ as the question, Alice measures the observables $\{A_i^{(\ell)} : i \in V_\ell\}$ and reports the outcomes, where V_ℓ is the set of indices of the variables in constraint ℓ. Similarly when Bob receives j as the question he measures the observable B_j and reports the outcome.

Alice's and Bob's observables are required to commute, i.e. $[A_i^{(\ell)}, B_j] = 0$ for all i, ℓ, j, but the Hilbert space $\mathcal{H}$ is not required to have a tensor product structure. Such correlations are natural in algebraic quantum field theory, where local observables of space-like separated regions are required to commute but the quantum state is global (in the sense that the Hilbert space $\mathcal{H}$ does not factorize as $\mathcal{H}_A \otimes \mathcal{H}_B$).

Statement (3) asks us to determine whether the defining relations of the solution group definition 15 imply that $J = e$. Checking this property is difficult in general (it is an instance of the word problem, which is undecidable), although for small enough instances a computer software, such as GAP, can compute it fairly quickly.

6.5 Generalized Pauli observables

For $d = 2$, the canonical examples of a satisfiability gap rely on the commutation relations of qubit Pauli observables. Pauli observables can be generalized to higher dimensions in a number of different ways. The Weyl-Heisenberg group is arguably the most natural generalization, since it preserves the commutativity structure and error-correction properties of the qubit Pauli group [27]. Elements of this group are commonly referred to as generalized Pauli operators, and are natural candidates for constructing quantum solutions for LCS over $\mathbb{Z}_d$ for $d > 2$.

The single-qubit Pauli operators can be generalized to arbitrary dimensions D[3] via the generalized *shift* and *boost* operators

$$\begin{aligned} X &= \sum_{j \in \mathbb{Z}_D} |j+1\rangle\langle j| \\ Z &= \sum_{j \in \mathbb{Z}_D} \omega^j |j\rangle\langle j| \,. \end{aligned} \tag{6.4}$$

[3] We use D to refer to the dimension since the more usual d is reserved for the modulus of the LCS.

The generalized shift and boost operators satisfy the commutation relation

$$Z^p X^q = \omega^{-pq} X^q Z^p \,. \tag{6.5}$$

A generalized Pauli operator for a single qu$D$it is any operator $A \propto Z^p X^q$ for some $p, q \in \mathbb{Z}_D$. For n qu$D$its, a generalized Pauli operator is defined as a tensor product of single qu$D$it generalized Paulis. From eq. (6.5), any two generalized Pauli operators commute up to a Dth root of unity.

Chapter 7

No-go theorems for a satisfiability gap in linear constraint systems

7.1 Introduction

In this chapter we prove a few results for linear constraint system modulo $d > 2$. First, we give a concise proof that the existence of a quantum solution which uses generalized Pauli observables in some odd dimension implies the existence of a classical solution. This fact follows from the non-negative discrete Wigner function for Hilbert spaces of odd dimension [48, 13, 49], but, interestingly, does not require the full Wigner function, only its definition at a single point in phase-space. While a discrete Wigner function can be defined in even dimensions based on generalized Pauli observables [50], it is easy to show that a satisfiability gap can be demonstrated using generalized Pauli observables in every even dimension (see Table 7.1).

For a LCS over $\mathbb{Z}_d$ given by the equation $Mx = b \mod d$, checking whether a classical solution exists is easy, for example by Gaussian elimination if d is prime, or other efficient methods if d is not prime [51, 52]. On the other hand, checking whether a quantum solution exists is a difficult task. One characterization is given in [23] in terms of the so-called solution group. The solution group associated with $Mx = b$ is a finitely presented group which encodes the algebraic relations satisfied by every quantum solution to $Mx = b$. It is generated by symbols $\{g_i\}$, which can be thought of as a non-commutative relaxation of the variables $\{x_i\}$, in addition to a special generator J. The characterization in [23] is based on properties of the element J. Namely, for a LCS over $\mathbb{Z}_d$ with d prime, a quantum solution exists if and only if J is not equivalent to the identity element in the group.

Checking whether an element in a finitely presented group is equivalent to the identity is an instance of the word problem, which is undecidable in general. Nevertheless, for small enough instances or in the case of highly structured groups one can use computer software, such as GAP [53], to find a so-called confluent rewriting system, thus solving the word problem for the group.

In this chapter, we note a simple extension of the characterization in [23] which covers arbitrary d. Specifically, a quantum solution exists for a LCS over $\mathbb{Z}_d$ if and only if J has order d. When d is prime, this characterization reduces to the one in [23].

Furthermore, we note a property of some LCSs which we call phase-commutation. Roughly speaking, phase-commutation is an abstraction of matrix anti-commutation relations of the form $AB = -BA$. A LCS $Mx = b$ exhibits phase-commutation if its non-commutative relaxation (as defined by the solution group) contains two variables, g_1 and g_2, with commutation relation

$$g_1 g_2 = J^c g_2 g_1, \quad c \neq 0. \tag{7.1}$$

We show that, if d is odd, any relation of the form in eq. (7.1), or of the more general form in eq. (7.21), implies that quantum solutions do not exist in any (finite- or countably infinite-dimensional) Hilbert space. As a consequence, this implies that all natural generalizations of the Peres-Mermin magic square and pentagram to odd prime d do not exhibit a satisfiability gap.

This is in contrast to the case of even d, where the Peres-Mermin magic square and pentagram exhibit satisfiability gaps that rely on using Pauli operators to represent eq. (7.1). In fact, a much stronger statement was proven for $d = 2$ in Ref. [25], namely that any group generated by involutions can be embedded (in a strong sense) in the solution group of some binary LCS which admits a quantum solution. A significant consequence of this embedding for $d = 2$ is the separation, proven in the latter work, between commuting and tensor-product quantum nonlocal strategies. Our result shows that a similar statement cannot hold for odd d, as the group defined by the phase-commutation relation in eq. (7.1) cannot be embedded in a LCS that has a quantum solution.

The idea that phase-commutation can rule out a satisfiability gap appears in (the newer version of) Ref. [46] for particular instances of the magic square and pentagram LCSs, and a similar idea appears more recently in [28], as part of a homotopical treatment of quantum contextuality. In comparison, our exposition is more general (it applies to all LCSs, and we use it to cover all natural generalizations of the square and pentagram for odd d), and perhaps more accessible to the non-expert reader (it does not require familiarity with topological ideas like cell-complexes and group pictures).

7.2 A no-go theorem for a satisfiability gap using generalized Pauli observables

Theorem 18. *Let D be odd. Then the LCS (M, b) has a quantum solution using n-qu$D$it generalized Pauli operators if and only if it is classically satisfiable.*

Proof. The first direction is easy: suppose that $Mx = b$ has a classical solution $x \to a \equiv (a_1, \dots, a_n)$. Then $A_j = \omega^{a_j} I$ defines a quantum solution in which every operator is a Pauli operator. Conversely, let us assume that the LCS admits a quantum solution $\{A_j\}$ consisting of generalized n-qu$D$it Pauli operators. We will construct a classical solution by consistently assigning a root of unity $\omega^{v(\sigma)}$ to each generalized n-qu$D$it Pauli operator σ. The map v will be defined in terms of the parity operator $\Pi_D \equiv \sum_{j=0}^{D-1} |-j\rangle\langle j|$. Namely, we set

$$\omega^{v(\sigma)} = \operatorname{Tr}\left(\Pi_D^{\otimes n} \sigma\right). \tag{7.2}$$

For a single-qu$D$it Pauli $\sigma = \omega^k Z^p X^q$, eq. (7.2) gives $v(\sigma) = k + 2^{-1}pq$, where the inverse is modulo D. It can be readily verified that the following identities hold for any n-qu$D$it generalized Paulis σ, τ

1. $v(\omega^j I) = j$,
2. $v(\sigma \otimes \tau) = v(\sigma) + v(\tau) \mod D$,
3. $v(\sigma^j) = j v(\sigma) \mod D$,

To prove that the map $A_j \to v(A_j)$ gives a classical solution to the LCS in question, it is enough to show that, if σ and τ are commuting n-qu$D$it generalized Pauli operators, then $v(\sigma\tau) = v(\sigma) + v(\tau)$. Indeed if that were true then using identities 1 & 3 we get, for every row i of M,

$$\sum_{j=1}^{n} M_{i,j} v\left(A_j\right) = \sum_{j=1}^{n} v\left(A_j^{M_{i,j}}\right) = v\left(\prod_{j=1}^{n} A_j^{M_{i,j}}\right) = v\left(\omega^{b_i} I\right) = b_i. \tag{7.3}$$

Let $\sigma \propto Z[\vec{p}]X[\vec{q}]$ and $\tau \propto Z[\vec{r}]X[\vec{s}]$ be tensor products of generalized Pauli operates. Here we are denoting $A[\vec{a}] = \bigotimes_{k=1}^{n} A^{a_k}$ for an n-element vector $\vec{a} = (a_1, \dots, a_n)$, and, for simplicity, ignoring the phase factors in the definition of generalized Paulis, which does not affect the analysis. Note that

$$v(Z^p X^q) + v(Z^r X^s) = 2^{-1}(pq + rs)\ \forall\, p, q, r, s. \tag{7.4}$$

On the other hand, by eq. (6.5),

$$v(Z^p X^q Z^r X^s) = v(\omega^{-qr} Z^{p+r} X^{q+s}) = -qr + 2^{-1}(p+r)(q+s)\,. \tag{7.5}$$

Comparing Eq. (7.4) and (7.5) and using identity 2,

$$v(Z[\vec{p}]X[\vec{q}]Z[\vec{r}]X[\vec{s}]) = v(Z[\vec{p}]X[\vec{q}]) + v(Z[\vec{r}]X[\vec{s}]) \iff \vec{p}\cdot\vec{s} - \vec{q}\cdot\vec{r} = 0 \pmod{d}\,. \tag{7.6}$$

The condition $\vec{p}\cdot\vec{s} - \vec{q}\cdot\vec{r} = 0$ is precisely the requirement that $Z[\vec{p}]X[\vec{q}]$ and $Z[\vec{r}]X[\vec{s}]$ commute. Indeed, by eq. (6.5)

$$Z[\vec{p}]X[\vec{q}]Z[\vec{r}]X[\vec{s}] = \omega^{\vec{q}\cdot\vec{r}} Z[\vec{p}]Z[\vec{r}]X[\vec{s}]X[\vec{q}] = \omega^{\vec{q}\cdot\vec{r}-\vec{p}\cdot\vec{s}} Z[\vec{r}]X[\vec{s}]Z[\vec{p}]X[\vec{q}]\,. \tag{7.7}$$

Therefore $v(\sigma\tau) = v(\sigma) + v(\tau)$ for any commuting σ and τ, finishing the proof. □

The above proof fails for even D. The proof hinges on the ability to consistently assign a root of unity to every generalized Pauli operator. For $\sigma = \omega^k Z^p X^q$, eq. (7.2) reads

$$\omega^{v(\sigma)} = \mathrm{Tr}\,(\Pi)_D\,\sigma = \sum_j \langle j|\sigma|-j\rangle = \sum_{j:2j=q} \omega^{pj+k}. \tag{7.8}$$

When D is even, the sum evaluates to 0 whenever q is odd, making it impossible to define the map in this way. One may wonder whether the map v can be defined in a different way for even D. This is also impossible, as shown by the family of quantum solutions in Table 7.1. Note that the parity operator Π_D in the proof of Theorem 18 is one of the phase point operators for the discrete Wigner function in odd dimensions [48]. While it is possible to define a discrete Wigner function in even dimensions based on generalized Pauli operators [50], important properties, such as non-negativity of stabilizer states and Clifford-covariance, are lost in that case [54].

7.3 A no-go theorem for a satisfiability gap with Pauli-like commutation relations

Theorem 18 asserts that a quantum/classical satisfiability gap cannot be illustrated using generalized Pauli operators in odd dimension. We now show that the canonical commutation relations of these operators cannot be embedded in the solution group of a LCS that has a quantum solution. Some of this discussion is inspired by (the corrected version of) Ref. [46], but it applies more generally than what is considered in that work.

A key characterization of quantum satisfiability when d is prime is the following:

$Z^t \otimes X^t$	$X^t \otimes Z^t$	$X^t Z^t \otimes Z^t X^t$
$Z^t \otimes I$	$I \otimes Z^t$	$Z^t \otimes Z^t$
$I \otimes X^t$	$X^t \otimes I$	$X^t \otimes X^t$

$Z^{t+1} \otimes X^{t+1}$	$X^t \otimes Z^t$	$X^t Z^{t-1} \otimes Z^t X^{t-1}$
$Z^{t-1} \otimes I$	$I \otimes Z^t$	$Z^{t+1} \otimes Z^t$
$I \otimes X^{t-1}$	$X^t \otimes I$	$X^t \otimes X^{t+1}$

Table 7.1: Quantum solutions to a generalized magic square LCS modulo $D = 2t$, for odd (even) t on the left (right). In both cases, the generalized two-qu$D$it Pauli operators along every row and column mutually commute, their product along the third column is $\omega^t I = -I$, and their product along the rows and remaining columns is I. This shows that a satisfiability gap exists using n-qu$D$it generalized Pauli operators for every even dimension D.

Theorem 19. *[23]. Let (M, b) be a LCS over $\mathbb{Z}_d$ where d is prime. Then (M, b) admits a quantum solution if and only if the solution group $\mathcal{G}(M, b)$ has the property that $J \neq e$.*

The proof of Theorem 19 relies on the fact that, when $J \neq e$, the group algebra of $\mathcal{G}(M, b)$ contains a subspace on which it is easy to construct a quantum solution. A similar proof works for non-prime d as well, provided we make an adjustment to the statement of the theorem.

Theorem 20. *Let (M, b) be a LCS over $\mathbb{Z}_d$ with $d > 1$. Then (M, b) admits a quantum solution if and only if the order of J is d.*

Proof (A minor modification of the proof in [23]). Suppose (M, b) has a quantum solution $\{A_i\}$. By Proposition 16, there is a representation ϕ of $\mathcal{G}(M, b)$ such that $\phi(J) = \omega I$. For $c \neq 0 \mod d$, $\phi(J^c) = \omega^c I \neq I$, and therefore $J^c \neq e$. Thus the order of J is d. Conversely, suppose that $\mathcal{G}(M, b)$ is such that the order of J is d. Define the complex vector space $\mathcal{H}$ by

$$\mathcal{H} = \left\{ \sum_{g \in \mathcal{G}} \alpha_g |g\rangle : \|\alpha\|^2 < \infty \right\}. \tag{7.9}$$

Define the operators $\{L_g : g \in \mathcal{G}\}$ via $L_g|h\rangle = |gh\rangle$. The proof proceeds in two steps. First we list certain properties of the L_g. Second, we define a certain subspace of $\mathcal{H}$ on which the restriction of $L_{g_1}, \ldots, L_{g_n}$ is a quantum solution. The relevant properties of the L_g are the following

1. The eigenvalues of L_{g_i} are dth roots of unity:

$$L_{g_i}^d = L_{g_i^d} = L_e = I. \tag{7.10}$$

2. Let i, j and k be such that $M_{ij} \neq 0$ and $M_{ik} \neq 0$, then, for all $g \in \mathcal{G}(M, b)$,

$$L_{g_j} L_{g_k} |g\rangle = |g_j g_k g\rangle = |g_k g_j g\rangle = L_{g_k} L_{g_j} |g\rangle, \tag{7.11}$$

   and therefore $[L_{g_j}, L_{g_k}] = 0$.

3. For any $i \in \{1, \dots, m\}$ and $g \in \mathcal{G}(M, b)$,

$$\begin{aligned} L_{g_1}^{M_{i1}} \dots L_{g_n}^{M_{in}} |g\rangle &= |g_1^{M_{i1}} \dots g_n^{M_{in}} g\rangle \\ &= |J^{b_i} g\rangle \\ &= L_J^{b_i} |g\rangle. \end{aligned} \tag{7.12}$$

4. Define a fiducial vector $|\psi\rangle = d^{-1/2} \sum_{c \in \mathbb{Z}_d} \omega^c |J^c\rangle$. Then

$$L_J^{-b_i} |\psi\rangle = \sum_{c \in \mathbb{Z}_d} \omega^c |J^{c-b_i}\rangle = \omega^{b_i} |\psi\rangle. \tag{7.13}$$

To construct the quantum solution, first we define the subspace $\mathcal{H}' \subset \mathcal{H}$

$$\mathcal{H}' \equiv \left\{ E|\psi\rangle : E \in \mathfrak{A} \right\}, \tag{7.14}$$

where $\mathfrak{A}$ is the algebra generated by $L_{g_1}, \dots, L_{g_n}$ (i.e. all linear combinations of products of the L_{g_i}). Note that the powers of J are all distinct by assumption: if $J^a = J^b$ then $J^{a-b} = e$, and hence $a - b = 0 \mod d$. In particular this implies that $|\psi\rangle \neq 0$, and, as $|\psi\rangle \in \mathcal{H}'$, $\mathcal{H}' \neq \{0\}$.

Define the quantum solution A_i as the restriction of $(L_{g_i})^{-1}$ to $\mathcal{H}'$, i.e. $A_i = L_{g_i}^{-1}|_{\mathcal{H}'}$. Let us verify that this satisfies the definition of a quantum solution:

1. $A_i^d = I$ follows from eq. (7.10),

2. if $M_{ij} \neq 0$ and $M_{ik} \neq 0$ then $[A_j, A_k] = 0$ follows from eq. (7.11),

3. First use eq. (7.13) and the fact that $L_J^{b_i}$ commutes with every element of $E \in \mathfrak{A}$ to deduce that $L_J^{-b_i} E|\psi\rangle = E L_J^{-b_i} |\psi\rangle = \omega^{b_i} E|\psi\rangle$. From eq. (7.12) we then have

$$\prod_j A_j^{M_{ij}} = \prod_j L_{g_j}^{-M_{ij}}|_{\mathcal{H}'} = L_J^{-b_i}|_{\mathcal{H}'} = \omega^{b_i} I. \tag{7.15}$$

□

Note that Theorem 19 is a special case of Theorem 20 since, for prime d, cyclicity of the powers of J imply that $J \neq e$ if and only the order of J is d.

Generalized Pauli operators obey commutation relations of the form $AB = \omega^c BA$, and are in a sense uniquely determined by these relations. More specifcally, consider the finite group $\mathcal{P}_d$ with generators $\{\mathcal{X}, \mathcal{Z}, \eta\}$ and relations

$$\mathcal{X}\mathcal{Z} = \eta\mathcal{Z}\mathcal{X}, \quad \mathcal{X}^d = \mathcal{Z}^d = \eta^d = e, \quad \eta\mathcal{X} = \mathcal{X}\eta, \quad \eta\mathcal{Z} = \mathcal{Z}\eta. \tag{7.16}$$

By Schur's lemma and the fact that $\eta^d = e$, every irreducible representation of $\mathcal{P}_d$ assigns a dth root of unity $\omega^j I$ to η. Furthermore, it was shown by Weyl that up to unitary equivalence every irreducible representation of $\mathcal{P}_d$ in which $j \neq 0$ maps $\mathcal{X}$ and $\mathcal{Z}$ to (phase-multiples of) the clock and shift operators in some dimension d' that divides d [55].

We consider a particular notion of an embedding of eq. (7.16) into the solution group of a LCS. Namely, suppose that the solution group contains a relation of the form

$$g^a h^b = J^c h^b g^a \tag{7.17}$$

for a pair of generators g and h and some nonzero $a, b, c \in \mathbb{Z}_d$. We refer to such a relation as a phase-commutation relation in the solution group. A phase-commutation relation implies that the LCS is classically unsatisfiable; no integers x and y can satisfy the equation $ax + by = c + by + ax \mod d$ for $c \neq 0 \mod d$.

This notion of an embedding can be motivated by considering the situation for $d = 2$. Given a group K, a collection of elements $k_1, \ldots, k_m \in K$ and a non-identity central element η, such that $k_i^2 = \eta^2 = e$, it was shown in [25] that there exists a LCS $Mx = b$ and an embedding $\psi : K \to \mathcal{G}(M, b)$, such that $\psi(k_i) = g_i$ and $\psi(\eta) = J$. As ψ is injective, this implies that $J \neq e$, and so every such K embeds in the solution group of some LCS which has a (possibly infinite-dimensional) quantum solution.

In contrast, we show next that, for odd d, a phase-commutation relation implies that quantum solutions to the LCS in question do not exist in any dimension, which rules out a satisfiability gap for any LCS in which eq. (7.16) embeds in the solution group.

Theorem 21. *Let d be odd. Suppose that a LCS over $\mathbb{Z}_d$ has a phase-commutation relation. Then it has no quantum solutions.*

Theorem 21 can be proven diagrammatically using symmetries of the so-called group-picture associated with solution groups [25, 46]. We avoid group-pictures for the sake of simplicity, and present a more direct proof instead.

Proof. Let $S \equiv \{J, g_1, \dots, g_n\}$ denote the generators of the solution group $\mathcal{G}$, and $S' \equiv \{J^{-1}, g_1^{-1}, \dots, g_n^{-1}\}$. Let $R \subset \mathcal{F}(S)$ be the set of relations defining $\mathcal{G}$, so that $\mathcal{G} = \mathcal{F}(S)/N(R)$. Let $g \in S$ and $h \in S$ be a phase-commuting pair, so that, in $\mathcal{G}$,

$$g^a h^b = J^c h^b g^a, \quad c \neq 0. \tag{7.18}$$

Let w be the word $J^{-c} g^a h^b g^{-a} h^{-b}$. Since $w = e$ in $\mathcal{G}$, it holds that, in $\mathcal{F}(S)$, w is equivalent to an element of the normal closure $N(R)$. This implies that there exist $u_1, \dots, u_\ell \in \mathcal{F}(S)$ and $q_1, \dots, q_\ell \in R$, such that the word

$$W \equiv (u_1 q_1 u_1^{-1})(u_2 q_2 u_2^{-1}) \dots (u_\ell q_\ell u_\ell^{-1}), \tag{7.19}$$

is equivalent to w in $\mathcal{F}(S)$, that is, W can be reduced to w using only generator contractions of the form $g_i g_i^{-1} = e$. Let $\tilde{W}$ be the "reflected" version of W

$$\tilde{W} \equiv (\tilde{u}_\ell^{-1} \tilde{q}_\ell \tilde{u}_\ell) \dots (\tilde{u}_2^{-1} \tilde{q}_2 \tilde{u}_2)(\tilde{u}_1^{-1} \tilde{q}_1 \tilde{u}_1), \tag{7.20}$$

where $\tilde{x}$ is obtained from x by reversing the order of multiplication of the letters making up x. For example, if $x = g_1 g_2 g_3$ then $\tilde{x} = g_3 g_2 g_1$.

Recall, from the definition of solution groups, that each of the relations in R (and therefore each of the q_i) consists of mutually commuting letters from $S \cup S'$. It therefore holds that $\tilde{q}_i = q_i = e$ in $\mathcal{G}$. This implies that $\tilde{W} = e$ in $\mathcal{G}$. Furthermore, we have $\tilde{W} = \tilde{w}$ in $\mathcal{G}$, since the same sequence of generator contractions that reduces W to w can be applied to reduce $\tilde{W}$ to $\tilde{w}$, so $\tilde{W}$ is equivalent to $\tilde{w}$ in $\mathcal{F}(S)$, and therefore also in $\mathcal{G}$.

Thus we have $\tilde{w} = e$ in $\mathcal{G}$, and therefore $h^b g^a = J^c g^a h^b$ in $\mathcal{G}$. Combining this with $g^a h^b = J^c h^b g^a$ implies that $J^{2c} = e$ in $\mathcal{G}$. Since d is odd, we have $2c \neq 0 \mod d$. Therefore, by Theorem 20, there are no quantum solutions. □

The statement and proof of Theorem 21 can be generalized by considering relations of the form

$$g_1^{a_1} \dots g_k^{a_k} = J^c g_k^{a_k} \dots g_1^{a_1}. \tag{7.21}$$

Following the same steps in the proof of Theorem 21, we can deduce that $J^{2c} = e$, which, for $c \neq 0 \mod d$, implies that the order of J is less than d, and hence that there are no quantum solutions.

7.4 Examples

7.4.1 Magic square modulo prime $d > 2$

Let $d > 2$ be prime. Here we consider a broad class of generalizations of the magic square LCS. Namely, we consider any LCS, over $\mathbb{Z}_d$ of the form in Figure 6.1 (a) with arbitrary coefficients, i.e. any system of the form

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= b_1, \\ a_4x_4 + a_5x_5 + a_6x_6 &= b_2, \\ a_7x_7 + a_8x_8 + a_9x_9 &= b_3, \\ a_1'x_1 + a_4'x_4 + a_7'x_7 &= b_4, \\ a_2'x_2 + a_5'x_5 + a_8'x_8 &= b_5, \\ a_3'x_3 + a_6'x_6 + a_9'x_9 &= b_6, \end{aligned} \tag{7.22}$$

for some $b \in \mathbb{Z}_d^6$, and some set of coefficients $a_i, a_i' \in \mathbb{Z}_d - \{0\}$.

We can simplify by making some reductions. First, we assume that the coefficients in the first three constraints are all equal to one, if necessary by relabeling $x_i \to a_ix_i$. Second, as d is prime, every nonzero element of $\mathbb{Z}_d$ has a multiplicative inverse, so we can divide the last three constraints respectively by the coefficients of x_1, x_2, and x_3. These two steps yield a LCS of the form

$$\begin{aligned} x_1 + x_2 + x_3 &= b_1 \\ x_4 + x_5 + x_6 &= b_2 \\ x_7 + x_8 + x_9 &= b_3 \\ x_1 + \gamma_4x_4 + \gamma_7x_7 &= b_4 \\ x_2 + \gamma_5x_5 + \gamma_8x_8 &= b_5 \\ x_3 + \gamma_6x_6 + \gamma_9x_9 &= b_6, \end{aligned} \tag{7.23}$$

for some $\gamma_i \in \mathbb{Z}_d - \{0\}$, and (generally different) $b \in \mathbb{Z}_d^6$. Note that these operations preserve classical and quantum satisfiability (whereas more general row operations do not). Denote the above reduced LCS by $Mx = b$, where

$$M = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & \gamma_4 & 0 & 0 & \gamma_7 & 0 & 0 \\ 0 & 1 & 0 & 0 & \gamma_5 & 0 & 0 & \gamma_8 & 0 \\ 0 & 0 & 1 & 0 & 0 & \gamma_6 & 0 & 0 & \gamma_9 \end{pmatrix}. \tag{7.24}$$

It is straightforward to check that

$$\operatorname{rank}(M) = \begin{cases} 5 & \text{if } \gamma_4 = \gamma_5 = \gamma_6 \text{ and } \gamma_7 = \gamma_8 = \gamma_9, \\ 6 & \text{otherwise,} \end{cases} \tag{7.25}$$

where the rank is computed modulo d (which is well-defined as d is prime). Thus the system $Mx = b$ is classically unsatisfiable only if $\gamma_4 = \gamma_5 = \gamma_6$ and $\gamma_7 = \gamma_8 = \gamma_9$. Essentially this means that we only need to check $(d-1)^2$ cases, corresponding to the values of the $\gamma_i \in \mathbb{Z}_d - \{0\}$ that could make the LCS classically unsatisfiable. As we will see next, in each of these cases a phase-commutation relation prohibits the existence of a satisfiability gap.

Theorem 22. *Let $d > 2$ be a prime integer. A LCS of the form in eq. (7.22) over $\mathbb{Z}_d$ has a quantum solution if and only if it is classically satisfiable.*

Proof. If a classical solution exists then a quantum solution exists. Conversely, assume that a quantum solution exists, and denote the associated reduced LCS in eq. (7.23) by $Mx = b$. Assume that $\gamma_4 = \gamma_5 = \gamma_6 \equiv \gamma$ and $\gamma_7 = \gamma_8 = \gamma_9 \equiv \delta$, otherwise there is a classical solution and we are done. In the solution group $\mathcal{G}(M, b)$, we have

$$\begin{aligned} g_1 g_2 g_3 &= J^{b_1} \quad & g_4 g_5 g_6 &= J^{b_2} \quad & g_7 g_8 g_9 &= J^{b_3} \\ g_1 g_4^{\gamma} g_7^{\delta} &= J^{b_4} \quad & g_2 g_5^{\gamma} g_8^{\delta} &= J^{b_5} \quad & g_3 g_6^{\gamma} g_9^{\delta} &= J^{b_6}. \end{aligned} \tag{7.26}$$

We note the group identity

$$\begin{aligned} g_1 g_4^{\gamma} g_2 g_5^{\gamma} &= J^{b_4+b_5} g_7^{-\delta} g_8^{-\delta} \\ &= J^{-\delta b_3+b_4+b_5} g_9^{\delta} \\ &= J^{-\delta b_3+b_4+b_5+b_6} g_3^{-1} g_6^{-\gamma} \\ &= J^{b_1-\gamma b_2-\delta b_3+b_4+b_5+b_6} g_1 g_2 g_4^{\gamma} g_5^{\gamma}. \end{aligned} \tag{7.27}$$

Multiplying both sides on the left by g_1^{-1} and on the right by $g_5^{-\gamma}$ reveals the phase-commutation

$$g_2 g_4^\gamma = J^{-b_1-\gamma b_2-\delta b_3+b_4+b_5+b_6} g_4^\gamma g_2. \tag{7.28}$$

Since d is odd, by Theorem 21 and eq. (7.28), a necessary condition for the existence of a quantum solution is that $-b_1-\gamma b_2-\delta b_3+b_4+b_5+b_6 = 0 \mod d$. This implies that the LCS is classically satisfiable: set $x_b = (-\gamma b_2 - \delta b_3 + b_4, b_5, b_1 + \gamma b_2 + \delta b_3 - b_4 - b_5, b_2, 0, 0, b_3, 0, 0)$, then $Mx_b = (b_1, b_2, b_3, b_4, b_5, b_1 + \gamma b_2 + \delta b_3 - b_4 - b_5) = b$. This shows that the original LCS in eq. (7.22) has a classical solution. □

7.4.2 Magic pentagram modulo prime $d > 2$

Similar to the magic square LCS, we consider any LCS, over $\mathbb{Z}_d$ of the pentagram form in Figure 6.1 (b) with arbitrary coefficients, i.e. any system of the form

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 &= b_1 \\ a_1'x_1 + a_5x_5 + a_6x_6 + a_7x_7 &= b_2 \\ a_2'x_2 + a_5'x_5 + a_8x_8 + a_9x_9 &= b_3 \\ a_3'x_3 + a_6'x_6 + a_8'x_8 + a_{10}x_{10} &= b_4 \\ a_4'x_4 + a_7'x_7 + a_9'x_9 + a_{10}'x_{10} &= b_5. \end{aligned} \tag{7.29}$$

We can perform similar reductions to the magic square case. First, we set all coefficients equal to 1 in the first constraint by relabeling $x_i \to a_i x_i$ if necessary. Next, we divide the remaining four constraints by the coefficients of the first variable from the left. Finally, for $i = 5, \ldots, 10$, we relabel $x_i \to a_i x_i$ if necessary to ensure that the coefficient of each x_i equals 1 in the first constraint it appears in. This yields the LCS

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= b_1 \\ x_1 + x_5 + x_6 + x_7 &= b_2 \\ x_2 + \gamma_5x_5 + x_8 + x_9 &= b_3 \\ x_3 + \gamma_6x_6 + \gamma_8x_8 + x_{10} &= b_4 \\ x_4 + \gamma_7x_7 + \gamma_9x_9 + \gamma_{10}x_{10} &= b_5, \end{aligned} \tag{7.30}$$

for some $\gamma_i \in \mathbb{Z}_d - \{0\}$, and (generally different) $b \in \mathbb{Z}_d^5$. As before, the reduction preserves quantum and classical satisfiability. Denote the reduced LCS above by $Mx = b$, where

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \gamma_5 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & \gamma_6 & 0 & \gamma_8 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & \gamma_7 & 0 & \gamma_9 & \gamma_{10} \end{pmatrix}. \tag{7.31}$$

It is straightforward to verify that

$$\operatorname{rank}(M) = \begin{cases} 4 & \text{if } \gamma_i = -1 \text{ for all } i, \\ 5 & \text{otherwise,} \end{cases} \tag{7.32}$$

where the rank is computed modulo d. Thus the system $Mx = b$ is classically unsatisfiable only if $\gamma_i = -1$ for all i. Again, we find that phase-commutation prohibits the existence of a satisfiability gap in this case.

Theorem 23. *Let $d > 2$ be a prime integer. A LCS of the form in eq. (7.29) over $\mathbb{Z}_d$ has a quantum solution if and only if it is classically satisfiable.*

Proof. If a classical solution exists then a quantum solution exists. Conversely, assume that a quantum solution exists, and denote the associated reduced LCS in eq. (7.30) by $Mx = b$. Assume that $\gamma = (-1, -1, -1, -1, -1, -1)$, otherwise there is a classical solution and we are done. In the solution group $\mathcal{G}(M, b)$, we have

$$\begin{gathered} g_1 g_2 g_3 g_4 = J^{b_1} \quad g_1 g_5 g_6 g_7 = J^{b_2} \quad g_2 g_5^{-1} g_8 g_9 = J^{b_3} \\ g_3 g_6^{-1} g_8^{-1} g_{10} = J^{b_4} \quad g_4 g_7^{-1} g_9^{-1} g_{10}^{-1} = J^{b_5}. \end{gathered} \tag{7.33}$$

We can deduce a phase-commutation relation by multiplying the last four constraints:

$$\begin{aligned} J^{b_2+b_3+b_4+b_5} &= (g_1 g_5 g_6 g_7)(g_2 g_5^{-1} g_8 g_9)(g_3 g_6^{-1} g_8^{-1} g_{10})(g_4 g_7^{-1} g_9^{-1} g_{10}^{-1}) \\ &= (g_1 g_6 g_7)(g_2 g_9)(g_3 g_6^{-1})(g_4 g_7^{-1} g_9^{-1}) \\ &= (g_1 g_7)(g_2 g_9)(g_3)(g_4 g_7^{-1} g_9^{-1}) \\ &= (g_1 g_7)(g_2 g_9)(g_7^{-1} g_9^{-1})(g_3 g_4) \\ &= (g_1 g_7) g_2 g_7^{-1} (g_3 g_4) \\ &= g_7 (g_1 g_2 g_4) g_7^{-1} g_3 \end{aligned} \tag{7.34}$$

where in the third and fourth lines we use the facts that g_6 commutes with g_2g_9, and g_3 commutes with g_7g_9, respectively. Substituting $g_1g_2g_4 = J^{b_1}g_3^{-1}$ reveals the phase-commutation relation

$$g_7g_3^{-1} = J^{-b_1+b_2+b_3+b_4+b_5}g_3^{-1}g_7. \tag{7.35}$$

Since d is odd, by Theorem 21 and eq. (7.35), a necessary condition for the existence of a quantum solution is that $-b_1 + b_2 + b_3 + b_4 + b_5 = 0 \mod d$. This implies that the LCS is classically satisfiable: set $x_b = (b_2, b_3, b_4, b_1 - b_2 - b_3 - b_4, 0, 0, 0, 0, 0, 0)$, then $Mx_b = (b_1, b_2, b_3, b_4, b_1 - b_2 - b_3 - b_4) = b$. This shows that the original LCS in eq. (7.29) has a classical solution. □

7.4.3 Magic square and pentagram modulo odd, composite d

When d is a composite integer, $\mathbb{Z}_d$ is not a field, and one cannot define a vector space over $\mathbb{Z}_d$. This makes it difficult to generalize the proofs of the previous two subsections to this setting. In particular, zero-divisors can appear as coefficients in the LCS, making the reductions used in eqs. (7.23) and (7.30) inapplicable. Linear algebraic concepts, such as matrix rank and column space, are also invalid in this case, so we cannot use eqs. (7.25) and (7.32).

Nevertheless, if the coefficients in the LCS are restricted to $\{\pm 1\}$, the arguments used in the prime d case go through almost unchanged. Namely, we can use the same reductions to obtain the LCSs eqs. (7.23) and (7.30), and these LCSs satisfy the following.

Lemma 24. *If the LCS in eq. (7.23) is classically unsatisfiable modulo an odd integer and $\gamma_i \in \{\pm 1\}$, then $\gamma_4 = \gamma_5 = \gamma_6$ and $\gamma_7 = \gamma_8 = \gamma_9$.*

Proof of Lemma 24. Due to symmetry under the exchange $(x_4, x_5, x_6) \leftrightarrow (x_7, x_8, x_9)$, it is enough to show that if $(\gamma_4 = \gamma_5 = \gamma_6)$ does not hold then a classical solution exists. Suppose that $\gamma_4 = -\gamma_5$. Then a classical solution is given by

$$\begin{aligned}
&x_1 \to \frac{1}{2}(b_1 - \gamma_4 b_2 - \gamma_7 b_3 + b_4 - b_5 - b_6) && x_2 \to \frac{1}{2}(b_1 + \gamma_4 b_2 + \gamma_7 b_3 - b_4 + b_5 - b_6)\\
&x_4 \to \frac{1}{2}\gamma_4(-b_1 + \gamma_4 b_2 - \gamma_7 b_3 + b_4 + b_5 + b_6) && x_5 \to \frac{1}{2}\gamma_4(b_1 + \gamma_4 b_2 + \gamma_7 b_3 - b_4 - b_5 - b_6)\\
&x_3 \to b_6, \quad x_7 \to b_3, \quad x_6, x_8, x_9 \to 0.
\end{aligned} \tag{7.36}$$

Suppose now that $\gamma_4 = \gamma_5$ and $\gamma_5 = -\gamma_6$. Then a classical solution is given by

$$\begin{aligned} &x_1 \to \frac{1}{2}\left(b_1 + \gamma_4 b_2 - \gamma_7 b_3 + b_4 - b_5 - b_6\right) && x_3 \to \frac{1}{2}\left(b_1 + \gamma_4 b_2 + \gamma_7 b_3 - b_4 - b_5 + b_6\right) \\ &x_4 \to \frac{1}{2}\gamma_4\left(-b_1 + \gamma_4 b_2 - \gamma_7 b_3 + b_4 + b_5 + b_6\right) && x_6 \to \frac{1}{2}\gamma_4\left(b_1 + \gamma_4 b_2 + \gamma_7 b_3 - b_4 - b_5 + b_6\right) \\ &x_2 \to b_5, \quad x_7 \to b_3, \quad x_5, x_8, x_9 \to 0. \end{aligned} \tag{7.37}$$

These solutions are valid modulo d as long as 2 has a multiplicative inverse modulo d, which is the case when d is odd. □

We can also prove a similar statement for the pentagram LCS:

Lemma 25. *If the LCS in eq. (7.30) is classically unsatisfiable modulo an odd integer and $\gamma_i \in \{\pm 1\}$, then $\gamma = (-1, -1, -1, -1, -1, -1)$.*

Proof of Lemma 25. The following explicit solutions

$$\begin{aligned} x &\to \left(\frac{1}{2}\left(b_1 + b_2 - b_3 - b_4 - b_5\right), \frac{1}{2}\left(b_1 - b_2 + b_3 - b_4 - b_5\right), b_4, b_5, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right), 0, 0, 0, 0, 0\right) && \text{if } \gamma_5 = 1, \\ x &\to \left(\frac{1}{2}\left(b_1 + b_2 - b_3 - b_4 - b_5\right), b_3, \frac{1}{2}\left(b_1 - b_2 - b_3 + b_4 - b_5\right), b_5, 0, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right), 0, 0, 0, 0\right) && \text{if } \gamma_6 = 1, \\ x &\to \left(\frac{1}{2}\left(b_1 + b_2 - b_3 - b_4 - b_5\right), b_3, b_4, \frac{1}{2}\left(b_1 - b_2 - b_3 - b_4 + b_5\right), 0, 0, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right), 0, 0, 0\right) && \text{if } \gamma_7 = 1, \\ x &\to \left(b_2, \frac{1}{2}\left(b_1 - b_2 + b_3 - b_4 - b_5\right), \frac{1}{2}\left(b_1 - b_2 - b_3 + b_4 - b_5\right), b_5, 0, 0, 0, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right), 0, 0\right) && \text{if } \gamma_8 = 1, \\ x &\to \left(b_2, \frac{1}{2}\left(b_1 - b_2 + b_3 - b_4 - b_5\right), b_4, \frac{1}{2}\left(b_1 - b_2 - b_3 - b_4 + b_5\right), 0, 0, 0, 0, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right), 0\right) && \text{if } \gamma_9 = 1, \\ x &\to \left(b_2, b_3, \frac{1}{2}\left(b_1 - b_2 - b_3 + b_4 - b_5\right), \frac{1}{2}\left(b_1 - b_2 - b_3 - b_4 + b_5\right), 0, 0, 0, 0, 0, \frac{1}{2}\left(-b_1 + b_2 + b_3 + b_4 + b_5\right)\right) && \text{if } \gamma_{10} = 1, \end{aligned} \tag{7.38}$$

prove the statement, provided that 2 has a multiplicative inverse modulo d, which is the case if d is odd. □

We can now immediately generalize Theorems 22 and 23 to the case of odd d and coefficients in $\{\pm 1\}$. For the square LCS we get

Theorem 26. *Let $d > 1$ be odd. A LCS of the form in eq. (7.22) over $\mathbb{Z}_d$ with coefficients in $\{\pm 1\}$ has a quantum solution if and only if it is classically satisfiable.*

and for the pentagram LCS we get

Theorem 27. *Let $d > 1$ be odd. A LCS of the form in eq. (7.29) over $\mathbb{Z}_d$ with coefficients in $\{\pm 1\}$ has a quantum solution if and only if it is classically satisfiable.*

The proofs of Theorems 26 and 27 are identical to those of Theorems 22 and 23, respectively.

Chapter 8

Concluding remarks and open problems

One of the major frontiers of current scientific research is understanding the separation in computational power between quantum and classical computers. An essential tool in these investigations is the use of classical simulation algorithms as a theoretical paradigm. The main problem this thesis tackles, i.e. finding low-rank stabilizer decompositions of quantum states, is very unusual in that it is quite concise and yet has profound theoretical and practical implications. The second topic in this work, linear constraint systems, is a highly structured manifestation of certain physical phenomena that have been argued to be the main reason quantum computers have an advantage, namely nonlocality and contextuality. This thesis leaves a number of interesting open questions in both topics, and in this chapter we discuss some of them.

In Chapter 3 we showed that constructing approximate stabilizer decompositions is significantly facilitated by rewriting the circuit in a canonical form using Clifford recompilation. An interesting open question is how much further we can compress the stabilizer decomposition. We have some evidence that using Clifford decompositions of matrix products of multi-qubit gates can lead to dramatic improvements. Furthermore, randomized sparsification bears a close resemblance to well-known Monte Carlo methods in many-body physics. In these methods, significant improvements are possible by employing clever sampling schemes, such as loop and cluster updates. Whether similar ideas can work for stabilizer rank simulations is an interesting and practical question.

In Chapter 4 we found significantly more sparse exact stabilizer decompositions of m copies of a magic state, $|\psi\rangle^{\otimes m}$, and further conjectured that better decompositions exist.

A natural question is whether there exists a lower bound on $\chi(\psi^{\otimes m})$; the stabilizer rank of $|\psi\rangle^{\otimes m}$. If $\chi(\psi^{\otimes m})$ is polynomial or even sub-exponential in m, and we could efficiently compute any desired term of the optimal stabilizer decomposition, that would lead to extremely unlikely complexity theoretic consequences, such as collapse of the polynomial hierarchy [8]. Conversely this means that, either $\chi(\psi^{\otimes m})$ is exponential in m, or that computing the terms of the optimal decomposition is as hard as simulating quantum circuits. Distinguishing between the two possibilities is an interesting open question. Refs. [6, 7, 8] attempt to prove a lower bound on $\chi(T^{\otimes m})$ independently of complexity theoretic conjectures, and succeed in proving partial results, for example proving that $\chi(T^{\otimes m})$ is $\Omega(\sqrt{m})$ [6], and that any approximate decomposition of $\chi(T^{\otimes m})$ that uses tensor product stabilizer states require exponentially many terms [8].

The sparse decompositions derived in Chapter 4 are based on stabilizer states with a certain entanglement structure, but the exact role entanglement plays in this framework remains unclear. Entangled stabilizer states are in a sense classical, since they can be efficiently simulated using the Gottesman-Knill theorem. Yet using entangled stabilizer states seems to facilitate classical simulation of non-stabilizer states by allowing for more sparse decompositions. Finding a generalization of the results in Chapter 4 which elucidates this relationship between entanglement structure and the stabilizer rank is an interesting open problem.

Such a generalization may be possible by considering linear binary codes, i.e. subspaces of the vector space $\mathbb{F}_2^n$. For a given linear code $L \subset \mathbb{F}_2^n$ and a product basis $\{|x\rangle : x \in \mathbb{F}_2^n\}$, one can ask what the stabilizer rank of the uniform superposition $\sum_{x\in L}|x\rangle$ is. The cat states considered in Chapter 4 correspond to the repetition code in the basis $|\psi\rangle, |\psi^{\perp}\rangle$. Other linear codes have a different entanglement structure, and allow similar tensor contractions as in Chapter 4 to be used. Determining the optimal stabilizer decomposition of $|\psi\rangle^{\otimes m}$ obtained in this way is a natural next step.

In Chapter 5 we considered approximate stabilizer decompositions of the time evolved state $e^{iHt}|0\rangle^{\otimes n}$, and concluded that one can find significantly more sparse decompositions in comparison to more standard approaches. We stop short of claiming that this method is advantageous to known methods of simulating short time dynamics, since that would require extensive numerical evidence. This is left for a future work.

For linear constraint systems (LCS), the existence of a quantum/classical satisfiability gap in a LCS modulo d is deeply related to whether d is even or odd. Part of this dependence is elucidated by considering generalized Pauli operators. These operators enter the picture in two ways. First, as shown in Theorem 18, generalized Pauli operators in odd d cannot be used to demonstrate a satisfiability gap, since every quantum solution consisting of these operators can be reduced to a classical solution. Second, the canonical commutation relation $AB = \omega^c BA$, obeyed by generalized Pauli operators, is itself sufficient to rule out a satisfiability gap when d is odd. Specifically, Theorem 21 shows that if this relation is embedded in the solution group of a LCS over odd d, then quantum solutions do not exist in any dimension.

Both of these facts are in contrast to the case of even d; table 7.1 shows a family of LCSs for every even d, each of which has a phase-commutation relation, as well as a satisfiability gap which can be demonstrated using generalized Pauli operators in even d. Theorem 21 can be seen as an obstruction to generalizing the embedding theorem of [25], which holds for $d = 2$, to odd d.

In Section 7.4 we use the relationship between phase-commutation and satisfiability to prove that certain generalizations of the magic square and pentagram to odd d do not have a satisfiability gap. The LCSs we consider in these examples can be seen as arising from the incidence matrices of (weighted) graphs. Prior work for $d = 2$ characterizes the existence of a gap in terms of planarity of such graphs [17], but this characterization does not carry over to the case of $d > 2$. The techniques based on phase-commutation, outlined in the examples in Chapter 7, may become useful in finding such a generalization. It is, however, unclear how simple such a characterization might be. Planarity of a graph can be checked efficiently, see for example [56], but phase-commutation seems to be a more difficult property to check.

Finally, the results in Chapter 7 raise the question of whether there exists a LCS over odd d with a classical/quantum satisfiability gap. While no such cases are currently known, an upcoming work by Slofstra and L. Zhang [57] answers this question in the affirmative; there exists a LCS with a classical/quantum gap modulo d for every prime d, with some evidence that the gap is achievable using finite-dimensional quantum solutions.

References

[1] Daniel Gottesman. The Heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

[2] Leslie G Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing*, 31(4):1229–1254, 2002.

[3] Barbara M Terhal and David P DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Physical Review A*, 65(3):032325, mar 2002.

[4] Michael A. Nielsen and Isaac L. Chuang. Quantum information theory. In *Quantum Computation and Quantum Information*, pages 528–607. Cambridge University Press.

[5] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, nov 2004.

[6] Sergey Bravyi, Graeme Smith, and John A. Smolin. Trading classical and quantum computational resources. *Physical Review X*, 6(2), jun 2016.

[7] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Phys. Rev. Lett.*, 116:250501, Jun 2016.

[8] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, sep 2019.

[9] Dan Stahlke. Quantum interference as a resource for quantum speedup. *Phys. Rev. A*, 90:022302, Aug 2014.

[10] Hakop Pashayan, Joel J Wallman, and Stephen D Bartlett. Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Physical review letters*, 115(7):070501, 2015.

[11] Mark Howard and Earl Campbell. Application of a resource theory for magic states to fault-tolerant quantum computing. *Physical Review Letters*, 118(9):090501, 2017.

[12] Ryan S. Bennink, Erik M. Ferragut, Travis S. Humble, Jason A. Laska, James J. Nutaro, Mark G. Pleszkoch, and Raphael C. Pooser. Unbiased simulation of near-clifford quantum circuits. *Physical Review A*, 95(6), jun 2017.

[13] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics*, 14(11):113011, 2012. DOI: 10.1088/1367-2630/14/11/113011 .

[14] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson. Contextuality supplies the 'magic' for quantum computation. *Nature*, 510(7505):351–355, jun 2014.

[15] Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, oct 2018.

[16] Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, jul 2020.

[17] Alex Aleksandr Arkhipov. *Extending and characterizing quantum magic games*. PhD thesis, Massachusetts Institute of Technology, 2012.

[18] Richard Cleve and Rajat Mittal. Characterization of binary constraint system games. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, pages 320–331, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-43948-727.

[19] N. David Mermin. Hidden variables and the two theorems of john bell. *Reviews of Modern Physics*, 65(3):803–815, jul 1993.

[20] Asher Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991. DOI: 10.1088/0305-4470/24/4/003.

[21] Gilles Brassard, Anne Broadbent, and Alain Tapp. Quantum pseudo-telepathy. *Foundations of Physics*, 35(11):1877–1907, 2005. DOI: 10.1007/s10701-005-7353-4.

[22] M. Junge, M. Navascues, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner. Connes' embedding problem and tsirelson's problem. *Journal of Mathematical Physics*, 52(1):012102, jan 2011.

[23] Richard Cleve, Li Liu, and William Slofstra. Perfect commuting-operator strategies for linear system games. *Journal of Mathematical Physics*, 58(1):012202, 2017. DOI: 10.1063/1.4973422.

[24] William Slofstra. The set of quantum correlations is not closed. In *Forum of Mathematics, Pi*, volume 7. Cambridge University Press, 2019. DOI: 10.1017/fmp.2018.3.

[25] William Slofstra. Tsirelson's problem and an embedding theorem for groups arising from non-local games. *Journal of the American Mathematical Society*, 2019. DOI: 10.1090/jams/929.

[26] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Mip*= re. *arXiv preprint arXiv:2001.04383*, 2020.

[27] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Physical Review A*, 64(1):012310, 2001. 10.1103/PhysRevA.64.012310.

[28] Cihan Okay and Robert Raussendorf. Homotopical approach to quantum contextuality. *arXiv preprint arXiv:1905.03822*, 2019. https://arxiv.org/abs/1905.03822.

[29] Hammam Qassim, Joel J Wallman, and Joseph Emerson. Clifford recompilation for faster classical simulation of quantum circuits. *Quantum*, 3:170, 2019.

[30] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A*, 71(2):022316, 2005.

[31] Hammam Qassim and Joel J Wallman. Classical vs quantum satisfiability in linear constraint systems modulo an integer. *Journal of Physics A: Mathematical and Theoretical*, 53(38):385304, aug 2020.

[32] M Nest. Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. *arXiv preprint arXiv:0811.0898*, 2008.

[33] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[34] Sergey Bravyi and David Gosset. Complexity of quantum impurity problems. *Communications in Mathematical Physics*, 356(2):451–500, aug 2017.

[35] Daniel J. Brod and Ernesto F. Galvão. Extending matchgates into universal quantum computation. *Physical Review A*, 84(2), aug 2011.

[36] Sergey Bravyi, David P Divincenzo, Roberto I Oliveira, and Barbara M Terhal. The complexity of stoquastic local Hamiltonian problems. *arXiv preprint quant-ph/0606140*, 2006.

[37] Andrew M. Childs. Universal computation by quantum walk. *Physical Review Letters*, 102(18), may 2009.

[38] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, mar 2008.

[39] N Schuch, M M Wolf, K G H Vollbrecht, and J I Cirac. On entropy growth and the hardness of simulating time evolution. *New Journal of Physics*, 10(3):033032, mar 2008.

[40] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Physical Review A*, 90(2), aug 2014.

[41] Dominic W. Berry, Andrew M. Childs, Richard Cleve, Robin Kothari, and Rolando D. Somma. Simulating Hamiltonian dynamics with a truncated Taylor series. *Physical Review Letters*, 114(9), mar 2015.

[42] Masuo Suzuki. Transfer-matrix method and monte carlo simulation in quantum spin systems. *Physical Review B*, 31(5):2957–2965, mar 1985.

[43] Simon Kochen and Ernst P Specker. The problem of hidden variables in quantum mechanics. In *The logico-algebraic approach to quantum mechanics*, pages 293–328. Springer, 1975. DOI: 10.1007/978-94-010-1795-417.

[44] Zhengfeng Ji. Binary constraint system games and locally commutative reductions. *arXiv preprint:1310.3794*, 2013. https://arxiv.org/abs/1310.3794.

[45] Ken Dykema, Vern I Paulsen, and Jitendra Prakash. Non-closure of the set of quantum correlations via graphs. *Communications in Mathematical Physics*, 365(3):1125–1142, 2019. DOI: 10.1007/s00220-019-03301-1.

[46] Andrea Coladangelo and Jalex Stark. Robust self-testing for linear constraint system games. *arXiv preprint arXiv:1709.09267*, 2017. https://arxiv.org/abs/1709.09267.

[47] Pierre Antoine Grillet. *Abstract algebra*, volume 242. Springer Science & Business Media, 2007.

[48] David Gross. Hudson's theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics*, 47:122107, 2006. DOI: 0.1063/1.2393152.

[49] Nicolas Delfosse, Cihan Okay, Juan Bermejo-Vega, Dan E Browne, and Robert Raussendorf. Equivalence between contextuality and negativity of the Wigner function for qudits. *New Journal of Physics*, 19(12):123024, 2017. DOI: 10.1088/1367-2630/aa8fe3.

[50] Kathleen S. Gibbons, Matthew J. Hoffman, and William K. Wootters. Discrete phase space based on finite fields. *Physical Review A*, 70(6), dec 2004. DOI: 10.1103/PhysRevA.70.062101.

[51] Wai-Sin Ching. Linear equations over commutative rings. *Linear Algebra and its Applications*, 18(3):257–266, 1977. DOI: 10.1016/0024-3795(77)90055-6.

[52] Arne Storjohann and Thom Mulders. Fast algorithms for linear algebra modulo n. In *European Symposium on Algorithms*, pages 139–150. Springer, 1998. DOI: 10.1007/3-540-68530-812.

[53] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.10.2*, 2019. https://www.gap-system.org.

[54] Huangjun Zhu. Permutation symmetry determines the discrete wigner function. *Physical Review Letters*, 116(4), jan 2016. DOI: 10.1103/physrevlett.116.040501.

[55] Hermann Weyl. *The theory of groups and quantum mechanics (page 277)*. Courier Corporation, 1950.

[56] John M Boyer and Wendy J Myrvold. On the cutting edge: simplified O(n) planarity by edge addition. *J. Graph Algorithms Appl.*, 8(3):241–273, 2004. DOI: 10.1142/9789812773289-0014 .

[57] William Slofstra and L. Zhang. Private communication, 2019.

[58] Richard Cleve, P. Hoyer, Ben Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Proceedings. 19th IEEE Annual Conference on Computational Complexity, 2004.*, page 236. IEEE, 2004.

[59] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Reviews of Modern Physics*, 82(1):665–698, mar 2010.

[60] Elizabeth Crosson and Aram W Harrow. Rapid mixing of path integral monte carlo for 1d stoquastic hamiltonians. *arXiv preprint, arXiv:1812.02144*, 2018.

[61] Ramaswamy Jagannathan. On generalized clifford algebras and their physical applications. In *The legacy of Alladi Ramakrishnan in the mathematical sciences*, pages 465–489. Springer, 2010.

[62] Ulf Leonhardt. Discrete wigner function and quantum-state tomography. *Physical Review A*, 53(5):2998, 1996.

[63] Tobias J. Osborne. Simulating adiabatic evolution of gapped spin systems. *Physical Review A*, 75(3), mar 2007.

[64] Sebastian Paeckel, Thomas Köhler, Andreas Swoboda, Salvatore R. Manmana, Ulrich Schollwöck, and Claudius Hubig. Time-evolution methods for matrix-product states. *Annals of Physics*, 411:167998, dec 2019.

[65] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5), may 2005.

[66] K F Thompson, C Gokler, S Lloyd, and P W Shor. Time independent universal computing with spin chains: quantum plinko machine. *New Journal of Physics*, 18(7):073044, jul 2016.

[67] Michael S. Underwood and David L. Feder. Universal quantum computation by discontinuous quantum walk. *Physical Review A*, 82(4), oct 2010.