# Advancing Classical Simulators by Measuring the Magic of Quantum Computation

*James Robert Seddon*

Supervisors:

*Dr. Earl Campbell*

and

*Prof. Dan Browne*

Submitted in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy**

of

**University College London**.

Department of Physics and Astronomy

UCL

February 19, 2022

I, James Robert Seddon, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

Stabiliser operations and state preparations are efficiently simulable by classical computers. Stabiliser circuits play a key role in quantum error correction and fault-tolerance, and can be promoted to universal quantum computation by the addition of "magic" resource states or non-Clifford gates. It is believed that classically simulating stabiliser circuits supplemented by magic must incur a performance overhead scaling exponentially with the amount of magic. Early simulation methods were limited to circuits with very few Clifford gates, but the need to simulate larger quantum circuits has motivated the development of new methods with reduced overhead. A common theme is that algorithm performance can often be linked to quantifiers of computational resource known as magic monotones. Previous methods have typically been restricted to specific types of circuit, such as unitary or gadgetised circuits. In this thesis we develop a framework for quantifying the resourcefulness of general qubit quantum circuits, and present improved classical simulation methods. We first introduce a family of magic state monotones that reveal a previously unknown formal connection between stabiliser rank and quasiprobability methods. We extend this family by presenting channel monotones that measure the magic of general qubit quantum operations. Next, we introduce a suite of classical algorithms for simulating quantum circuits, which improve on and extend previous methods. Each classical simulator has performance quantified by a related resource measure. We extend the stabiliser rank simulation method to admit mixed states and noisy operations, and refine a previously known sparsification method to yield improved performance. We present a generalisation of quasiprobability sampling techniques with significantly reduced exponential scaling. Finally, we evaluate the simulation cost per use for practically relevant quantum operations, and illustrate how to use our framework to realistically estimate resource costs for particular ideal or noisy quantum circuit instances.

# Impact statement

For several years global technology companies such as Google and IBM, as well as a number of specialised quantum computing start-ups, have been engaged in serious efforts toward building working quantum computers. The long-term goal is to develop devices capable of carrying out useful computations to solve real-world problems that are intractable for classical (meaning conventional) computers. Prospective applications include materials and chemistry simulations, drug discovery, machine learning and optimisation. In the academic community, much work has been devoted to the development of quantum algorithms, and to the theory of quantum error correction and fault tolerance, which will be vital if scalable quantum computation is ever to become a reality. The question of whether and to what extent quantum circuits can be tractably simulated using classical computers is of central importance to research programmes both within and outside academia. Related to this is the question of what quantum resources are necessary to provide a quantum speedup.

This thesis is largely concerned with the quantum resource known as "magic". This resource was previously known to be necessary for universal quantum computation, difficult to simulate classically, and costly to manufacture experimentally in fault-tolerant constructions. Quantum circuits using this resource are set in contrast to stabiliser or Clifford circuits, which can be efficiently simulated on classical computers. The results presented in this thesis contribute to the resource theory of magic by presenting new well-behaved measures of magic for general multi-qubit states and operations, and formalising the link between magic resources and hardness of classical simulation. We present a family of classical simulation algorithms that extend the class of circuits that can be efficiently simulated, and improve simulator performance in the general case. We anticipate this having application, for example, in numerical studies of error correction and magic state distillation protocols. The results of this work have been disseminated to the academic community through articles published in the Proceedings of the Royal Society A [1] and Physical Review X Quantum [2], and via talks and poster presentations at a number of international quantum computation and quantum information conferences and workshops.

Beyond academia, we expect our work to have impact in the ongoing efforts to build prototype quantum computing devices and develop applications for them. The field is now entering the so-called Noisy Intermediate-Scale Quantum (NISQ) era, where devices comprising tens to hundreds of qubits are being constructed and tested. We anticipate that it will become increasingly important to benchmark and verify these devices against classical simulators. The algorithms presented in this thesis are ideally suited to simulating quantum circuits with a wide range of models. By using our methods researchers could potentially reduce the computing time needed to produce numerical data, and access classical simulation results for larger circuit instances with greater depths than would be possible otherwise.

# Acknowledgements

Firstly, I would like to thank my external supervisor and main collaborator, Earl Campbell. His guidance and insight during my PhD project has been invaluable, whether on the science itself, communicating ideas effectively, or other aspects of academic life. Earl has been very generous with his time throughout, and I have always felt supported in my studies. I would also like to thank my UCL supervisor Dan Browne and the members of his group, who have provided my academic home over the past few years. I am indebted to Dan both for insightful discussions on science and his support in navigating the PhD journey.

I am grateful to Mark Howard, with whom I had many helpful discussions during the early stages of my research. A big thank you is due to Bartosz Regula, Hakop Pashayan, and Yingkai Ouyang, who along with Earl Campbell were my timezone-spanning collaborators and co-authors during the later stages of my PhD. I also want to thank those whose interest in our work has sparked many fruitful discussions, in particular Markus Heinrich, David Gross, Arne Heimendahl, and Ryuji Takagi.

It has been a great opportunity to be a part of the wider UCL Quantum community, and in particular the Quantum Technologies CDT. This was made possible by my funding award from the Engineering and Physical Sciences Research Council (grant number EP/P510270/1). I would like to make special mention of the members of CDT cohort 3, Diego Aparicio, Gioele Consani, Abdulah Fawaz, Ed Grant, Matt Flinders, Tom Hird, Gareth Jones, Tamara Kohler, TK Le, Conor Mc Keever, Jamie Potter, Marietta Stasinou and James Watson. Your friendship and humour helped get me through the intense first year of the CDT, and it was a privilege to share the experience with you all. A big thank you goes to Lopa Murgai for her tireless support.

I am very thankful for the encouragement of my family and friends. Too many of you I have not seen nearly enough of during the intensity of the past few years, in

particular my brother Tom, his wife Laura and their son Louis. I want to thank you all for your understanding. Thank you to my parents for always supporting me in everything I have done. Finally, a huge thank you to my wife Vanessa. I am so grateful for your love and support through the whole of my postgraduate study, and especially for everything you have taken on while I have been completing my research and writing. I dedicate this thesis to you and to our daughter Zoe.

# Statement of contribution

Part I of the thesis, comprising Chapters 1-3, largely reviews prior art. In Section 1.3 we give criteria for completely stabiliser-preserving operations, and in Section 2.3 and Appendix C.1 we briefly discuss Choi state robustness and its failure as a well-behaved measure of magic. Similar material previously appeared in the author's MRes thesis, and in our paper Ref. [1]. We include the material in this thesis not as novel work, but as necessary background to novel results presented in Part II. Some of the discussion on sparsification in Section 3.4 and Appendix E.1 was adapted from our paper Ref. [2].

In Part II, Chapters 4-6, we present our novel contribution. Most of the results in Section 4.1 first appeared in Ref. [2], which was joint work with Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl Campbell. The present author's main contributions to that work were in developing the classical simulation algorithms, sparsification results, and associated proofs appearing in Section VII and Appendices C-H of Ref. [2], with input from Hakop Pashayan and Earl Campbell. These simulators and results are discussed in detail in Section 5.2 and Appendices E.2 and F of this thesis. Our co-authors were the principal originators of the magic state monotone results presented in Section 4.1, but we include them as they are closely connected to the classical algorithms just mentioned, and to the novel channel monotones defined by the present author (Section 4.2). The more general form of the monotone equality lemma (Lemma 4.13) is new to this thesis.

The results relating directly to channel robustness (Section 4.2.2) and its formulation as a linear program (Appendix D.1), diagonal channels (Section 4.4), the stabiliser-Kraus teleportation scheme (Section 5.1), the static channel simulator (Section 5.3.1), the numerical results in Section 6.1, and the simple example in Appendix B.1 first appeared in Ref. [1], co-authored with Earl Campbell. There we also defined and proved properties of the magic capacity with respect to robustness of magic (Section 4.3.1), and proposed the associated dynamic simulator (Section 5.3.5). The generalisation of magic capacity to other magic state monotones presented in Section 4.3 is new to this thesis and is the author's own work. In Section 6.1 we include numerical results adapted from our Ref. [1]. A suggestion by an anonymous reviewer of Ref. [2] resulted in the refinement to the stabiliser-Kraus subroutine that we call the polar scheme.

The rest of the novel material presented in Part II is the author's own work and has not been published elsewhere at the time of submission.

# Contents

**List of Figures** 18

**List of Tables** 19

**List of Symbols and Abbreviations** 21

**I Introduction, background and literature review** 23

**1 Quantum computation and the stabiliser formalism** 25

1.1 Preliminaries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

1.2 The Gottesman-Knill theorem . . . . . . . . . . . . . . . . . . . . . . 47

1.3 Completely stabiliser-preserving operations . . . . . . . . . . . . . . . . 55

1.4 Universality, fault tolerance and state injection . . . . . . . . . . . . . . . 56

1.5 Simulating circuits with magic . . . . . . . . . . . . . . . . . . . . . . 62

1.6 Resource theories and monotones . . . . . . . . . . . . . . . . . . . . . 65

1.7 Summary and outlook . . . . . . . . . . . . . . . . . . . . . . . . . . 67

**2 Quasiprobabilities and phase space methods** 69

2.1 The phase space picture of quantum mechanics . . . . . . . . . . . . . . 69

2.2 Wigner function negativity as a computational resource . . . . . . . . . . 71

2.3 Robustness of magic . . . . . . . . . . . . . . . . . . . . . . . . . . . 77

2.4 Phase space picture for qubits . . . . . . . . . . . . . . . . . . . . . . 82

2.5 Simulating CPTP maps . . . . . . . . . . . . . . . . . . . . . . . . . 84

2.6 Summary and outlook . . . . . . . . . . . . . . . . . . . . . . . . . . 87

3 Stabiliser rank methods 89

3.1 State vector decompositions . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

3.2 Exact stabiliser rank . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

3.3 Approximate stabiliser rank and stabiliser extent . . . . . . . . . . . . . . 92

3.4 Sparsification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94

3.5 Inner products, exponential sums and norms . . . . . . . . . . . . . . . . . 98

3.6 Emulating quantum circuits . . . . . . . . . . . . . . . . . . . . . . . . . . 101

3.7 Summary and outlook . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105

II Research 107

4 Quantifying magic for qubit states and operations 109

4.1 Magic monotones for states . . . . . . . . . . . . . . . . . . . . . . . . . . 109

4.2 Measures of magic for quantum channels . . . . . . . . . . . . . . . . . . . 121

4.3 Magic-increasing capacity of quantum channels . . . . . . . . . . . . . . . 140

4.4 Computing monotones for diagonal channels . . . . . . . . . . . . . . . . . 152

4.5 Summary and conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . 164

5 Classical simulation algorithms for non-stabiliser circuits 167

5.1 Stabiliser-Kraus subroutine . . . . . . . . . . . . . . . . . . . . . . . . . . 168

5.2 Algorithms for stabiliser circuits with magic state inputs . . . . . . . . . 176

5.3 Algorithms for magic-generating channels . . . . . . . . . . . . . . . . . . 209

5.4 Summary and conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . 231

6 Resource costs for elements of quantum circuits 233

6.1 Robustness-based channel simulators . . . . . . . . . . . . . . . . . . . . . 236

6.2 Channel decompositions for bit-string sampling . . . . . . . . . . . . . . . 242

6.3 Runtime comparison between dyadic frame and stabiliser rank . . . . . . 251

6.4 Summary and conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . 254

7 Conclusions 257

7.1 Outlook and future work . . . . . . . . . . . . . . . . . . . . . . . . . . . . 259

Appendices 262

**A Clifford circuits 263**

A.1 Generating the Clifford group . . . 263

A.2 Canonical forms . . . 265

A.3 CH-form and Clifford simulation . . . 266

**B Further results on stabiliser-preserving maps 269**

B.1 SP but not completely SP maps . . . 269

B.2 Stinespring dilation of stabiliser channels . . . 271

B.3 Incomplete Kraus representations . . . 274

**C Robustness of magic technical details 277**

C.1 Robustness of the Choi state . . . 277

C.2 Robustness of magic and stabiliser frames . . . 278

**D Additional channel monotone technical details 281**

D.1 Linear program for channel robustness . . . 281

D.2 Dyadic Clifford negativity properties . . . 285

D.3 Dyadic stabiliser channels are norm-contractive . . . 287

D.4 Towards Choi matrix criteria for dyadic channels . . . 287

D.5 Counting projective dyadic stabiliser channels . . . 290

**E Sparsification and bit-string sampling technical details 295**

E.1 Trace norm error for BBCCGH sparsification . . . 295

E.2 Bit-string sampling simulator technical details . . . 296

E.3 Observable estimation for sparsified mixed states . . . 302

E.4 Computing Z-rotation angle for given unitary extent . . . 305

**F Validity and runtime analysis for dyadic channel simulator 307**

**Bibliography 311**

# List of Figures

1.1 The Bloch sphere . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

1.2 The single-qubit stabiliser polytope. . . . . . . . . . . . . . . . . . . . . . 40

1.3 State injection gadget for general third-level $n$-qubit gate $U$. . . . . . . . 61

1.4 Gadget for “teleporting” a state $|\psi\rangle$ between qubit lines. . . . . . . . . . 61

1.5 State injection gadget for the $T$-gate. . . . . . . . . . . . . . . . . . . . . 62

1.6 State injection gadget for diagonal third-level $n$-qubit gate $U$. . . . . . . . 62

2.1 Stabiliser decomposition of a density matrix . . . . . . . . . . . . . . . . 77

4.1 Region $P_Y$ within positive octant of the Bloch sphere . . . . . . . . . . . 116

4.2 Slice $\mathbb{S}_f$ through the Bloch sphere . . . . . . . . . . . . . . . . . . . . 119

4.3 Comparison of robustness of magic with new magic monotones for many copies of noisy Hadamard eigenstate. . . . . . . . . . . . . . . . . . . . 120

5.1 Trace norm error for improved sparsification procedure compared with prior art. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197

5.2 Behaviour of the variable $C$ for single-qubit rotation gates. . . . . . . . . 204

6.1 Variational Quantum Eigensolver schematic . . . . . . . . . . . . . . . . 235

6.2 Classical overhead for amplitude-damped $X$-rotations, compared for different channel simulators. . . . . . . . . . . . . . . . . . . . . . . . . 237

6.3 Normalised channel robustness for $Z$-rotations up to 4 qubits. . . . . . . . 238

6.4 Comparison of channel monotones for multicontrol phase gates . . . . . . 239

6.5 Comparison of channel robustness with robustness of the Choi state for random diagonal gates. . . . . . . . . . . . . . . . . . . . . . . . . . 241

6.6 Stabiliser channel extent for dephased $Z$-rotations. . . . . . . . . . . . . . 247

6.7 Overhead saving for simulating $N$ uses of dephased vs. ideal gate using channel extent method. . . . . . . . . . . . . . . . . . . . . . . . . . . . 247

6.8 Channel extent for depolarised $Z$-rotations. . . . . . . . . . . . . . . . . . . 248

6.9 Bounds on channel extent for depolarised $THT$-gate. . . . . . . . . . . . 249

6.10 Stabiliser channel extent and equimagical upper bound for single-qubit gates subject to depolarising noise. . . . . . . . . . . . . . . . . . . . . . . . 251

6.11 Runtime comparison for observable estimation using dyadic frame and stabiliser rank methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . 253

# List of Tables

1.1 Truth table for phase gate update in stabiliser tableau method. . . . . . . . 49

3.1 Selection of known upper bounds for stabiliser rank for tensor products of Hadamard states and Haar random states. . . . . . . . . . . . . . . . . . 91

4.1 Number of pure stabiliser states $N_S$ for number of qubits $n$. . . . . . . . . . 152

4.2 Number of $n$-qubit stabiliser states compared with number of affine spaces. 160

4.3 Families of magic monotones . . . . . . . . . . . . . . . . . . . . . . . . 164

5.1 Error bounds for constrained path simulator in different parameter regimes 178

6.1 Final robustness following multicontrol-$T$ gate for initial stabiliser states against linear subspace dimension. . . . . . . . . . . . . . . . . . . . . . . 240

D.1 Counting the number of elements $\mathcal{T}$ corresponding to the set of orthogonal projectors $\left\{\Pi_j^{(Z)}\right\}$, neglecting phase. . . . . . . . . . . . . . . . . . . 292

# List of Symbols and Abbreviations

| | |
|---|---|
| $\mathbb{F}_k$ | The finite field with $k$ elements |
| $\mathbb{F}_k^n$ | The $n$-dimensional vector space over the field $\mathbb{F}_k$ |
| $\mathcal{P}_n$ | The $n$-qubit Pauli group |
| $\mathcal{P}_{n\pm}$ | The set of Hermitian $n$-qubit Pauli operators |
| $\mathcal{P}_{n+}$ | The set of unsigned $n$-qubit Pauli operators |
| $\mathrm{STAB}_n$ | The set of pure $n$-qubit stabiliser state vectors |
| $\overline{\mathrm{STAB}_n}$ | The convex hull of $n$-qubit stabiliser state density operators |
| CP | Completely positive |
| CPTP | Completely positive trace-preserving |
| $\Phi_{\mathcal{E}}$ | Choi state for the channel $\mathcal{E}$ (Thm. 1.3) |
| $\mathrm{SP}_{n,n}$ | The set of completely stabiliser-preserving $n$-qubit channels |
| $\lVert A \rVert_1$ | Schatten 1-norm (trace norm) of the operator $A$ |
| $\lVert \mathbf{q} \rVert_1$ | $\ell_1$-norm of the vector $\mathbf{q}$ |
| RoM, $\mathcal{R}$ | Robustness of magic, (Def. 2.3) |
| CPR | The set of Clifford gates and Pauli-reset channels |
| $\mathcal{R}_{\mathrm{CPR}}$ | Robustness with respect to the convex hull of CPR |
| $\chi$ | Stabiliser rank (Def. 3.1) |
| $\xi$ | Pure-state stabiliser extent (Def. 3.3) |
| BBCCGH | Bravyi, Browne, Calpin, Campbell, Gosset and Howard [3] |
| $\mathcal{R}_*$ | Channel robustness (Def. 4.18) |
| gRoM, $\Lambda^+$ | Generalised robustness of magic (Def. 4.4) |

| | |
|---|---|
| $\Lambda$ | Dyadic negativity (Def. 4.2) |
| $\Xi$ | Density-operator stabiliser extent (Def. 4.1) |
| $\mathcal{C}_{\mathcal{M}}$ | Magic capacity with respect to monotone $\mathcal{M}$ (Def. 4.25) |
| i.i.d. | Independent and identically distributed |

# Part I

# Introduction, background and literature review

## Chapter 1

# Quantum computation and the stabiliser formalism

Computers can be modelled as devices that take input data, perform a set of instructions we call an algorithm and return output data [4]. The circuit model of classical computation is a strictly less general one, but arguably more instructive. In this model logic circuits [5] are finite-sized constructions that can be used to compute Boolean functions $f : \{0,1\}^n \to \{0,1\}^m$. In a given circuit instance, the input is encoded as an $n$-bit binary number, the output is some other $m$-bit string, and the algorithm is implemented via a finite sequence of elementary logic gates such as AND, OR, NOT and CNOT (controlled-NOT), which perform Boolean operations on a subset of bits. Gates are connected by "wires" which pass the output from one gate to the input of another. This is not a universal model for classical computation, in the sense of the famous Turing machine model [6, 7] since while logic circuits always halt for any inputs, Turing machines do not. Nevertheless, it has been shown that any function that can be computed by a Turing machine, can be computed by some circuit family [4]. Crucially, this kind of classical logic can be realised reliably by physical components such as transistors, enabling their exploitation in the conventional computers ubiquitous today.

Since logical gates must be implemented by physical components, logical data takes up physical space, and any logical operation takes place in some finite time. An important consideration in algorithm design is therefore the amount of memory or *space* a given algorithm requires, in terms of number of data and work bits, and the number of logical gates that must be carried out in sequence, which translates to the algorithm runtime. Classical complexity theory is the branch of computer science concerned with fundamental limits on time and space requirements for different families of computational task [8]. Computational tasks are often formulated as decision problems, where a problem instance is specified by an input bit string of length $n$, and the task of the computer is to decide whether the output of the algorithm should be “YES” or “NO”, or equivalently, whether the output bit should be 0 or 1. Problems are sorted into complexity classes by studying their space and time costs with respect to some computational model, and how they scale with problem size, as quantified by the length of the input string. For example, a problem is in **P** if it can be decided in time polynomial in $n$ by a deterministic Turing machine (i.e. one which always gives the same output for a given input), whereas **PSPACE** is the class of problems that can be solved using a polynomial amount of memory.

Other complexity classes are concerned with how easily the solution to a problem can be checked. A problem is in **NP** if for every “YES” instance, it is possible for a deterministic Turing machine to check that a solution is valid in polynomial time. A famous open question in complexity theory is whether **P** is equal to **NP**. It is clear that $\mathbf{P} \subseteq \mathbf{NP}$ [9]. In the other direction, it is widely conjectured that $\mathbf{NP} \nsubseteq \mathbf{P}$, though this has never been proven definitively. The archetypal illustration of this is the problem of prime factorisation. There is no known classical algorithm that is able to factor arbitrary $n$-bit integers into primes in $\text{poly}(n)$ time, but it is straightforward to check a given factorisation simply by computing the product of the given factors. Therefore the problem is in **NP** but is not known to be in **P** [4].

Many other practically important problems are not known to have an efficient classical algorithm. Of particular interest for this thesis is the problem of simulating quantum systems, which has wide application in many areas of fundamental and applied science and industry [10, 11]. Simulating quantum systems is hard in general, since the amount of information needed to describe the wave function grows exponentially with the number of physical particles [11]. In the early 1980s Feynman [12] and Manin [13] independently hit on the idea that rather than simulating a quantum system with a conventional computer, which is in essence a physical system that evolves classically, it would be more natural if the simulator itself evolved quantum mechanically at the logical level. This proposition eventually led to the field of quantum computation [4]. The model of quantum computation we focus on in this thesis is the circuit model. This model is analogous to the logic circuit in classical computation; quantum bits, or *qubits*, replace classical bits as the fundamental unit of information, and reversible logic gates are replaced by unitary operations. Other models include measurement-based [14, 15] and continuous variable [16] quantum computation. Constructing a quantum computer represents a major engineering challenge, and it is only in recent years that serious efforts toward building one have got underway. It has recently been suggested that we are now entering the era of Noisy Intermediate-Scale Quantum (NISQ) technologies, and can expect to see devices comprising 50 to several hundred noisy qubits appearing over the next few years [17].

There is a widespread expectation that quantum computers will at some point outstrip the ability of classical computers to simulate them, and there is a body of work dedicated toward achieving this milestone [18–21]. Quantum complexity theory [22] introduces new complexity classes based on quantum models of computation. **BQP**, which stands for bounded-error quantum polynomial time, is the class of problems solvable with a quantum computer in polynomial time, allowing a bounded probability error. This is the analogue of a classical complexity class called **BPP**, comprising problems that can be solved classically with bounded probability of error. **BQP** can be shown to contain **BPP**, which in turn contains **P**, and it is hoped that there exist problems in the gap between **BQP** and **BPP**, as this is where the potential for quantum advantage exists. Various candidate problems have been proposed for which an efficient quantum algorithm exists, but for which there is no known classical algorithm [4, 23]. Bravyi, Gosset and Koenig showed that the 2D Hidden Linear Function problem requires logarithmic depth circuits in the classical case, but can be solved in constant depth by a quantum computer [24]. The search for unconditional quantum advantage in scenarios where the classical circuit size grows superpolynomially remains an active field of research.

Meanwhile, a number of techniques have been developed for classically simulating restricted classes of quantum circuit [25–29]. The Gottesman-Knill theorem was an important early result that tells us that there is a large subclass of circuits, called stabiliser circuits, that can be efficiently simulated classically [25, 30, 31]. The stabiliser circuits are usually taken to be generated by Clifford gates and measurements of Pauli observables, supplemented by classical randomness and adaptivity. In order to achieve universal quantum computation (and thereby quantum speedup), it is necessary to add another ingredient, and the standard choice is the single-qubit unitary operation known as the $T$-gate. It turns out that in many error correction schemes for scalable quantum computation, the stabiliser circuits can be implemented fault-tolerantly, but $T$-gates cannot [32, 33]. The usual solution is to "inject" a resource state known as a magic state; by consuming the magic state, a stabiliser circuit can in effect be promoted to an implementation of the $T$-gate [34, 35]. The catch is that in order to avoid introducing uncorrectable noise into the computation, these magic states must be prepared to extremely high fidelity through magic state distillation protocols, introducing a significant operational overhead every time a $T$-gate is required in the circuit [33–44]. While the term *magic state* is often used to mean the specific resource states used in fault-tolerant quantum computation, in this thesis we adopt a broader definition. We simply say that a state is a magic state if it is not a stabiliser state. Similarly, an operation is termed magic if it can convert a stabiliser state to a magic state. We make precise the definitions of stabiliser states and operations later in this chapter.

Non-stabiliser circuits can in principle be classically simulated *inefficiently* by extending the Gottesman-Knill framework [31], with a simulation overhead that is expected to grow exponentially with the amount of magic resource used by the circuit. In this thesis, we are primarily motivated by the intermediate regime of noisy "near-Clifford" circuits, where the magic resource involved is modest enough that simulation may at least be practical within a reasonable timeframe, given a powerful enough classical computer [45]. One of the main aims of the thesis is to make precise what is meant by magic resource, so that we can rigorously quantify the resource cost of a given quantum operation. Identifying the required resource for a given quantum channel is of interest both from the point of view of classical simulation, and with respect to the implementation of prototype fault-tolerant devices where magic state resource must be consumed in order to complete a sequence of operations. The former application is our focus in this thesis. There are two main strands to the novel results presented in later chapters:

1. The development of a framework for meaningfully quantifying magic as a resource, including understanding the resourcefulness of general non-stabiliser operations on $n$-qubit systems. This forms the topic of Chapter 4.

2. The design and improvement of classical simulation algorithms for emulating general quantum circuits on qubits, based on extensions of stabiliser simulation tech-

niques. Our improved algorithms are presented in Chapter 5, where we also prove results relating to their performance.

In Chapter 6, we draw on the formal results of Chapters 4 and 5 to numerically estimate classical simulation costs for sequences of gates relevant to practical quantum algorithms suitable for implementation on near-term devices. We conclude in Chapter 7. Before turning to this novel work, in the rest of this chapter we introduce fundamental concepts, tools and notation that will be important throughout the thesis, including the representation and classical simulation of stabiliser states and operations. We will see how the stabiliser formalism stands in relation to universal quantum computation and the theory of fault tolerance. We discuss the idea that “non-stabiliserness” or magic is a resource for quantum advantage; classical simulation of non-stabiliser circuits is thought to require computational resources that grow exponentially with the “amount” of magic. We close the chapter by reviewing how quantum resource theories can be used to formalise what is meant by the “amount” of some resource. In Chapters 2 and 3 we will review the prior state of the art in the theory of magic resources and methods for their classical simulation.

## 1.1 Preliminaries

In this section we briefly review some important basic concepts and notation. We follow treatments given in standard texts on quantum computation and quantum information, such as Refs. [4, 30, 46], and the interested reader can find more comprehensive introductions to these topics therein.

### 1.1.1 Definitions and notation

For vectors in a finite dimensional Hilbert space $\mathcal{H}$, we use $\|\cdot\|$ to denote the usual 2-norm, i.e. $\|\psi\| = \||\psi\rangle\| = \sqrt{\sum_j |c_j|^2}$, for $|\psi\rangle = \sum_j c_j |j\rangle$ decomposed in some orthonormal basis $|j\rangle$. We will also make use of the following family of operator norms.

**Definition 1.1** (Schatten $p$-norms)**.** *For operator $A$, the Schatten $p$-norm for positive integer $p$ is defined*

$$\|A\|_p = \left( \sum_k (s_k(A))^p \right)^{\frac{1}{p}}, \tag{1.1}$$

*where $s_k(A)$ are the singular values of $A$.*

The Schatten norms are unitarily invariant, $\|UAV\|_p = \|A\|_p$ for any unitary operators $U$ and $V$. The Schatten 1-norm will be particularly important in later chapters. This is also known as the trace norm and can be equivalently defined,

$$\|A\|_1 = \mathrm{Tr}[|A|] = \sum_k s_k(A), \quad \text{where} \quad |A| = \sqrt{A^\dagger A} \tag{1.2}$$

We will encounter many *convex sets*. Given a space $\mathbb{V}$, a linear combination of points $\mathbf{a}' = \sum_j p_j \mathbf{a}_j$, $\mathbf{a}_j \in \mathbb{V}$, is called a *convex combination* if $\sum_j p_j = 1$ and $p_j \geq 0$, $\forall j$. A set $C \subseteq \mathbb{V}$ is *convex* if any convex combination of points in $C$ is also an element of $C$. For a generic subset $S \subseteq \mathbb{V}$, we can obtain a convex set by taking the convex hull $\mathrm{conv}(S)$, defined as the set of all convex combinations of points in $S$,

$$\mathrm{conv}(S) = \left\{ \mathbf{a} : \mathbf{a} = \sum_j p_j \mathbf{s}_j, \quad \sum_j p_j = 1, p_j \geq 0, \mathbf{s}_j \in S, \forall j \right\}. \tag{1.3}$$

Later in this thesis we will describe algorithms using pseudocode. For the most part the notation used is standard, but for the avoidance of doubt, the notation $a \leftarrow b$ means "update variable $a$ to the value $b$". For example $a \leftarrow a+1$ should be read as "increment $a$ by 1".

### 1.1.2 The Pauli group

Throughout this thesis we will use the standard convention of representing pure single-qubit states as $\mathbb{C}^2$ vectors, expressed in the computational basis:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{1.4}$$

where $|0\rangle$ and $|1\rangle$ are respectively the +1 and -1 eigenstates of the Pauli $Z$ operator. In this basis the single-qubit Pauli operators are represented:

$$\sigma_0 = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{1.5}$$

$$\sigma_2 = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.6}$$

These operators are self-inverse ($\sigma_j^2 = \mathbb{1}$), Hermitian ($\sigma_j^\dagger = \sigma_j$), and are traceless except for the identity ($\mathrm{Tr}(\sigma_j) = 2\delta_{j,0}$). We denote the eigenstates of $X$ and $Y$ as $|\pm\rangle$ and $|\pm_i\rangle$ respectively:

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \pm 1\end{pmatrix}, \quad |\pm_i\rangle = \frac{|0\rangle \pm i|1\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \pm i\end{pmatrix}. \tag{1.7}$$

We will shortly define stabiliser states formally, but for now we simply note that the pure single-qubit stabiliser states are precisely the six Pauli eigenstates just described. Distinct non-identity Pauli operators anticommute and compose as

$$XY = iZ, \quad YZ = iX, \quad ZX = iY. \tag{1.8}$$

It follows that the single-qubit Pauli operators with phase ±1 or $\pm i$ form a group, $\mathcal{P}_1 = \langle i\mathbb{1}, Z, X\rangle$. Generalised Pauli matrices can be defined as $n$-fold tensor products of elements from $\mathcal{P}_1$. We use a subscript to indicate which qubit a single-qubit Pauli acts on non-trivially; for example, for a 4-qubit system, $Z_3 = \mathbb{1}\otimes\mathbb{1}\otimes Z\otimes\mathbb{1}$. We will use the notation $\mathbb{1}_n$ to denote the identity operator for an $n$-qubit Hilbert space, represented by the $2^n \times 2^n$ identity matrix. Let $\mathcal{P}_n$ be the group of all $n$-qubit Pauli matrices, let $\mathcal{P}_{n\pm}$ be the subset of Pauli operators with real phase, and let $\mathcal{P}_{n+}$ be the unsigned Pauli operators. Each $P \in \mathcal{P}_{n\pm}$ inherits the standard properties from the single-qubit case. Namely, they are self-inverse and Hermitian, and traceless unless equal to the identity, $\mathrm{Tr}\,P = 2^n\delta_{P,\mathbb{1}_n}$. Since $ZX = iY$, we can represent any $P \in \mathcal{P}_{n\pm}$ in the form,

$$P = (-1)^{a_0} T_{\mathbf{a}}, \quad T_{\mathbf{a}} = (-i)^{\mathbf{z}\cdot\mathbf{x}} \prod_{k=1}^{n} Z_k^{z_k} X_k^{x_k}, \tag{1.9}$$

where $a_0 \in 0, 1$, and $\mathbf{a} = (\mathbf{z}, \mathbf{x})$ is an element of the $2n$-dimensional linear vector space $\mathbb{F}_2^{2n}$. Note that the factor $(-i)^{\mathbf{z}\cdot\mathbf{x}}$ ensures that any imaginary phase arising from a product of $Z$ and $X$ operators on the same qubit is cancelled.

Given a pair of unsigned Pauli operators, $T_{\mathbf{a}}, T_{\mathbf{b}} \in \mathcal{P}_{n+}$ we can check whether they commute as follows. Let $P = P_1 \otimes \ldots \otimes P_n$ and $Q = Q_1 \otimes \ldots \otimes Q_n$, and let $N_a$ be the number of indices such that $\{P_j, Q_j\} = 0$. If $N_a$ is even, $P$ and $Q$ commute; if odd, they anticommute. This can be expressed in terms of the symplectic inner product,

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}_x \cdot \mathbf{b}_z \oplus \mathbf{a}_z \cdot \mathbf{b}_x, \tag{1.10}$$

where $\oplus$ indicates that addition is modulo 2. This is consistent with the definition of the symplectic inner product in higher dimensions, as in $\mathbb{F}_2$, addition and subtraction are equivalent. So, two Pauli operators commute if and only if $\langle \mathbf{a}, \mathbf{b} \rangle = 0$,

$$T_{\mathbf{a}} T_{\mathbf{b}} = (-1)^{\langle \mathbf{a}, \mathbf{b} \rangle} T_{\mathbf{b}} T_{\mathbf{a}}. \tag{1.11}$$

Using the relation (1.9), and observing that we acquire a factor of $(-1)$ for every place a $Z_k$ operator must be commuted past an $X_k$ operator, it is straightforward to check that signed Pauli operators $P = (-1)^{a_0} T_{\mathbf{a}}$ and $Q = (-1)^{b_0} T_{\mathbf{b}}$ compose as

$$PQ = (-1)^{(a_0 + b_0)} T_{\mathbf{a}} T_{\mathbf{b}} = (-1)^{(a_0 + b_0)} i^{(\mathbf{z}_a \cdot \mathbf{x}_b - \mathbf{x}_a \cdot \mathbf{z}_b)} T_{\mathbf{a}+\mathbf{b}}. \tag{1.12}$$

Note that the expression $(\mathbf{z}_a \cdot \mathbf{x}_b - \mathbf{x}_a \cdot \mathbf{z}_b)$ is computed modulo 4, so differs from the symplectic inner product. However when $[P, Q] = 0$, the expression evaluates to either 0 or 2, so the product of commuting $P$ and $Q$ remains Hermitian.

### 1.1.3 Single-qubit rotations

Next we review some basic properties of single-qubit systems. The unsigned Pauli matrices form a real basis for the space of Hermitian operators on $\mathbb{C}^2$. In particular, the density operator for any single-qubit state can be written,

$$\rho = \frac{\mathbb{1} + \mathbf{a} \cdot \sigma}{2} \tag{1.13}$$

where $\mathbf{a}$ is a real-valued three-dimensional vector, called a Bloch vector, and $\sigma$ is the vector of non-trivial Pauli matrices. The operator $\mathbf{a} \cdot \sigma$ has eigenvalues $\pm ||\mathbf{a}||$ where $||\cdot||$ is the usual Euclidean norm. The positivity of $\rho$ imposes $||\mathbf{a}|| \leq 1$, while the purity condition $\mathrm{Tr}\left[\rho^2\right] = 1$ is equivalent to $||\mathbf{a}|| = 1$. This leads to the Bloch sphere picture (Figure 1.1), where pure states are represented as unit vectors on the unit sphere, while mixed states are represented as vectors in the interior of the unit ball. The axes $\hat{x}$, $\hat{y}$ and $\hat{z}$ are associated with the Pauli operators $X$, $Y$ and $Z$ respectively, while an arbitrary

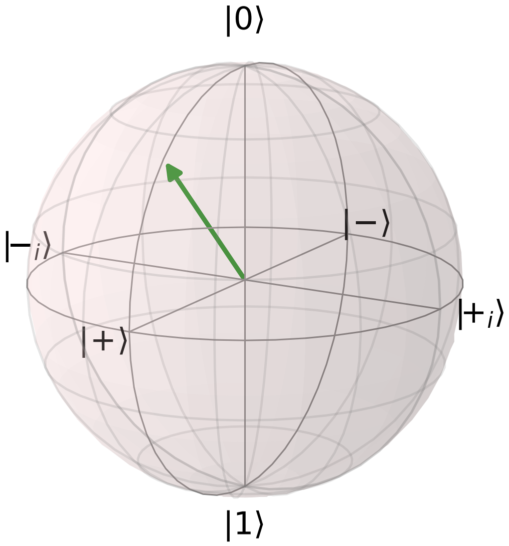


**Figure 1.1:** The Bloch sphere picture for single-qubit states. As an example, the green vector represents the Hadamard eigenstate $|H\rangle = (|0\rangle + |+\rangle)/\sqrt{2}$, which has Bloch vector $\mathbf{a} = [1/\sqrt{2}, 0, 1/\sqrt{2}]^T$. Image produced using the QuTIP package [47].

axis defined by the unit vector $\hat{a}$ is associated with the operator $\hat{a} \cdot \sigma$. Pure states can alternatively be parameterised as $|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\phi}\sin(\theta/2)\,|1\rangle$, up to a global phase, where $\theta$ and $\phi$ are respectively the polar and azimuthal angle on the Bloch sphere.

Reversible operations on qubits are represented by elements of the unitary group U(2). Since global phase can be ignored, the restriction to SU(2) is sufficient to represent any physical unitary operation. Elements in SU(2) can be parameterised

$$U(\alpha,\beta,\gamma) = \begin{pmatrix} e^{i\beta}\cos(\alpha) & -e^{-i\gamma}\sin(\alpha) \\ e^{i\gamma}\sin(\alpha) & e^{-i\beta}\cos(\alpha) \end{pmatrix}. \tag{1.14}$$

The isomorphism from SU(2) to SO(3) means that any reversible single-qubit operation can be represented by a 3-dimensional rotation matrix $R$ in the Bloch sphere. This transforms vectors $\mathbf{a}$ in the Bloch sphere as $\mathbf{a}' = R\mathbf{a}$. Up to overall phase, any unitary operation can be parameterised as $U_{\hat{r}}(\theta) = \exp(-i\theta(\hat{r}\cdot\sigma)/2)$ for some angle $\theta$ and unit vector $\hat{r}$. For example, a rotation about the $\hat{z}$-axis can be represented

$$U_{\hat{z}}(\theta) = \begin{pmatrix} e^{\frac{-i\theta}{2}} & 0 \\ 0 & e^{\frac{i\theta}{2}} \end{pmatrix} \quad \text{or} \quad R_{\hat{z}}(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.15}$$

Any rotation $R_{\hat{a}}(\theta)$ can be generated by a sequence of rotations about a pair of orthogonal axes $\hat{b}$ and $\hat{c}$, such that we have:

$$R_{\hat{a}}(\theta) = R_{\hat{b}}(\theta_3)R_{\hat{c}}(\theta_2)R_{\hat{b}}(\theta_1) \quad \longleftrightarrow \quad U_{\hat{a}}(\theta) = U_{\hat{b}}(\theta_3)U_{\hat{c}}(\theta_2)U_{\hat{b}}(\theta_1) \tag{1.16}$$

for some triple of angles $(\theta_1, \theta_2, \theta_3)$. This covers most of the machinery we will need for single-qubit unitary evolution, but before moving onto more general quantum operations, it is worth introducing several important single-qubit gates.

#### 1.1.3.1 Single-qubit Clifford gates

The Clifford gates are the unitary operators that map Pauli operators to Pauli operators. We will discuss these in detail later, here we give some important examples for the single-qubit case. The SO(3) matrices for Cliffords are signed permutation matrices. Pauli operators are themselves Clifford operators, and can be represented as $\pi$-rotations about the relevant axis (up to a phase), e.g. $Z \propto U_{\hat{z}}(\pi)$. Their SO(3) matrices are diagonal, flipping the sign of the elements corresponding to the other two Pauli operators:

$$R_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, R_Y = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, R_Z = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.17}$$

This is equivalent to swapping basis states, and indeed the $X$ gate is sometimes called the *NOT* or bit-flip gate, $X|0\rangle = |1\rangle$ and $X|1\rangle = |0\rangle$.

The Hadamard gate exchanges the $Z$ and $X$ operators, and flips the sign of $Y$. It can be represented as a $\pi$-rotation about the axis $\hat{h} = (\hat{x} + \hat{z})/\sqrt{2}$.

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = iU_{\hat{h}}(\pi), \quad R_H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{1.18}$$

The phase gate $S$, sometimes referred to as $P$ or $\sqrt{Z}$ in the literature, is a rotation through $\pi/2$ about the $Z$-axis:

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} = e^{\frac{i\pi}{2}} U_{\hat{z}}\left(\frac{\pi}{2}\right), \quad R_S = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.19}$$

All single-qubit Clifford gates, of which there are 24, can be generated by composition of the Hadamard and phase gates [30].

#### 1.1.3.2 Non-Clifford gates

Another gate that we will discuss many times in the course of this thesis is the $T$-gate, which we will see has an important role in fault-tolerant quantum computation. This is represented by a rotation through $\pi/4$ about the $\hat{z}$-axis, though it is sometimes called the $\pi/8$-gate in the literature due to the phase factors appearing on the diagonal in the SU(2) representation.

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{\frac{i\pi}{4}} \end{pmatrix} = e^{\frac{i\pi}{8}} \begin{pmatrix} e^{-\frac{i\pi}{8}} & 0 \\ 0 & e^{\frac{i\pi}{8}} \end{pmatrix}, \quad R_T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1.20}$$

### 1.1.4 Representing quantum operations

While unitary operations are central in the circuit model of quantum computation, non-unitary processes will be of particular interest in this thesis. These include adaptive maps, where operations are conditioned on the outcome of previous measurements, and noise channels. The former are often a vital component in architectures for quantum computation, while the latter can be a serious obstacle. Noise channels are usually understood as resulting from global unitary dynamics that induce an uncontrolled interaction between the system combined with its environment; ignorance of the environment manifests as non-unitary dynamics in the system of interest. Modelling such channels is vitally important in the theories of error correction and mitigation [30, 48, 49].

Quantum operations are represented by linear maps from an input space $\mathcal{H}_1$ to an output space $\mathcal{H}_2$. We will typically assume the input and output space have the same dimension. The most general maps that are physically meaningful in the sense that they map physical states to physical states, up to normalisation, are the *completely positive*

(CP) maps. These are the maps $\mathcal{T}$ that preserve the positivity of operators, even when extended to a larger system: $\mathcal{T}\otimes\mathbb{1}(A)\geq 0$ for all $A\geq 0$. The CP maps $\mathcal{E}$ that also preserve normalisation are called *completely positive trace-preserving* (CPTP) maps; these satisfy $\mathrm{Tr}(\mathcal{E}(A))=\mathrm{Tr}(A), \forall A$. Throughout this thesis we will refer to *quantum channels*, and this should be taken to mean CPTP maps, unless otherwise specified. However, non-trace-preserving CP maps can have a meaningful operational interpretation. For example, we will sometimes want to consider maps corresponding to post-selection on a particular measurement outcome. We now review several representations for quantum operations.

#### 1.1.4.1 Kraus representation

Any quantum channel $\mathcal{E}$ on a system $S$ can be modelled as resulting from unitary dynamics $U$ a larger system $SE$ [46, 50], where the environment $E$ can, without loss of generality, be assumed to be initialised in some pure state $|\psi_0\rangle$. To obtain the action of this process on the state of system $S$ alone, we trace out the environment $E$.

$$\begin{aligned}\mathcal{E}(\rho) &= \mathrm{Tr}_E[U(\rho_S\otimes|\psi_0\rangle\langle\psi_0|_E)U^\dagger] &(1.21)\\ &=\sum_j\langle j|_E U|\psi_0\rangle_E\,\rho_S\,\langle\psi_0|_E U|j\rangle_E=\sum_j K_j\rho K_j^\dagger. &(1.22)\end{aligned}$$

Here $\{|j\rangle\}$ is any orthonormal basis for the space $\mathcal{H}_E$. In the last line, $K_j=\langle j|_E U|\psi_0\rangle_E$ is called a Kraus operator, and (1.22) is called a *Kraus representation* or *operator-sum representation* of the channel. The representation is in general non-unique because we can choose any orthonormal basis for the partial trace. If we assume the channel is trace-preserving, then for any operator $A$ we have

$$\mathrm{Tr}[\mathcal{E}(A)]=\mathrm{Tr}\left[\sum_j K_jAK_j^\dagger\right]=\mathrm{Tr}\left[\sum_j K_j^\dagger K_jA\right]=\mathrm{Tr}[A] \qquad (1.23)$$

where we used linearity and cyclicity of the trace. But $\mathrm{Tr}\left[\sum_j K_j^\dagger K_jA\right]=\mathrm{Tr}[A]$ holds for arbitrary $A$ if and only if $\sum_j K_j^\dagger K_j=\mathbb{1}$. In fact, given any set of operators $\{K_j\}_j$, we can be sure that $\mathcal{E}(\cdot)=\sum_j K_j(\cdot)K_j^\dagger$ represents a valid quantum channel provided that the operators satisfy the *completeness relation*,

$$\sum_j K_j^\dagger K_j=\mathbb{1}. \qquad (1.24)$$

It can be shown that any map satisfying this relation is completely positive, hermiticity-preserving and trace-preserving [46]. We can also model quantum operations that are trace non-increasing by changing the equality in the completeness relation (1.24) to an operator inequality,

$$\sum_j K_j^\dagger K_j \leq \mathbb{1}. \tag{1.25}$$

Maps represented by Kraus operators satisying (1.25) remain linear, completely positive and Hermiticity-preserving. At this point it is useful to define the *dual* of a map [46], sometimes called the adjoint map.

**Definition 1.2** (Dual map)**.** *Given a linear map $\mathcal{T}$, the dual map $\mathcal{T}^*$ is the map that satisfies, for arbitrary operators $A$ and $B$,*

$$\mathrm{Tr}[A\mathcal{T}^*(B)] = \mathrm{Tr}[\mathcal{T}(A)B]. \tag{1.26}$$

The dual map can easily be obtained from the Kraus representation. By cyclicity of the trace, $\mathrm{Tr}\left[\sum_j K_j A K_j^\dagger B\right] = \mathrm{Tr}\left[A \sum_j K_j^\dagger B K_j\right]$. So, we can write $\mathcal{T}^*(\cdot) = \sum_j K_j^\dagger(\cdot)K_j$. Note that it follows that the map $\mathcal{T}$ is trace-preserving if and only if the dual map satisfies $\mathcal{T}^*(\mathbb{1}) = \mathbb{1}$.

#### 1.1.4.2 The Choi-Jamiolkowski isomorphism

The Choi-Jamiolkowski isomorphism [51–53] reveals a duality between states and channels that will be of central importance in this thesis. Let $\mathcal{B}(\mathcal{H})$ be the space of bounded linear operators on Hilbert space $\mathcal{H}$. Then the isomorphism is encapsulated in the following theorem, where we follow the conventions used in Ref. [46].

**Theorem 1.3** (Choi-Jamiolkowski isomorphism [51, 52])**.** *Suppose $\mathcal{T}$ is a linear map from $\mathcal{B}(\mathcal{H}_B)$ to $\mathcal{B}(\mathcal{H}_A)$. Let $d_B = \dim(\mathcal{H}_B)$. Then there exists a unique operator $\Phi_{\mathcal{T}} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that*

$$\Phi_{\mathcal{T}} = (\mathcal{T} \otimes \mathbb{1})(|\Phi\rangle\langle\Phi|^{B|B'}), \quad \mathrm{Tr}[P\mathcal{T}(Q)] = d_B \mathrm{Tr}\left[\Phi_{\mathcal{T}} P \otimes Q^T\right] \tag{1.27}$$

*for any $P \in \mathcal{B}(\mathcal{H}_A)$ and $Q \in \mathcal{B}(\mathcal{H}_B)$, where*

$$|\Phi\rangle^{B|B'} = \frac{1}{\sqrt{d_B}} \sum_{j=0}^{d_B-1} |j\rangle^B \otimes |j\rangle^{B'}, \tag{1.28}$$

*and* $\{|j\rangle\}$ *is an orthonormal basis for* $\mathcal{H}_A$*, and the transpose in* (1.27) *is taken with respect to this basis. The map* $\mathcal{T}$ *and matrix* $\Phi_{\mathcal{T}}$ *have the following equivalences:*

1. *Trace-preservation:* $\mathcal{T}^*(\mathbb{1}) = \mathbb{1} \quad \Longleftrightarrow \quad \mathrm{Tr}_A(\Phi_{\mathcal{T}}) = \mathbb{1}/d\,.$
2. *Unitality:* $\mathcal{T}(\mathbb{1}) = \mathbb{1} \quad \Longleftrightarrow \quad \mathrm{Tr}_B(\Phi_{\mathcal{T}}) = \mathbb{1}/d.$
3. *Complete positivity:* $\mathcal{T}$ *is CP* $\quad \Longleftrightarrow \quad \Phi_{\mathcal{T}} \geq 0.$
4. *Hermiticity:* $\mathcal{T}(A^\dagger) = (\mathcal{T}(A))^\dagger \quad \Longleftrightarrow \quad \Phi_{\mathcal{T}}^\dagger = \Phi_{\mathcal{T}}.$
5. *Normalisation:* $\mathrm{Tr}[\Phi_{\mathcal{T}}] = \mathrm{Tr}(\mathcal{T}^*(\mathbb{1}))/d.$

*It follows that* $\Phi_{\mathcal{T}}$ *is a valid density operator if and only if* $\mathcal{T}$ *is a CPTP map. In this case we call* $\Phi_{\mathcal{E}}$ *the* ***Choi state*** *for the channel.*

For $n$-qubit systems, we take $\{|j\rangle\}_j$ to be the computational basis. We will usually consider channels $\mathcal{E}$ from an $n$-qubit space to an $n$-qubit space, so that the Choi state $\Phi_{\mathcal{E}}$ is a $2n$-qubit density operator. We will use the convention of labelling subsystems with superscripts when necessary (see equation (1.27)), so as to avoid confusion with other indices used in the subscript. Another useful property is as follows.

**Lemma 1.4** (Tensor product Choi state)**.** *The Choi state for a tensor product of two maps* $\mathcal{T}_1 \otimes \mathcal{T}_2$ *is a tensor product of the respective Choi states for each map,*

$$\Phi_{\mathcal{T}_1 \otimes \mathcal{T}_2} = \Phi_{\mathcal{T}_1} \otimes \Phi_{\mathcal{T}_2}. \tag{1.29}$$

This follows from the fact that the maximally entangled state for a $2n$-qubit system can be re-expressed as an $n$-fold tensor product of two-qubit Bell pairs. Let $B_k$ label the $k$-th qubit in the partition $B$. We use $|\Phi_n\rangle$ to denote a $2n$-qubit maximally entangled state,

$$|\Phi_n\rangle^{B|B'} = \frac{1}{2^{n/2}} \sum_{j=0}^{2^n-1} |j\rangle^B |j\rangle^{B'} \tag{1.30}$$

$$= \frac{1}{2^{n/2}} \sum_{j_1=0}^{1} \cdots \sum_{j_n=0}^{1} |j_1\rangle^{B_1} \ldots |j_n\rangle^{B_n} |j_1\rangle^{B'_1} \ldots |j_n\rangle^{B'_n} \tag{1.31}$$

$$= \bigotimes_{k=1}^{n} \left( \frac{|0\rangle^{B_k} |0\rangle^{B'_k} + |1\rangle^{B_k} |1\rangle^{B'_k}}{\sqrt{2}} \right) = \bigotimes_{k=1}^{n} |\Phi_1\rangle^{B_k|B'_k}. \tag{1.32}$$

This implies that for the $2n$-qubit state $|\Phi_n\rangle^{B|B'}$, for any $1 \leq m < n$, we can write $B = CD$ as a partitioning into an $m$-qubit subsystem $C$ and $(n-m)$-qubit subsystem $D$, and similarly for $B'$ and then write $|\Phi_n\rangle^{CD|C'D'} = |\Phi_m\rangle^{C|C'} \otimes |\Phi_{n-m}\rangle^{D|D'}$. Lemma 1.4 then follows.

### 1.1.5 Stabiliser states and the Clifford group

The stabiliser states will be central to this thesis. These are the set of states that can be efficiently represented, and therefore simulated, within the Gottesman-Knill framework [25, 31]. They are also central to the theory of quantum error correction [30]. The details of the stabiliser formalism can be found in standard texts on quantum computation [4, 30]. For the majority of this thesis, we will focus on the stabiliser formalism as defined for systems of $n$ qubits, i.e. for systems with Hilbert space of dimension $2^n$, and we will introduce this setting first. The stabiliser formalism is also defined for systems of qudits, with odd local dimension $d$, but we defer discussion of this to Chapter 2. For qubits, stabiliser groups are defined in relation to the Pauli group.

**Definition 1.5** (Stabiliser group)**.** *Any abelian subgroup $\mathcal{S}$ of the Pauli group $\mathcal{P}_n$ such that $-\mathbb{1}_n \notin \mathcal{S}$ is called a stabiliser group, or a stabiliser.*

Note that $n$-qubit stabiliser groups cannot contain Paulis with imaginary phase, as these would generate $-\mathbb{1}_n$. Recall that any signed Pauli operator $P \in \mathcal{P}_{n\pm}$ can be represented $P = (-1)^{a_0} T_{\mathbf{a}}$, for $a_0 \in \{0,1\}$ and $\mathbf{a} \in \mathbb{F}_2^{2n}$, so we can specify any element of a stabiliser group with a string of $2n+1$ classical bits, and a stabiliser group with $k$ independent generators can be fully specified by $k(2n+1)$ bits.

We say that a state $|\psi\rangle$ is stabilised by a stabiliser group $\mathcal{S}$ if $S|\psi\rangle = |\psi\rangle$ for all $S \in \mathcal{S}$ [30]. A stabiliser group with $m$ independent generators will have $2^m$ elements in total. As we add each independent generator to the stabiliser, in effect we halve the dimension of the subspace stabilised by the group. When $m < n$ for an $n$-qubit system, this subspace is non-trivial, having dimension $2^{n-m}$. In the context of quantum error correction, the elements of such subspaces form the codewords in stabiliser codes [30, 48]. For the purposes of this thesis, we are instead interested in stabilisers $\mathcal{S}$ with $m = n$ independent generators. In this case, there is a unique state $|\phi\rangle$ that is stabilised by $\mathcal{S}$. It is these states that we call stabiliser states. We can therefore specify a stabiliser state by giving an $n$-element generating set.

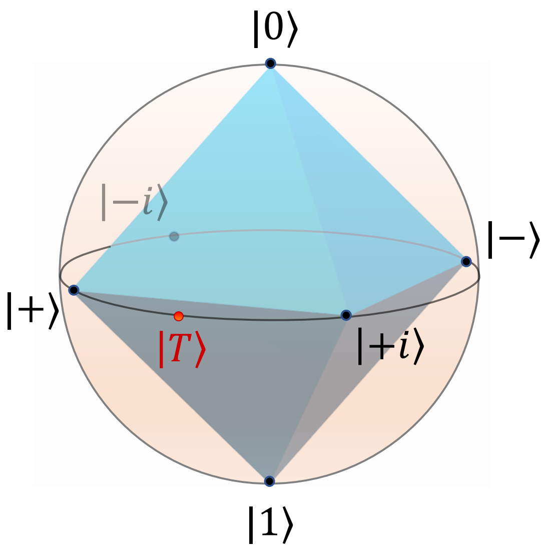


**Figure 1.2:** The single-qubit stabiliser polytope embedded within the Bloch sphere.

We can now justify our earlier claim that there are just six pure single-qubit stabiliser states, and that these are the eigenstates of the Pauli operators. The computational basis states $|0\rangle$ and $|1\rangle$ are stabilised by the subgroups $\{\mathbb{1}, Z\}$ and $\{\mathbb{1}, -Z\}$, respectively. Similarly $\{\mathbb{1}, \pm X\}$ stabilises $|\pm\rangle$ and $\{\mathbb{1}, \pm Y\}$ stabilises $|\pm_i\rangle$. There are no other possibilities for non-trivial stabiliser groups, since distinct single-qubit Pauli operators anticommute. In the Bloch sphere picture, the six pure stabiliser states form the vertices of an octahedron (Figure 1.2). Points on the facets or within the interior of the octahedron represent probabilistic mixtures of stabiliser states, and we call these *mixed stabiliser states*. This octahedron is the $n = 1$ instance of a $(4^n - 1)$-dimensional object called the $n$-qubit stabiliser polytope, which is the convex hull of the pure $n$-qubit stabiliser states,

$$\overline{\mathrm{STAB}}_n = \mathrm{conv}(\{|\phi\rangle\langle\phi| : |\phi\rangle \in \mathrm{STAB}_n\}). \tag{1.33}$$

The structure becomes much richer for multiple qubits, as it includes both tensor products and superpositions of the single-qubit states, and we must now specify $n$ generators to uniquely determine a state. For example, the state $|0\rangle \otimes |0\rangle$ is stabilised by the group with generators $\mathbb{1} \otimes Z$ and $Z \otimes \mathbb{1}$.

$$\langle \mathbb{1} \otimes Z, Z \otimes \mathbb{1} \rangle = \{\mathbb{1} \otimes \mathbb{1}, \mathbb{1} \otimes Z, Z \otimes \mathbb{1}, Z \otimes Z\} \tag{1.34}$$

The Bell states are also stabiliser states:

- The Bell states $|\Phi_\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$ are stabilised by $\langle Z\otimes Z, \pm X\otimes X\rangle$;

- The Bell states $|\Psi_\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$ are stabilised by $\langle -Z\otimes Z, \pm X\otimes X\rangle$.

Another set of operators fundamental to the study of the stabiliser formalism is the set of Clifford gates. These unitary operations preserve stabiliser structure.

**Definition 1.6** (Clifford group)**.** *The n-qubit Clifford group* $\mathrm{Cl}_n$ *is the normaliser of the Pauli group* $\mathcal{P}_n$ *in the unitary group. That is,*

$$\mathrm{Cl}_n = \Big\{ U \in U(2^n) : UPU^\dagger \in \mathcal{P}_n, \quad \forall P \in \mathcal{P}_n \Big\}. \tag{1.35}$$

Equivalently, the Clifford gates are those that map stabiliser states to stabiliser states. One can show that the Clifford group can be generated by the gate set

$$\mathcal{G} = \{H, S, CNOT\}, \tag{1.36}$$

comprising the single-qubit Hadamard $H$ and phase $S$ gates already defined, supplemented by two-qubit $CNOT$ (controlled-NOT) gates. The list of gates on the right-hand side of equation (1.36) should be understood as shorthand for $\{H_1, \dots, H_n\}$ where $H_j$ is the Hadamard on the $j$-th qubit, and similarly for the $S$ and $CNOT$ gates. The $H$ and $S$ gates transform single-qubit Pauli operators as

$$HXH^\dagger = Z, \quad HYH^\dagger = -Y, \quad HZH^\dagger = X, \tag{1.37}$$

$$SXS^\dagger = Y, \quad SYS^\dagger = -X, \quad SZS^\dagger = Z. \tag{1.38}$$

Note that the transformation of $Y$ can always be inferred from that of $X$ and $Z$, from the relation $ZX = iY$. Let $C_{j\to k}$ denote the $CNOT$ controlled on qubit $j$ and targeted on qubit $k$. For two qubits, in block form we have

$$C_{1\to 2} = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & X \end{pmatrix}. \tag{1.39}$$

More generally we define the *CNOT* by its action on $Z$ and $X$ operators.

$$C_{j\to k}Z_jC^{\dagger}_{j\to k} = Z_j, \quad C_{j\to k}Z_kC^{\dagger}_{j\to k} = Z_jZ_k \tag{1.40}$$

$$C_{j\to k}X_jC^{\dagger}_{j\to k} = X_jX_k \quad C_{j\to k}X_kC^{\dagger}_{j\to k} = X_k. \tag{1.41}$$

We will see later how these update rules lead to efficient simulation of Clifford circuits [25]. The gate set $\mathcal{G}$ generates a group that preserves the Pauli group, since

- if $UPU^{\dagger} \in \mathcal{P}_n$ and $VPV^{\dagger} \in \mathcal{P}_n$ and for all $P \in \mathcal{P}_n$, then $UVP(UV)^{\dagger} \in \mathcal{P}_n$, so we have closure.
- $H^2 = CNOT^2 = S^4 = \mathbb{1}$, so the identity and inverses can be found in the generated set.

For brevity, we have not proved that the set $\mathcal{G} = \{H, S, CNOT\}$ generates the whole Clifford group, but the interested reader can find a proof in Appendix A.1.

#### 1.1.5.1 Operator representation

Consider the $n$-qubit state $|\phi\rangle$ stabilised by the group $\mathcal{S}$ with independent generating set $\{g_1, g_2, \ldots, g_n\}$, i.e. $g_j |\phi\rangle = |\phi\rangle$ for all $j$. Notice that the operator $\Pi_j = (\mathbb{1}_n + g_j)/2$ is the projection onto the +1 eigenspace of the Pauli operator $g_j$, and we have $\Pi_j |\phi\rangle = |\phi_j\rangle$. Suppose we construct the product of all $n$ projectors $\Pi_j$:

$$\Pi_G = \prod_{j=1}^{n} \frac{\mathbb{1}_n + g_j}{2} = \frac{1}{2^n} \sum_{\mathbf{x}} g_1^{x_1} g_2^{x_2} \ldots g_n^{x_n}, \tag{1.42}$$

where the summation on the right-hand side is over all length-$n$ bit strings $\mathbf{x} = (x_1, \ldots, x_n)$. This summation is obtained by enumerating over all possible multiples of $g_j$ and $g_j^0 = \mathbb{1}_n$. Now, notice that any element of the stabiliser $S \in \mathcal{S}$ must have some decomposition as a product of generators $S = g_{j_1} g_{j_2} \ldots g_{j_m}$ for some number $m$. But recalling that a stabiliser group is always abelian, we are free to reorder this product so that we can always write $S = g_1^{y_1} g_2^{y_2} \ldots g_n^{y_n}$, where $\mathbf{y}$ is a list of integers of length $n$. Furthermore, since any Pauli operator is self-inverse, for any generator $g$ we have $g^k = g^{k \bmod 2}$, so that without loss of generality we can assume that $\mathbf{y}$ is a binary string. But the right-hand side of equation (1.42) enumerates over all possible such binary strings, therefore it is a summation over all elements of the stabiliser, $\Pi_G = 2^{-n} \sum_{S \in \mathcal{S}} S$.

We can now check some properties of $\Pi_G$. First, note that since all Pauli operators except the identity are traceless , we have,

$$\mathrm{Tr}[\Pi_G] = 2^{-n}\sum_S \mathrm{Tr}[S] = 2^{-n}\sum_S 2^n \delta_{S,\mathbb{1}_n} = 1 \tag{1.43}$$

since exactly one element of $\mathcal{S}$ is the identity. Next, $\Pi_G^\dagger = \Pi_G$ follows from the Hermiticity of $S \in \mathcal{S}$. Finally, $\Pi_G$ is positive semidefinite since it is a projector. Therefore $\Pi_G$ is a normalised density operator, and furthermore since $\Pi_G^2 = \Pi_G$, we have $\mathrm{Tr}\left[\Pi_G^2\right] = 1$. This tells us that it represents a pure state, so there exists some $|\psi\rangle$ such that $\Pi_G = |\psi\rangle\langle\psi|$. But recall that $\mathcal{S}$ stabilises the state $|\phi\rangle$. Therefore $|\langle\phi|\psi\rangle|^2 = \langle\phi|\Pi_G|\phi\rangle = \langle\phi|\phi\rangle = 1$. Thus $|\psi\rangle = |\phi\rangle$ is the unique state stabilised by $\mathcal{S}$, and we can write the corresponding density operator as,

$$|\phi\rangle\langle\phi| = \prod_{j=1}^{n} \frac{\mathbb{1}_n + g_j}{2} = \frac{1}{2^n}\sum_{S\in\mathcal{S}} S \tag{1.44}$$

This relation is useful in proving many results relating to stabiliser states. For example, it is manifest from this equation that the transformation of one stabiliser state to another by a Clifford circuit $U$ is fully specified by the transformation of the generators,

$$U|\phi\rangle\langle\phi|U^\dagger = \prod_{j=1}^{n} \frac{UU^\dagger + Ug_jU^\dagger}{2} = \prod_{j=1}^{n} \frac{\mathbb{1}_n + g'_j}{2} = |\phi'\rangle\langle\phi'|, \tag{1.45}$$

where $g'_j = Ug_jU^\dagger$, and $\{g'_1, g'_2, \ldots, g'_n\}$ generates the stabiliser of $|\phi'\rangle = U|\phi\rangle$.

### 1.1.5.2 Stabiliser tableaux

We saw above that $2n+1$ classical bits are required to specify each generator, so that $\mathcal{O}(n^2)$ bits are needed to specify the stabiliser group. This gives us a much more efficient way to represent a given stabiliser state, as compared to having to write down $2^n$ complex amplitudes. It will be useful to represent the generators of the stabiliser in so-called stabiliser tableau form [31]. Recalling that each Pauli operator can be specified by a binary vector $\mathbf{a}_j = (\mathbf{z}_j, \mathbf{x}_j, s_j)$, where $P_j = (-1)^{s_j}(-i)^{\mathbf{z}_j\cdot\mathbf{x}_j}\prod_{k=1}^{n} Z_k^{z_{j,k}} X_k^{x_{j,k}}$, the tableau for an $n$-qubit stabiliser state is written:

$$\left(\begin{array}{c|c|c} \mathbf{z}_1^T & \mathbf{x}_1^T & s_1 \\ \vdots & \vdots & \vdots \\ \mathbf{z}_n^T & \mathbf{x}_n^T & s_n \end{array}\right) \tag{1.46}$$

For example, the stabiliser state $|000\rangle$ has stabiliser generated by $Z_1$, $Z_2$ and $Z_3$, so has tableaux form:

$$\left(\begin{array}{ccc|ccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right). \tag{1.47}$$

Similarly $|001\rangle$ has stabiliser $\langle Z_1, Z_2, -Z_3\rangle$, so can be specified as:

$$\left(\begin{array}{ccc|ccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right). \tag{1.48}$$

Note that these representations are not unique for a given state, as we have freedom in choosing the generating set. Given any pair of independent generators $g_j$ and $g_k$, we can obtain a new generating set by making the substitution $g_j \to g'_j = g_j g_k$. In the tableaux picture this is equivalent to adding two rows modulo 2, up to some additional arithmetic to compute the sign bit $s'_j$ in the last column.

### 1.1.5.3 Affine space representation

We next introduce one more representation of stabiliser states, due to Dehaene and De Moor, that will be instrumental in proving some of our results in Chapter 4.

**Theorem 1.7** (Affine space representation of stabiliser states [54, 55])**.** *We can write any pure n-qubit stabiliser state in the form:*

$$|\mathcal{K}, q, d\rangle = \frac{1}{|\mathcal{K}|^{1/2}} \sum_{\mathbf{x}\in\mathcal{K}} i^{\mathbf{d}^T\mathbf{x}} (-1)^{q(x)} |x\rangle, \tag{1.49}$$

*where $\mathbf{x} \in \mathbb{F}_2^n$ is a binary vector, $\mathcal{K} \subseteq \mathbb{F}_2^n$ is an affine space, $\mathbf{d}$ is some fixed binary vector, and $q(x)$ is a quadratic form,*

$$q(x) = \mathbf{x}^T Q\mathbf{x} + \lambda^T \mathbf{x}, \tag{1.50}$$

*where $Q$ is a binary, strictly upper triangular matrix, and $\lambda$ is a vector. Conversely, any state written in this form must be a stabiliser state.*

An affine space $\mathcal{K}$ is obtained by adding some constant shift vector $\mathbf{h}$ to all elements of a linear subspace $\mathcal{L} \subseteq \mathbb{F}_2^n$. For example, suppose we take the $n = 2$ linear subspace

$\mathcal{L} = \{(0,0)^T, (0,1)^T\}$. Then we can get a new affine space by adding modulo 2 some shift vector, eg. $\mathbf{h} = (1,1)^T$, giving us $\mathcal{K} = \mathcal{L} + \mathbf{h} = \{(1,0)^T, (1,1)^T\}$. Now suppose we set the following parameters:

$$Q = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.51}$$

These are all valid choices according to Theorem 1.7, so they fully specify a stabiliser state:

$$|\mathcal{K}, q, \mathbf{d}\rangle = \frac{1}{\sqrt{2}}\left(i^0(-1)^{0\oplus 0}|10\rangle + i^1(-1)^{1\oplus 1}|11\rangle\right) \tag{1.52}$$

$$= \frac{1}{\sqrt{2}}(|10\rangle + i|11\rangle) = |1\rangle \otimes |+_i\rangle, \tag{1.53}$$

which in this case happens to be a tensor product of single-qubit stabiliser states.

### 1.1.6 Types of classical simulation

We will shortly discuss a famous result called the Gottesman-Knill theorem[25, 31], which establishes that stabiliser circuits can be efficiently classically simulated. Before doing so it is worth reflecting on what it means to classically simulate a quantum circuit. In this thesis, we primarily deal with the following circuit-based model of quantum computation: (1) a system of $n$ qubits is initialised in a standard state $|\phi_0\rangle$; (2) the system undergoes a sequence of quantum operations, which can be mediated by a classical control computer, and can be represented by a quantum channel $\mathcal{E}$; (3) some subset of $w$ qubits are measured in the computational basis and the resulting binary string is relayed back to the classical device. The quantum computer can therefore be viewed as a device that draws a string $\mathbf{x}$ randomly from a distribution $p(\mathbf{x})$, which we call the quantum output distribution, and which is determined by the Born rule:

$$p(\mathbf{x}) = \mathrm{Tr}[(|\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w})\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]. \tag{1.54}$$

So, a quantum computer does *not* in general have the ability to exactly compute any given Born rule probability $p(\mathbf{x})$, or indeed compute a general expectation value $\langle A\rangle$ for some observable $A$, in any single run of the circuit. If we cannot expect this of a quantum computer, we should not *require* it from a classical simulation algorithm. In this sense, a classical simulator that has the ability to efficiently draw a bit string $\mathbf{x}$ from $p(\mathbf{x})$ would have the same power as the quantum computer. This is sometimes called *weak* simulation in the literature [56], and is thought to be efficient only for restricted classes of circuit, otherwise super-polynomial quantum advantage would be an impossibility. Weak simulation can either be exact or approximate. For this type of simulation to be meaningful, the classical distribution should be quantifiably close to the quantum output distribution. In this thesis we use the $\ell_1$-norm, and we say that two distributions $p_1$ and $p_2$ are $\delta$-close if

$$\|p_1 - p_2\|_1 = \sum_{\mathbf{x}} |p_1(\mathbf{x}) - p_2(\mathbf{x})| \leq \delta. \tag{1.55}$$

On the other hand, certain classical algorithms *do* have the ability to compute or estimate Born rule probabilities or other expectation values when the quantum device is suitably restricted. This is often called *strong* simulation [56]. For example, as we shall see shortly, given a pure stabiliser state $|\phi\rangle$ and Clifford operation $U$, it is possible to efficiently compute both $\Pr(\mathbf{x}) = |\langle \mathbf{x}|U|\phi\rangle|^2$ and $\langle \phi|U^\dagger P U|\phi\rangle$ exactly, where $|\mathbf{x}\rangle$ is any computational basis state, and $P$ is any Pauli operator [25, 31]. In other cases estimation of observable expectation values is possible inefficiently and/or approximately. This thesis will consider two types of approximation error for direct expectation value estimation. We define these here for Born rule probability estimation, but we will use analogous definitions for more general expectation values. Given a projector $\Pi$, quantum channel $\mathcal{E}$, and initial state $\rho$, we say that a random variable $\widehat{P}$ output from a classical algorithm has additive error $\delta > 0$, where usually $\delta \ll 1$, if it satisfies

$$|\widehat{P} - \mathrm{Tr}[\Pi\mathcal{E}(\rho)]| \leq \delta. \tag{1.56}$$

We say we can estimate up to multiplicative error $\varepsilon > 0$ if the variable satisfies

$$(1-\varepsilon)\,\mathrm{Tr}[\Pi\mathcal{E}(\rho)] \leq \widehat{P} \leq (1+\varepsilon)\,\mathrm{Tr}[\Pi\mathcal{E}(\rho)]. \tag{1.57}$$

We will usually consider classical algorithms where the additive error $\delta$ or multiplicative error $\varepsilon$ can be made arbitrarily small subject to additional runtime overhead. Note that when Born rule probabilities can be computed exactly or up to multiplicative error, we can construct a bit-string sampling simulator by computing a chain of conditional probabilities and sampling the string **x** one bit at a time. We discuss how this is done in more detail in later chapters. To prevent ambiguity, we will usually avoid the terms weak and strong in this thesis, preferring instead to distinguish between expectation value estimation and simulated sampling from the output distribution, and we will be precise about whether a given method is exact or approximate, and which type of error is assumed.

## 1.2 The Gottesman-Knill theorem

There are multiple routes to proving the Gottesman-Knill theorem, the key result that establishes the efficiency of classical simulation of stabiliser circuits. We will work in the stabiliser tableau picture.

**Theorem 1.8** (Gottesman-Knill Theorem [25]). *Suppose that a quantum circuit involves at most $n$ qubits and can be described as a sequence of $m$ operations from the following list: (i) computational basis state preparations; (ii) the generators of the Clifford group: H, S, CNOT; (iii) measurement of Pauli observables; (iv) efficient classical processing. Then the circuit can be simulated by a classical computer in* $\text{poly}(n,m)$ *time and space.*

Before proving the theorem, we first make concrete what is meant by efficient classical processing. Any quantum computing device must be interfaced with via a classical control computer, which sends instructions to the quantum hardware telling it which physical operation to perform at each step. We make the following restrictions on the computational power of this classical device. We assume a classical register of size $\text{poly}(n,m)$. The classical register can store results of any Pauli measurements carried out by the quantum device, and the register can be used to decide which instructions to send to the quantum processor at any later step. This is referred to as conditioning or adaptivity in the literature. Any kind of classical algorithm can be used to decide instructions, provided that it takes no more than $\text{poly}(n,m)$ time at each step. The control computer can therefore introduce any kind of classical randomness, provided it can do so efficiently.

*Proof.* The proof will be constructive; for each operation listed in the theorem, we need to give an efficient procedure for updating the stabiliser tableau. Here we largely follow an argument due to Aaronson and Gottesman [31], using a stabiliser tableau augmented with an additional $n$ rows that represent generators of the *destabiliser*. This is a subgroup that generates the entire Pauli group when combined with the stabiliser generators. While the inclusion of these additional rows doubles the number of classical bits needed to store the state during the simulation, we will see that it speeds up the subroutine for simulating Pauli measurements from $\mathcal{O}(n^3)$ time to $\mathcal{O}(n^2)$. This difference can be significant when the number of qubits is in the thousands. With destabiliser included, the tableau takes the form,

$$\left(\begin{array}{c|c|c} \mathbf{z}_1^T & \mathbf{x}_1^T & s_1 \\ \vdots & \vdots & \vdots \\ \mathbf{z}_n^T & \mathbf{x}_n^T & s_n \\ \hline \mathbf{z}_{n+1}^T & \mathbf{x}_{n+1}^T & s_{n+1} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_{2n}^T & \mathbf{x}_{2n}^T & s_{2n} \end{array}\right). \tag{1.58}$$

where the first $n$ rows specify the stabiliser, and the last $n$ give the destabiliser. We assume that the state will always be represented by a single tableau, so the algorithm requires $2n(2n+2) = \mathcal{O}(n^2)$ bits of classical memory to store the state. We will denote $z_{j,a}$ to mean the $a$-th element of vector $\mathbf{z}_i$, and so on. We also assume the tableau always starts in a standard initial form

$$\left(\begin{array}{c|c|c} \mathbb{1}_{n\times n} & 0 & 0 \\ \hline 0 & \mathbb{1}_{n\times n} & 0 \end{array}\right). \tag{1.59}$$

where $\mathbb{1}_{n\times n}$ means the $n \times n$ identity matrix. Before describing the algorithm, we note several properties of the tableau that can be checked by inspection, and that remain invariant even after the updates described below: (1) all elements of the destabiliser commute; (2) the stabiliser generator in row $j$ anticommutes with the destabiliser generator in row $n+j$, but (3) commutes with all other destabiliser generators.

**(i) Computational basis state preparations.** This step is straightforward as the stabiliser group for any $n$-qubit computational basis state $|\mathbf{a}\rangle$ is comprised solely of $Z$-operators, differing only in the phase on each generator:

$$\mathcal{S} = \langle (-1)^{a_1} Z_1, \ldots, (-1)^{a_n} Z_n \rangle. \tag{1.60}$$

For the $|0^n\rangle$ state, the stabiliser group is generated by the single-qubit unsigned $Z_j$ op-

| Initial | | Final | |
|---|---|---|---|
| $z_j$ | $x_j$ | $z_j$ | $x_j$ |
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 |

**Table 1.1:** Truth table for phase gate update in stabiliser tableau method [31].

erators, for $j \in \{1,2,\ldots,n\}$. It follows that the destabiliser comprises the $n$ single-qubit $X_j$ operators. The standard initial tableau (1.59) is therefore already in the correct form to represent $|0^n\rangle$. To generalise to specific computational basis states, we merely need to set $s_j$ to the appropriate value, as per equation (1.60), so if we assume that the tableau is always initialised in a standard state (1.59), this takes $\mathcal{O}(n)$ steps.

**(ii) Clifford gates, $H$, $S$ and $CNOT$.** For unitary operations, we treat both stabiliser and destabiliser generators alike, as this will ensure that combined they generate the whole Pauli group. First consider the Hadamard, $H$. Recall that this gate exchanges the $X$ and $Z$ operators and changes the sign of $Y$. So if the Hadamard acts on the $j$-th qubit, in each row $i$ we swap $z_{i,j}$ with $x_{i,j}$. Then, the sign changes only if $x_{i,j} = z_{i,j} = 1$, so we update $s_i \leftarrow s_i \oplus x_{i,j}z_{i,j}$. Since the entry for each of the $2n$ rows must be updated, this requires $\mathcal{O}(n)$ steps.

Next consider the phase gate $S$. This gate maps $X$ to $Y$, $Y$ to $-X$, and leaves $Z$ unchanged. Suppose $S$ acts on qubit $j$. Again the sign changes only if initially $x_{i,j} = z_{i,j} = 1$, so the $s_i$ is updated in the same way as for $H$. The truth table for this update is shown in Table 1.1. Therefore we simply need to update $z_j$ to $z_j \oplus x_j$.

Finally consider the CNOT $C_{j\to k}$ from qubit $j$ to qubit $k$. This leaves $Z_j$ unchanged, but exchanges $Z_k$ with $Z_jZ_k$; and leaves $X_k$ unchanged but exchanges $X_j$ with $X_jX_k$. This immediately determines how all the $X$ and $Z$ entries will update, but we need to take care of any sign changes. In fact one can check that the sign changes only in the exchange $Y_jY_k \leftrightarrow X_jZ_k$. In terms of the tableau entries, this means that in row $i$, the sign changes only when

$$(x_{i,j} = z_{i,k} = 1) \quad \text{AND} \quad \big(x_{i,k} = z_{i,j} = 0 \quad \text{OR} \quad x_{i,k} = z_{i,j} = 1\big). \tag{1.61}$$

In arithmetic, this means we set:

$$s_i \leftarrow s_i \oplus x_{i,j} \cdot z_{i,k} \cdot (x_{i,k} \oplus z_{i,j} \oplus 1). \tag{1.62}$$

Then we can update the $X$ and $Z$ entries by setting $z_{i,j} \leftarrow z_{i,j} \oplus z_{i,k}$ and $x_{i,k} \leftarrow x_{i,k} \oplus x_{i,j}$. Therefore the CNOT updates takes $\mathcal{O}(n)$ steps.

**(iii) Pauli measurement.** There are three possible scenarios when simulating measurement of a Pauli observable $Q$ given a state $|\phi\rangle$ represented by stabiliser group $\mathcal{S}$. Either $Q \in \mathcal{S}$, $-Q \in \mathcal{S}$, or neither is in $\mathcal{S}$. The first task of the simulator is to efficiently determine which case holds. For simplicity, following Aaronson and Gottesman, we will assume that $Q = Z_j$ for some $j \in \{1, \ldots, n\}$ [31], but the procedure can be generalised to any Pauli operator. Recall that $\mathcal{S}$ is a maximal Abelian subgroup of the Pauli group such that $-\mathbb{1} \notin \mathcal{S}$. If $\pm Q \notin \mathcal{S}$ there must exist a generator of $\mathcal{S}$ that anticommutes with $Q$; otherwise we could construct a larger subgroup by adding $Q$ to the list of generators, which is a contradiction since $\mathcal{S}$ is maximal. Therefore $\pm Z_j \in \mathcal{S}$ if and only if all generators commute with $Z_j$, so we just need to check whether any row $i \in \{1, \ldots, n\}$ has $x_{i,j} = 1$, which can be done in $\mathcal{O}(n)$ steps. If such an $i$ exists then the corresponding stabiliser generator anticommutes with $Z_j$, so $\pm Z_j \notin \mathcal{S}$. Otherwise, either $Z_j \in \mathcal{S}$ or $-Z_j \in \mathcal{S}$. We will see that $\pm Z_j \in \mathcal{S}$ implies a deterministic measurement result, whereas if $\pm Z_j \notin \mathcal{S}$ the measurement outcome will be random. We deal with each case in turn.

**Deterministic outcome,** $\pm Z_j \in \mathcal{S}$. When $Z_j \in \mathcal{S}$, then $Z_j |\phi\rangle = |\phi\rangle$, so the measurement outcome is always "+1". If $-Z_j \in \mathcal{S}$ then the outcome is "−1". In each of these cases, we do not need to update the state, we only need determine the sign of $Z_j$ in $\mathcal{S}$. This is non-trivial, as $\pm Z_j$ is not necessarily one of the generators encoded in the tableau, so we cannot simply read off the sign bit from the corresponding row. Naively, one could use Gaussian elimination to put the stabiliser generator tableau in a standard form so that $\pm Z_j$ appears in the tableau, but in practice this would take $\mathcal{O}(n^3)$ steps. Instead the destabiliser tableau can be used for a more efficient procedure [31].

We first need a subroutine to simulate group operations. Aaronson and Gottesman [31] defined ROWSUM$(h,i)$ which simulates a change in generating set (but generating the same group) by replacing the generator $g_h$ represented by row $h$ with the generator $g_i g_h$. For the $X$ and $Z$ entries this simply involves making the update:

$$\mathbf{x}_h \leftarrow \mathbf{x}_i + \mathbf{x}_h, \quad \mathbf{z}_h \leftarrow \mathbf{z}_i + \mathbf{z}_h. \tag{1.63}$$

To update the entry $s_h$ representing the phase, we can use the relation (1.12), so that the new phase will be given by,

$$(-1)^{(s_h+s_i)} i^{(\mathbf{z}_h \cdot \mathbf{x}_i - \mathbf{x}_h \cdot \mathbf{z}_i)} = i^{(2s_h + 2s_i + \mathbf{z}_h \cdot \mathbf{x}_i - \mathbf{x}_h \cdot \mathbf{z}_i)}, \tag{1.64}$$

so that

$$s_h \leftarrow s_h + s_i + \frac{\mathbf{z}_h \cdot \mathbf{x}_i - \mathbf{x}_h \cdot \mathbf{z}_i}{2} \pmod 2. \tag{1.65}$$

This arithmetic can be evaluated in time $\mathcal{O}(n)$.

Let $\mathbf{r}_h$ be the $h$-th row of the stabiliser tableau, representing generator $g_h$, excluding the sign bit, and let $\mathbf{r}_Z$ represent $\pm Z_j$. Then since some product of generators generates $\pm Z_j$, there is some linear combination of rows that generates $\mathbf{r}_Z$:

$$\sum_{h=1}^{n} c_h \mathbf{r}_h = \mathbf{r}_Z, \quad c_h \in \{0,1\}. \tag{1.66}$$

We need to determine which $c_h$ are non-zero, as then by using ROWSUM we can determine the sign bit for $\mathbf{r}_Z$. Recall that $\langle \mathbf{r}_h, \mathbf{r}_i \rangle = 0$ if and only if the $g_h$ and $g_i$ commute, otherwise they anticommute, and that a stabiliser generator $g_h$ anticommutes with a destabiliser generator $g_{i+n}$ if and only $i = h$. Then, since the symplectic inner product is linear,

$$\langle \mathbf{r}_{i+n}, \mathbf{r}_Z \rangle = \sum_{h=1}^{n} c_h \langle \mathbf{r}_{i+n}, \mathbf{r}_h \rangle = c_i. \tag{1.67}$$

So $c_i = 1$ if and only if $Z_j$ anticommutes with $g_{i+n}$, which occurs only if $x_{i+n,j} = 1$. Therefore if we use ROWSUM to combine all stabiliser rows $i$ such that $x_{i+n,j} = 1$, then read off the sign bit, we can determine the outcome of the measurement. This takes time $\mathcal{O}(n^2)$.

**Random outcome, $\pm Z_j \notin \mathcal{S}$.** We argued that when neither $Z_j$ nor $-Z_j$ is in the stabiliser group, there exists some $P \in \mathcal{S}$ that anticommutes with $Z_j$. Then we have

$$\langle \phi | Z_j | \phi \rangle = \langle \phi | Z_j P | \phi \rangle = - \langle \phi | P Z_j | \phi \rangle = - \langle \phi | Z_j | \phi \rangle, \tag{1.68}$$

where we used the fact that $|\phi\rangle = P|\phi\rangle$, and $P$ is Hermitian. Equation (1.68) can hold only if $\langle\psi|Z_j|\psi\rangle = 0$, so the outcome of the measurement will be either "+1" or "-1" with equal probability. So, once the algorithm has checked that $\pm Z_j \notin \mathcal{S}$, an outcome can simply be chosen randomly by flipping a fair coin. The state should be projected onto the $\pm 1$ eigenspace depending on the outcome of this coin toss. We will now show it remains a stabiliser state, and we need only replace one of the generators by $\pm Z_j$, given a suitable choice of generating set. Let $Q = Z_j$ if the outcome was "+1", and $Q = -Z_j$ otherwise. Let $\mathcal{G} = \{g_1, g_2, \dots, g_n\}$ be the generating set for $\mathcal{S}$ represented by the current stabiliser tableau, and let $\mathbf{r}_i$ be the row of the tableau corresponding to generator $g_i$. We first show that the tableau can efficiently be put in a form where there is a single generator $g_a$ that anticommutes with $Q$, and all other generators commute. Let $\mathbf{r}_a$ be any row of the tableau with $x_{a,j} = 1$, so that $g_a$ anticommutes with $Z_j$. Then for all other rows $\mathbf{r}_h$, $h \neq a$, such that $x_{h,j} = 1$, use ROWSUM to update $\mathbf{r}_h$ to represent a new generator $g'_h = g_h g_a$. Then,

$$g'_h Q = g_h g_a Q = -g_h Q g_a = Q g_h g_a = Q g'_h, \tag{1.69}$$

so $g'_h$ now commutes with $Q$. This must also be done for all elements of the destabiliser to ensure consistency of the tableau. Therefore in $\mathcal{O}(n^2)$ steps we obtain a tableau representing a change to generating set $\mathcal{G}' = \{g'_1, g'_2, \dots, g'_n\}$ with exactly one element $g'_a = g_a$ that anticommutes with $Q$.

The set $\mathcal{G}'$ still represents the stabiliser $\mathcal{S}$, but the update to the post-measurement state will now be very straightforward. Recall that any pure stabiliser state can be expressed as a product of the projection operators for each generator, so we can write

$$|\phi\rangle\langle\phi| = \frac{1}{2^n}(\mathbb{1} + g'_a)\prod_{j\neq a}(\mathbb{1} + g'_j). \tag{1.70}$$

Consider the projector onto the eigenspace corresponding to the measurement outcome, $\Pi_Q = (\mathbb{1} + Q)/2$. This commutes with all projectors $(\mathbb{1} + g'_j)/2$ except $(\mathbb{1} + g'_a)/2$, for which we obtain,

$$\left(\frac{\mathbb{1}+Q}{2}\right)\left(\frac{\mathbb{1}+g'_a}{2}\right)\left(\frac{\mathbb{1}+Q}{2}\right) = \frac{1}{2}\frac{\mathbb{1}+Q}{2}, \tag{1.71}$$

so that after normalising, the updated state is

$$|\psi'\rangle\langle\psi'| = 2\Pi_Q |\psi\rangle\langle\psi| \Pi_Q = \frac{1}{2^n}(\mathbb{1} \pm Q)\prod_{j=2}^{n}(\mathbb{1} + g_j). \tag{1.72}$$

Therefore we simply need to update the stabiliser tableau to replace $g'_a$ with $Q$ in the generating set, and update the $a$-th row in the destabiliser tableau so that it anticommutes with Q. A suitable choice is to simply set $\mathbf{r}_{a+n} = \mathbf{r}_a$, and then complete the update of the stabiliser tableau by setting the $a$-th row to the bit string for $Q = \pm Z_j$ (zeroes in all entries except $z_{a,j} = 1$, and sign bit $s_a$ such that $Q = (-1)^{s_a} Z_j$). The total procedure to update the tableau after random measurement outcome therefore takes $\mathcal{O}\left(n^2\right)$ steps.

**(iv) Classical processing.** By assumption any classical processing in between quantum operations requires no more than poly$(n,m)$ time and space. Having established that each of the listed quantum processes can be simulated using stabiliser tableaux in polynomial time, it follows that combining this with classical processing steps remains efficient. □

The proof of the Gottesman-Knill theorem shows that we can efficiently simulate sampling from the output distribution of a stabiliser circuit. Moreover, there is no $\ell_1$-norm error in this sampling algorithm; since we know that the outcomes of Pauli measurements on pure stabiliser states are either deterministic or occur with equal probability, we sample from exactly the same distribution as the quantum circuit. The proof can also be adapted to show that exact estimation is possible in certain settings. For example, it is clear from the argument above that we can exactly compute the expectation value of any Pauli operator following a sequence of Clifford gates. In Ref. [31] it is also shown that the magnitude of any inner product $|\langle\phi|\psi\rangle|$ for a pair of stabiliser states $|\phi\rangle, |\psi\rangle \in \mathrm{STAB}_n$ can be computed in time $\mathcal{O}(n^3)$; we omit the details here as inner products will be discussed in more detail in Section 3.5 of Chapter 3. One can also adapt the proof to show that additional subsystems in stabiliser states can be appended to the tableau efficiently, and that when a bipartite stabiliser state is known to be in a product state, tableaux for the tensor factors can be obtained efficiently [57]. Finally, Aaronson and Gottesman also generalised the method to simulate projectors onto stabiliser codes [31].

### 1.2.1 Phase-sensitive Clifford simulators

The stabiliser tableau representation tracks the stabiliser group corresponding to a pure quantum state, so that strictly speaking it is able to track the evolution of the state vector only up to a global phase. Clearly this is unimportant if one is only interested in the measurement statistics for a stabiliser state, but we will see later that it can be useful to track this phase when a Clifford simulator is used as a subroutine for a more general algorithm. For example, given a vector $|\psi\rangle = \sum_k c_k |\mathcal{S}_k\rangle$ where $|\mathcal{S}_k\rangle$ is a stabiliser state with stabiliser group $\mathcal{S}_k$, if we want to make the transformation $U|\psi\rangle = \sum c_k U|\phi_k\rangle$, where $U$ is Clifford, using a stabiliser update method, it is vital that we correctly compute the relative phase $\alpha_k$ for each term $U|\mathcal{S}_k\rangle = \alpha_k |\mathcal{S}_k\rangle$. Indeed such a phase-sensitive Clifford subroutine will be essential for some of the new algorithms we present in Chapter 5. Fortunately several subroutines of this type have appeared in the literature [3, 58–60]. Typically these are based on representing the Clifford preparation circuit for the target state in canonical form [3, 31, 58, 61, 62]. For concreteness, we will assume that the phase-sensitive subroutines employed later in the thesis use the CH-form described by Bravyi et al. in [3]. Since the details are rather technical, and we have already proved the Gottesman-Knill theorem using the tableau method, here we will simply state the capabilities of the simulator. We refer the interested reader to Appendix A.2 for a brief overview of canonical forms and a description of the CH data format, and to Section 4.1 of Ref. [3] for a comprehensive technical description.

The CH-form relies on the fact that any stabiliser state vector can be represented $|\phi\rangle = \omega U_C U_H |\mathbf{s}\rangle$, where $\omega$ is a complex number, $U_H$ is a layer of Hadamard gates, and $U_C$ is a block of "control-type" Clifford gates satisfying $U_C|0^n\rangle = |0^n\rangle$. In Ref. [3] it is shown that this form can be fully specified by a data format requiring $3n^2 + 4n + \mathcal{O}(1)$ classical bits. Given this data format, the following simulation tasks can be performed in time at most $\mathcal{O}(n^2)$:

1. Compute the CH-form for $\Gamma|\phi\rangle$, where $\Gamma$ is either an elementary Clifford gate from $\{S, CZ, CNOT, H\}$, or a Pauli projector $\Gamma = (\mathbb{1}_n + P)/2$, for $P \in \mathcal{P}_{n\pm}$.

2. Evaluate the inner product $\langle \mathbf{x}|\phi\rangle$.

3. Draw a string $\mathbf{x}$ from the distribution $P(\mathbf{x}) = |\langle \mathbf{x}|\phi\rangle|^2$.

## 1.3 Completely stabiliser-preserving operations

The Gottesman-Knill theorem establishes the classical simulability of a class of quantum circuits defined operationally, namely those composed from (i) computational basis state preparation, (ii) Clifford gates, (iii) Pauli measurements and (iv) classical processing. We call CPTP maps composed from these operations the *stabiliser operations*, SO. We denote by $\mathrm{SO}_n$ this class restricted to $n$-qubit operations. The proof of the Gottesman-Knill theorem depends on the fact that these operations preserve stabiliser structure. Computational basis states are stabiliser states, Clifford gates and Pauli measurements map stabiliser states to stabiliser states, and the addition of classical adaptivity and randomness can be modelled by channels that map stabiliser states to convex mixtures of stabiliser states. In short, any circuit in $\mathrm{SO}_n$ preserves the $n$-qubit stabiliser polytope.

Later in this thesis we will be concerned with the resource theory of magic, and we will study *magic monotones*, which quantify the resourcefulness of magic states and operations. From this point of view, the stabiliser operations are considered free operations, since they cannot generate magic resource. In resource theories, it is useful to have an axiomatic definition for the class of free operations. In Ref. [1], we defined the *completely stabiliser-preserving* channels as follows.

**Definition 1.9** (Completely stabiliser-preserving channels [1]). *Let $\mathcal{E}$ be an n-qubit CPTP map. We call $\mathcal{E}$ a **completely stabiliser-preserving channel**, if it satisfies*

$$(\mathcal{E}\otimes\mathbb{1}_n)(\rho)\in\overline{\mathrm{STAB}}_{2n},\quad \forall\rho\in\overline{\mathrm{STAB}}_{2n}. \tag{1.73}$$

*We denote the set of all n-qubit completely stabiliser-preserving channels by* $\mathrm{SP}_{n,n}$.

The notation $\mathrm{SP}_{n,n}$ is motivated by the following more general definition:

**Definition 1.10** (Stabiliser-preserving channels [1]). *Let $\mathcal{E}$ be CPTP map that acts non-trivially on n qubits. We say that $\mathcal{E}\in\mathrm{SP}_{n,k}$ if it satisfies*

$$(\mathcal{E}\otimes\mathbb{1}_k)(\rho)\in\overline{\mathrm{STAB}}_{n+k},\quad \forall\rho\in\overline{\mathrm{STAB}}_{n+k}, \tag{1.74}$$

*that is, if the channel remains stabiliser-preserving when applied to an n-qubit subsystem of an $n+k$-qubit system.*

This distinction is significant because there exist channels that are in $\mathrm{SP}_{n,n-k}$ but not in $\mathrm{SP}_{n,n}$, for $k \leq n$. That is, there are $n$-qubit maps that are stabiliser-preserving when acting on part of a $2n-k$-qubit system, but are able to generate magic resources when applied to a $2n$-qubit system. On the other hand, it can be shown that $\mathrm{SP}_{n,n} \equiv \mathrm{SP}_{n,n+k}$ for all $k \geq 0$ [1]. This justifies Definition 1.9 as the appropriate definition for the largest class of channels that can be considered free, in that they do not generate magic resources on any size of system. For the interested reader, further discussion justifying this choice can be found in Appendix B. In Ref. [1] we also showed that there is a convenient characterisation of completely stabiliser-preserving maps in terms of the Choi state (see Section 1.1.4.2).

**Theorem 1.11** (Choi states for completely stabiliser-preserving maps [1].). *Suppose $\mathcal{T}$ is a linear n-qubit map. Then $\mathcal{T}$ is a completely stabiliser-preserving CP map if and only if its Choi state is a stabiliser state, up to normalisation,*

$$\frac{(\mathcal{T}\otimes\mathbb{1}_n)\rho}{\mathrm{Tr}[(\mathcal{T}\otimes\mathbb{1}_n)\rho]} \in \overline{\mathrm{STAB}}_{2n}, \quad \forall\rho \in \overline{\mathrm{STAB}}_{2n} \iff \frac{\Phi_{\mathcal{T}}}{\mathrm{Tr}[\Phi_{\mathcal{T}}]} \in \overline{\mathrm{STAB}}_{2n} \tag{1.75}$$

*Moreover, $\mathcal{T}$ is a completely stabiliser-preserving CPTP map, $\mathcal{T} \in \mathrm{SP}_{n,n}$ if its Choi state is a stabiliser state, and* $\mathrm{Tr}_A[\Phi_{\mathcal{T}}^{AB}] = \mathbb{1}/2^n$.

Since channels from the operationally defined class $\mathrm{SO}_n$ are manifestly stabiliser-preserving, it is clear that $\mathrm{SO}_n \subseteq \mathrm{SP}_{n,n}$. It is then natural to ask whether the two classes are identical, or if $\mathrm{SP}_{n,n}$ is a strictly larger class. It has very recently been proven by Heimendahl, Heinrich and Gross [63] that $\mathrm{SO}_1 \equiv \mathrm{SP}_{1,1}$, but $\mathrm{SP}_{n,n}$ is strictly larger than $\mathrm{SO}_n$ for $n \geq 2$. In Chapter 5, Section 5.1, we present an algorithm showing that all maps in $\mathrm{SP}_{n,n}$ can be efficiently classically simulated, up to some caveats related to the number of Kraus operators. The recent result due to Heimendahl, Heinrich and Gross shows that our $\mathrm{SP}_{n,n}$ simulator is strictly more powerful than the standard Gottesman-Knill algorithm [63].

## 1.4 Universality, fault tolerance and state injection

Aaronson and Gottesman have shown that the problem of simulating a stabiliser circuit without classical adaptivity is in a complexity class called $\oplus\mathbf{L}$ (or parity-$\mathbf{L}$) [64], which is conjectured to be strictly contained in $\mathbf{P}$ [31, 64]. This suggests that stabiliser operations are not universal even for classical computation (i.e. there exist algorithms that one could run on a classical computer that cannot be simulated by stabiliser circuits without the aid of a classical co-processor). On the other hand it is believed that a universal quantum computer could perform certain tasks that are currently intractable for classical computers [4, 23]. Certainly, stabiliser circuits are not universal for quantum computation, so it is natural to ask which sets of primitive quantum operations are needed to construct a universal device. In this section, we address this question and illustrate how this naturally leads to the notion that magic states are a resource for quantum computation.

### 1.4.1 Circuit synthesis

Real quantum devices are subject to physical constraints so that only certain classes of quantum operation may be carried out in a single operation cycle. For example, superconducting architectures typically have a 2D layout where only nearest neighbour interactions are available, so that the native gate set only includes single-qubit and local two-qubit gates. More complicated multi-qubit interactions must be synthesised as a sequence of layers of few-qubit gates. We say that a gate set $\mathcal{G}$ is universal for quantum computation if any $n$-qubit unitary operation $U \in U(2^n)$ can be realised (up to arbitrarily small error) by a sequence of gates from $\mathcal{G}$. A sequence of early results due to Deutsch [65] and DiVicenzo [66] showed that any $n$-qubit unitary operation $U \in U(2^n)$ can be synthesised by gates from the set $\mathcal{G}_1 = \{CNOT, U_{\hat{r}}(\theta)\}$ is universal, where $U_{\hat{r}}(\theta) = \exp[-i(\hat{r} \cdot \sigma)\theta/2]$.

This is not the end of the story, as it is not possible to correct arbitrary errors on gates from a continuous set. For scalable quantum computation on real quantum hardware, it is necessary to seek discrete gate sets with known fault-tolerant constructions [33, 67]. This means we must sacrifice the ability to synthesise arbitrary unitary operations exactly. For the purposes of quantum computation it is sufficient to have the ability to implement $U$ up to some small error $\delta$ which can be made arbitrarily small by increasing the depth of the circuit [68–71]. Clearly, a Clifford generating set such as $\mathcal{G}_C = \{H, S, CNOT\}$ will not suffice, since for any fixed $n$ the Clifford group $\mathrm{Cl}_n$ is finite. Remarkably, however, only a minimal addition needs to be made to achieve universality. Boykin et al. [72] showed that only one more single-qubit gate need be added to reach universality. Specifically, the Clifford+T set, generated by $\mathcal{G}_{C+T} = \{H, S, CNOT, T\}$, is known to be universal, where $T$ is defined $T = \mathrm{diag}(1, e^{i\pi/4})$. In fact, the addition of any non-Clifford single-qubit gate is enough to reach universality, though $T$ is the usual choice in fault-tolerant schemes

[33].

The next question is whether a given $n$-qubit circuit can be synthesised *efficiently*. It can be shown that generic unitary operations in $SU(2^n)$ cannot be approximated efficiently by a finite library of gates that act on a bounded number of qubits, and there exist $U \in SU(2^n)$ that require at least $\Omega(2^n \log(1/\varepsilon)/\log(n))$ library gates [4, 73]. For a quantum algorithm to be efficient, there must exist polynomial-sized decomposition into few-qubit gates. Fortunately, it is not necessary for each of these gates to be elements of any particular standard gate set, due to the famous Solovay-Kitaev theorem [74, 75]. To state the theorem we need the notions of a dense set, and of an $\varepsilon$-net. We say that a set $\langle \mathcal{G} \rangle$ is dense in some other set $S$ if for all $W \in S$, and for arbitrary error $\varepsilon > 0$, there always exists some $U \in \langle \mathcal{G} \rangle$ such that $\|U - W\|_1 < \varepsilon$. As shown by Boykin et al., the Clifford + T set satisfies this criterion. The condition for an $\varepsilon$-net is a weaker one; we say that a set $R$ is an $\varepsilon$-net for $S$, for some *fixed* $\varepsilon$, if for every $U \in S$, there exists $W \in R$ such that $\|U - W\|_1 < \varepsilon$. The theorem can then be stated as follows:

**Theorem 1.12** (Solovay-Kitaev theorem for SU(2) [74])**.** *Let $\mathcal{G}$ be a finite subset of $SU(2)$ such that the set $\langle \mathcal{G} \rangle$ generated by $\mathcal{G}$ is dense in $SU(2)$, and for each $W \in \langle \mathcal{G} \rangle$, $W^\dagger \in \langle \mathcal{G} \rangle$. Then $\langle \mathcal{G} \rangle$. Then for any $\varepsilon > 0$, the set $\langle \mathcal{G} \rangle_M$ consisting of sequences of at most $M = O(\log^c(1/\varepsilon))$ gates from $\mathcal{G}$ is an $\varepsilon$-net for SU(2), where c is a constant.*

The Solovay-Kitaev theorem was first proven for $SU(2)$, as stated above, and subsequently generalised to $SU(d)$ for arbitrary $d$ [74]. It is natural to ask what is the optimal value for $c$. In early constructions it was found that the theorem held for $c \approx 3.97$, and it can be shown [4, 73, 76] that $c \geq 1$. In recent work [68–71], notably due to Kliuchnikov, Maskov and Mosca [68], and to Ross and Selinger [71], it has been proven that the lower bound of $c = 1$ is achievable for synthesis using the Clifford + T set. In summary, if a circuit can be efficiently implemented using any finite few-qubit gate set, then one can construct an efficient implementation in terms of Clifford + T.

### 1.4.2 Fault tolerance and magic state injection

Two broad strategies can be deployed against the noise that inevitably arises on real quantum hardware. Recently, much attention has been focused on noisy intermediate-scale quantum (NISQ) devices [17], comprised of up to several hundred imperfect physical qubits. It is acknowledged that these near-term devices, currently in development by various groups in academia and industry, will be incapable of full-scale quantum computation. Instead, many groups have proposed techniques tailor-made to these noisy devices [77], including algorithms that are inherently robust against noise [78], or through error mitigation protocols employing classical post-processing [79–81], randomised circuit constructions [49, 80, 82], or other active error recovery techniques [83].

In the longer term, scalable universal quantum computers will require quantum error correction and fault tolerance. These ideas have a longer history than the NISQ-era proposals just mentioned, and there exists an extensive literature on the topic [33, 48, 67, 84–87]. A detailed review is beyond the scope of this introduction, but we give a brief sketch. The central idea of quantum error correction is that logical qubits are redundantly encoded in a larger number of physical qubits, so that an error on any individual qubit is not fatal to the computation. In stabiliser codes [30, 88], a logical code space of dimension $2^k$ is fixed by specifying a stabiliser group with $n-k$ independent generators. Measurement of the stabiliser generators will yield a pattern of "$\pm 1$" outcomes called a syndrome. At least one "$-1$" outcome indicates that an error has caused the state to migrate outside the codespace. Classical decoders can be used to infer the most likely physical error corresponding to a given syndrome, and therefore the necessary unitary correction.

The simple model just described implicitly assumes that while errors could occur in the logical quantum circuit, the process of syndrome extraction and correction could be done perfectly. More realistically, we have to assume that any quantum operation, including all gates and measurements involved in syndrome extraction and error correction, can be faulty. The theory of fault tolerance [33, 85] is concerned with engineering protocols so that logical quantum information remains protected, even when all operations are faulty. A central result is the fact that quantum codes can have a *threshold* [67]. Loosely speaking, threshold theorems show that provided the maximum single-qubit error rate $p$ for any operation is below a certain threshold $p_c$, logical errors can be efficiently suppressed by increasing the number of physical qubits, so that a noisy quantum device with error correction can efficiently simulate an ideal quantum computer. Conversely, if the physical error rate is too high, increasing the size of the code only amplifies the logical error rate. Therefore engineering gates to operate below threshold is a key consideration in experimental efforts to build quantum computers.

Each logical operation in a circuit must have an implementation in the codespace which is itself fault-tolerant. One way to make a gate fault-tolerant is to find a *transversal* implementation, where each physical gate in the protocol acts non-trivially on no more than one physical qubit in each code block. For example, logical $Z$ in the toric code is fault-tolerantly implemented by multiple single-qubit $Z$ gates. The intuition behind this is that when errors do occur they need to remain localised in order to be correctable. Multi-qubit physical gates can spread errors throughout the code block, corrupting the logical quantum information. However, the Eastin-Knill theorem [32] states that for a given quantum error-correcting code, the set of gates that can be implemented transversally cannot be universal. For many stabiliser codes, only certain Clifford gates have a transversal implementation, while non-Clifford gates do not. Various techniques can be used to sidestep this problem [33], though they typically come with significant overhead.

The magic state model [33–44] is one of the leading schemes for implementing fault-tolerant non-Clifford gates. Certain non-Clifford unitary operations can be emulated deterministically by a non-unitary circuit, called a state injection or gate teleportation gadget, that consumes a magic state along with the logical input state, but where all other circuit elements employed are stabiliser operations [89]. At this point it is instructive introduce the Clifford hierarchy [89]. We take the Pauli group to be the first level of the hierarchy. Then we say that the $k$-th level of the hierarchy is the set of gates that map the Pauli operators to the $(k-1)$-th level by conjugation. For example, the Clifford gates make up the second level, since for any $U \in \mathrm{Cl}_n$ and Pauli $P \in \mathcal{P}_n$, $UPU^\dagger \in \mathcal{P}_n$. Of interest for the present discussion are the third level gates, which map Pauli operators onto Clifford operators, $UPU^\dagger \in \mathrm{Cl}_n$ for all $P \in \mathcal{P}_n$. We will shortly see that any third-level gate can be implemented deterministically by state injection.

The key idea is to take a teleportation circuit consuming a entangled state $|\Phi_+\rangle$, apply the third-level gate $U$ to the teleported state, but then commute $U$ back through the circuit onto a secondary input state. This is illustrated for single-qubit $U$ in Figure 1.3. Commuting $U$ past the conditional $X$ and $Z$ gates transforms them to $C_X = UXU^\dagger$ and $C_Z = UZU^\dagger$, which are Clifford gates by virtue of the fact that $U$ is a third-level gate. This shows that implementation of the non-Clifford gate $U$ can be traded for preparation of the magic state $(U \otimes \mathbb{1})\,|\Phi_+\rangle$. This can be straightforwardly generalised to arbitrary $n$-qubit third-level gates by parallelising the Bell state preparation and measurement circuit.

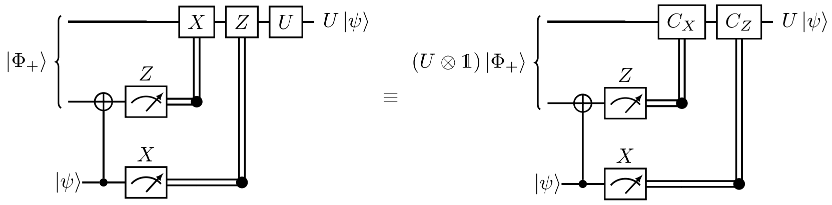


**Figure 1.3:** State injection gadget for general third-level $n$-qubit gate $U$ [89]. When the resource state is the Bell state $|\Phi_+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, the protocol simply teleports a state $|\psi\rangle$ unchanged from qubit 3 to qubit 1, after which the gate $U$ is applied. But when $U$ is a third-level gate, we can propagate it back through the circuit onto the initial resource state, converting the Pauli corrections $X$, $Z$ into Clifford corrections $C_X, C_Z$.

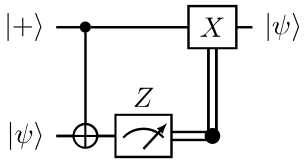


**Figure 1.4:** Gadget for “teleporting” a state $|\psi\rangle$ between qubit lines.

High fidelity preparation of an entangled resource state may in itself present challenges, but the construction can be streamlined for third-level gates diagonal in the computational basis [35, 89]. The “half-teleportation” gadget shown in Figure 1.4 transfers an input state $|\psi\rangle$ from qubit 2 to qubit 1, by means of an entangling gate followed by a conditional Clifford correction. Now suppose we replace $|+\rangle$ with a resource state $U|+\rangle$ for diagonal third-level $U$. The *CNOT* commutes with any single-qubit diagonal gate applied to the control qubit, so as in the the general case we can obtain an identity with a circuit that teleports $|\psi\rangle$ from qubit 2 to 1 and then applies $U$. This is illustrated for the $T$-gate in Figure 1.5, where we use $TXT^\dagger = \alpha XS$, where $\alpha$ is an unimportant phase. Again one can generalise this for general $n$-qubit diagonal third level gates (Figure 1.6).

In this way, the burden of a fault-tolerant construction is shifted from direct implementation with physical gates to high fidelity preparation of the magic state. This can be achieved by distillation protocols that consume large numbers of noisy magic state copies, and output a much smaller number of higher fidelity magic states [33–44]. This process is costly experimentally, both in terms of the storage space required for the reservoir of noisy magic states, and the time needed to distil each magic state ready for injec-

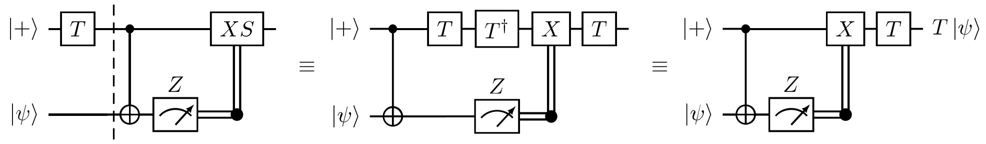


**Figure 1.5:** State injection gadget for the $T$-gate [90].

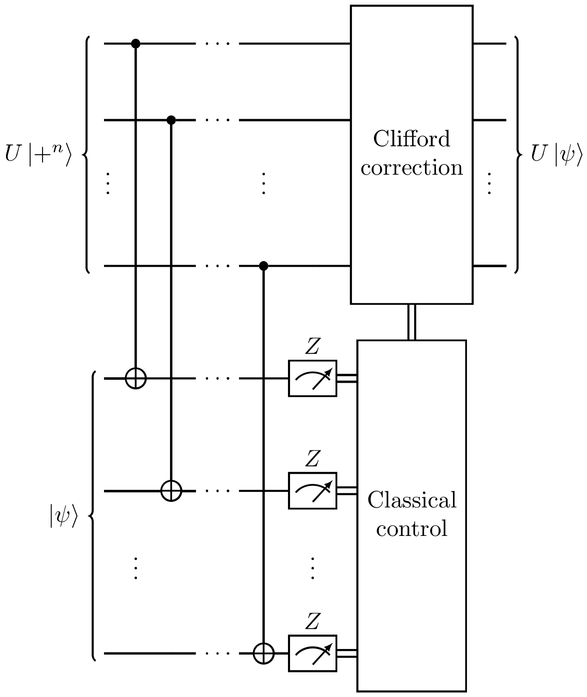


**Figure 1.6:** State injection gadget for diagonal third-level $n$-qubit gate $U$ [90].

tion. There is ongoing work to develop highly optimised magic state factories in order to drive down these space and time requirements [91, 92]. More importantly for this thesis, the arguments above motivate the idea of magic states as a computational resource that can be traded for non-Clifford gates. A classical simulator that admits adaptive Clifford circuits initialised with magic states, is in effect also a classical simulator for non-Clifford gates. In the next section we start to consider schemes that extend stabiliser simulators to circuits involving magic.

## 1.5 Simulating circuits with magic

It has been known since at least 2004 that one can simulate universal circuits using extensions of the Gottesman-Knill algorithm, but with an overhead exponential in the number of magic resources required [31]. We sketch two such algorithms here, again due to Aaronson and Gottesman [31]. The first simulates stabiliser circuits initialised with magic states, the second deals with arbitrary non-Clifford gates.

The method for magic states assumes that the initial state $\rho$ is a tensor product state comprising $m$ blocks of up to $b$ qubits, $\rho = \bigotimes_{j=1}^{m} \rho_j$, and exploits the fact that Pauli operators always have tensor product structure for any partitioning of qubits, $P = \bigotimes_{j=1}^{m} P_j$. The circuit is assumed to be a sequence of polynomial-sized Clifford circuits $U_i$ alternating with single-qubit measurements $Z_i$. We stipulate at most $M$ measurements. Note that this model allows for state injection, but at least one measurement is needed for each injection gadget, so that $t$ injected magic states implies at least $t$ measurements. Rather than evolve the density operator, we can move to the Heisenberg picture and evolve the Pauli operators. For example, the probability of obtaining each outcome for the first measurement is,

$$\Pr(\pm 1) = \mathrm{Tr}\left[\frac{\mathbb{1} \pm Z_1}{2} U_1 \rho U_1^\dagger\right] = \mathrm{Tr}\left[\frac{\mathbb{1} \pm U_1^\dagger Z_1 U_1}{2} \rho\right]. \tag{1.76}$$

Since $U_1$ is a Clifford operator, $Z_1$ is mapped to some new Pauli operator $U_1^\dagger Z_1 U_1 = P = \bigotimes_{j=1}^{m} P_j$. Then we can rewrite as:

$$\Pr(\pm 1) = \frac{1}{2} \pm \frac{1}{2} \mathrm{Tr}\left[\bigotimes_{j=1}^{m} [P_j \rho_j]\right] = \frac{1}{2} \pm \frac{1}{2} \prod_{j=1}^{m} \mathrm{Tr}\left[P_j \rho_j\right] \tag{1.77}$$

Each trace in this product involves at most $b$ qubits, so can simply be computed by matrix multiplication in time $\mathcal{O}\left(2^{2b}\right)$, which is tractable if $b$ is small. A biased coin is flipped based on the computed probabilities, and based on the result, the appropriate projector is applied to the state,

$$\rho' = \frac{1}{4} \frac{(\mathbb{1} \pm Z_1) U_1 \rho U_1^\dagger (\mathbb{1} \pm Z_1)}{\Pr(\pm 1)}. \tag{1.78}$$

The process is repeated for the measurement of the second qubit, except that equation (1.78) is now a linear combination of four terms, so the measurement step now involves $2 \cdot 4 = 8$ terms, each taking time $\mathcal{O}\left(m 2^{2b}\right)$ to evaluate. By the time we get the final measurement, there are $2^{2M-1}$ terms, so the total runtime is $\mathcal{O}\left(m 2^{2M+2b}\right)$. For the case where we inject $t$ magic states, there are at least $t$ blocks of qubits and at least $t$ measurements, so the runtime is $\mathcal{O}\left(t2^{2t}\right)$.

The method for non-Clifford gates is based on decompositions of $b$-qubit gates $U$ in the Pauli basis, which can have up to $4^b$ terms, $U = \sum_{i=1}^{4^b} c_i P_i$. Let $\mathbf{r}(P)$ be the $2n$-element vector labelling the positions of $Z$ and $X$ operators for Pauli $P$. Then a pure stabiliser state $\rho$ with generators $g_k$ is updated as

$$U\rho U^\dagger = \sum_{i,j} c_i c_j^* P_i \left( \prod_{k=1}^{n} \frac{\mathbb{1} + g_k}{2} \right) P_j \tag{1.79}$$

$$= \sum_{i,j} c_{ij} P_{ij} \left( \prod_{k=1}^{n} \frac{\mathbb{1} + (-1)^{\langle \mathbf{r}(g_k), \mathbf{r}(P_j) \rangle} g_k}{2} \right), \tag{1.80}$$

where $c_{ij} P_{ij} = c_i c_j^* P_i P_j$. With each additional non-Clifford gate, each term branches a further $4^{2b}$ times. In the worst case, we need to keep track of $4^{2bt}$ coefficients, and $\mathcal{O}\left(4^{bt}\right)$ strings of length $\mathcal{O}(n)$, representing Pauli matrices and inner products.

Measurements are simulated by letting projectors $\Pi_Q = (\mathbb{1} + Q)/2$ act by conjugation on the evolved state $\rho'$. We omit the details here, but the result is that the decomposition retains the same form as equation (1.80), and the number of terms does not increase. To evaluate the trace of $\Pi_Q \rho_t \Pi_Q$, where $\rho_t$ is the state at the $t$-th step of the circuit, Aaronson and Gottesman note that the trace of each term is either $\pm 2^n c_{ij}$ if $P_{ij}$ is in the stabiliser, and 0 otherwise, and $\mathcal{O}\left(4^{2bt}\right)$ terms must be summed. Putting this all together, the algorithm requires time and space $\mathcal{O}\left(4^{2bt}\right)$.

Both of the techniques just described have runtime exponential in the number of magic resources. We draw attention to the fact that both those methods are state/gate-agnostic, in the sense that the strength of the scaling does not depend on the details of the particular resource consumed, only on coarser parameters such as the total number of non-Clifford gates or magic states, and the number of qubits on which they are defined. It is natural to ask whether there are not, for example, gates that are easier to simulate because they more closely approximate stabiliser states, and indeed this will be one of the central questions considered in this thesis. To this end, it is useful to formalise how quantum resources can be quantified. To close this chapter, we introduce some of the basic machinery of quantum resource theories.

## 1.6 Resource theories and monotones

Performing certain quantum information processing tasks requires access to resourceful states and operations. Resource theories have become an important tool for the study of many phenomena in quantum information theory [93–95], including entanglement [96, 97], coherence [98], contextuality [99], athermality [100], and numerous others. Our main interest in this thesis is in the resource theory of magic [1–3, 101–104]. Resource theories can be used to study questions such as which resourceful states can be interconverted [103, 105, 106], and the rate at which higher quality resource states can be distilled from many lower quality copies [33–44].

A resource theory can be defined by specifying a class of free operations $\mathcal{F}$ [93], such as local operations and classical communication (LOCC) in entanglement theory [107]. One can then identify the set of free states $\mathcal{S}$ as those states that can be prepared from a standard fiducial state using free operations alone. By virtue of being non-free, states outside $\mathcal{S}$ therefore acquire value as resourceful states. Likewise non-free operations are considered resourceful. Beyond the binary distinction of free versus non-free, it is useful to assign a resource value to a state of interest. This can be done using a *resource monotone*.

**Definition 1.13** (Resource monotones)**.** *Given a resource theory equipped with a class of free operations $\mathcal{F}$, we say that a mapping $\mathcal{M}$ from states to the real numbers is a monotone for that resource if it is non-increasing under the class of free operations. That is, for all states $\rho$:*

$$\mathcal{M}(\mathcal{E}(\rho)) \leq \mathcal{M}(\rho), \quad \forall \mathcal{E} \in \mathcal{F}. \tag{1.81}$$

*We call this property monotonicity.*

For example, for any well-defined resource theory, once we have specified the set of free states $\mathcal{S}$, we can always define the relative entropy distance to the set of free states [93]. The relative entropy of a state $\rho$ with respect to state $\sigma$ is defined $S(\rho\|\sigma) = \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \sigma]$. Then the relative entropy distance with respect to the set of free states $\mathcal{S}$ is given by $\mathcal{D}_{\mathcal{S}}(\rho) = \inf_{\sigma \in \mathcal{S}} S(\rho\|\sigma)$. The relative entropy is contractive under CPTP operations [108] (i.e. $S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq S(\rho\|\sigma)$ for any CPTP $\mathcal{E}$), and it follows that $\mathcal{D}_{\mathcal{S}}$ must be non-increasing under CPTP free operations. Several other properties are advantageous in proving results in magic theory. We note that monotones defined in the literature do not always have the properties defined below, but we will make use of them many times in this thesis.

**Definition 1.14** (Faithfulness)**.** *We say that a monotone $\mathcal{M}$ is* ***faithful*** *if there is a constant c such that $\mathcal{M}(\rho) = c$ if $\rho \in \mathcal{S}$, and $\mathcal{M}(\rho) > c$ otherwise.*

For entropy-like monotones such as relative entropy distance, the constant $c$ is zero. We can see this easily for the relative entropy distance, since $S(\rho\|\sigma) \geq 0$, but if $\rho \in \mathcal{S}$ then $\inf_{\sigma\in\mathcal{S}} S(\rho\|\sigma) = S(\rho\|\rho) = 0$. Many of the magic monotones we consider later in this thesis are defined in such a way that $c = 1$. However, this is a matter of convention, as each of our quantifiers can be converted into a monotone such that $c = 0$, simply by taking the logarithm. For the purposes of this thesis, we find the non-logarithmic variant to be more convenient.

**Definition 1.15** (Submultiplicativity/subadditivity)**.** *We say that a monotone is submultiplicative under tensor product if for any pair of states $\rho$ and $\sigma$, we have:*

$$\mathcal{M}(\rho \otimes \sigma) \leq \mathcal{M}(\rho)\mathcal{M}(\sigma). \tag{1.82}$$

*Submultiplicativity applies to non-logarithmic monotones. Equivalently, if $\mathcal{M}$ is submultiplicative, then the logarithmic version is subadditive:*

$$\log \mathcal{M}(\rho \otimes \sigma) \leq \log \mathcal{M}(\rho) + \log \mathcal{M}(\sigma). \tag{1.83}$$

When submultiplicativity/subadditivity holds, it ensures that one cannot increase the apparent resourcefulness of a state by preparing a free state on an auxiliary subsystem. We note that submultiplicativity is not a strict requirement for all quantifiers in all contexts, as it is conceivable that measures that do not have this property may be useful in the study of processes such as catalysis, and activation of bound resources [105, 109]. Indeed, some authors have defined quantifiers of magic that can be supermultiplicative for some states [110], but we will later argue that this property can be problematic in attempting to quantify classical simulation costs. Indeed, submultiplicativity is key to proving several novel results presented later in this thesis. Returning to the notion of faithfulness, there is a natural choice of the constant $c$ depending on whether a monotone is of the (sub)multiplicative or (sub)additive type. For subadditive monotones, we expect that composing one state with another via the tensor product adds some non-negative real number to the monotone value. But we want to ensure that composing with another stabiliser state does not increase the magic content. It therefore makes sense to set $c = 0$, since 0 is the additive identity for the set of real numbers. For submultiplicative monotones, composing two states at most multiplies their value, so it is natural to choose the multiplicative identity, and define monotones so that $c = 1$.

**Definition 1.16** (Convexity). *A monotone $\mathcal{M}$ is convex if for any state $\rho = \sum_j q_j \rho_j$, we have:*

$$\mathcal{M}(\rho) \leq \sum_j |q_j| \mathcal{M}(\rho_j). \tag{1.84}$$

Again, convexity is not a strict requirement for a resource monotone; indeed, logarithmic monotones are typically non-convex [111]. However, it holds for many of the monotones we introduce later, and is instrumental in proving further properties.

## 1.7 Summary and outlook

In this chapter, we introduced the stabiliser formalism [30] and other key concepts in quantum information and computation. We showed that stabiliser circuits can be simulated efficiently by a classical computer [25, 31], but that promotion to universality is achieved by the addition of one type of resource, namely non-stabiliser, or magic, resources [35]. Exploiting these resources fault-tolerantly is a major engineering challenge that must be overcome if scalable quantum computation is to become a reality [33]. The computational power of stabiliser circuits supplemented by $t$ magic resources, as quantified by the runtime of a classical simulation, appears to be exponential in $t$, though the precise scaling depends on the particular algorithm employed. For the early simulators of Section 1.5, the scaling was $\mathcal{O}(4^t)$ or worse, so that the algorithms quickly become impractical even for relatively small numbers of magic state copies. In more modern algorithms, we still expect asymptotic exponential scaling $\mathcal{O}(2^{\alpha t})$, but recent work [3, 59, 60, 103] has enabled the coefficient $\alpha$ to be significantly reduced. With these new methods the performance scaling is sensitive to the "amount" of magic per gate (or state), where "amount" is made precise by the resource theory of magic [101–103, 112]. In this way, the exponential scaling is tamed so that classical simulation remains tractable for larger non-stabiliser circuits. The main aim of this thesis is to contribute to this effort to improve classical simulators, both by further developing the resource theory of magic, and by designing new classical algorithms with improved performance. In the next two chapters we review related prior work where stabiliser simulation methods have been extended to more general classes of quantum circuit, broadly divided into two strands. In the next chapter, we will review quasiprobability methods, including phase space methods and related techniques, which can be applied for general density operators. In contrast, the stabiliser rank techniques reviewed in Chapter 3 are usually restricted to pure state evolution, but can lead to faster classical simulators. In the main part of the thesis, following these chapters, we will present our novel work that improves and extends both quasiprobability and stabiliser rank techniques.

Chapter 2

# Quasiprobabilities and phase space methods

## 2.1 The phase space picture of quantum mechanics

The Wigner function has a long history in quantum theory, having first been introduced by Wigner in 1932 as a quantum analogue for probability distributions over classical phase space [113]. Given a classical system with coordinates $x_i$ and momenta $p_i$, one can write down a probability density function $P(x_i; p_i)$ for finding the system in a particular infinitesimal region of phase space. In contrast, in quantum mechanics the possibility of writing down a joint probability distribution for position and momentum is famously prohibited by the uncertainty principle. Nevertheless, Wigner realised that one can formulate a real-valued function $W(x_i, p_i)$ that shares many features of a probability distribution. A crucial difference is that its values can go negative, and it is in this sense that the Wigner function yields a quasiprobability distribution. When this Wigner function is integrated over either position or momentum, the correct quantum mechanical probability density for the conjugate variable is recovered.

More recently, phase space techniques have seen application in the field of quantum optics, where negativity of the Wigner function has been proposed as a signature of nonclassicality [114, 115]. Conversely, the Gaussian states are those with non-negative Wigner functions. This result is known as Hudson's theorem [116, 117], and these are often seen as continuous-variable analogues of stabiliser states [118]. As quantum information theory emerged as a field in its own right, various authors sought to define and study Wigner functions for discrete, finite-dimensional quantum systems [119]. In particular, Gibbons et al. [119] proposed axioms that candidates for discrete Wigner function $W$ should possess in order to be consistent with the continuous-variable counterpart. These include the stipulation that for systems of dimension $d$, $W$ should be a mapping from the space of operators to functions $W(p,q)$ on a phase space, which can be modelled as a $d \times d$ lattice; and the Wigner function should be covariant with respect to the action of Heisenberg-Weyl operators, meaning that the values $W(p,q)$ are permuted between phase points $(p,q)$ but the set of values does not change [118]. It turns out that for dimension $d$, there are $d^{d+1}$ distinct definitions for generalized Wigner functions satisfying the constraints of Gibbons et al. [119]. It was subsequently shown that stabiliser states in prime-power dimension are precisely those that are non-negative for all valid choices of Wigner function [120, 121]. However, Gross [118] showed that in any odd dimension there is a natural choice of Wigner function.

Gross's discrete Wigner function is based on the stabiliser formalism for odd-dimension qudits. Since this thesis is largely concerned with systems of even dimension, we will only give a brief outline here, and refer the interested reader to Ref. [118]. Given the standard single-qudit basis $\{|x\rangle\}_{x\in\mathbb{F}_d} = \{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$, the shift and boost operators can be defined analogous to Pauli X and Z,

$$X(q)\,|x\rangle = |x+q \mod d\rangle, \quad Z(p)\,|x\rangle = \omega^{xp}\,|x\rangle, \tag{2.1}$$

where $\omega$ is a primitive $d$-th root of unity, $\omega = \exp\left(\frac{2\pi i}{d}\right)$. These then generate the Heisenberg-Weyl operators $T_{\mathbf{a}} = T_{(p,q)} = \omega^{-2^{-1}pq} Z(p)X(q)$, where $\mathbf{a} = (p,q)^T \in \mathbb{F}_d \times \mathbb{F}_d$. Notice that $2^{-1}$ is not defined for $d = 2$, so that this definition cannot be applied to qubits. Multi-qudit Heisenberg-Weyl operators are constructed by forming tensor products of the single-qudit operators, and these can similarly be labelled by vectors $(\mathbf{p},\mathbf{q}) \in \mathbb{F}_d^n \times \mathbb{F}_d^n$ where $\mathbf{p}$ and $\mathbf{q}$ are the direct sum of the entries $(p,q)$ for each qudit. Heisenberg-Weyl operators compose according to $T_{\mathbf{a}}T_{\mathbf{b}} = \omega^{2^{-1}\langle\mathbf{a},\mathbf{b}\rangle} T_{\mathbf{a}+\mathbf{b}}$ where $\langle\cdot,\cdot\rangle$ is the appropriate symplectic inner product for the space $V = \mathbb{F}_d^n \times \mathbb{F}_d^n$. Thus equipped with the symplectic inner product, $V$ is called the phase space. Qudit stabiliser states are specified by maximally commuting subgroups of the Heisenberg-Weyl operators. The qudit Clifford operations correspond to symplectic affine transformations on the phase space [118], and it follows that their action on elements of stabiliser groups can be simulated efficiently, as in the qubit case.

The Heisenberg-Weyl operators form a complete basis orthonormal with respect to the Hilbert-Schmidt inner product, up to normalisation, i.e. $d^{-n}\mathrm{Tr}\left[T_{\mathbf{a}}^{\dagger}T_{\mathbf{b}}\right] = \delta_{\mathbf{a},\mathbf{b}}$. Any density operator therefore has a unique expansion in this basis. The characteristic function specifies the expansion coefficient for each Heisenberg-Weyl operator. The discrete Wigner function due to Gross [118] is obtained by taking the symplectic Fourier transform of the characteristic function,

$$\rho = \sum_{\mathbf{a}\in V} W_{\rho}(\mathbf{a})A_{\mathbf{a}}, \quad A_{\mathbf{a}} = d^{-n}\sum_{\mathbf{b}\in V} \omega^{-\langle\mathbf{a},\mathbf{b}\rangle}T_{\mathbf{b}}^{\dagger}, \tag{2.2}$$

where $A_{\mathbf{a}}$ are a set of orthonormal operators with unit trace called phase point operators.The Wigner function for Hermitian $\rho$ is always real-valued, but in general can take negative values. This is key to the interpretation of equation 2.2 as a quasiprobability distribution. The discrete Hudson's theorem due to Gross [118] states that for pure states, the stabiliser states are precisely those with a non-negative Wigner function, which means their density operator description can be modelled as a proper probability distribution over phase point operators. It follows that any convex combination of stabiliser states must also have non-negative Wigner function, but it can be shown that the converse is not true, and there exist non-stabiliser mixed states with non-negative Wigner function. These are known as *bound magic states* [101, 122]. While a supply of noisy magic states that are not bound can be refined to magic states with higher purity via magic state distillation, this cannot be done for bound magic states using stabiliser operations alone [101].

## 2.2 Wigner function negativity as a computational resource

The fact that states with positive discrete Wigner function can be cast as probability distributions over phase point operators suggests that they may be simulated by random sampling of these operators. Indeed, Veitch et al. [101] showed by explicit construction of a classical algorithm that states and measurements with non-negative Wigner function can be simulated efficiently if the intervening operations are limited to Clifford gates. This result is therefore strictly stronger than the Gottesman-Knill theorem for qudits, as it shows that efficient simulation remains possible for a subclass of magic states, and for measurements beyond projective measurements in stabiliser bases. Independently, Mari and Eisert [123] developed a similar classical simulation scheme for odd qudit systems, which additionally allows efficient simulation of quantum channels with positive phase space representation. Thus, to achieve any speedup over classical computation, a quantum circuit on a system of odd dimension must involve Wigner negativity in either state preparations, transformations or measurements.

Later work made this notion of Wigner negativity as a resource quantitative by defining magic monotones [102, 124]. In a follow-up paper Veitch et al. [102] introduced the *sum negativity* sn and the related quantity *mana* $\mathcal{M}_W$, defined,

$$\mathrm{sn}(\rho) = \sum_{\mathbf{u}:W_\rho(\mathbf{u})<0} \left|W_\rho(\mathbf{u})\right| \tag{2.3}$$

$$\mathcal{M}_W(\rho) = \log[\|W_\rho\|_1] = \log[2\mathrm{sn}(\rho)+1]. \tag{2.4}$$

Veitch et al. showed that the sum negativity is unique as a quantifier of magic for qudits, in the sense that any reasonable magic monotone based on the negativity of the Wigner function must be a function of $\mathrm{sn}(\rho)$ alone. Both sum negativity and mana are monotone under stabiliser operations, but the mana has the convenient property of being exactly additive. The mana distinguishes bound and distillable magic states since positively represented states have $\mathcal{M}_W(\rho) = 0$, whereas $\mathcal{M}_W(\rho) > 0$ for distillable states. Indeed, Veitch et al. use their monotones to numerically analyse magic state distillation protocols, comparing conversion rates for known protocols with the upper bound derived from magic monotones [102]. Wigner negativity can therefore be seen as a necessary resource for quantum advantage.

In this thesis our primary interest is in the cost of classical simulation, and it has been shown that Wigner negativity can be directly related to runtime for a class of Monte Carlo-type algorithms based on quasiprobabilities [124]. Wigner functions can be situated within a broader class of quasiprobability representation through the notion of a *frame* [124–127]. In this picture we first define a pair of operator bases $\{F(\lambda)\}_{\lambda\in\Lambda}$ and $\{G(\lambda)\}_{\lambda\in\Lambda}$, called the frame and dual frame respectively, which satisfy the relation $A = \sum_{\lambda\in\Lambda} G(\lambda)\,\mathrm{Tr}[AF(\lambda)]$ for any Hermitian operator $A$, where the label $\lambda$ runs over some defined space $\Lambda$. For example, this can be satisfied by choosing $\Lambda$ to be the qudit phase space already described, choosing the frame to be the phase point operators normalised by dimension $F(\mathbf{u}) = A_{\mathbf{u}}/d^n$, and the dual frame to be the unnormalised phase point operators, $G(\mathbf{u}) = A_{\mathbf{u}}$. However, the frame concept is more general, as the frame and dual frame bases are allowed to be overcomplete, and $\Lambda$ is not constrained to have the structure of a phase space.

Having defined the frame and dual frame, one can define a quasiprobability distribution for any density operator, $W_\rho(\lambda) = \mathrm{Tr}[F(\lambda)\rho]$. Of course, when the frame is chosen to be the set of normalised phase point operators, this is precisely the discrete Wigner function identified by Gross [118]. One can also define distributions for unitaries $U$ and measurement effect operators $E$:

$$W_U(\lambda'|\lambda) = \mathrm{Tr}\left[F(\lambda')UG(\lambda)U^\dagger\right], \quad W(E|\lambda) = \mathrm{Tr}[EG(\lambda)]. \tag{2.5}$$

Notice that for the representation of $U$, this notation is suggestive of a conditional probability for transition from point $\lambda$ to point $\lambda'$, and $W(E|\lambda)$ suggests the probability of obtaining some the measurement corresponding to $E$ given a system at point $\lambda$. Indeed by applying the Born rule along with the definitions above,

$$\mathrm{Pr}\left(E|U\rho U^\dagger\right) = \sum_{\lambda',\lambda} W(E|\lambda')W_U(\lambda'|\lambda)W_\rho(\lambda). \tag{2.6}$$

However, we must keep in mind that these functions should be understood as *quasi*-probability distributions rather than proper probability distributions, since they can take negative values. Nevertheless, Pashayan et al. [124] proposed a Monte Carlo simulation method involving a renormalisation of the quasiprobability distribution, so that a genuine probability distribution can be recovered. The negativity of $W_\rho$ is defined to be the $\ell_1$-norm, $||W_\rho||_1 = \sum_{\lambda\in\Lambda}\left|W_\rho(\lambda)\right|$. Pashayan et al. define the point-negativity of a unitary $U$ at point $\lambda$ to be $||W_U(\cdot|\lambda)||_1 = \sum_{\lambda'\in\Lambda}|W_U(\lambda'|\lambda)|$. To define probability distributions, we normalise the distributions as

$$\mathrm{Pr}(\lambda|\rho) = \frac{\left|W_\rho(\lambda)\right|}{||W_\rho||_1}, \quad \mathrm{Pr}\left(\lambda'|U,\lambda\right) = \frac{|W_U(\lambda'|\lambda)|}{||W_U(\cdot|\lambda)||_1}. \tag{2.7}$$

We can then rewrite the Born rule probability from equation (2.6) as

$$\Pr\Big(E|U\rho U^\dagger\Big) = \sum_{\lambda',\lambda} \widehat{E}_{\lambda',\lambda} \Pr\big(\lambda'|U,\lambda\big)\Pr(\lambda|\rho), \tag{2.8}$$

$$\text{where} \quad \widehat{E}_{\lambda',\lambda} = W(E|\lambda')\cdot \mathrm{sign}\big(W_U(\lambda'|\lambda)W_\rho(\lambda)\big)\cdot \|W_U(\cdot|\lambda)\|_1 \cdot \|W_\rho\|_1 \tag{2.9}$$

The simulation strategy is therefore to randomly sample a point $\lambda$ from the distribution $\{\Pr(\lambda|\rho)\}_{\lambda\in\Lambda}$, before sampling a transition to a second point $\lambda'$ from the distribution corresponding to the unitary $U$. This can be straightforwardly extended to a length $T$ sequence of unitary operations $U_t$, where each $U_t$ is described by its own quasiprobability distribution. Pseudocode for the procedure is given in Algorithm 1.

**Algorithm 1** Pashayan quasiprobability simulator [124]

```
Input: Initial state $\rho$ described by quasiprobability function $W_\rho$; Sequence of unitary
    operations $U = U_T \circ U_{T-1} \circ \ldots \circ U_1$, each described by function $W_{U_t}$; measurement
    effect operator described by function $W(E|\cdot)$; number of samples $M$.
Output: Estimate $\widehat{E}$ for $\mathrm{Tr}\big[EU\rho U^\dagger\big]$.
 1: $\widetilde{E} \leftarrow 0$
 2: for $m \leftarrow 1$ to $M$ do
 3:     Sample $\lambda_0$ from $\{\Pr(\lambda|\rho)\}_{\lambda\in\Lambda}$.
 4:     $Q \leftarrow \mathrm{sign}\big(W_\rho(\lambda_0)\big) \times \|W_\rho\|_1$
 5:     for $t \leftarrow 1$ to $T$ do
 6:         Sample $\lambda_t$ from $\{\Pr(\lambda'|U,\lambda_{t-1})\}_{\lambda'\in\Lambda}$.
 7:         $Q \leftarrow Q \times \mathrm{sign}(W_{U_t}(\lambda_t|\lambda_{t-1})) \times \|W_{U_t}(\cdot|\lambda_{t-1})\|_1$
 8:     end for
 9:     $\widetilde{E}_m \leftarrow Q \times W(E|\lambda_T)$
10:     $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_m$
11: end for
12: return $\widehat{E} = \widetilde{E}/M$.
```

Let $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_T)$ be a vector representing a trajectory of sampled points $\lambda_t$ in Algorithm 1. Then let

$$\begin{aligned} E_{\boldsymbol{\lambda}} = W(E|\lambda_T)\cdot \mathrm{sign}\big(W_\rho(\lambda_0)\big)\cdot \|W_\rho\|_1 \\ \times \prod_{t=1}^{T} \mathrm{sign}(W_{U_t}(\lambda_t|\lambda_{t-1}))\cdot \|W_{U_t}(\cdot|\lambda_{t-1})\|_1 \end{aligned} \tag{2.10}$$

We can then compute the expected value of the random variable $\widetilde{E}_m$ evaluated in step 9.

We have

$$\mathbb{E}(\widetilde{E}_m) = \sum_{\boldsymbol{\lambda}} \Pr(\lambda_0|\rho) \left( \prod_{t=1}^{T} \Pr(\lambda_t|U, \lambda_{t=1}) \right) E_{\boldsymbol{\lambda}} \tag{2.11}$$

$$= \sum_{\boldsymbol{\lambda}} W(E|\lambda_T) \left( \prod_{t=1}^{T} W_{U_t}(\lambda_t|\lambda_{t-1}) \right) W_\rho(\lambda_0), \tag{2.12}$$

where in the last line we combined the definition of the probability distributions in equation (2.7) with the random variable (2.10) to rewrite

$$\operatorname{sign}\big(W_\rho(\lambda_0)\big) \cdot \|W_\rho\|_1 \Pr(\lambda|\rho) = \operatorname{sign}\big(W_\rho(\lambda_0)\big) \cdot \|W_\rho\|_1 \frac{|W_\rho(\lambda_0)|}{\|W_\rho\|_1} = W_\rho(\lambda_0), \tag{2.13}$$

and similarly for each unitary operation $U_t$. But given the quasiprobability distributions defined using the frame and dual frame above, line (2.12) is an expression for the Born rule probability, $\Pr\big(E|U\rho U^\dagger\big) = \operatorname{Tr}\big[EU\rho U^\dagger\big] = \mathbb{E}(\widetilde{E}_m)$. In other words, the random variable output from Algorithm 1 is an unbiased estimator for the quantum mean value $\operatorname{Tr}\big[EU\rho U^\dagger\big]$.Therefore $\widehat{E}$ will converge to the quantum mean value given a large enough number of samples $M$. The critical question is how large $M$ needs to be to achieve a given precision, as this governs the runtime. Pashayan et al. make use of the Hoeffding inequality.

**Theorem 2.1** (Hoeffding inequality [128]). *Given a sequence of $M$ i.i.d. random variables $X_j$ bounded by $|X_j| \leq b$ and with expected mean value $\mathbb{E}(X)$, the probability that $\sum_{j=1}^{M} X_j/M$ deviates from the mean by more than $\varepsilon$ is upper bounded by:*

$$\Pr\left\{ \left| \mathbb{E}(X) - \sum_{j=1}^{M} \frac{X_j}{M} \right| \geq \varepsilon \right\} \leq 2\exp\left( -\frac{M\varepsilon^2}{2b^2} \right). \tag{2.14}$$

Supposing that we want to achieve a fixed precision and success probability, we can use the following corollary, which follows immediately from Theorem 2.1.

**Corollary 2.2** (Hoeffding inequality - sample complexity). *Suppose that we can generate a sequence of $M$ bounded random variables $\{X_j\}$ as described in Theorem 2.1. We can achieve the precision $\left|\mathbb{E}(X) - \sum_{j=1}^{M} X_j/M\right| \leq \varepsilon$ with probability at least $(1 - p_{\text{fail}})$ by* *setting the number of samples as follows:*

$$M = \left\lceil 2b^2 \frac{1}{\varepsilon^2} \ln\left(\frac{2}{p_{\text{fail}}}\right) \right\rceil \tag{2.15}$$

We have already seen that $\mathbb{E}(\widetilde{E}_m) = \mathrm{Tr}\left[EU\rho U^\dagger\right]$. Pashayan et al. [124] define the total forward negativity $\mathcal{M}_\rightarrow$ for a circuit to be the maximum negativity over all possible trajectories:

$$\mathcal{M}_\rightarrow = \|W_\rho\|_1 \left(\prod_{t=1} \mathcal{M}_{U_t}\right) \max_{\lambda_T} |W(E|\lambda_T)|, \tag{2.16}$$

where $\mathcal{M}_{U_t} = \max_\lambda \|W_U(\cdot|\lambda)\|_1$ is referred to as the negativity of $U_t$. Then, comparing with equation (2.10) one can check that the random variables $\widetilde{E}_m$ generated in Algorithm 1 is bounded as $\left|\widetilde{E}_m\right| \leq \max|E_{\boldsymbol{\lambda}}| = \mathcal{M}_\rightarrow$. Therefore if $M$ samples are generated by the algorithm, by Hoeffding's inequality (Corollary 2.2) the output of the classical algorithm will be within additive error $\varepsilon$ of the quantum Born rule value with probability at least $(1 - p_{\text{fail}})$ provided that $M$ is set to be at least:

$$M \geq 2\mathcal{M}_\rightarrow^2 \frac{1}{\varepsilon^2} \ln\left(\frac{2}{p_{\text{fail}}}\right). \tag{2.17}$$

Pashayan et al. [124] demonstrated the method by classically simulating 100-qutrit random Clifford circuits, where $k$ qutrits were initialised in a magic state, and the remainder initialised in the $|0\rangle$ state, choosing the canonical phase space frame, so that the quasiprobability distribution for the initial state is the discrete Wigner function. Recall that in Ref. [102] the mana was defined $\mathcal{M}_W(\rho) = \log[\|W_\rho\|_1]$, and is an additive monotone, so that for $k$ copies of a magic state $\rho$, the mana is $\mathcal{M}_W(\rho^{\otimes k}) = k\mathcal{M}_W(\rho)$. Therefore the runtime for this instance of the algorithm is $\mathcal{O}\left(\|W_{\rho^{\otimes k}}\|_1^2\right) = \mathcal{O}\left(2^{k\mathcal{M}_W(\rho)}\right)$, so exponential in the number of magic states $k$. This feature is shared with the Aaronson-Gottesman algorithm [31], but the key difference is that whereas the Aaronson-Gottesman algorithm treats all single-qubit magic states on an equal footing, the strength of the exponential scaling in the Pashayan simulation is sensitive to the single-qubit mana $\mathcal{M}_W(\rho)$ appearing in the exponent.

The phase space formalism for odd-dimension qudits, combined with the quasiprobability simulation framework therefore provides an elegant connection between the quantification of magic as a resource and the hardness of classical simulation. However, our primary interest for this thesis is the setting of qubit quantum circuits, and here we find that the convenient phase space picture breaks down [129, 130]. It is certainly possible to develop a meaningful phase space framework for even dimension [110], but one finds that many of the properties that lead to the striking connection between Wigner function negativity and magic theory in odd dimension are lost. We will discuss the reasons for this breakdown in Section 2.4. First, it will be useful to introduce alternative approaches that allow us to construct practical quasiprobability simulators for the qubit case.

## 2.3 Robustness of magic

It is clear that in the state vector picture, any pure state can be written in a basis of stabiliser states, and we will consider decompositions of this type in Chapter 3. However, it is also the case that the pure stabiliser state projectors $|\phi\rangle\langle\phi|$, where $|\phi\rangle \in \mathrm{STAB}_n$ form an overcomplete basis for the set of $2^n$-dimensional density matrices [103]. One can see this by recalling that any pure stabiliser state projector can be written as a product of Pauli projections (Section 1.1.5.1). For example in the base case for a single qubit, we have $|0\rangle\langle 0| = (\mathbb{1} + Z)/2$ and $|1\rangle\langle 1| = (\mathbb{1} - Z)/2$. But then we can write $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$. We can similarly decompose any Pauli operator as a linear combination of pure stabiliser states. In turn, it is well known the Pauli operators form a complete basis for the space of Hermitian operators. We can therefore write the normalised density matrix for *any* state as an affine combination of pure stabiliser state projectors $\rho = \sum_j q_j \left|\phi_j\right\rangle\!\left\langle\phi_j\right|$ where $\left|\phi_j\right\rangle \in \mathrm{STAB}_n$, and $\sum_j q_j = 1$ (Figure 2.1). In general, $q_j$ can be negative, so this is again a quasiprobability distribution. It is clear that the basis of stabiliser projectors is overcomplete, since the number of $n$-qubit stabiliser states grows faster than $4^n$. It follows that there are typically many possible stabiliser decompositions for a given density oper-

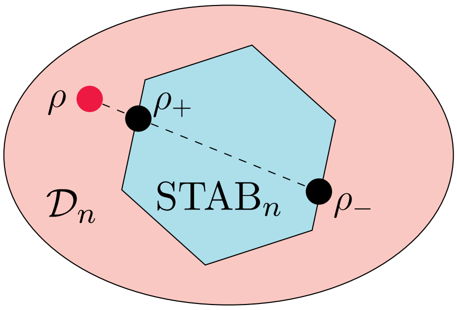


**Figure 2.1:** Schematic illustration of a density matrix $\rho \in \mathcal{D}_n$ decomposed as an affine combination of elements from the stabiliser polytope $\mathrm{STAB}_n$. Figure reproduced from our paper Ref. [1].

ator. The robustness of magic (RoM) was defined by Howard and Campbell [103] to be the minimal $\ell_1$-norm $\|\vec{q}\|_1 = \sum_j |q_j|$ over all valid decompositions [103].

**Definition 2.3** (Robustness of magic (RoM) [103])**.** *Given the density operator $\rho$ for an n-qubit state, the **robustness of magic** $\mathcal{R}$ is defined:*

$$\mathcal{R}(\rho) = \min_{\vec{q}} \left\{ \|\vec{q}\|_1 : \sum_j q_j |\phi_j\rangle\langle\phi_j| = \rho,\, |\phi_j\rangle \in \mathrm{STAB}_n \right\} \tag{2.18}$$

$$= \min\left\{1+2p : (1+p)\rho_+ - p\rho_- = \rho, \quad p \geq 0, \rho_\pm \in \overline{\mathrm{STAB}_n}\right\} \tag{2.19}$$

The robustness of magic is a well-behaved magic monotone, having the following properties:

1. *Convexity*: $\mathcal{R}\left(\sum_j q_j \rho_j\right) \leq \sum_j |q_j| \mathcal{R}(\rho_j)$;

2. *Faithfulness*: If $\rho \in \mathrm{STAB}_n$, then $\mathcal{R}(\rho) = 1$. Otherwise $\mathcal{R}(\rho) > 1$;

3. *Monotonicity under stabiliser operations*: If $\mathcal{E}$ is a stabiliser operation, then $\mathcal{R}(\mathcal{E}(\rho)) \leq \mathcal{R}(\rho)$;

4. *Submultiplicativity under tensor product*: $\mathcal{R}(\rho_A \otimes \rho_B) \leq \mathcal{R}(\rho_A)\mathcal{R}(\rho_B)$.

The RoM can be calculated using standard linear programming techniques [131] (for example using the MATLAB package CVX [132]). The naive formulation of the linear program is practical on a desktop computer for up to five qubits (the number of stabiliser states increases super-exponentially with $n$). It was recently shown by Heinrich and Gross [133] that when states possess certain symmetries, the original optimisation problem can be mapped to a more tractable one, so that RoM can be calculated for up to 10 copies of a state. For example, this strategy can be applied when the state is a $k$-fold tensor product of $T$-states, $\rho = |T\rangle\langle T|^{\otimes k}$. There is a known lower bound which can be directly computed, namely the *stabiliser norm*, defined as follows.

**Definition 2.4** (Stabiliser norm)**.** *Given a normalised n-qubit state $\rho$ the stabiliser norm is defined:*

$$\|\rho\|_{\mathrm{st}} = \frac{1}{2^n} \sum_{P \in \mathcal{P}_+} |\mathrm{Tr}[P\rho]|, \tag{2.20}$$

*where $\mathcal{P}_+$ is the set of unsigned Pauli operators.*

The stabiliser norm is multiplicative, convex, and is a one-way magic witness in the sense that $\|\rho\|_{\text{st}} > 1$ implies that $\rho$ is a magic state, but the converse is not true. It is always the case that $\|\rho\|_{\text{st}} \leq \mathcal{R}(\rho)$, and a tighter lower bound is obtained from the formula:

$$\mathcal{R}(\rho) \geq \frac{\|\rho\|_{\text{st}} - 2^{-n}}{1 - 2^{-n}}. \tag{2.21}$$

Although the stabiliser norm gives rather a loose bound, it is significant that it is multiplicative, since it means that for $n$ copies of a magic state $\rho$ with $\|\rho\|_{\text{st}} > 1$, one has $\mathcal{R}(\rho^{\otimes n}) \geq \|\rho\|_{\text{st}}^{n}$. In other words, though RoM is itself submultiplicative, we cannot expect it to be dramatically so for many-qubit states, as it is lower bounded by a quantity that increases exponentially with $n$.

A further key result of Ref. [103] is that robustness of magic directly quantifies simulation cost for a classical algorithm. Howard and Campbell adapt the approach of Pashayan et al. [124] to classically estimate observables for stabiliser circuits with mixed magic state inputs; rather than sampling over points in phase space, one samples a stabiliser state from the distribution which can then be evolved efficiently using the Gottesman-Knill theorem. Pseudocode is shown in Algorithm 2.

**Algorithm 2** RoM simulator [103]

```
Input: Vector $\mathbf{q}$ describing stabiliser decomposition of an $n$-qubit state, $\rho = \sum_k q_k |\phi_k\rangle\langle\phi_k|$, $|\phi_k\rangle \in \text{STAB}_n$; description of a stabiliser circuit $\mathcal{E}$; stabiliser observable $E$; number of samples $M$
Output: Estimate $\widehat{E}$ for $\text{Tr}[E\mathcal{E}(\rho)]$.
$\widetilde{E} \leftarrow 0$
for $m \leftarrow 1$ to $M$ do
    Sample $|\phi_k\rangle\langle\phi_k|$ from $\{\Pr(k) = \frac{|q_k|}{\|\mathbf{q}\|_1}\}$.
    $\widetilde{E}_m \leftarrow \text{sign}(q_k)\|\mathbf{q}\|_1 \text{Tr}[E\mathcal{E}(|\phi_k\rangle\langle\phi_k|)]$        ▷ Computed using Gottesman-Knill.
    $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_m$
end for
return $\widehat{E} = \widetilde{E}/M$.
```

Again, one can demonstrate the validity of the simulator by showing that the output variable is an unbiased estimator, and that each $\widetilde{E}_m$ is bounded by the $\ell_1$-norm of the distribution, then applying Hoeffding's inequality. We have:

$$\mathbb{E}(\widetilde{E}_m) = \sum_k \Pr(\mathbf{k}) \cdot \operatorname{sign}(q_k) \|\mathbf{q}\|_1 \operatorname{Tr}[E\mathcal{E}(|\phi_k\rangle\langle\phi_k|)] \tag{2.22}$$

$$= \sum_k q_k \operatorname{Tr}[E\mathcal{E}(|\phi_k\rangle\langle\phi_k|)] = \operatorname{Tr}[E\mathcal{E}(\rho)]. \tag{2.23}$$

This shows $\widetilde{E}_m$ is an unbiased estimator for the quantum mean value. To bound $\widetilde{E}_m$, notice that $\operatorname{Tr}[E\mathcal{E}(|\phi_k\rangle\langle\phi_k|)] \in [-1,1]$ when $E$ is a Pauli observable, or in the range $[0,1]$ if a stabiliser projector. Therefore, by inspecting step 4,

$$|\widetilde{E}_m| \leq \max_k |\operatorname{sign}(q_k)\|\mathbf{q}\|_1 \operatorname{Tr}[E\mathcal{E}(|\phi_k\rangle\langle\phi_k|)]| \leq \|\mathbf{q}\|_1. \tag{2.24}$$

Using the Hoeffding inequality (Theorem 2.1), we obtain the following result.

**Theorem 2.5** (RoM simulator performance [103])**.** *Given an $n$-qubit magic state with known decomposition $\rho = \sum_k q_k |\phi_k\rangle\langle\phi_k|$, for $|\phi_k\rangle \in \mathrm{STAB}_n$, where $\{|q_k|/\|\mathbf{q}\|_1\}$ can be sampled from in time* $\mathrm{poly}(n)$*, we can estimate the mean value of a stabiliser observable $E$ up to additive error $\varepsilon$ and with success probability at least $(1-p_{\mathrm{fail}})$ with polynomial space and with runtime at most,*

$$\tau = 2\frac{\|q\|_1^2}{\varepsilon^2}\log\left(\frac{2}{p_{\mathrm{fail}}}\right)T\mathrm{poly}(n). \tag{2.25}$$

*If the known decomposition of $\rho$ is optimal then the runtime is $\mathcal{O}\big(\mathcal{R}(\rho)^2\big)$.*

In practice, when $n$ is large, the exact optimal decomposition is hard to compute. However, it is often the case that the initial state to be simulated is a tensor product of few-qubit magic states. In this case, supposing without loss of generality that $n = bk$ for some small $b$, and arbitrary $k$, one can break the state into blocks of $b$ qubits, $\rho = \bigotimes_{j=1}^{k} \rho_j$ and compute the optimal decomposition for each block. Typically, the composite decomposition will be suboptimal for $n$-qubit state, but will be a valid quasiprobability distribution that can be used in the algorithm. Importantly, this also ensures that the probability distribution $\{|q_k|/\|\mathbf{q}\|_1\}$ is a product distribution so can be efficiently sampled from. In this case the runtime of the algorithm will be $\mathcal{O}\Big(\Pi_{j=1}^{k}\mathcal{R}(\rho_j)^2\Big)$. Since RoM is submultiplicative, it is beneficial to make $b$ as large as is tractable for numerical optimisation.

The framework naturally extends to a subclass of non-stabiliser circuits: those that may be implemented by deterministic state injection [34, 103], including all gates from the third level of the Clifford hierarchy (recall Section 1.4). For such cases, the non-Clifford gate is replaced by a sequence of stabiliser operations taking a resource state $\rho$ as input. This admits classical simulation using the method described above, and each consumed resource $\rho$ contributes a factor $\mathcal{R}(\rho)^2$ to the runtime. It has recently been shown [134] that gadgetisation can be extended to third level diagonal gates subject to depolarising noise. By modifying the state injection gadget, the $n$-qubit pure resource state is replaced by a noisy $2n$-qubit resource state. Taking the $T$-gate as an example, the resource state $|T\rangle$ is first replaced by a two-qubit ideal resource $|T+\rangle = |T\rangle \otimes |+\rangle$ state with modified injection circuit. Depolarising noise is then simulated by applying Pauli errors to ideal resource state $\rho_T = |T+\rangle\langle T+|$, that is,

$$\rho = \left(1 - \frac{3p}{4}\right)\rho_T + \frac{p}{4}(Z_1\rho_T Z_1 + Z_2\rho_T Z_2 + Z_1 Z_2\rho_T Z_1 Z_2). \tag{2.26}$$

The cost of simulation then scales with $\mathcal{R}(\rho)$, exploiting the reduced robustness of the noisy resource state. However, more generally, magic channels cannot necessarily be implemented using *deterministic* state injection, and in this case it is not clear that the robustness of some resource state is the appropriate measure. Recall from Section 1.3 that a channel $\mathcal{E}$ is completely stabiliser-preserving if and only if its Choi state $\Phi_{\mathcal{E}}$ is a stabiliser state. It follows immediately that the RoM of the Choi state is faithful in the sense that $\mathcal{R}(\Phi_{\mathcal{E}}) = 1$ only for stabiliser channels, and is strictly larger otherwise. One might wonder whether this quantity is suitable as a measure of magic for channels. It turns out that it fails to be a well-behaved measure of magic, for reasons that will become clear when we make precise the notion of a channel monotone in Chapter 4. For completeness, we prove some properties in Appendix C.1.

The quasiprobability simulation method of Howard and Campbell [103] is clearly similar in spirit to that of Pashayan et al. [124]. It is natural to ask whether decomposition over stabiliser projectors can be understood through the formalism of frames. Recall that given a fixed frame and a dual frame, a unique quasiprobability distribution is computed for any $\rho$. This is not the case for robustness of magic, where many possible decompositions are possible, and the optimal decomposition must be found by linear optimisation. Therefore if there exists a frame $\{F(\lambda)\}$ that leads to the function $W_\rho(\lambda) = \mathrm{Tr}[F(\lambda)\rho]$

such that $\|W_\rho\|_1 = \mathcal{R}(\rho)$, the choice of frame must be state-dependent. For the interested reader, in Appendix C.2 we show how such an optimal frame can be found starting from any known stabiliser projector frame [135]. However, it is not clear whether much insight or practical advantage is gained by viewing stabiliser projector decompositions from this perspective. One may ask if there exists a more natural phase space picture which provides a suitable frame for the $n$-qubit case. In the next section we argue that while there are ways to modify the phase space construction to this end, their usefulness with respect to practical classical simulation of quantum circuits is limited.

## 2.4 Phase space picture for qubits

At this point the reader may ask why the elegant phase space construction has direct application in quasiprobability simulators for qudits [101, 102, 124], whereas for even dimension we must resort to overcomplete bases such as the set of stabiliser state projectors [103]. Fully understanding the differences between odd- and even-dimension systems is a field of active research [110, 129, 130] outside our scope, but we will sketch some of the issues pertaining to classical simulation. One useful perspective is via contextuality [130]. Given the set of binary observables $O = \{O_j\}_j$ that can be measured on a given system, we call a subset $S \subseteq O$ non-contextual if one can simultaneously make an assignment $\lambda(O_j) \in \pm 1$ for all $O_j \in S$ that is consistent with respect to compatible observables. That is, $\lambda(O_j)\lambda(O_k) = \lambda(O_j O_k)$ for all commuting $O_j, O_k, O_j O_k \in S$. If such an assignment is not possible, then the presence of contextuality is indicated. Contextuality has often been proposed as a signature of non-classicality, and can be studied in a state-dependent or -independent setting [130, 136–139].

For qudits, a non-contextual value assignment is possible for any subset of the Heisenberg-Weyl operators by setting $\lambda(T_{\mathbf{u}}) = 1$ for all $\mathbf{u}$. This follows from the composition law $T_{\mathbf{a}} T_{\mathbf{b}} = \omega^{2^{-1}\langle \mathbf{a},\mathbf{b}\rangle} T_{\mathbf{a}+\mathbf{b}}$, which reduces to $T_{\mathbf{b}} T_{\mathbf{b}} = T_{\mathbf{a}+\mathbf{b}}$ for any pair of *compatible* Heisenberg-Weyl operators, $\langle \mathbf{a},\mathbf{b}\rangle = 0$. In this sense, there is no state-independent contextuality in stabiliser computation for odd $d$ qudits. Indeed, for odd dimension the emergence of state-dependent contextuality has been shown to align with Wigner function negativity and hardness of classical simulation [140, 141]. This argument fails for qubit systems due to the non-existence of $2^{-1}$ in $\mathbb{F}_2$. In fact, state-independent contextuality can be demonstrated for Pauli measurements $n$-qubit systems with $n \geq 2$; this can be proved for two qubits, for example, via the Mermin square [137, 138]. This problematises efforts to portray contextuality as a resource for quantum speedup in the qubit case. It has also been proved that for any definition of the Wigner function following the framework of Gibbons et al. [119], there always exist stabiliser states with negative Wigner function [130]. Several authors have shown that one can recover a consistent picture connecting contextuality, Wigner function positivity and classical simulability for qubits, but only by imposing strict restrictions on the free states and operations in some way. For example, Delfosse et al. [129] introduced a model of computation restricted to rebits at all steps of the quantum circuit, while Raussendorf et al. constrain the free measurements to those preserving positivity [130].

Recent work by Raussendorf et al. [110] has attempted to re-align Wigner function positivity with classical simulability for all stabiliser states by modifying the definition of phase point operators. Standard phase point operators are sums over all Heisenberg-Weyl operators, and differentiated by a phase assignment on each term. The altered definition involves a sum over so-called closed non-contextual (cnc) subsets of the phasespace. Thus the new phase point operators are labelled by a subset and a phase assignment. It turns out that since a non-contextual assigment is always possible in the qudit operators, the only non-trivial cnc set comprises the whole phase space, and the associated Wigner function coincides with the standard definition. In stark contrast, in even dimension there is an explosion in the number of maximal cnc sets, leading to a complicated constellation of new phase point operators. The new basis is massively overcomplete, so that the corresponding Wigner function is non-unique and must be optimised. Raussendorf et al. define a robustness measure for this set analogous to the RoM, and show that their measure is strictly smaller. However, their new definition of phase point operators is not closed under tensor product, which means one cannot easily construct product quasiprobability distributions as was the case in the RoM simulator [103] (Section 2.3). It is therefore not known how to use the modified phase space picture to obtain more than a small constant factor speedup over Howard and Campbell, and the optimisation problem appears to be strictly harder.

## 2.5 Simulating CPTP maps

While the set of operations accessible in the Pashayan simulator [124] is universal, the natural choice of frame for qudits does not carry over to qubits, where the optimal choice of stabiliser frame is state-dependent. This leads to ambiguity when considering a sequence of unitary operations, as the frame that is optimal for a given state or gate is suboptimal for others. Meanwhile, the RoM simulator [103] is able to admit universal quantum circuits by gadgetisation, since, for example, any Clifford+$T$ circuit can be simulated by a stabiliser circuit with injected magic states. From the point of view of classical simulation, this may lead to inefficiency, since typically synthesis of an arbitrary single-qubit rotation requires multiple $T$ gates. It is likely advantageous to simulate each single-qubit gate more directly. Below we review quasiprobability techniques extended to deal with general CPTP maps. In addition to the goal of driving down the simulation cost for arbitrary gates, this is motivated in part by the need to simulate noisy quantum devices, which may involve imperfect gates, intermediate measurements and non-stabiliser noise models.

### 2.5.1 Clifford and Pauli-reset decomposition

In Ref. [45], Bennink et al. proposed an algorithm, which we will call the Oak Ridge simulator, which takes an approach similar in spirit to Howard and Campbell [103]. In this simulator initial states are decomposed as an $\ell_1$-norm-optimised linear combination of stabiliser states, just as in Ref. [103]. However, the Oak Ridge simulator admits a larger class of quantum circuit, since magic-increasing CPTP maps can be directly decomposed without gadgetisation, as a quasiprobability distribution over stabiliser channels from a class we will refer to as CPR. This is the set of Clifford gates, supplemented by Pauli reset channels. The latter are stabiliser-preserving CPTP maps that perform the following sequence of adaptive operations (i) measure the Pauli observable $P$; (ii) if outcome is "$-1$", perform a Clifford correction to map the state to the "$+1$" eigenspace, otherwise do nothing. For a single qubit, the Pauli reset channels are simply the stabiliser circuits that prepare the six pure stabiliser states. Bennink et al. show that the affine span of CPR contains all CPTP maps, so for any quantum channel $\mathcal{E}$ it is always possible to write $\mathcal{E} = \sum_k q_k \mathcal{T}_k$, where $q_k \in \mathbb{R}$ and $\mathcal{T}_k \in$ CPR. We can define a cost function analogous to

RoM,

$$\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}) = \min_{\mathbf{q}} \left\{ \|\mathbf{q}\|_1 : \sum_k q_k \mathcal{T}_k = \mathcal{E}, \quad \mathcal{T}_k \in \mathrm{CPR} \right\} \tag{2.27}$$

In practice the linear program to find an optimal decomposition is only tractable for up to two qubits. The authors therefore assume an $n$-qubit quantum circuit $\mathcal{E}$ is specified as a composition of $L$ two-qubit circuit elements $\mathcal{E}_j$, $\mathcal{E} = \mathcal{E}_L \circ \mathcal{E}_{L-1} \circ \ldots \circ \mathcal{E}_2 \circ \mathcal{E}_1$. A quasiprobability decomposition is found for each circuit element $\mathcal{E}_j = \sum_k q_k^{(j)} \mathcal{T}_k$, and for the initial state $\rho = \sum_k q_k^{(0)} |\phi_k\rangle\langle\phi_k|$. This gives an overall quasiprobability distribution for the output state $\mathcal{E}(\rho) = \sum_{\mathbf{k}} q_{\mathbf{k}} \mathcal{T}_{\mathbf{k}}\left(|\phi_{k_0}\rangle\langle\phi_{k_0}|\right)$, where $\mathbf{k}$ is an $(L+1)$-element vector labelling an initial stabiliser state and trajectory through the circuit,

$$q_{\mathbf{k}} = \prod_{j=0}^{L} q_{k_j}^{(j)}, \quad \text{and} \quad \mathcal{T}_{\mathbf{k}} = \mathcal{T}_{k_L} \circ \mathcal{T}_{k_{L-1}} \circ \ldots \circ \mathcal{T}_{k_1}. \tag{2.28}$$

Renormalising this quasiprobability distribution gives us a product probability distribution which can be sampled from in the same way as for RoM, but the sample cost is now proportional to $\prod_{j=1}^{L} [\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}_j)]^2$. This is consistent with the notion that the simulation cost should be exponential in the number of gates, but also captures the fact that some channels can be represented more efficiently than others.

The caveat here is that CPR is a strict subset of the stabiliser operations; indeed, Bennink et al. acknowledge that they do not consider general adaptive stabiliser operations [45]. We will show explicitly later in this thesis (Section 5.3.1) that there exist stabiliser channels that are not in the convex hull of CPR. Therefore $\mathcal{R}_{\mathrm{CPR}}$ cannot be considered a measure of magic, as it is not faithful with respect to the set of stabiliser operations, and there is no guarantee that a given stabiliser circuit can be efficiently simulated with the Oak Ridge algorithm.

### 2.5.2 Pauli propagation

We close this chapter by discussing a pair of complementary simulation algorithms proposed by Rall et al. [142]. These make use of quasiprobability techniques, but are also reminiscent of the non-Clifford simulator due to Aaronson and Gottesman [31] (Section 1.5), as they involve expansion in the Pauli basis. Rall et al. treat the Pauli expansion of a generic operator $A$ as a quasiprobability distribution. The normalising factor is the stabiliser norm introduced earlier as a lower bound for RoM. That is, we can always write $A = \sum_{P \in \mathcal{P}_+} q_P P$, where $q_P = \mathrm{Tr}[PA]2^{-n}$, and $\mathcal{P}_+$ is the set of unsigned Pauli operators. Recalling that the stabiliser norm is defined $\|A\|_{\mathrm{st}} = 2^{-n} \sum_{P \in \mathcal{P}_+} |\mathrm{Tr}[PA]|$, it is clear that we can normalise the coefficients form a probability distribution. There are two variants of the technique. The first samples a Pauli operator from the distribution, and then propagates it through a sequence of channels $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(T)}$, sampling a new Pauli operator whenever it is mapped to a linear combination of more than one. Finally a sample value for a target observable is obtained, and the process is repeated many times, as usual. In the second variant of the method, an observable is propagated backward through the circuit, and each sample is obtained by evaluating the transformed observable with the initial state. Observables need not be Pauli operators in general, but are assumed to have tensor product structure. The algorithms admit general CPTP maps on bounded number of qubits, subject to overhead related to the stabiliser norm. Rall et al. call the forward and backward propagation variants Schrödinger and Heisenberg propagation, respectively.

For both variants, the algorithm leads to an unbiased estimator for the quantum mean value. Since $\|\rho\|_{\mathrm{st}}$ is a loose lower bound on the RoM for any density operator $\rho$, one might imagine the associated algorithm improves on the RoM simulator. However, the simulators have some unusual properties that mitigate against this. In both cases, the worst-case runtime depends on the overall stabiliser norm maximised over all trajectories. Considering Schrödinger propagation first, note that $|\mathrm{Tr}[EP]|$ can be as large as $2^n$ if $E$ is a Pauli observable. If $E$ is a local observable acting on $k$ qubits, $E' \otimes \mathbb{1}_{n-k}$ then the runtime will be $\mathcal{O}\left(2^{n-k}\right)$, even if $E$ is a stabiliser projector. It therefore seems that the Schrödinger variant is likely only practically useful when all but a constant number of qubits are measured. In addition,when $\mathcal{E}$ is an adaptive (i.e. non-unital) channel, such as a state injection gadget, $\|\mathcal{E}(P)\|_{\mathrm{st}}$ can be larger than 1, even for stabiliser operations.

The Heisenberg variant does not suffer from limitations as severe as Schrödinger propagation. Remarkably, the sampling cost is independent of the input state, since $\mathrm{Tr}[P\rho] \in [-1, +1]$ for any normalised state $\rho$. In this respect, the hardness of simulating magic states is sidestepped in the Heisenberg variant by back-propagating observables to meet the initial magic state, which is assumed to have tensor product structure. It is still in general inefficient for non-unital stabiliser maps, but the possibility is left open that the Heisenberg simulator may perform better than the RoM simulator for certain circuits, for example circuits comprised solely of unitary gates and unital stabiliser-preserving noise channels. Hakkaku and Fujii have recently compared Heisenberg propagation with the RoM simulator for random Clifford + T circuits subject to depolarising noise of strength $p$ [134]. It was found that the technique based on RoM tends to perform better for modest amounts of noise, but there is a crossover where Heisenberg propagation gains the advantage for circuits with large $p$. For random Clifford +T circuits for example, the RoM simulator has the advantage for circuits with error rate smaller than $p \approx 0.06 - 0.11$, depending on the precise configuration. Broadly speaking, the RoM simulator appears to gain the advantage for noisy random circuits where the amount of magic per unit cell is higher, either due to higher density of $T$ gates, or lower error rates.

## 2.6 Summary and outlook

We have reviewed prior work on quasiprobability distributions over "free" elements, and saw that the negativity of such decompositions can be viewed both as a well-behaved measure of magic resource and a quantifier of hardness of classical simulation. We argued that the phase space approach that has been applied successfully for odd-dimension qudits is unsuitable for the study of magic resource and classical simulation in the qubit case. Instead, the robustness of magic was proposed as a magic monotone for general $n$-qubit states [103]. The RoM simulator admits universal quantum circuits via gadgetisation, but this restricts the gate set to Cliffords supplemented by third-level hierarchy gates such as the $T$ gate. This may introduce unnecessary classical simulation overhead, as multiple, and sometimes many, third-level gates are required to synthesise each generic single-qubit rotation, and each state injection gadget requires the initialisation of at least one auxiliary qubit. In the last section of the chapter, we discussed direct decomposition of general CPTP maps, avoiding the cumbersome gadgetisation step. Both the Oak Ridge [45] and Pauli propagation [142] methods were efficient only for a restricted subclass of stabiliser operation, however. This hints at the possibility that further performance improvements can be gleaned by expanding the class of simulable operations.

In Part II of this thesis, we will present novel work showing how faithful measures of magic can be defined for quantum channels, linking these to classical simulation overhead for non-stabiliser circuits. We will also introduce generalisations of the quasiprobability methods reviewed in this chapter leading to provable performance gains [2]. Before presenting this novel work, in the next chapter we will review a class of simulators based on stabiliser rank [3, 59, 60]. Whereas quasiprobability simulators can provide additive error estimates for individual Born rule probabilities, stabiliser rank simulators allow the estimation of Born rule probabilities up to *multiplicative* error. This makes them suitable for emulating the ability of a quantum computer to randomly draw a bit-string from the full distribution. We will see that classical simulation cost can be linked to a magic resource monotone in this context as well.

## Chapter 3

# Stabiliser rank methods

## 3.1 State vector decompositions

Just as a density operator can be decomposed as a linear combination of pure stabiliser state projectors [103], any state vector can be expressed as a superposition of stabiliser states [143]. The exact stabiliser rank $\chi(\psi)$ [3, 59, 60] for a given state $\psi$ is the smallest number of stabiliser terms needed to specify the state. Loosely speaking, stabiliser rank simulators work by combining Clifford simulation results computed separately for each stabiliser term, so that the runtime is a function of the number of terms $\chi$. Various techniques have been proposed for reducing this runtime by replacing the exact decomposition with a sparsified proxy [3, 60]. In particular, Bravyi et al. [3] introduced a sparsification technique that leads to a classical simulation algorithm where the runtime is quantified by a well-behaved continuous monotone called stabiliser extent. We will refer back to Ref. [3] many times throughout this thesis, especially in Chapter 5, where we extend the sparsification techniques, so we will use the shorthand BBCCGH to denote this paper (after the authors Bravyi, Browne, Campbell, Calpin, Gosset and Howard). In this chapter, we first review work on stabiliser decompositions of pure magic states, before turning to the associated simulators.

## 3.2 Exact stabiliser rank

The question of whether arbitrary pure states on $n$ qubits can be written as a subexponential superposition of quantum states was posed by Garcia, Markov and Cross [143], motivated by the fact that while the dimension of Hilbert space grows as $2^n$, the number of $n$-qubit stabiliser states grows faster ($\Omega(2^{n^2/2})$). However, they found that one can construct families of tensor product states $|\Psi\rangle = |\psi\rangle^{\otimes n}$, such that the overlap with any polynomial-sized superposition of orthogonal stabiliser states vanishes as $n \to \infty$. For these "stabiliser-evading" families of states, $|\Psi\rangle$ cannot be expressed with a polynomial number of terms in any stabiliser basis, even approximately. However, whether the number of terms could be reduced by allowing superpositions of non-orthogonal stabiliser states was left an open question. Bravyi, Smith and Smolin used this idea to define a quantifier of non-stabiliserness.

**Definition 3.1** (Stabiliser rank). *Given an n-qubit pure state $|\psi\rangle$, the* ***stabiliser rank*** *is defined:*

$$\chi(\psi) = \min\left\{k : |\psi\rangle = \sum_{j=1}^{k} c_j |\phi_j\rangle, \quad |\phi_j\rangle \in \mathrm{STAB}_n\right\}. \tag{3.1}$$

Since stabiliser rank is defined only for pure states, the notion of monotonicity is only well-defined for purity-preserving operations. Purity-preserving stabiliser operations such as Clifford gates will update each stabiliser term, but will not increase the number of them, so in this sense stabiliser rank is a pure-state magic monotone. It is also submultiplicative, since given exact decompositions of states $\psi$ and $\phi$, with $\chi(\psi)$ and $\chi(\phi)$, we can always form a valid, but possibly sub-optimal decomposition of $|\psi\rangle \otimes |\phi\rangle$ with $\chi(\psi)\chi(\phi)$ terms, so that $\chi(\psi \otimes \phi) \leq \chi(\psi)\chi(\phi)$.

Upper bounds for the stabiliser rank can be obtained by finding explicit decompositions for given states. In Ref. [59], Bravyi, Smith and Smolin give a numerical method for finding stabiliser decompositions. The method is based on a random walk over $k$-tuples of stabiliser states, $(\phi_1, \ldots, \phi_k)$. If a state is left invariant by projection onto the subspace spanned by some $k$-tuple, then $k$ must upper bound the stabiliser rank. Upper bounds for tensor powers $|H\rangle^{\otimes n}$ of the Hadamard state $|H\rangle = \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$ (Clifford-equivalent to the $T$ state), were initially found up to the $n = 6$ case [59]. This was extended to $n = 7$ by BBCCGH [3], where numerical searches were also carried out for tensor products of Haar random single qubit states $|\psi\rangle^{\otimes n}$ (Table 3.1). Much more recently Kocia [144] found an improved upper bound for the $n = 12$ case. Computing stabiliser rank exactly becomes intractable for many-qubit states, but one can compute upper bounds by breaking tensor product states up into blocks of $m$ or fewer qubits, where $|\psi\rangle^{\otimes m}$ has a known decomposition. For example, using Kocia's 12-qubit upper bound $\chi(|H\rangle^{\otimes 12}) \leq 47$, then for $n = 12m$ one obtains $\chi(|H\rangle^{\otimes n}) \leq \chi(|H\rangle^{\otimes 12})^{\frac{n}{12}} = \mathcal{O}(2^{0.463n})$.

| Number of qubits, $n$ | Hadamard state $\vert H\rangle^{\otimes n}$ | Generic $n$-qubit magic state $\vert\psi\rangle^{\otimes n}$ |
|---|---|---|
| 1 | 2 | 2 |
| 2 | 2 | 3 |
| 3 | 3 | 4 |
| 4 | 4 | 5 |
| 5 | 6 | 6 |
| 6 | 6 | 8 |
| 7 | 12 | 14 |
| 8 | 12 | - |
| 12 | 47 | - |

**Table 3.1:** Selection of known upper bounds for stabiliser rank for tensor products of Hadamard states and Haar random states. [3, 59]

However, Qassim et al. [145] have very recently discovered a method for directly obtaining decompositions for $m$-qubit product states, giving the state-of-the-art upper bound of $\chi(|H\rangle^{\otimes m}) = \mathcal{O}\big(2^{0.3963m}\big)$. BBCCGH obtained an analytic upper bound for generic $nm$-qubit states of the form $|\psi\rangle^{\otimes m}$ for $m \leq 5$, and where $|\psi\rangle$ is an $n$-qubit state

$$\chi(|\psi\rangle^{\otimes m}) \leq \begin{pmatrix} 2^n+m-1 \\ m \end{pmatrix}. \tag{3.2}$$

In their recent paper, Qassim et al. gave a new asymptotic upper bound for the $n = 1$ case, $\chi(|\psi\rangle^{\otimes m}) = \mathcal{O}\big((m+1)2^{\frac{m}{2}}\big)$.

Tight lower bounds have proved difficult to obtain. Bravyi, Smith and Smolin [59] previously showed that $\chi(|H\rangle^{\otimes m}) = \Omega(m^{1/2})$, and Peleg et al. [146] have recently improved this bound to $\Omega(m)$. However, it is expected that in the limit of large $m$ the stabiliser rank should scale superpolynomially. In Section 3.5 we explain how one can simulate sampling from a stabiliser circuit with magic state injection in time $\mathcal{O}(\chi(\psi))$, where $|\psi\rangle$ is the injected magic state [3, 60]. Since this class of circuits is universal for quantum computation, polynomial scaling of stabiliser rank would imply **BQP** $=$ **BPP** [3]. In any case, the presently known suboptimal stabiliser decompositions for $m$ copies of the Hadamard state described above lead to a simulation method for Clifford + T circuits with runtime $\mathcal{O}(\chi) = \mathcal{O}\big(2^{0.3963m}\big)$, where $m$ is the number of T-gates [59]. The desire to drive down simulation overhead for these circuits has motivated several authors to address the question of whether and to what extent the number of stabiliser terms can be reduced by relaxing the requirement for an exact decomposition, and instead allowing superpositions of stabiliser states that approximate the target state. We discuss this next.

## 3.3 Approximate stabiliser rank and stabiliser extent

The notion of an *approximate stabiliser rank* was first introduced by Bravvi and Gosset [60], and developed further by BBCCGH [3]. Bravyi and Gosset [60] gave a randomised algorithm for generating an approximation $|\mathcal{L}^*\rangle$ to the magic state $|H\rangle^{\otimes n}$ up to arbitrary error $\delta$ in the fidelity, $\left|\langle\mathcal{L}^*|H\rangle^{\otimes t}\right|^2 \geq 1-\delta$. The method is based on writing $|H\rangle^{\otimes t}$ as a summation over terms $|\widetilde{x}_1 \otimes \ldots \otimes \widetilde{x}_t\rangle$ in a non-orthogonal basis, where $|\widetilde{0}\rangle = |0\rangle$ and $|\widetilde{1}\rangle = |+\rangle$. The exact summation is over all $2^t$ elements of the linear space $\mathbb{F}_2^t$; the approximation strategy is to randomly generate $k$-dimensional subspaces $\mathcal{L} \subset \mathbb{F}_2^t$ until one with the desired fidelity is found. They showed that an approximation of rank $\mathcal{O}\big(2^{0.23t}\delta^{-1}\big)$ can be obtained in time $\mathcal{O}\big((\cos\big(\frac{\pi}{8}\big))^{-2t}\delta^{-2}\big)$. In Ref. [3], BBCCGH defined the approximate stabiliser rank for the general case.

**Definition 3.2** (Approximate stabiliser rank [3])**.** *Given n-qubit state $\psi$, and precision parameter $\delta > 0$, the* ***approximate stabiliser rank*** *$\chi_\delta$ is defined,*

$$\chi_\delta(\psi) = \min_{\widetilde{\psi}}\{\chi(\widetilde{\psi}) : \|\psi - \widetilde{\psi}\| \leq \delta\}. \tag{3.3}$$

Note that this differs slightly from an earlier definition due to Bravyi and Gosset [60], where the error $\delta$ was specified with respect to fidelity rather than norm $\|\psi - \widetilde{\psi}\|$. In Section 3.4 we will see that the revised definition is convenient for a sparsification procedure introduced by BBCCGH, applicable to generic pure states with known stabiliser decomposition [3]. Computing the approximate stabiliser rank exactly is hard, but it can be upper bounded using sparsification. Using this method, for a pure state with decomposition $|\psi\rangle = \sum c_j \left|\phi_j\right\rangle$, for $\left|\phi_j\right\rangle \in \mathrm{STAB}_n$ the approximate stabiliser rank satisifies,

$$\chi_\delta(\psi) \leq 1 + \frac{\|\mathbf{c}\|_1^2}{\delta^2}, \tag{3.4}$$

where $\|\mathbf{c}\|_1 = \sum_j \left|c_j\right|$ is the $\ell_1$-norm of the vector of coefficients $\mathbf{c}$. This motivates the definition of *stabiliser extent*, a continuous pure-state measure of magic that turns out to be easier to work with than stabiliser rank.

**Definition 3.3** (Pure-state stabiliser extent [3])**.** *Given an n-qubit pure state $|\psi\rangle$, the*

***pure-state stabiliser extent*** *$\xi$ is given by*

$$\xi(\psi) = \min\{\|\mathbf{c}\|_1^2 : \sum_j c_j |\phi_j\rangle = |\psi\rangle, \quad |\phi_j\rangle \in \mathrm{STAB}_n\} \tag{3.5}$$

The stabiliser extent is invariant under Clifford operations, and can be computed as a convex optimisation problem. By strong duality, we can recast the problem in terms of witnesses. We must first define *stabiliser fidelity*, the maximum fidelity between a target state and any stabiliser state [3], $F(\psi) = \max_{|\phi\rangle \in \mathrm{STAB}_n} |\langle\psi|\phi\rangle|^2$. The dual form of the problem can then be expressed as

$$\xi(\psi) = \max_{|\omega\rangle} \frac{|\langle\psi|\omega\rangle|^2}{F(\omega)} \tag{3.6}$$

This immediately leads to a lower bound; for any $\psi$ we must have $\xi(\psi) \geq F(\psi)^{-1}$. Solving the optimisation problem directly is tractable only for a few qubits. In contrast to robustness of magic, however, this is ameliorated greatly by the multiplicativity of stabiliser extent for an important class of states, as follows.

**Theorem 3.4** (Multiplicativity of pure-state stabiliser extent [3])**.** *Suppose $|\psi\rangle = \bigotimes_{j=1}^{L} |\psi_j\rangle$ is a tensor product state such that each $|\psi_j\rangle$ is a state on at most three qubits. Then:*

$$\xi(\psi) = \prod_{j=1}^{L} \xi(\psi_j). \tag{3.7}$$

This means that the stabiliser extent can be efficiently computed for any product state, provided each tensor factor is on no more than three qubits. For example, we can exactly compute the extent for the magic resource state needed to implement any Clifford + T circuit. This is all the more remarkable, since it has recently been proven that stabiliser extent cannot be multiplicative in general [147]; the property must vanish at some number of qubits greater than three.

Importantly, a number of the tools developed in Ref. [3] can be carried over to decompositions of unitary operations. This will be useful when we consider stabiliser rank simulation techniques for general quantum circuits, as it means that gadgetisation can often be avoided, eliminating the need to re-synthesise arbitrary qubit rotations in terms of $T$ gates, and potentially reducing simulation overhead. The extent for unitary operators is defined in terms of Clifford gates rather than stabiliser states.

**Theorem 3.5** (Stabiliser extent for unitary operators [3]). *Given an n-qubit unitary $U$, the **unitary stabiliser extent** is defined,*

$$\xi(U) = \min\{\|\mathbf{c}\|_1^2 : \sum_j c_j V_j = U,\ V_j \in \mathrm{Cl}_n\}. \tag{3.8}$$

This can be connected to the stabiliser extent for resource states $|V\rangle = V|+^t\rangle$ for diagonal gates $V$, via the *lifting lemma.*

**Lemma 3.6** (Lifting lemma [3]). *Let $|V\rangle$ be the resource state for a diagonal $t$-qubit unitary $V$, with decomposition $|V\rangle = \sum_j c_j |\phi_j\rangle$ over equatorial stabiliser states $|\phi_j\rangle = W_j|+^t\rangle$, where $W_j$ are diagonal Clifford gates. Then $V = \sum_j c_j W_j$, so that $\xi(V) \leq \|\mathbf{c}\|_1^2$. If $\|\mathbf{c}\|_1^2 = \xi(|V\rangle)$ then $\xi(|V\rangle) = \xi(V)$.*

This allows stabiliser rank simulation methods to be applied directly to circuits involving arbitrary phase gates, without the need for Clifford + T compilation and gadgetisation. We will shortly discuss these classical simulation techniques, but first we consider the sparsification procedure already mentioned.

## 3.4 Sparsification

In the context of stabiliser rank methods, sparsification refers to techniques for reducing the number of terms in a stabiliser decomposition. We already discussed the method of approximating $|H\rangle^{\otimes t}$ by randomly sampling linear subspaces of dimension smaller than $t$, introduced by Bravyi and Gosset [60]. That method was tailored to the specific structure of that particular type of magic state. Other classes of state lend themselves to similarly tailored approximation techniques. For example, consider tensor product states $|\theta\rangle^{\otimes t}$, where $|\theta\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. If $\theta$ is sufficiently close to zero, a reasonable approximation strategy is to expand $|\theta\rangle^{\otimes t}$ in the standard basis and truncate amplitudes below a certain cutoff. This yields an upper bound for approximate stabiliser rank of order $2^{h_2(\cos^2(\theta/2))}$, with $h_2$ being binary entropy [3].

We will now discuss a sparsification method, introduced by BBCCGH [3], which can be applied to any known exact decomposition, and is particularly effective when used with decompositions that are optimal with respect to the stabiliser extent. The sparsification procedure is described in Algorithm 3.

**Algorithm 3** BBCCGH sparsification procedure. [3]

```
Input: Decomposition of $n$-qubit pure state $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, where $|\phi_j\rangle \in \mathrm{STAB}_n$; target number of terms $k$.
Output: Sparsified vector $|\Omega\rangle$ with $k$ terms.
function SPARSIFY($|\psi\rangle$,$k$)
    for $\alpha \leftarrow 1$ to $k$ do
        Sample index $j_\alpha$ from probability distribution $\{P_j = |c_j|/\|c\|_1\}$
        $|\omega_\alpha\rangle \leftarrow (c_{j_\alpha}/|c_{j_\alpha}|)\,|\phi_{j_\alpha}\rangle$.
    end for
    $|\Omega\rangle \leftarrow (\|\mathbf{c}\|_1/k)\sum_{\alpha=1}^{k}|\omega_\alpha\rangle$
    return $|\Omega\rangle$
end function
```

Thus the subroutine SPARSIFY outputs randomised $k$-term vectors $|\Omega\rangle$, which may in general be unnormalised. The utility of this procedure as an approximation method is established by the *sparsification lemma* [3], which we restate here.

**Lemma 3.7** (BBCCGH sparsification lemma)**.** *Given an n-qubit state with known decomposition $|\psi\rangle = \sum_j c_j |\phi_j\rangle$ where $|\phi_j\rangle \in \mathrm{STAB}_n$, the distribution of random vectors $|\Omega\rangle$ output by* SPARSIFY$(|\psi\rangle, k)$ *satisfies:*

$$\mathbb{E}(\||\psi\rangle - |\Omega\rangle\|^2) \leq \frac{\|\mathbf{c}\|_1^2}{k} \tag{3.9}$$

In Chapter 5 we show how sparsification can be extended to density operators, and for comparison it is intructive to evaluate error with respect to the trace-norm. In Appendix E.1 we prove the following simple corollary to Lemma 3.7.

**Corollary 3.8** (BBCCGH sparsification trace-norm error)**.** *Given a normalised n-qubit state $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, for any $k > 0$, one can sample from a distribution of sparsified vectors $|\Omega\rangle$ using the procedure* SPARSIFY$(|\psi\rangle, k)$*, such that:*

$$\mathbb{E}(\| |\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega| \|_1) \leq 2\frac{\|\mathbf{c}\|_1}{\sqrt{k}} + \frac{\|\mathbf{c}\|_1^2}{k}. \tag{3.10}$$

Lemma 3.7 implies that the average squared norm error in vectors output from SPARSIFY can be made arbitrarily small by choosing a large enough number of terms $k$. Specifically, $\mathbb{E}(\||\psi\rangle - |\Omega\rangle\|^2) \leq \delta^2$ can be satisfied for any $\delta > 0$ by setting $k = \lceil(\|\mathbf{c}\|_1/\delta)^2\rceil$. This immediately yields the upper bound for approximate stabiliser rank $\chi_\delta(\psi)$ given in equation (3.4). With the above choice for $k$, Lemma 3.7 guarantees that there exists at least one $|\Omega\rangle$ such that $\||\psi\rangle - |\Omega\rangle\| \leq \delta$. It follows that $\chi_\delta(\psi) \leq k \leq 1 + \|\mathbf{c}\|_1^2\delta^{-2}$. This is true for any decomposition optimal with respect to extent, so $\chi_\delta(\psi) \leq 1 + \xi(\psi)\delta^{-2}$. This provides a computable upper bound for the approximate stabiliser rank, though it is known not to be tight in all cases [3]. Since we will use similar proof techniques in Section 5.2.3.1, it is useful to reproduce the proof of the sparsification lemma [3] here.

*Proof.* Given a state with decomposition $\sum_j c_j |\phi_j\rangle$, which we assume to be normalised, the for-loop of Algorithm 3 generates i.i.d. random vectors $|\omega_\alpha\rangle$ such that $P_j = \Pr\big(|\omega_\alpha\rangle = (c_j/|c_j|)\,|\phi_j\rangle\big) = |c_j|/\|\mathbf{c}\|_1$. This vector has expected value,

$$\mathbb{E}(|\omega_\alpha\rangle) = \sum_j P_j \frac{c_j}{|c_j|}|\phi_j\rangle = \sum_j \frac{c_j}{\|\mathbf{c}\|_1}|\phi_j\rangle = \frac{|\psi\rangle}{\|\mathbf{c}\|_1}. \tag{3.11}$$

Then for a random sparsification , $\mathbb{E}(|\Omega\rangle) = \frac{\|\mathbf{c}\|_1}{k}\sum_{\alpha=1}^k \mathbb{E}(|\omega_\alpha\rangle) = |\psi\rangle$. We want to compute the expected value of

$$\||\psi\rangle - |\Omega\rangle\|^2 = \langle\psi|\psi\rangle - \langle\psi|\Omega\rangle - \langle\Omega|\psi\rangle + \langle\Omega|\Omega\rangle. \tag{3.12}$$

Since $|\psi\rangle$ is normalised, $\mathbb{E}(\langle\psi|\psi\rangle) = 1$. We also have $\mathbb{E}(\langle\psi|\Omega\rangle) = \langle\psi|\mathbb{E}(|\Omega\rangle) = \langle\psi|\psi\rangle = 1$. To evaluate $\mathbb{E}(\langle\Omega|\Omega\rangle)$ we expand,

$$\begin{aligned}\langle\Omega|\Omega\rangle &= \frac{\|\mathbf{c}\|_1^2}{k^2}\sum_{\alpha=1}^k\sum_{\beta=1}^k \langle\omega_\alpha|\omega_\beta\rangle & (3.13)\\ &= \frac{\|\mathbf{c}\|_1^2}{k^2}\sum_{\alpha=1}^k \langle\omega_\alpha|\omega_\alpha\rangle + \frac{\|\mathbf{c}\|_1^2}{k^2}\sum_{\alpha=1}^k\sum_{\beta\neq\alpha}\langle\omega_\alpha|\omega_\beta\rangle & (3.14)\end{aligned}$$

Since $|\omega_\alpha\rangle$is normalised, the first summation contributes a constant value of $\|\mathbf{c}\|_1^2/k$. In the second term, $|\omega_\alpha\rangle$ and $|\omega_\beta\rangle$ are independent, so

$$\mathbb{E}(\langle\omega_\alpha|\omega_\beta\rangle) = \mathbb{E}(\langle\omega_\alpha|)\mathbb{E}(|\omega_\beta\rangle) = \frac{\langle\psi|\psi\rangle}{\|\mathbf{c}\|_1^2} = \frac{1}{\|\mathbf{c}\|_1^2}. \tag{3.15}$$

Since there are $k(k-1)$ terms in the second summation, we obtain

$$\mathbb{E}(\langle\Omega|\Omega\rangle) = \frac{\|\mathbf{c}\|_1^2}{k} + \frac{k(k-1)}{k^2} = \frac{\|\mathbf{c}\|_1^2}{k} + 1 - \frac{1}{k}. \tag{3.16}$$

Then we have:

$$\mathbb{E}(\||\psi\rangle - |\Omega\rangle\|^2) = -1 + \mathbb{E}(\langle\Omega|\Omega\rangle) = \frac{\|\mathbf{c}\|_1^2}{k} - \frac{1}{k} \leq \frac{\|\mathbf{c}\|_1^2}{k} \tag{3.17}$$

This completes the proof. □

The BBCCGH sparsification lemma shows that sparsified vectors are close to a target state on average. In itself, this does not preclude the possibility of occasionally obtaining $|\Omega\rangle$ that are poor estimates for $|\psi\rangle$. One can check numerically that this is rare, but it is preferable to analytically bound the probability of obtaining such outliers, not least because we need to understand how this could affect the accuracy of simulation algorithms using a sparsification step. Bravyi *et al.* addressed this issue with the sparsification tail bound [3].

**Lemma 3.9** (Sparsification tail bound [3]). *Suppose we set the number of terms $k$ in* SPARSIFY *so that $k \geq \|\mathbf{c}\|_1^2/\delta^2$. Then the following inequalities hold.*

$$\mathbb{E}(\langle\Omega|\Omega\rangle - 1) \leq \delta^2 \tag{3.18}$$

$$\Pr\left\{\||\psi\rangle - |\Omega\rangle\|^2 \leq \langle\Omega|\Omega\rangle - 1 + \delta^2\right\} \geq 1 - 2\exp\left(-\frac{\delta^2}{8F(\psi)}\right) \tag{3.19}$$

*Here $F(\psi) = \max_{|\phi\rangle\in\mathrm{STAB}_n} |\langle\psi|\phi\rangle|^2$ is the stabiliser fidelity.*

Note that the usefulness of the tail bound depends on $\langle\Omega|\Omega\rangle$ being close to (or smaller than) 1. Though equation (3.18) makes $\langle\Omega|\Omega\rangle \gg 1$ rather unlikely, in principle it can be as large $\|\mathbf{c}\|_1^2$. Normalising $|\Omega\rangle$ does not solve the problem, as the BBCCGH sparsification results only give guarantees about the closeness of the unnormalised vector. The issue can be avoided by a post-selection step where we estimate $\langle\Omega|\Omega\rangle$ (e.g. using the techniques described below in Section 3.5) and then discard if we find $\langle\Omega|\Omega\rangle - 1 \gg \delta^2$.

A second difficulty can arise in cases where the stabiliser fidelity $F(\psi)$ is not small. According to inequality (3.19), for fixed $\delta$, the probability of failure (that is, of obtaining $|\Omega\rangle$ that fails to satisfy the condition in braces in (3.19)) is at most $2e^{-\frac{1}{\mathcal{O}(F)}}$. This upper bound can become non-negligible if the stabiliser fidelity is not sufficiently small. Bravyi et al. [3] argue that for important classes of states, the stabiliser fidelity is exponentially small in the number of qubits $n$. Let us unpack this by considering a specific case. Assume that $\langle\Omega|\Omega\rangle \approx 1$, so that the expression in square brackets in (3.19) is $\||\psi\rangle - |\Omega\rangle\|^2 \lesssim \delta^2$. From (3.19), a target error $\delta$ is compatible with a given success probability $p$ only if the stabiliser fidelity satisfies

$$F(\psi) \leq \frac{\delta^2}{8\log\left(\frac{2}{1-p}\right)}. \tag{3.20}$$

To make this concrete, let us use the modest assumptions that we want error $\delta$ no larger than 0.1, and success probability better than 1/2. This can be achieved only when $F(\psi) \lesssim 0.0009$. For $t$-fold tensor products of the $T$-state $|T\rangle$, stabiliser fidelity is multiplicative [3], so $F(|T\rangle^{\otimes t}) \approx (0.854)^t$. It follows that (3.20) is satisfied only when we have at least 45 copies of $|T\rangle$. The minimum value of $t$ needed to satisfy (3.20) increases for more stringent $\delta$ and $p$. Curiously, the sparsification tail bound also seems to suggest worse performance for states containing *less* magic, as quantified by stabiliser fidelity. For example, if instead of the $\pi/8$-state $|T\rangle$ we consider $t$-fold tensor products of the $\pi/32$-state $|\pi/32\rangle = (|0\rangle + e^{i\pi/16}|1\rangle)/\sqrt{2}$, we need at least $t \approx 1200$ before (3.19) gives a non-trivial lower bound on $p$. Therefore there is a large class of interesting intermediate-sized quantum circuits for which the BBCCGH sparsification tail bound cannot be applied. In Chapter 5, Section 5.2.3.1, we will show how to sidestep these difficulties by considering the difference in the trace norm between $|\psi\rangle\langle\psi|$ and the *ensemble* $\rho_1 = \sum_\Omega \Pr(\Omega)\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}$ from which sparsified vectors are drawn, rather than fidelity of the $|\psi\rangle$ with any particular sampled $|\Omega\rangle$.

## 3.5 Inner products, exponential sums and norms

We next review methods that take advantage of stabiliser rank decompositions to simulate quantum processes involving non-stabiliser resources, at the cost of some overhead. In Section 3.6 we discuss integration of these techniques into quantum circuit simulations. Here, we first consider methods that have been developed for estimating inner products and norms for states with stabiliser decompositions.

Bravyi, Smith and Smolin [59] proposed an algorithm for classically simulating Pauli-based computations (PBC), a model where a system is initialised with an $m$-fold product magic state $|H\rangle^{\otimes m}$, and a sequence of adaptive Pauli measurements are carried out. Polynomial-time classical processing is applied to the measured eigenvalues to output a single bit. They showed that any $n$-qubit Clifford + T circuit with $m$ T-gates can first be converted into an $n+m$-qubit generalised PBC initialised in the state $|0\rangle^{\otimes n}|H\rangle^{\otimes m}$. This PBC can in turn be simulated by repeating a PBC on $m$ qubits supplemented by classical processing. The stated aim of Ref. [59] was to show that an $(n+k)$-qubit quantum circuit can be simulated by an $n$-qubit quantum circuit with additional classical processing taking time $2^{\mathcal{O}(k)}\mathrm{poly}(n)$, but our main interest here is in their algorithm for simulating a PBC using classical computation alone, which is able to simulate any $m$-qubit PBC in time $2^{\alpha m}\mathrm{poly}(n)$, where $\alpha \approx 0.94$. Given a $\chi$-term stabiliser decomposition of the initial state, $|\psi\rangle = \sum_{j=1}^{\chi} c_j |\phi_j\rangle$, one can express the probability of obtaining a string of measurement outcomes $\{o_1,\ldots,o_t\}$, where $o_i = \pm 1$, as $\Pr(o_1,\ldots,o_t) = \sum_{j,k} c_j^* c_k \langle\phi_j|\Pi_{o_1\ldots o_t}|\phi_k\rangle$, where $\Pi_{o_1\ldots o_t}$ is a product of $t$ Pauli projectors. Since there are $\chi^2$ terms in the summation, the total runtime will be $\mathcal{O}(\chi^2\tau)$, where $\tau$ is the time to evaluate each $\langle\phi_j|\Pi_{o_1\ldots o_t}|\phi_k\rangle$. It was already known that a stabiliser inner product could be evaluated in $\mathcal{O}(n^3)$ time. Bravyi, Smith and Smolin [59] gave a procedure to compute $\langle\phi_j|\Pi_{o_1\ldots o_t}|\phi_k\rangle$ with same $n^3$ scaling, but avoiding having to simulate all $t$ Pauli projections explicitly. Instead they used the affine space formalism developed by Dehaene, De Moor and Van den Nest [54, 61] to show that for a pair of stabiliser states $|\phi\rangle$ and $|\phi'\rangle$, and a stabiliser group $\mathcal{S}$, one can write $\langle\psi|\Pi_{\mathcal{S}}|\phi\rangle = 2^{-n+t}\sum_{\mathbf{x}\in\mathbb{F}_2^n}\omega^{F(\mathbf{x})}$, where $F$ is a degree-two polynomial which depends only on parameters specifying $\Pi_{\mathcal{S}}$, $\psi$ and $\phi$. They showed that the sum is equal to either 0 or $2^{p/2}\omega^m$, for $m \in \mathbb{F}_8$ and $p \in \{n, n+1, \ldots, 2n\}$, and that this can be computed in time $\mathcal{O}(n^3)$. The time to compute the outcome probability for a single string will be $\mathcal{O}(\chi^2 n^3)$. The probability of obtaining a given outcome at the $t$-th step of a PBC is conditional on the $t-1$ previous outcomes, so to exactly simulate one run of a PBC requires computing $\mathcal{O}(n)$ probabilities, with total runtime $\mathcal{O}(\chi^2 n^4)$. In practice the exact stabiliser rank of $|\psi\rangle = |H\rangle^{\otimes n}$ will not be known, so a sub-optimal decomposition can be used. Using the result of Qassim et al. [145], the runtime to simulate a PBC with $n$ copies of $|H\rangle$ will be $\mathcal{O}((2^{0.3963n})^2 n^4) \approx \mathcal{O}(2^{0.79n} n^4)$.

The algorithm just described computes probabilities exactly. Bravyi and Gosset [60] showed that the runtime for computing the probability $\langle\psi|\Pi|\psi\rangle = ||\Pi|\psi\rangle||^2$ can be reduced from quadratic to linear in stabiliser rank $\chi$, by tolerating a multiplicative error $\varepsilon$. The error can be made arbitrarily small, at the cost of runtime factor $\mathcal{O}(\varepsilon^{-2})$. Since the runtime for circuits with significant magic resource will typically be dominated by the expected exponential scaling of $\chi$, in many scenarios $\mathcal{O}\left(\chi\varepsilon^{-2}\right)$ scaling will be cheaper than the $\mathcal{O}\left(\chi^2\right)$ scaling for exact computation, even at relatively high precision. Bravyi and Gosset exploit the fact that stabiliser states form a projective 2-design [148, 149],

$$\frac{1}{|\mathrm{STAB}_n|}\sum_{|\phi\rangle\in\mathrm{STAB}_n}|\phi\rangle\langle\phi|^{\otimes 2}=\int d\mu(\psi)\,|\psi\rangle\langle\psi|^{\otimes 2}, \tag{3.21}$$

where $\mu$ denotes the Haar measure. It follows that, if we let $|\phi_i\rangle$ denote a stabiliser state drawn uniformly at random from $\mathrm{STAB}_n$, we have:

$$\mathbb{E}(|\langle\phi_i|\psi\rangle|^2)=\frac{\||\psi\rangle\|^2}{2^n},\quad \mathbb{E}(|\langle\phi_i|\psi\rangle|^4)=\frac{2\||\psi\rangle\|^4}{2^n(2^n-1)}. \tag{3.22}$$

One can therefore estimate the norm $\||\psi\rangle\|^2$ by randomly sampling stabiliser states $|\phi_i\rangle$ to generate the random variable $\eta=\frac{2^n}{L}\sum_{i=1}^{L}|\langle\phi_i|\psi\rangle|^2$ . From (3.22), it follows that $\mathbb{E}(\eta)=\||\psi\rangle\|^2$ and $\mathrm{Var}(\eta)<\frac{\|\psi\|^4}{L}$. Using Chebyshev's inequality, by choosing $L=\lceil 4\varepsilon^{-2}\rceil$ the following holds with probability at least $3/4$.

$$(1-\varepsilon)\|\psi\|^2\le\eta\le(1+\varepsilon)\|\psi\|^2. \tag{3.23}$$

One can then amplify the success probability by taking the median of means. Taking $M$ estimates $\{\eta_1,\eta_2,\ldots,\eta_M\}$, the median satisfies equation (3.23) with probability at least $1-p_{\text{fail}}$, provided we choose $M=\mathcal{O}(\log(1/p_{\text{fail}}))$ [150]. In total we need to compute $LM=\mathcal{O}\left(\varepsilon^{-2}\log\left(p_{\text{fail}}^{-1}\right)\right)$ terms $|\langle\phi_i|\psi\rangle|^2$. Assuming a $k$-term decomposition of $|\psi\rangle$, this means computing $\mathcal{O}\left(k\varepsilon^{-2}\log\left(p_{\text{fail}}^{-1}\right)\right)$ stabiliser inner products $\langle\phi_a|\phi_b\rangle$. Bravyi and Gosset proposed another exponential sum method for doing so, outputting a triple of integers $(r,s,t)$ such that $\langle\phi_a|\phi_b\rangle=r\cdot 2^{s/2}\cdot e^{i\pi t/4}$ in time $\mathcal{O}\left(n^3\right)$. They also gave a method for choosing stabiliser states uniformly at random using the affine space description, taking time $\mathcal{O}\left(n^2\right)$ on average, but $\mathcal{O}\left(n^3\right)$ in the worst case. Overall, the total worst-case runtime for the fast norm estimation procedure is given by

$$\tau_{\text{FN}}=\mathcal{O}\left(kMLn^3\right)=\mathcal{O}\left(\frac{kn^3}{\varepsilon^2}\log\frac{1}{p_{\text{fail}}}\right). \tag{3.24}$$

If the exact stabiliser rank decomposition were used ($k=\chi(\psi)$) the runtime would be $\mathcal{O}\left(\frac{\chi n^3}{\varepsilon^2}\log\frac{1}{p_{\text{fail}}}\right)$.

BBCCGH refined this procedure in two ways [3]. First, they showed that it is unnecessary to sample from the full set of stabiliser states, as one can use the restricted set of *equatorial states*. These are the stabiliser states with full support in the standard basis. That is, they can be written $|\phi_A\rangle = 2^{-n/2}\sum_{\mathbf{x}\in\mathbb{F}_2^n} i^{\mathbf{x}^T A\mathbf{x}}|\mathbf{x}\rangle$, where $A \in \mathcal{A}_n$, and $\mathcal{A}_n$ is the set of $n\times n$ symmetric matrices with off-diagonal entries in $\mathbb{F}_2$ and diagonal entries in $\mathbb{F}_4$. Sampling uniformly from this subset is significantly more straightforward than sampling from $\mathrm{STAB}_n$, as it amounts to uniformly sampling a matrix of $\mathcal{A}_n$. This always takes $\mathcal{O}(n^2)$ time, rather than only on average, as was the case in the previous procedure. BBCCGH [3] that if we define a random variable $\eta_A = 2^n|\langle\phi_A|\psi\rangle|^2$, where $|\phi_A\rangle$ is sampled uniformly from the equatorial state $|\psi\rangle$ is any unnormalised state vector, then $\mathbb{E}(\eta_A) = \|\psi\|^2$ and $\mathrm{Var}(\eta) \leq \|\psi\|^4$ which allows the same norm estimation routine as before, but with a simplified sampling step.

BBCCGH also improved the inner product subroutine [3]. Again, an exponential sum algorithm is employed, but the method is adapted for the CH-simulator (see Section 1.2.1). The modified subroutine has the same asymptotic $n^3$-scaling as in Ref. [60], but turns out to perform significantly better for intermediate sized systems. For example, benchmarking on a Linux PC with 3.2GHz i5-6500 CPU, the BBCCGH algorithm was shown in Ref. [3] to have average runtime 0.036 ms for $n = 60$, compared with 1.72ms for the method due to Bravyi and Gosset [60]. In combination with the improved sampling step, this leads to significant improvement to the fast norm estimation technique.

## 3.6 Emulating quantum circuits

There are two main paradigms for simulating quantum circuits within the stabiliser rank literature: the gadget-based model where non-Clifford gates are implemented by state injection [3, 59, 60], and gate decomposition methods [3], where each non-Clifford gate is itself decomposed in some way. We will first describe these simulation models, and then sketch how each of these methods can be used to simulate sampling from the output distribution of a quantum circuit.

With respect to gadget-based techniques, two strategies can be identified, referred to as fixed-sample and random-sample in BBCCGH [3]. For both strategies assume an $n$-qubit circuit $U = C_tU_tC_{t-1}\dots C_1U_1C_0$ initialised in the state $|0^n\rangle$ with $t$ non-Clifford gates $U_i$ interleaved with Clifford layers $C_i$. First each non-Clifford gate $U_i$ is replaced by a state injection gadget consuming a corresponding resource state $|U_i\rangle$. We assume for simplicity that each gate acts on a single qubit, but in general this assumption is not required. The initial state is now an $(n+t)$-qubit state $|0^n\rangle \otimes |\psi\rangle$, where $|\psi\rangle = \otimes_{i=1}^{t} |U_i\rangle$. The first strategy, the fixed-sample method, is to bypass the adaptivity inherent to gadgetisation by post-selecting on the "+1" outcome of each measurement, so that the circuit corresponds to a well-defined Clifford unitary $C$, up to a normalising factor, $U|0^n\rangle = 2^{t/2}(\mathbb{1}_n \otimes \langle 0^t|)C|0^n\rangle|\psi\rangle$ . By using an exact decomposition of $|\psi\rangle$ and applying phase-sensitive Clifford simulation with post-selection and normalisation, one obtains a stabiliser decomposition of the final state of the circuit, $|\psi'\rangle = \sum_j c_j |\phi_j\rangle$ for some vector of coefficients $\mathbf{c}$. One can then estimate Born rule probabilities as described earlier. BBCCGH [3] noted that sparsification cannot be used in this fixed-sample method, as any error in the initial state decomposition may be magnified by the normalisation factor $2^{t/2}$. The runtime for norm estimation will be at least $\Omega(\chi(\psi)\varepsilon^{-2})$. In practice the optimal decomposition will not be known, so the runtime will typically be $\mathcal{O}\big(\varepsilon^{-2}\prod_m \chi(\psi_m)\big)$, where $|\psi\rangle = \otimes_m |\psi_j\rangle$ is a suitable partitioning of $|\psi\rangle$ into few-qubit blocks for which the exact stabiliser rank can be computed.

The random-sample strategy [3, 60] can be applied in the case where each non-Clifford gate is taken from the third level of the Clifford hierarchy. As discussed in the last chapter, when this holds, state injection can be performed deterministically using Clifford corrections alone, so that the necessary correction can be efficiently simulated for any outcome. For each run of the circuit, an outcome is chosen at random for each gadget. Then one simulates the circuit $U|0^n\rangle = 2^{t/2}(\mathbb{1}_n \otimes \langle \mathbf{x}|)C_{\mathbf{x}}|0^n\rangle|\psi\rangle$ where $\mathbf{x}$ is the vector of outcomes sampled for each gadget and $C_{\mathbf{x}}$ is the Clifford circuit adapted with the corresponding Clifford corrections. Repeating this sampling many times, provided a $\delta$-close approximation is used for $|\psi\rangle$, the error in the outcome distribution is $\mathcal{O}(\delta)$ on average [60]. Using sparsification followed by norm estimation, the runtime is $\mathcal{O}\big(\xi(\psi)\delta^{-2}\varepsilon^{-2}\big)$, where $\xi$ is stabiliser extent and $\delta$ is the sparsification error. Repeated sampling cannot be avoided, as any individual sample $\mathbf{x}$ would suffer from the same error amplification as the fixed-sample method. Aside from the reduced number of terms, a second advantage of sparsification is that, by Theorem 3.4, $\xi$ is exactly multiplicative, provided each resource state injected is on no more than three qubits. Therefore, one can employ a decomposition that is provably optimal.

Gadgetisation can be avoided altogether by employing direct decomposition methods. BBCCGH introduced the sum-over-Cliffords method [3]. As discussed earlier, one can decompose non-Clifford gates as a linear combination of Cliffords $V = \sum_j c_j K_j$. Assume a circuit $U = C_t U_t C_{t-1} \dots C_1 U_1 C_0$, where $C_i$ are Clifford gates and $U_i$ are not. By repeated use of the lifting lemma (Lemma 3.6) for each $U_i$, this translates to stabiliser decomposition of the final state of the circuit. There are two key advantages over gadgetisation. First, the $t$-qubit auxiliary system on which resource states are prepared is not needed, so Clifford simulation subroutines take poly$(n)$ time instead of poly$(n+t)$. More importantly, sum-over-Clifford admits gates outside the Clifford hierarchy, so recompilation for a specific gate set is not necessary. Typically multiple third-level gates are needed to synthesise an arbitrary rotation, so the number of non-Clifford gates $t$ can be significantly reduced if the circuit can be decomposed directly. Indeed, optimal decompositions are known for some important cases. Arbitrary single-qubit $Z$-rotations $U_{\hat{z}}(\theta) = \exp[-i\theta Z/2]$ can be decomposed as [3],

$$U_{\hat{z}}(\theta) = \left[\cos\left(\frac{\theta}{2}\right) - \sin\left(\frac{\theta}{2}\right)\right]\mathbb{1} + \sqrt{2}\sin\left(\frac{\theta}{2}\right)S. \tag{3.25}$$

Using the lifting lemma, this can be shown to be optimal, so that $\xi(U_{\hat{z}}(\theta)) = \xi(U_{\hat{z}}(\theta)\left|+\right\rangle) = \left[\cos\left(\frac{\theta}{2}\right) + \tan\left(\frac{\pi}{8}\right)\sin\left(\frac{\theta}{2}\right)\right]^2$. BBCCGH [3] also gave an optimal decomposition for the $CCZ$-gate, showing that $\xi(CCZ) = \xi(CCZ\left|+^3\right\rangle) = 16/9$. This leads to runtime $\propto \prod_{i=1}^{t}\xi(U_i)$. This can lead to very significant savings compared to the case where gates must be recompiled for the Clifford+T set.

Before closing this section, we sketch how we can perform simulated bit-string sampling from the output of a quantum circuit in time $\mathcal{O}(\xi)$ [3]. We will defer consideration of the full technical details until Chapter 5, where we develop our own sampling algorithms. Consider the setting where we have an $n$-qubit quantum circuit $U$ with initial state $|\phi\rangle$, and final computational basis measurement on a subset of $w$ qubits. Without loss of generality we can assume that the first $w$ qubits are measured. If we represent the output of the circuit as a bit-string $\mathbf{x}$ of length $w$, then the circuit outputs $\mathbf{x}$ with probability $P(\mathbf{x}) = ||\Pi_{\mathbf{x}} U |\phi\rangle||^2$, where $\Pi_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w}$. Suppose we want a “black box” classical algorithm that approximately emulates the operation of the quantum circuit, so that it outputs a string $\mathbf{x}$ with probability $P_{\text{sim}}(\mathbf{x})$, such that $||P - P_{\text{sim}}||_1 \leq \delta$, for $\delta > 0$.

Computing each $P_{\text{sim}}(\mathbf{x})$ individually and sampling from the resulting distribution is inefficient, as in general this will involve $2^w$ calculations. Instead, the strategy is to sample each bit one by one by a chain of conditional probabilities. Let $\mathbf{x}_j$ be the $j$-bit string obtained by truncating the last $w-j$ bits of $\mathbf{x}$; we use $\mathbf{x}_0$ to denote an empty string. Let

$$\Pr\big(y_j|\mathbf{x}_{j-1}\big) = \frac{\Pr\big([\mathbf{x}_{j-1}, y_j]\big)}{\Pr\big(\mathbf{x}_{j-1}\big)} \tag{3.26}$$

be the probability of sampling the $j$-th bit to be $y_j$, having sampled $\mathbf{x}_{j-1}$ for the first $j-1$ bits. Here $[\mathbf{x}_{j-1}, y_j]$ means the concatenation of $\mathbf{x}_{j-1}$ with $y_j$ to form a $j$-bit string. Then the probability of obtaining string $\mathbf{x}$ can be decomposed as

$$\Pr(\mathbf{x}) = \Pr(x_w|\mathbf{x}_{w-1})\Pr(x_{w-1}|\mathbf{x}_{w-2})\ldots\Pr(x_2|x_1)\Pr(x_1) \tag{3.27}$$

Let $|\psi\rangle = U|\phi\rangle$ be the final state of the circuit, and assume we have obtained a stabiliser decomposition of this state by one of the methods outlined earlier, with $\ell_1$-norm equal to $\sqrt{\xi(\psi)}$. Let SPARSIFY be the sparsification subroutine described in Algorithm 3, and let FASTNORM be the fast norm estimation method described in Section 3.5. To sample a string of measurement outcomes, we first perform one call to SPARSIFY to obtain a decomposition with $k = \mathcal{O}\big(\xi(\psi)\delta_S^{-2}\big)$ terms, where $\delta_S$ is the sparsification error. Then we call FASTNORM to estimate the marginal distribution for the first qubit, and flip a biased coin to select either $x_1 = 0$ or $x_1 = 1$. Next, we use FASTNORM and the relation (3.26) to compute $\Pr(x_2 = 0|x_1)$ and $\Pr(x_2 = 1|x_1)$ and sample from this distribution. We repeat this process for each conditional probability in the chain (3.27) until all qubits have been "measured". There are $w$ calls to FASTNORM, each of which has runtime $\mathcal{O}\big(k\varepsilon_{FN}^{-2}n^3\big)$, where $\varepsilon_{FN}$ is the relative error. Therefore the total runtime to generate one $w$-bit string is $\mathcal{O}\big(wk\varepsilon_{FN}^{-2}n^3\big)$. It can be shown that to achieve an $\ell_1$-norm error of no more than $\delta$ in the final distribution we must set both $\varepsilon_{FN}$ and $\delta_S$ proportional to $\delta/w$, and that this leads to the runtime given in the following theorem.

**Theorem 3.10** (BBCCGH bit-string sampling algorithm [3]). *Given an $n$-qubit pure state with decomposition $|\psi\rangle = \sum_j c_j|\phi_j\rangle$, for stabiliser states $|\phi_j\rangle$, there is a classical algorithm that samples from a distribution $P_{\text{sim}}(\mathbf{x})$ that is $\delta$-close in $\ell_1$-norm to the quantum distribution $P(\mathbf{x}) = \|\Pi_{\mathbf{x}}|\psi\rangle\|^2$, with probability at least $(1-p_{\text{fail}})$, for any $p_{\text{fail}}, \delta > 0$.*

*The runtime of this algorithm is given by*

$$\tau = \mathcal{O}\big(\|\mathbf{c}\|_1^2 \delta^{-4} w^3 n^3 \log\big(w p_{\text{fail}}^{-1}\big)\big). \tag{3.28}$$

*If the decomposition of $|\psi\rangle$ is optimal with respect to stabiliser extent, the runtime is proportional to $\xi(\psi)$.*

In this sketch we have neglected several technical details, but we will give a full analysis when we introduce our improved algorithm in Chapter 5.

## 3.7 Summary and outlook

We have seen in this chapter that decomposing magic states as superpositions of stabiliser states can lead to fast classical algorithms that are able to simulate sampling from the output of near-Clifford quantum circuits. The runtime of these algorithms can be reduced dramatically by approximating the target state vector by one with fewer stabiliser terms. Of particular relevance to our work in later chapters are the sum-over-Clifford and sparsification techniques introduced by BBCCGH [3]. For ideal quantum circuits, these methods are typically much faster than the quasiprobability simulators discussed in the previous chapter. For example, the robustness of magic $\mathcal{R}$ can be shown to be strictly larger than stabiliser extent $\xi$ for magic states, often by a significant margin [2]. Furthermore, the runtime of the Howard and Campbell simulator scales with $\mathcal{R}^2$, while for fast norm estimation the scaling is only linear. Quasiprobability simulators also have the disadvantage that they are not suitable for simulated sampling as they can only estimate Born rule probabilities up to additive error.

The main limitation of the stabiliser rank techniques described so far is that whereas the Howard-Campbell simulator is able to simulate noisy stabiliser operations applied to imperfect magic states, and the Oak Ridge [45] simulator admits noisy non-stabiliser operations, stabiliser rank methods have previously only been defined for ideal circuits initialised with pure states. While the stabiliser extent and robustness of magic are somewhat similar in flavour as they are both defined in terms of the $\ell_1$-norm of a decomposition into stabiliser states, the decompositions are of different type, and the simulation techniques appear to be unrelated, save for their use of Gottesman-Knill-like subroutines. The precise connection between the two classes of simulator and their associated monotones has previously not been fully understood. In Chapter 4 we will show that by relaxing the definition of robustness of magic on the one hand, and redefining stabiliser extent for density operators on the other, the connections between these two avenues of research start to be revealed. In the process, we can obtain improved performance for both types of simulator.

# Part II

# Research

## Chapter 4

# Quantifying magic for qubit states and operations

In this chapter we develop a comprehensive framework for quantifying the magic of general $n$-qubit states and operations, which encompasses and extends the quasiprobability and stabiliser rank approaches reviewed in Chapters 2 and 3. In Chapter 5 we will make use of this framework to design improved algorithms for performing various classical simulation tasks. For this chapter, we instead focus on resource-theoretic aspects. We first introduce a triplet of new magic monotones for states, which are efficiently computable for important classes of many-qubit states, and elucidate connections between the stabiliser rank and quasiprobability pictures. In Section 4.2, we introduce well-behaved monotones that quantify the magic embodied by quantum channels outside $\mathrm{SP}_{n,n}$, the set of completely stabiliser-preserving maps. Finally, Section 4.4 provides techniques for more easily calculating these quantities for the case of diagonal channels.

## 4.1 Magic monotones for states

In this section we define three new magic monotones for states: density-operator extent, generalised robustness of magic, and dyadic negativity. These monotones were first introduced for magic theory in our article Ref. [2]. We will argue that each can be seen as an extension of the stabiliser extent, previously only defined for state vectors, to general density operators. Meanwhile, two of the new monotones can be cast as relaxations of the definition of robustness of magic. In Chapter 5, we will show that each of these magic monotones characterises the performance of a different simulation task. In this way we find that there is a deep connection between the previously disparate stabiliser rank and quasiprobability simulation methods. For now, we will define the monotones, give some important properties, and show how they are related formally.

### 4.1.1 Stabiliser extent for density operators

Recall that the pure-state extent [3] is defined in its primal formulation as,

$$\xi(\psi) := \min\{\|c\|_1^2 : |\psi\rangle = \sum_j c_j |\phi_j\rangle ; |\phi_j\rangle \in \mathrm{STAB}_n\}, \tag{4.1}$$

or in the dual formulation, $\xi(\psi) := \max\left\{|\langle\omega|\psi\rangle|^2 : |\langle\omega|\phi\rangle| \leq 1 \,\forall |\phi\rangle \in \mathrm{STAB}_n\right\}$. We extend this definition to density operators $\rho$ via the convex roof extension [151, 152], which yields the average pure-state extent minimised over all ensemble decompositions.

**Definition 4.1** (Density-operator stabiliser extent)**.** *Given an n-qubit density operator $\rho$, the density-operator extent $\Xi$ is defined,*

$$\Xi(\rho) := \min\left\{\sum_j p_j \xi(\psi_j) : \rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|, p_j \geq 0\right\}. \tag{4.2}$$

*When there exists an optimal decomposition such that $\xi(\psi_j) = \Xi(\rho)$ for all $\psi_j$, we say that the decomposition is **equimagical**.*

In general, the convex roof extension is difficult to compute, but we will see that an exact optimal decomposition can be found for any single-qubit state. In combination with the multiplicativity properties discussed in Section 4.1.5, this means that the density-operator extent can be computed efficiently for any product state $\rho = \otimes_j \rho_j$, where each $\rho_j$ is either a single-qubit state, or a pure state on up to three qubits. We note that the density-operator extent is manifestly a direct extension of pure-state extent to the domain of general density operators, since trivially the density operator for any pure state $|\psi\rangle$ has a unique ensemble decomposition $\rho = |\psi\rangle\langle\psi|$. It then follows immediately that $\Xi(|\psi\rangle\langle\psi|) = \xi(\psi)$.

### 4.1.2 Generalised notions of robustness

Recall that robustness of magic (RoM) [103] was defined using projectors,

$$\mathcal{R}(\rho) := \min_{\mathbf{q}}\left\{\|q\|_1 : \rho = \sum_j q_j |\phi_j\rangle\langle\phi_j| ; \ |\phi_j\rangle \in \mathrm{STAB}_n\right\}, \tag{4.3}$$

or alternatively in terms of a pair of mixed stabiliser states,

$$\mathcal{R}(\rho) := \min\left\{2p+1 : \rho = (1+p)\rho_+ - p\rho_-; \rho_\pm \in \overline{\mathrm{STAB}}_n, p \geq 0\right\}. \tag{4.4}$$

We obtain two new monotones by relaxing these definitions in two different ways. First, in place of the projectors in equation (4.3), we allow decomposition in terms of outer products $|L\rangle\langle R|$, which we call stabiliser dyads, where $|L\rangle$ and $|R\rangle$ can be different stabiliser states, and the coefficients for each dyad are allowed to be complex. We call this new quantity the dyadic negativity.

**Definition 4.2** (Dyadic negativity (primal form)). *Given an n-qubit density operator $\rho$, the dyadic negativity is defined:*

$$\Lambda(\rho) := \min\left\{ \|\alpha\|_1 : \rho = \sum_j \alpha_j \left|L_j\right\rangle\!\left\langle R_j\right|;\ \left|L_j\right\rangle, \left|R_j\right\rangle \in \mathrm{STAB}_n, \alpha_j \in \mathbb{C} \right\}. \tag{4.5}$$

Similar quantities have been defined for other resource theories; for example the projective tensor norm in the theory of entanglement [153, 154], and the $\ell_1$-norm of coherence [155]. Since projectors are a special case of stabiliser dyads, the dyadic negativity is optimised over a strictly larger set than RoM, so it is always the case that $\mathcal{R}(\rho) \geq \Lambda(\rho)$. Indeed, we will see that the gap is often significant. The dual formulation of the dyadic negativity [95] can be defined in terms of a class of witnesses $\mathcal{W}_\Lambda$.

**Definition 4.3** (Dyadic negativity (dual form)). *Given an n-qubit state $\rho$, the dyadic negativity is given by:*

$$\Lambda(\rho) = \max\{\mathrm{Tr}[W\rho] : W \in \mathcal{W}_\Lambda\}, \tag{4.6}$$

$$\textit{where} \quad \mathcal{W}_\Lambda = \{W : |\langle L| W |R\rangle| \leq 1 \quad \forall\, |L\rangle, |R\rangle \in \mathrm{STAB}_n,\ W \text{ Hermitian}\}. \tag{4.7}$$

The final monotone we define in this section is the *generalised robustness of magic* (gRoM). Analogous generalised robustness quantities have previously been defined for the resource theory of entanglement and others [156, 157]. The generalised robustness is a relaxation of the definition (4.4), up to a change in variable. Whereas in equation (4.4) both $\rho_+$ and $\rho_-$ were required to be mixed stabiliser states, here we only require this of the positive part of the decomposition.

**Definition 4.4** (Generalised robustness of magic (primal form)). *The generalised robustness $\Lambda^+$ for an n-qubit state $\rho$ is defined:*

$$\Lambda^+(\rho) := \min\{\lambda \geq 1 : \lambda\rho_+ - (\lambda - 1)\rho_- = \rho, \quad \rho_+ \in \mathrm{STAB}_n\}, \tag{4.8}$$

$$= \min\{\lambda : \rho \leq \lambda\rho_+, \quad \rho_+ \in \mathrm{STAB}_n\}, \tag{4.9}$$

*where $\rho_-$ can be any normalised density operator.*

For a given state $\rho$, comparing the optimal $\lambda = \lambda_*$ in equation (4.8) with the optimal $p = p_*$ for equation (4.4), and noting that we obtained the gRoM by relaxing the constraints on the optimisation, it must be the case that $\lambda_* \leq p_* + 1$. Then since we have $\Lambda^+(\rho) = \lambda_*$ and $\mathcal{R}(\rho) = 2p_* + 1$, the standard and generalised RoM are related by

$$\mathcal{R}(\rho) \geq 2\Lambda^+(\rho) - 1 \geq \Lambda^+(\rho) \tag{4.10}$$

The dual form of the gRoM is obtained by restricting the set of witnesses $\mathcal{W}_\Lambda$ to be positive semidefinite, which has the effect of simplifying the condition $|\langle L| W |R\rangle| \leq 1$ to $\langle\phi|W|\phi\rangle \leq 1$ for stabiliser states $|\phi\rangle \in \mathrm{STAB}_n$. Let $\mathcal{W}^+$ be this class of witnesses. We can then give the dual formulation as follows.

**Definition 4.5** (Generalised robustness of magic (dual form)). *Given an n-qubit state $\rho$, its generalised robustness $\Lambda^+$ can be defined:*

$$\Lambda^+(\rho) = \max\{\mathrm{Tr}[W\rho] : W \in \mathcal{W}^+\}, \tag{4.11}$$

$$\textit{where} \quad \mathcal{W}^+ = \{W : \langle\phi|W|\phi\rangle \leq 1, W \geq 0\}. \tag{4.12}$$

Note that $\mathcal{W}^+$ is strictly contained in $\mathcal{W}_\Lambda$. Therefore the maximal value of $\mathrm{Tr}[W\rho]$ over $W \in \mathcal{W}^+$ can be no larger than that over $W \in \mathcal{W}_\Lambda$, so $\Lambda^+(\rho) \leq \Lambda(\rho)$. Both gRoM and dyadic negativity are computable in the sense that they can be formulated as standard convex optimisation problems, but the optimisation quickly becomes intractable for more than a few qubits, due to the super-exponential increase in the number of $n$-qubit stabiliser states. However, we will see that the results of Section 4.1.5 make both quantities efficiently computable for $n$-fold tensor products of single-qubit states. This differs from the standard RoM, where we can construct non-optimal decompositions for $n$-fold tensor products, but optimal decompositions achieving $\ell_1$-norm equal to $\mathcal{R}$ are hard to compute.

### 4.1.3 Generic monotone properties

Results due to Regula on generalised resource theories [95] imply that each monotone $\mathcal{M}$ defined earlier in this section satisfies the following properties:

1. *faithfulness*: $\mathcal{M}(\rho) = 1$ if and only if $\rho \in \overline{\mathrm{STAB}}_n$;
2. *monotonicity*: $\mathcal{M}(\rho) \geq \mathcal{M}(O(\rho))$ for any $O \in \mathrm{SP}_{n,n}$, and any qubit state $\rho$;
3. *strong monotonicity* (monotonicity on average under selective free measurements):

   $$\mathcal{M}(\rho) \geq \sum_i p_i \mathcal{M}\left(\frac{K_i \rho K_i^\dagger}{p_i}\right), \tag{4.13}$$

   where $\rho$ is an $n$-qubit state, and $\{K_i\}_i$ are the Kraus operators of a quantum channel such that each Kraus operator is completely stabiliser-preserving, i.e. $K_i |\phi\rangle \propto |\phi'\rangle \in \mathrm{STAB}_{2n} \,\forall\, |\phi\rangle \in \mathrm{STAB}_{2n}$, and $p_i = \mathrm{Tr}\left(K_i \rho K_i^\dagger\right)$;
4. *convexity*: $\mathcal{M}(\sum_j p_j \rho_j) \leq \sum_j p_j \mathcal{M}(\rho_j)$;
5. *submultiplicativity*: $\mathcal{M}(\otimes_j \rho_j) \leq \prod_j \mathcal{M}(\rho_j)$.

We also have the following relations between the monotones. For the special case of pure states, the connection between the monotones is crystallised as follows.

**Lemma 4.6** ([95])**.** *For any pure state* $\psi$*,* $\Lambda^+(|\psi\rangle\langle\psi|) = \Lambda(|\psi\rangle\langle\psi|) = \Xi[|\psi\rangle\langle\psi|] = \xi(\psi)$.

This result was already proved by Regula in the context of generalised resource theories [95], but here we reproduce the proof given in the Appendix of Ref. [2], which is more intuitive for our setting.

*Proof.* By definition $\Xi(|\psi\rangle\langle\psi|) = \xi(\psi)$. We will prove that $\Lambda(|\psi\rangle\langle\psi|) = \Xi(|\psi\rangle\langle\psi|)$, then argue that the same proof strategy holds for $\Lambda^+$. For any state $|\psi\rangle$, there exists a decomposition optimal with respect to pure-state extent:

$$|\psi\rangle = \sum_k c_k |\phi_k\rangle \quad \text{s.t.} \quad \Xi(|\psi\rangle\langle\psi|) = \xi(\psi) = \left(\sum_k |c_k|\right)^2. \tag{4.14}$$

Expanding $|\psi\rangle\langle\psi|$, we obtain a (possibly non-optimal) decomposition over stabiliser dyads, $|\psi\rangle\langle\psi| = \sum_{j,k} c_j^* c_k \left|\phi_j\right\rangle\!\left\langle\phi_k\right|$, so that

$$\Lambda(|\psi\rangle\langle\psi|) \leq \sum_{j,k} |c_j|\,|c_k| = \sum_j |c_j| \sum_k |c_k| = (\sum_j |c_j|)^2 = \xi(\psi). \tag{4.15}$$

Now we prove the inequality $\Lambda(|\psi\rangle\langle\psi|) \geq \xi(\psi)$. From the dual formulation of pure-state extent, there exists some unnormalised witness $|\omega\rangle$ such that $|\langle\omega|\psi\rangle|^2 = \xi(\psi)$, and $|\langle\omega|\phi\rangle| \leq 1$ for all $|\phi\rangle \in \mathrm{STAB}_n$. But in the dual formulation of $\Lambda$, we have that $\Lambda(\rho) = \max_{W\in\mathcal{W}_\Lambda}\{\mathrm{Tr}[W\rho]\}$, where the witnesses in $W \in \mathcal{W}_\Lambda$ satisfy $|\langle L|W|R\rangle| \leq 1$ for all $|L\rangle, |R\rangle \in \mathrm{STAB}_n$. In particular this condition is satisfied by $W = |\omega\rangle\langle\omega|$ for the $\omega$-witness optimal w.r.t. $\xi(\psi)$. So $|\omega\rangle\langle\omega| \in \mathcal{W}_\Lambda$ but in general this might not be the maximal witness w.r.t. the dyadic negativity. So,

$$\Lambda(|\psi\rangle\langle\psi|) \geq \mathrm{Tr}[|\omega\rangle\langle\omega|\,|\psi\rangle\langle\psi|] = |\langle\omega|\psi\rangle|^2 = \xi(\psi). \tag{4.16}$$

Combined with inequality (4.15), this proves $\Lambda(|\psi\rangle\langle\psi|) = \Xi(|\psi\rangle\langle\psi|) = \xi(\psi)$.

Now, by definition the witness class $\mathcal{W}^+$ appearing in the dual form of generalised robustness is a restriction of $\mathcal{W}_\Lambda$, so $\Lambda^+(\rho) \leq \Lambda(\rho)$, as we argued when defining the monotones. Thus immediately we have $\Lambda^+(|\psi\rangle\langle\psi|) \leq \xi(\psi)$. Nevertheless we still find that $|\omega\rangle\langle\omega| \in \mathcal{W}^+$, so by the same argument as above we have $\Lambda^+(|\psi\rangle\langle\psi|) \geq \xi(\psi)$. Hence $\Lambda^+(|\psi\rangle\langle\psi|) = \Xi(|\psi\rangle\langle\psi|) = \xi(\psi)$ and equality between all four quantities has been established for the case of pure states. □

An ordering of the monotones is established by the following theorem.

**Theorem 4.7** ([95])**.** *For any density operator $\rho$ on any number of qubits, we have*

$$\Lambda^+(\rho) \leq \Lambda(\rho) \leq \Xi(\rho). \tag{4.17}$$

*Proof.* The first inequality we have immediately by comparing the dual forms of $\Lambda$ and $\Lambda^+$, as discussed above. Proving the second inequality requires convexity and the result of Lemma 4.6. Any state $\rho$ has some optimal decomposition $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$, such that $\Xi(\rho) = \sum_j p_j \xi(\psi_j)$. Now using the convexity of the dyadic negativity, $\Lambda(\rho) \leq \sum_j p_j \Lambda(\left|\psi_j\right\rangle\!\left\langle\psi_j\right|)$. But from Lemma 4.6, we know that for pure states $\Lambda(\left|\psi_j\right\rangle\!\left\langle\psi_j\right|) =$

$\xi(\psi)$, so we have, $\Lambda(\rho) \leq \sum_j p_j \xi(|\psi_j\rangle\langle\psi_j|) = \Xi(\rho)$ which completes the proof. □

### 4.1.4 Optimal decompositions for single-qubit states

Beyond the generic properties that hold for other resource settings, in Ref. [2] we proved results for single-qubit states specific to magic theory. These take the form of three lemmata leading to the following theorem.

**Theorem 4.8.** *For any single-qubit state $\rho$, we have $\Lambda^+(\rho) = \Lambda(\rho) = \Xi(\rho)$, and furthermore $\rho$ admits an optimal equimagical decomposition.*

The first lemma establishes the equality of the monotones for density operators in the convex hull of certain subsets of pure states. The second and third lemmata will show that any single-qubit density operator lies within such a convex hull.

**Lemma 4.9** (Monotone equality lemma)**.** *For any $\omega$-witness $|\omega\rangle$, we define the set $B_\omega$ to be the convex hull of all pure states $\psi$ for which $|\langle\omega|\psi\rangle|^2 = \xi(\psi)$. It follows that for all $\rho \in B_\omega$ we have $\Lambda^+(\rho) = \Lambda(\rho) = \Xi(\rho) = \langle\omega|\rho|\omega\rangle$.*

We omit the proof of this version of the lemma, as we will shortly prove a variant generalised to other classes of witness. For now we consider the other two lemmata required to prove Theorem 4.8. It remains to show that for any single-qubit density operator $\rho$, we can find a suitable set of pure states and witnesses such that $\rho \in B_\omega$. The second lemma gives an explicit form for all optimal single-qubit witnesses. This is achieved with the aid of the Bloch sphere, which can be divided into eight octants, defining the positive octant as follows.

**Definition 4.10.** *The positive octant is the set $P := \{\rho : \langle X\rangle, \langle Y\rangle, \langle Z\rangle \geq 0\}$. We further subdivide the positive octant as follows:*

$$\begin{aligned} P_X &:= \{\rho : \rho \in P, \langle X\rangle \leq \langle Y\rangle, \langle X\rangle \leq \langle Z\rangle\}, \\ P_Y &:= \{\rho : \rho \in P, \langle Y\rangle \leq \langle X\rangle, \langle Y\rangle \leq \langle Z\rangle\}, \\ P_Z &:= \{\rho : \rho \in P, \langle Z\rangle \leq \langle X\rangle, \langle Z\rangle \leq \langle Y\rangle]\}. \end{aligned} \tag{4.18}$$

*where we use the shorthand $\langle M\rangle := \mathrm{Tr}[\rho M]$.*

The region $P_Y$ is shown within the positive octant in Figure 4.1. The strategy is to prove the results for all states in the region $P_Y$. This is sufficient to prove Theorem 4.8

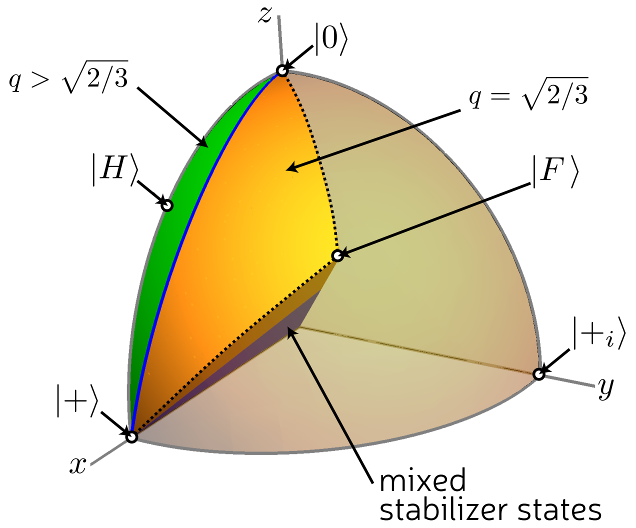


**Figure 4.1:** Region $P_Y$ within positive octant of the Bloch sphere. The variable $q$ specifies the optimal witness for pure states, as defined in Lemma 4.11. In the orange region of the surface, all states share the same witness, with $q = \sqrt{2/3}$. In the green region $q$ varies continuously between $\sqrt{2/3}$ and 1. Image by Earl Campbell, reproduced from Ref. [2] under CC-BY 4.0 license.

for all single-qubit states, since any point in the Bloch ball is Clifford-equivalent to some point in $P_Y$, and our monotones are Clifford-invariant. We can now state the second lemma.

**Lemma 4.11.** *Let $|\psi\rangle$ be any pure, single-qubit magic state in the set $P_Y$. Then the $\omega$-witness $|\omega\rangle$ that achieves $|\langle\psi|\omega\rangle|^2 = \xi(\psi)$ has an operator representation,*

$$|\omega\rangle\langle\omega| = \frac{\mathbb{1} + qH + \sqrt{1-q^2}Y}{1+q/\sqrt{2}}, \tag{4.19}$$

*where $\sqrt{2/3} \leq q \leq 1$ and $H = (X+Z)/\sqrt{2}$. Furthermore, if $|\psi\rangle$ is in the set $P_Y \cap P_X$ or $P_Y \cap P_Z$ then $q = \sqrt{2/3}$ and the $\omega$-witness takes the form*

$$|\omega\rangle\langle\omega| = \frac{\mathbb{1} + (X+Y+Z)/\sqrt{3}}{1+1/\sqrt{3}}. \tag{4.20}$$

The form of the optimal $\omega$-witness established here is used in Lemma 4.12, below, to prove that any single-qubit state is in the convex hull of a set of pure states sharing the same optimal witness. Combined with the equality lemma, (Lemma 4.9), this completes the proof of Theorem 4.8.

**Lemma 4.12.** *For any single-qubit non-stabilizer state $\rho$, there exists an $\omega$-witness $|\omega\rangle$*

*such that $\rho \in B_\omega$ (as defined in Lemma 4.9).*

The proofs of these lemmata are due to Earl Campbell, so we omit the full technical details for Lemmata 4.11 and 4.12; the interested reader can find these in Ref. [2]. Below we instead sketch the proofs, as the intuition from these arguments will later be useful as a stepping stone toward understanding decompositions of noisy single-qubit rotations. Note that while Lemma 4.11 and Lemma 4.12 relate specifically to the single-qubit case, Lemma 4.9 holds for general $n$-qubit systems. Before sketching the arguments we state and prove a slightly more general version of the lemma than that given in Ref. [2], which applies for any Hermitian witness, rather than just those of the form $|\omega\rangle\langle\omega|$.

**Lemma 4.13** (Generalised monotone equality)**.** *For any Hermitian witness $W \in \mathcal{W}$ we define the set $B_W$ to be the convex hull of all pure states $\psi$ for which $\langle\psi|W|\psi\rangle = \Lambda(|\psi\rangle\langle\psi|)$. It follows that for all $\rho \in B_W$ we have*

$$\Lambda(\rho) = \Xi(\rho) = \mathrm{Tr}[W\rho]. \tag{4.21}$$

*Similarly, for positive witnesses $W^+ \in \mathcal{W}^+$, if $B_{W^+}$ is the convex hull of all pure states such that $\langle\psi|W^+|\psi\rangle = \Lambda^+(|\psi\rangle\langle\psi|)$, then for all $\rho \in B_{W^+}$:*

$$\Lambda^+(\rho) = \Lambda(\rho) = \Xi(\rho) = \mathrm{Tr}\left[W^+\rho\right]. \tag{4.22}$$

*Proof.* If $\rho \in B_W$ there exists a decomposition $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$, such that $\left\langle\psi_j\middle|W\middle|\psi_j\right\rangle = \Lambda(|\psi\rangle\langle\psi|)$ for all $j$. Then by the definition of density-operator extent,

$$\begin{aligned} \Xi(\rho) &\leq \sum_j p_j \xi(\psi_j) = \sum_j p_j \Lambda(\psi_j) && (4.23)\\ &= \sum_j p_j \mathrm{Tr}\left[W \left|\psi_j\right\rangle\!\left\langle\psi_j\right|\right] = \mathrm{Tr}[W\rho], && (4.24) \end{aligned}$$

where we used the fact that the monotones are all equal for pure states. But since $W$ is a feasible witness, the dyadic negativity of $\rho$ is lower bounded by $\mathrm{Tr}[W\rho] \leq \Lambda(\rho)$, so $\Xi(\rho) \leq \Lambda(\rho)$. Whereas by Theorem 4.7 we always have $\Lambda(\rho) \leq \Xi(\rho)$. This proves the statement for witnesses $W \in \mathcal{W}$. An identical argument holds if we replace $\mathcal{W}$ with $\mathcal{W}^+$ and $\Lambda$ with $\Lambda^+$, proving the second statement. □

We can then give the following immediate corollary:

**Corollary 4.14** (Monotone inequality). *Suppose $\rho$ is an n-qubit state such that $\Lambda^{+}(\rho) < \Lambda(\rho)$, i.e. there is a gap between the generalised robustness and dyadic negativity. Then for any set of pure states $B'$ such that $\rho \in \mathrm{conv}(B')$, the elements $|\psi\rangle \in B'$ cannot all share the same optimal witness $|\omega\rangle$ for the pure-state extent, and neither can they share the same optimal witness $W^{+} \in \mathcal{W}$ for generalised robustness. Similarly, if dyadic negativity and density-operator extent are gapped, $\Lambda(\rho) < \Xi(\rho)$, then for any set $B'$ as defined above, $|\psi\rangle \in B'$ cannot all share the same $\omega$-, $W$- or $W^{+}$-witness.*

We now sketch the proofs for Lemmata 4.11 and 4.12. The first step to prove Lemma 4.11 is to notice that, given a pure state with optimal decomposition $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, we can equate the primal and dual solutions for the stabiliser extent to obtain the relation $(\sum_j |c_j|)^2 = \left|\sum_j c_j \langle\omega|\phi_j\rangle\right|^2$, where $|\omega\rangle$ is an optimal witness. Since every valid $\omega$-witness satisfies $|\langle\omega|\phi\rangle| \leq 1$ for any $|\phi\rangle \in \mathrm{STAB}_1$, it follows from this relation that the optimal witness must be tight against every state in the decomposition, $\left|\langle\omega|\phi_j\rangle\right| = 1$. Next we express the unknown witness $|\omega\rangle\langle\omega|$ in the Pauli basis,

$$|\omega\rangle\langle\omega| = \lambda(\mathbb{1} + q_x X + q_y Y + q_z Z). \tag{4.25}$$

To find the unknown variables, one can apply the following constraints: $|\omega\rangle\langle\omega|$ is rank-1 so $q_x^2 + q_y^2 + q_z^2 = 1$; it is positive so $\lambda > 0$; and $q_x = q_z$, which follows from membership of $P_Y$ and the tightness of the witness against $|+\rangle$ and $|0\rangle$. This leads to the witness form given in equation (4.19). Specialising to states in the intersections $P_Y \cap P_Z$ and $P_Y \cap P_X$, one finds that $q_x = q_y = q_z$, resulting in equation (4.20).

To prove Lemma 4.12 we consider slices $\mathbb{S}_f$ through the positive octant (Figure 4.2) parameterised by a constant $f = \mathrm{Tr}[\rho\sigma_F]$, where $\sigma_F = (X+Y+Z)/\sqrt{3}$ . Notice that every state in the positive octant must lie in one of these slices. The states $\psi_f^X$, $\psi_f^Y$ and $\psi_f^Z$ are the unique pure states that lie at the intersections between $\mathbb{S}_f$ and $P_Y \cap P_Z$, $P_Z \cap P_X$ and $P_X \cap P_Y$ respectively. By Lemma 4.11, these all share the same optimal witness $|\omega_*\rangle\langle\omega_*|$ (equation (4.20)). Then by Lemma 4.9, for all states $\rho$ in the purple shaded region, the three monotones are all equal to $\langle\omega_*|\rho|\omega_*\rangle$, since they are in the convex hull of $\{\psi_f^X, \psi_f^Y, \psi_f^Z\}$. States outside the triangle can be expanded $\rho = (\mathbb{1} + r_A\sigma_A + r_B\sigma_B + f\sigma_F)/2$ (see Figure 4.2). One can show that any such $\rho$ is a convex mixture of a pair of pure states $|\Phi_\rho^{\pm}\rangle\langle\Phi_\rho^{\pm}| = (\mathbb{1} + r_A\sigma_A \pm \sqrt{1 - r_A^2 + f^2}\sigma_B + f\sigma_F)/2$, which share an optimal

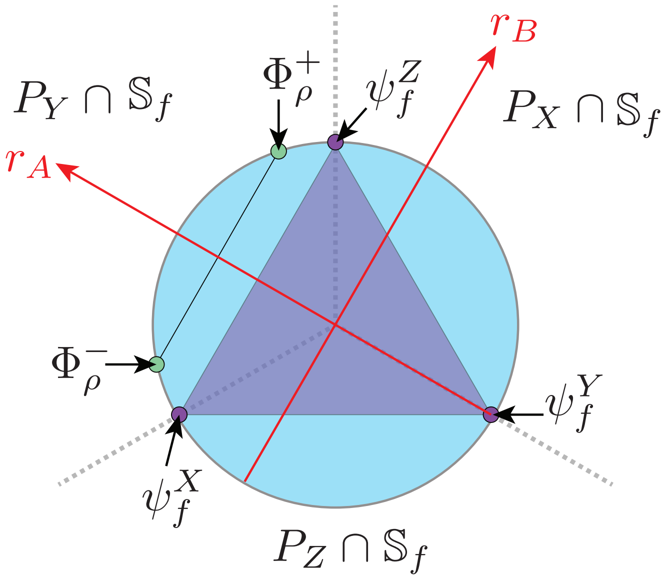


**Figure 4.2:** Slice $\mathbb{S}_f$ through the Bloch sphere, labelled by parameter $f = \text{Tr}[\rho\sigma_F]$, where $\sigma_F = (X+Y+Z)/\sqrt{3}$. The axes shownare defined by the coordinates $r_A = \text{Tr}[\rho\sigma_A]$ and $r_B = \text{Tr}[\rho\sigma_B]$, where $\sigma_A = (X+Z-2Y)/\sqrt{6}$, and $\sigma_B = (X-Z)/\sqrt{2}$. Figure by Earl Campbell, reproduced from Ref. [2] under CC-BY 4.0 license.

witness. This completes the argument.

### 4.1.5 Multiplicativity and comparison with robustness

In Chapter 2 we discussed the fact that the robustness of magic is submultiplicative and cannot be computed efficiently for many-qubit states. In contrast, we next prove that our monotones are multiplicative for tensor products of single-qubit states.

**Theorem 4.15.** *Let $\sigma_j$ be single-qubit states. Then*

$$\Lambda(\otimes_j\sigma_j) = \Xi(\otimes_j\sigma_j) = \Lambda^+(\otimes_j\sigma_j) = \prod_j \Lambda^+(\sigma_j), \tag{4.26}$$

*and furthermore $\rho = \otimes_j\sigma_j$ admits an optimal equimagical decomposition, $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, where $\Xi(\rho) = \xi(\psi_j)$ for all $j$.*

*Proof.* It is clear from the definition of density-operator extent $\Xi$ that it is submultiplicative, since a decomposition for $\rho = \otimes_j\sigma_j$ can always be constructed from optimal decompositions of the single-qubit states $\sigma_j$. Then, since Theorem 4.8 showed that $\Lambda^+(\sigma_j) = \Xi(\sigma_j)$ for all $j$, we have the upper bound $\Xi(\otimes_j\sigma_j) \le \prod_j \Xi(\sigma_j) = \prod_j \Lambda^+(\sigma_j)$. To prove equality, we need to show that $\prod_j \Lambda^+(\sigma_j) \le \Xi(\otimes_j\sigma_j)$. Another consequence of Theorem 4.8 is that every single-qubit state $\sigma_j$ has an optimal $W^+$ witness of the

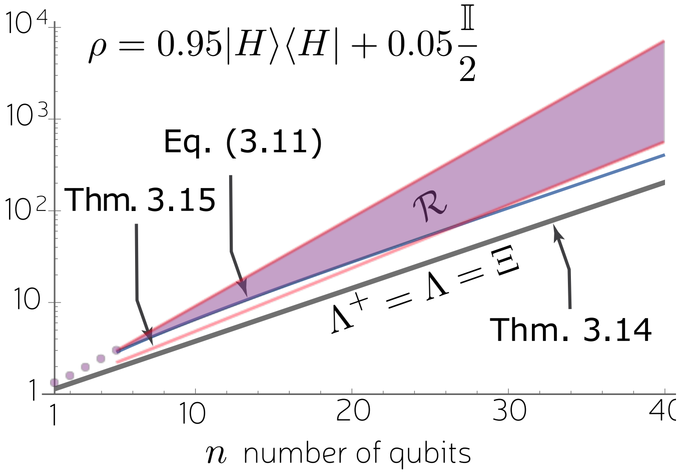


**Figure 4.3:** Comparison of robustness of magic with new magic monotones for many copies of the noisy Hadamard eigenstate $\rho$. Straight grey line shows the exact multiplicative scaling for our monotones, $\Lambda^+(\rho^{\otimes n}) = \Lambda(\rho^{\otimes n}) = \Xi(\rho^{\otimes n})$. Dots show robustness of magic $\mathcal{R}$ computed exactly, up to 5 qubits. For larger $n$, $\mathcal{R}$ can only be bounded (shaded region). The upper bound (red line) is based on submultiplicativity, relative to known decompositions. Two lower bounds are shown: (i) $\mathcal{R} \geq 2\Lambda^+ - 1$ given in equation (4.10), tighter for low numbers of qubits (blue line); and (ii) exponential bound from Theorem 4.16, which is tighter for larger numbers of qubits (lower red line). Adapted from Fig. 1 in Ref. [2] under CC-BY 4.0 license.

form $|\omega_j\rangle\langle\omega_j|$, so that $\Lambda^+(\sigma_j) = \langle\omega_j|\rho|\omega_j\rangle$. It was shown by Bravyi et al. [3] that if $\{|\omega_j\rangle\}$ are valid witnesses, then so is $|\Omega\rangle = \otimes_j |\omega_j\rangle$. This gives us a lower bound for the generalised robustness of $\rho = \otimes_j \sigma_j$. Therefore,

$$\prod_j \Lambda^+(\sigma_j) = \prod_j \langle\omega_j|\sigma_j|\omega_j\rangle = \langle\Omega|\otimes_j \sigma_j|\Omega\rangle \leq \Lambda^+(\otimes_j\sigma_j). \tag{4.27}$$

But it is always the case that $\Lambda^+(\otimes_j\sigma_j) \leq \Xi(\otimes_j\sigma_j)$ (see Theorem 4.7). This gives us the equality (4.26). To see that the $n$-qubit state admits an optimal equimagical decomposition, recall that Theorem 4.8 guarantees an optimal equimagical decomposition for any single-qubit state. This naturally gives a decomposition for the $n$-qubit state, where each term has the form $\prod_j |\psi_j\rangle\langle\psi_j|$, where $\xi(\psi_j) = \Xi(\sigma_j)$. It was already shown by Bravyi et al. [3] that the pure-state extent is multiplicative for single-qubit states (recall Theorem 3.4 in Section 3.3). It follows that for each pure-state term $\otimes_j |\psi_j\rangle\langle\psi_j|$ in the decomposition of $\rho$, we have $\xi(\prod_j |\psi_j\rangle\langle\psi_j|) = \prod_j \xi(\psi_j) = \prod_j \Xi(\sigma_j)$, so the decomposition is equimagical. □

The discussion of Theorem 4.8 showed how to obtain an optimal decomposition analytically for any single-qubit state. We have just seen that the monotones are multiplicative for single-qubit states. These two results combined show that our monotones can be computed efficiently for tensor products of single-qubit states. Robustness of magic is submultiplicative, but in Section 2.3 we saw that it has a multiplicative lower bound, namely the stabiliser-norm, $\|\cdot\|_{\text{st}}$. In Ref. [2] we used this to show that there is an exponential gap between our new monotones and the robustness of magic. We state the theorem here but omit the full proof, which can be found in our paper [2].

**Theorem 4.16.** *Given any single-qubit magic state $\rho$, there exist positive real constants $\alpha$ and $\beta$ where $\alpha > \beta$ and*

$$2^{\alpha n} \leq \mathcal{R}(\rho^{\otimes n}) \tag{4.28}$$

$$2^{\beta n} = \Lambda(\rho^{\otimes n}) = \Lambda^{+}(\rho^{\otimes n}) = \Xi(\rho^{\otimes n}), \tag{4.29}$$

As a concrete example, for the Hadamard $|H\rangle$ state this holds with $\alpha = 0.271553$ and $\beta = 0.228443$. Figure 4.3 shows how the monotones scale for many copies of a noisy Hadamard state. There is a significant gap for all values of $n$.

## 4.2 Measures of magic for quantum channels

So far, we have largely considered monotones that measure magic resource for $n$-qubit states. These monotones are useful in the setting of quantum circuits where all operations are in $\text{SP}_{n,n}$, but the initial state of the system is non-stabiliser. This is the case, for example, when all non-Clifford gates in a circuit are implemented by magic state injection (recall Section 1.4), as in the Clifford + T model. However gadgetisation is not necessarily the optimal strategy for simulating gates outside the Clifford hierarchy such as small-angle rotations, since the number of magic states required can be much larger than the number of non-Clifford gates. This motivates the development of measures for quantifying the magic of quantum channels directly, without recourse to gadgetisation. In this section we introduce several well-behaved magic monotones for quantum channels.

*Channel robustness* is an analogue of robustness of magic, and relates to quasiprobability decompositions of channels over maps from $\text{SP}_{n,n}$. *Generalised channel robustness* is defined analogously to the generalised robustness $\Lambda^{+}$ introduced earlier in this chapter. Next, we define *dyadic channel negativity*, based on generalised quasiprobability distributions over a class of linear maps we call *dyadic stabiliser channels*, $\text{DSP}_{n,n}$, which contains $\mathrm{SP}_{n,n}$ as a strict subset. Though it is not clear that computing dyadic channel negativity is tractable in general, we discuss methods for computing upper and lower bounds. We also introduce *channel extent* by analogy with density-operator extent. This last quantity is not strictly a full channel monotone, but we find application for it in later chapters. Each quantity we define is associated with overhead for at least one classical simulation task that we will describe and discuss in detail in Chapter 5. In this section we state and prove their properties. First, however, we discuss desiderata for channel monotones.

### 4.2.1 Channel monotone properties

For states, the key feature of a monotone is that it is non-increasing under completely stabiliser-preserving operations. For a channel monotone, the equivalent property is that the value of the monotone for a particular channel $\mathcal{E}$ should not increase if the channel is pre- or post-composed with a completely stabiliser-preserving channel. Thus, for a map $\mathcal{M}$ from channels to real numbers, if we are to call $\mathcal{M}$ a channel monotone, at a minimum we require the following property.

**Definition 4.17** (Monotonicity for channels)**.** *Let $\mathcal{M}$ be a function that assigns a real number to every n-qubit CPTP map. We say that $\mathcal{M}$ is a channel magic monotone, if and only if, for any n-qubit map $\mathcal{E}$:*

$$(\mathbf{P0}) \quad \mathcal{M}(\mathcal{E}\circ\mathcal{E}_{SP}) \leq \mathcal{M}(\mathcal{E}) \quad \textit{and} \quad \mathcal{M}(\mathcal{E}_{SP}\circ\mathcal{E}) \leq \mathcal{M}(\mathcal{E}) \tag{4.30}$$

*for all completely stabiliser-preserving maps $\mathcal{E}_{SP} \in \mathrm{SP}_{n,n}$.*

Typically though, for a useful channel monotone there are several other properties that are desirable. These are as follows:

**(P1)** *Faithfulness*: If $\mathcal{E}$ is a CPTP channel, then $\mathcal{M}(\mathcal{E}) = 1$ if $\mathcal{E}$ is completely stabiliser-preserving ($\mathcal{E} \in \mathrm{SP}_{n,n}$), and strictly larger than 1 otherwise;

**(P2)** *Convexity*: $\mathcal{M}(\sum_j q_j\mathcal{E}_j) \leq \sum_j |q_j|\mathcal{M}(\mathcal{E}_j)$;

**(P3)** *Invariance under extension*: $\mathcal{M}(\mathcal{E}\otimes\mathbb{1}_m) = \mathcal{M}(\mathcal{E})$, for any $m > 0$.

**(P4)** *Submultiplicativity under tensor product*: $\mathcal{M}(\mathcal{E}^A\otimes\mathcal{E}'^B) \leq \mathcal{M}(\mathcal{E}^A)\mathcal{M}(\mathcal{E}'^B)$;

**(P5)** *Submultiplicativity under composition*: $\mathcal{M}(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{M}(\mathcal{E}_2)\mathcal{M}(\mathcal{E}_1)$.

Properties (P1), (P2) and (P4) are familiar from the study of magic state monotones, but (P5), submultiplicativity under composition, is specific to the channel picture. Note that property (P3), invariance under extension, is a distinct condition from (P4), as while submultiplicativity under tensor product (P4) combined with faithfulness (P1) entails $\mathcal{M}(\mathcal{E} \otimes \mathbb{1}_m) \leq \mathcal{M}(\mathcal{E})$, they do not imply the inequality holds in the other direction. In practice it is usually straightforward to prove (P3) if the monotone has been defined with care, but we include it as a separate monotone property since it is useful as a stepping stone to prove submultiplicativity under tensor product.

Once faithfulness and submultiplicativity under composition are proven, channel monotonicity (P0) follows as a special case, because for any $\mathcal{E}_{SP} \in \mathrm{SP}_{n,n}$, and any CPTP map $\mathcal{E}$, we then have:

$$\mathcal{M}(\mathcal{E} \circ \mathcal{E}_{SP}) \leq \mathcal{M}(\mathcal{E})\mathcal{M}(\mathcal{E}_{SP}) = \mathcal{M}(\mathcal{E}), \tag{4.31}$$

since $\mathcal{M}(\mathcal{E}_{SP}) = 1$ for faithful $\mathcal{M}$, and similarly $\mathcal{M}(\mathcal{E}_{SP} \circ \mathcal{E}) \leq \mathcal{M}(\mathcal{E})$. Therefore we consider properties (P1)-(P5) sufficient to identify a well-behaved channel monotone.

### 4.2.2 Channel robustness

In Chapter 2, we discussed the Oak Ridge simulator [45], where general quantum channels were decomposed as quasiprobability distributions $\mathbf{q}$ over elements from the set CPR, which comprises the Clifford gates supplemented by Pauli-reset channels, and the simulation overhead was quantified by the $\ell_1$-norm squared of the decomposition $\|\mathbf{q}\|_1^2$. The associated cost function is $\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}) = \min_{\mathbf{q}}\{\|\mathbf{q}\|_1 : \mathcal{E} = \sum_j q_j \mathcal{E}_j, \mathcal{E}_j \in \mathrm{CPR}\}$. Bennink et al. acknowledged in [45] that CPR does not give a comprehensive enumeration of adaptive stabiliser channels. The upshot is that its convex hull CPR is strictly contained in $\mathrm{SP}_{n,n}$. Consequently, not all stabiliser channels can be positively represented in terms of CPR elements. This means that $\mathcal{R}_{\mathrm{CPR}}$ is neither a faithful measure of magic, nor monotone under composition with stabiliser operations. While Bennink et al. did not set out to design a well-behaved channel monotone, the fact that $\mathcal{R}_{\mathrm{CPR}}$ is not a monotone directly impacts the performance of their simulator. We will discuss these implications further in Chapter 6. Our goal for now is to define a well-behaved, computable channel monotone based on quasiprobability distributions over stabiliser channels. We define the *channel robustness* as follows:

**Definition 4.18** (Channel robustness ). *Given an n-qubit CPTP map, the channel robustness $\mathcal{R}_*$ is the optimal $\ell_1$-norm minimised over quasiprobability decompositions of $\mathcal{E}$ in terms of completely stabiliser-preserving CPTP channels.*

$$\mathcal{R}_*(\mathcal{E}) = \min_{\mathbf{q}}\{\|\mathbf{q}\|_1 : \sum_j q_j\mathcal{E}_j = \mathcal{E}, \mathcal{E}_j \in \mathrm{SP}_{n,n}\} \tag{4.32}$$

$$= \min\{2p+1 : (1+p)\mathcal{E}_+ - p\mathcal{E}_- = \mathcal{E}, p \geq 0, \mathcal{E}_\pm \in \mathrm{SP}_{n,n}\}. \tag{4.33}$$

We will see below that the channel robustness is a well-behaved channel monotone. First we address its computability. CPR-decompositions are computed by minimisation over a finite set of Clifford gates and Pauli-reset channels. Similarly, we need a finite set of vertices in order to compute $\mathcal{R}_*$ in practice. We can do this using the Choi state characterisation of completely stabiliser-preserving maps. Recall that an $n$-qubit linear map $\mathcal{T}$ is trace-preserving if and only if its Choi state satisfies $\mathrm{Tr}_A(\Phi_{\mathcal{T}}) = \mathbb{1}_n/2^n$, and is in $\mathrm{SP}_{n,n}$ if and only if $\Phi_{\mathcal{T}}$ is a stabiliser state. Therefore an equivalent definition is

$$\mathcal{R}_*(\mathcal{E}) = \min_{\rho_\pm \in \overline{\mathrm{STAB}}_{2n}} \left\{2p+1 : (1+p)\rho_+ - p\rho_- = \Phi_{\mathcal{E}}^{AB}, p \geq 0, \mathrm{Tr}_A(\rho_\pm) = \frac{\mathbb{1}_n}{2^n}\right\}.$$

This can be calculated by linear program given access to a list of all stabiliser states (see Appendix D.1). Note that we sidestep the problem of explicitly finding the vertices of the set $\mathrm{SP}_{n,n}$ by including the constraint on the partial trace. If we omit this condition, we recover $\mathcal{R}(\Phi_{\mathcal{E}})$, the RoM of the Choi state. We will show later in this chapter that for gates $U$ from the third level of the Clifford hierarchy, $\mathcal{R}_*(U) = \mathcal{R}(\Phi_U)$. This does not hold for more general quantum channels, and it can be shown that $\mathcal{R}(\Phi_{\mathcal{E}})$ is not a well-behaved channel monotone since it is not submultiplicative under composition (see Appendix C.1). Nevertheless, we will see later that $\mathcal{R}(\Phi_{\mathcal{E}})$ can be useful for lower bounding other monotones that are harder to compute.

Returning to the channel robustness, we note that for diagonal channels $\mathcal{E}$, the problem is equivalent to a decomposition of the state $\mathcal{E}(|+\rangle\langle+|^{\otimes n})$ of the form,

$$\mathcal{E}(|+\rangle\langle+|^{\otimes n}) = (1+p)\rho_+ - p\rho_-, \tag{4.34}$$

where $\rho_\pm \in \overline{\mathrm{STAB}}_n$, and the condition on the partial trace for the general case is replaced by the requirement that all diagonal elements of the states $\rho_\pm$ are equal to $\mathbb{1}/2^n$. In the special case where $\rho_\pm$ are pure, this implies they are diagonal Clifford-equivalent to graph states [158], though more generally they may be mixtures of states that individually do not have full support in the standard basis. Reduction to an $n$-qubit problem is advantageous for computation, since the number of vertices of the stabiliser polytope grows super-exponentially with $n$. Full technical details of this simplification will be discussed in Section 4.4.

We now prove that the channel robustness is faithful (P1), convex (P2), invariant under extension (P3) and submultiplicative under tensor product (P4) and composition (P5).

**(P1) Faithfulness:** Suppose $\mathcal{E}$ is an $n$-qubit CPTP map. There are two cases:

(i) $\mathcal{E} \in \mathrm{SP}_{n,n}$. In this case, $\Phi_\mathcal{E}$ is itself a mixed stabiliser state, and since $\mathcal{E}$ is trace-preserving it satisfies $\mathrm{Tr}(\Phi_\mathcal{E}) = \mathbb{1}_n/2^n$. So $\Phi_\mathcal{E}$ is already trivially a decomposition of the correct form, with $p = 0$, so that $R_*(\mathcal{E}) = 1 + 2p = 1$.

(ii) $\mathcal{E} \notin \mathrm{SP}_{n,n}$. Then by faithfulness of robustness of magic (Section 2.3[103]), $\Phi_\mathcal{E}$ has $\mathcal{R}(\Phi_\mathcal{E}) > 1$. Since the definition of $\mathcal{R}_*(\mathcal{E})$ is a restriction of $\mathcal{R}(\Phi_\mathcal{E})$, it must be the case that $\mathcal{R}(\Phi_\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$. Therefore $\mathcal{R}_*(\mathcal{E}) > 1$.

**(P2) Convexity:** Take a set of Choi states $\Phi_{\mathcal{E}_j}$ corresponding to channels $\mathcal{E}_j$, with optimal decompositions $\Phi_{\mathcal{E}_j} = (1 + p_j)\rho_{j_+} - p_j\rho_{j_-}$,, where each $\rho_{j_\pm}$ separately satisfies $\mathrm{Tr}_A(\rho_\pm) = \frac{\mathbb{1}_n}{2^n}$, so that $\mathcal{R}_*(\mathcal{E}_j) = 1 + 2p_j$. Now take a real linear combination of such channels,

$$\mathcal{E} = \sum_i q_i \mathcal{E}_i = \sum_{j\in P} q_j \mathcal{E}_j + \sum_{k\in N} q_k \mathcal{E}_k, \tag{4.35}$$

where $P$ is the set of indices such that $q_j \geq 0$, and $N$ is the set such that $q_k < 0$. We assume that $\sum_i q_i = 1$ so that $\mathrm{Tr}(\Phi_\mathcal{E}) = 1$. Then the Choi state for $\mathcal{E}$ is

$$\begin{aligned}\Phi_\mathcal{E} &= \sum_{j\in P} q_j\big[(1+p_j)\rho_{j_+} - p_j\rho_{j_-}\big] - \sum_{k\in N} |q_k|\big[(1+p_k)\rho_{k_+} - p_k\rho_{k_-}\big] \qquad (4.36)\\ &= \left(\sum_{j\in P} q_j(1+p_j)\rho_{j_+} + \sum_{k\in N}|q_k|p_k\rho_{k_-}\right) - \left(\sum_{j\in P} q_j p_j \rho_{j_-} + \sum_{k\in N}|q_k|(1+p_k)\rho_{k_+}\right).\end{aligned}$$

Note that the terms inside the brackets are all non-negative, so can be renormalised to obtain a probabilistic mixture over stabiliser states. We define

$$\widetilde{\rho}_+ = \frac{\sum_{j\in P} q_j(1+p_j)\rho_{j_+} + \sum_{k\in N} |q_k| p_k \rho_{k_-}}{1+\widetilde{p}}, \tag{4.37}$$

$$\text{and} \quad \widetilde{\rho}_- = \frac{\sum_{j\in P} q_j p_j \rho_{j_-} + \sum_{k\in N} |q_k|(1+p_k)\rho_{k_+}}{\widetilde{p}}, \tag{4.38}$$

where $\widetilde{p} = \sum_{j\in P} q_j p_j + \sum_{k\in N} |q_k|(1+p_k)$. This allows us to rewrite the Choi state as $\Phi_{\mathcal{E}} = (1+\widetilde{p})\widetilde{\rho}_+ - \widetilde{p}\widetilde{\rho}_-$. Since $\widetilde{\rho}_\pm$ are convex combinations of stabiliser states satisfying $\mathrm{Tr}_A(\rho_{j\pm}) = \frac{\mathbb{1}_n}{2^n}$, they must satisfy the same condition. Given that $\widetilde{p} \geq 0$, it is clear that the decomposition is in the required form, except that it is not necessarily optimised to minimise $1+2\widetilde{p}$. So,

$$\mathcal{R}_*\left(\sum_j q_j \mathcal{E}_j\right) \leq 1+2\widetilde{p} = \sum_{j\in P} |q_j|(1+2p_j) + \sum_{k\in N} |q_k|(1+2p_k) \tag{4.39}$$

$$= \sum_i |q_i| \mathcal{R}_*(\mathcal{E}_j). \tag{4.40}$$

**(P3) Invariance under extension:** We want to show that

$$\mathcal{R}_*(\mathbb{1}_m^A \otimes \mathcal{E}^B) = \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}_{m'}^B) = \mathcal{R}_*(\mathcal{E}). \tag{4.41}$$

Recall from Lemma 1.4 that tensor product channels have product Choi states,

$$\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} = \Phi_{\mathcal{E}}^{A|B} \otimes \Phi_{\mathbb{1}_m}^{A'|B'} = \Phi_{\mathcal{E}}^{A|B} \otimes |\Phi_m\rangle\langle\Phi_m|^{A'|B'}. \tag{4.42}$$

The state $\Phi_{\mathcal{E}}^{A|B}$ will have some optimal decomposition $\Phi_{\mathcal{E}}^{A|B} = (1+p)\rho_+^{AB} - p\rho_-^{AB}$, where $\mathrm{Tr}_A(\rho_\pm^{AB}) = \mathbb{1}_n/2^n$, with channel robustness $\mathcal{R}_*(\mathcal{E}_A) = 1+2p$, whereas $|\Phi_m\rangle\langle\Phi_m|^{A'|B'}$ is itself a stabiliser state so that:

$$\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} = (1+p)\rho_+^{AB} \otimes |\Phi_m\rangle\langle\Phi_m|^{A'|B'} - p\rho_-^{AB} \otimes |\Phi_m\rangle\langle\Phi_m|^{A'|B'}. \tag{4.43}$$

This is a valid stabiliser decomposition satisfying the trace condition, (since $\mathrm{Tr}_{AA'}[\rho_\pm^{AB} \otimes |\Phi_m\rangle\langle\Phi_m|^{A'|B'}] = \mathrm{Tr}_A(\rho_\pm^{AB}) \otimes \mathrm{Tr}_{A'}(|\Phi_m\rangle\langle\Phi_m|^{A'|B'})$ so we have

$$\mathcal{R}_*(\mathcal{E}\otimes\mathbb{1}_m) \leq \mathcal{R}_*(\mathcal{E}). \tag{4.44}$$

We now show that $\mathcal{R}_*(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E} \otimes \mathbb{1}_m)$. Consider an optimal decomposition for $\Phi^{AA'|BB'}_{\mathcal{E}\otimes\mathbb{1}_m} = (1+p')\rho'_+ - p'\rho'_-$, such that $\mathcal{R}_*(\mathcal{E} \otimes \mathbb{1}_m) = 1+2p'$, where $\mathrm{Tr}_{AA'}(\rho_\pm) = \mathbb{1}_{n+m}/2^{n+m}$. Here we do not assume that $\rho'_\pm$ are products across the partition $AB|A'B'$, as was the case in equation (4.43). However, we have just seen that $\Phi_{\mathcal{E}\otimes\mathbb{1}_m}$ factorises, so that by tracing out systems $A'B'$ we obtain

$$\Phi_{\mathcal{E}^A} = (1+p')\,\mathrm{Tr}_{A'B'}(\rho'_+) - p'\,\mathrm{Tr}_{A'B'}(\rho'_-). \tag{4.45}$$

Partial trace of a stabiliser state remains a stabiliser state, so this is a stabiliser decomposition. We need to check that the condition on partial trace condition holds,

$$\mathrm{Tr}_A(\mathrm{Tr}_{A'B'}(\rho'_\pm)) = \mathrm{Tr}_{AA'B'}(\rho'_\pm) \stackrel{?}{=} \frac{\mathbb{1}_n}{2^n}, \tag{4.46}$$

but this is clearly the case from the fact that $\rho'_\pm$ were constrained such that $\mathrm{Tr}_{AA'}(\rho_\pm) = \mathbb{1}_{n+m}/2^{n+m}$. So we again have a valid optimal decomposition and $\mathcal{R}_*(\mathcal{E}^A) \leq 1+2p' = \mathcal{R}_*(\mathcal{E}^A \otimes \mathbb{1}^B)$. Combining with the inequality in the other direction, (4.44), we obtain the equality $\mathcal{R}_*(\mathcal{E}\otimes\mathbb{1}) = \mathcal{R}_*(\mathcal{E}) = \mathcal{R}_*(\mathbb{1}\otimes\mathcal{E})$.

Next, we prove submultiplicativity under composition (P5) before tensor product (P4) as the former is useful in proving the latter.

**(P5) Submultiplicativity under composition:** $\mathcal{R}_*(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \mathcal{R}_*(\mathcal{E}_1)\mathcal{R}_*(\mathcal{E}_2)$. The channels $\mathcal{E}_1$ and $\mathcal{E}_2$ will each have an optimal decomposition $\mathcal{E}_j = (1+p_j)\mathcal{E}_{j,+} - p_j\mathcal{E}_{j,-}$, where $\mathcal{R}_*(\mathcal{E}_j) = 1+2p_j$ and $\mathcal{E}_{j,\pm}$ are CPTP maps and completely stabiliser preserving. We obtain $\mathcal{E}_2 \circ \mathcal{E}_1 = (1+q)\mathcal{E}'_+ - q\mathcal{E}'_-$, where

$$\mathcal{E}'_+ = (1+q)^{-1}[(1+p_2)(1+p_1)\mathcal{E}_{2,+}\circ\mathcal{E}_{1,+} + p_2p_1\mathcal{E}_{2,-}\circ\mathcal{E}_{1,-}], \tag{4.47}$$

$$\mathcal{E}'_- = q^{-1}[p_2(1+p_1)\mathcal{E}_{2,-}\circ\mathcal{E}_{1,+} + (1+p_2)p_1\mathcal{E}_{2,+}\circ\mathcal{E}_{1,-}], \tag{4.48}$$

$$\text{and} \quad q = p_1 + p_2 + 2p_1p_2. \tag{4.49}$$

The set $\mathrm{SP}_{n,n}$ is closed under composition and convex, so both $\mathcal{E}'_\pm$ are in this set. Therefore, we have a valid decomposition for $\mathcal{E}_2\circ\mathcal{E}_1$ that entails $\mathcal{R}_*(\mathcal{E}_2\circ\mathcal{E}_1) \leq 1+2q$. One finds $1+2q = (1+2p_1)(1+2p_2) = \mathcal{R}_*(\mathcal{E}_1)\mathcal{R}_*(\mathcal{E}_2)$ which completes the proof.

**(P4) Submultiplicativity under tensor product:** $\mathcal{R}_*(\mathcal{E}\otimes\mathcal{E}') \leq \mathcal{R}_*(\mathcal{E})\mathcal{R}_*(\mathcal{E}')$. We treat tensor product as a special case of composition. For $n$-qubit $\mathcal{E}$ and $m$-qubit $\mathcal{E}'$:

$$\mathcal{R}_*(\mathcal{E}\otimes\mathcal{E}') \leq \mathcal{R}_*(\mathcal{E}\otimes\mathbb{1}_m)\mathcal{R}_*(\mathbb{1}_n\otimes\mathcal{E}'). \tag{4.50}$$

But by the invariance under extension property,

$$\mathcal{R}_*(\mathcal{E}\otimes\mathbb{1}_m)\mathcal{R}_*(\mathbb{1}_n\otimes\mathcal{E}') = \mathcal{R}_*(\mathcal{E})\mathcal{R}_*(\mathcal{E}'). \tag{4.51}$$

This completes the proof.

Notice that the proof of submultiplicativity under tensor product (P4) followed immediately from submultiplicativity under composition (P5) and invariance under extension (P3), and did not depend on any specific features of channel robustness. Therefore, for other monotones we omit the proof of (P4).

### 4.2.3 Stabiliser extent for unital channels

We next define a quantity somewhat analogous to density-operator extent. Recall that stabiliser extent for unitaries [3] (Section 3.3) was defined as

$$\xi(U) = \min\left\{ \|\mathbf{c}\|_1^2 : U = \sum_j c_j K_j, K_j \in \mathrm{Cl}_n \right\} \tag{4.52}$$

where $\mathrm{Cl}_n$ is the set of Clifford gates. We can also define this using Choi states, $\xi(U) = \min\left\{ \|\mathbf{c}\|_1^2 : |U\rangle = \sum_j c_j \left|K_j\right\rangle \right\}$, where $|W\rangle = W\otimes\mathbb{1}\,|\Phi\rangle$, for maximally entangled state $|\Phi\rangle$, and the minimisation is over the Choi states of Clifford gates. It could be the case that $\xi(|U\rangle) < \xi(U)$, because there are many pure stabiliser states that do not correspond to Cliffords. By analogy with the move from pure-state to density-operator extent, we introduce the *stabiliser channel extent*.

**Definition 4.19** (Stabiliser channel extent). *Let* $\mathbb{U}_n = \{\mathcal{V} = V(\cdot)V^\dagger : V \in U(2^n)\}$. *Let* $\mathcal{E}$ *be a CPTP map in the convex hull of* $\mathbb{U}_n$, $\mathcal{E}\in\mathrm{conv}(\mathbb{U}_n)$. *The* ***stabiliser channel extent*** $\Xi_*$ *is defined*

$$\Xi_*(\mathcal{E}) = \min\left\{ \sum_j p_j \xi(V_j) : \mathcal{E} = \sum_j p_j \mathcal{V}_j, \quad \mathcal{V}_j \in \mathbb{U}_n \right\} \tag{4.53}$$

$$= \min\left\{ \sum_j p_j \xi(V_j) : \Phi_{\mathcal{E}} = \sum_j p_j \left|V_j\right\rangle\!\left\langle V_j\right|, \quad \mathcal{V}_j \in \mathbb{U}_n \right\}, \tag{4.54}$$

*where $\Phi_{\mathcal{E}}$ and $|V_j\rangle\langle V_j|$ are the Choi states for $\mathcal{E}$ and $\mathcal{V}_j$ respectively.*

Again, the definition based on the Choi-Jamiolkowski isomorphism is restricted so that the minimisation is only over those $|V_j\rangle$ that represent unitary operations. Note that the stabiliser channel extent is not a monotone in the general sense, as it is defined only for unital channels. Nervertheless, restricting to channels in the convex hull of $\mathbb{U}_n$, $\Xi_*$ is non-increasing under composition with Clifford gates and convex mixtures thereof. Channel extent can therefore be considered a monotone with respect to this restricted set. This includes important noise models such as the depolarising or dephasing channel. Despite this limitation, in Chapter 5 we will show that decompositions of the form (4.53) lend themselves to classical simulation, and we give techniques for finding decompositions in Chapter 6.

### 4.2.4 Generalised channel robustness

In this section we introduce a monotone which extends the generalised robustness $\Lambda^+$ (Section 4.1.2) to the channel picture. In the late stages of preparing this thesis we were made aware of parallel work by Saxena and Gour in which generalised channel robustness is defined in a similar way [159].

**Definition 4.20** (Generalised channel robustness)**.** *Let $\mathcal{E}$ be an n-qubit CPTP map. Then the generalised channel robustness of $\mathcal{E}$ is given by:*

$$\Lambda^+_*(\mathcal{E}) = \min\{\lambda \geq 1 : \mathcal{E} = \lambda\mathcal{E}_+ - (\lambda-1)\mathcal{E}_-, \quad \mathcal{E}_+ \in \mathrm{SP}_{n,n},\ \mathcal{E}_- \in \mathrm{CPTP}\} \tag{4.55}$$

$$= \min\left\{\lambda : \Phi_{\mathcal{E}} \leq \lambda\sigma, \sigma \in \mathrm{STAB}_{2n}, \mathrm{Tr}_A[\sigma] = \frac{\mathbb{1}_n}{2^n}\right\}, \tag{4.56}$$

*where $\Phi_{\mathcal{E}}$ is the Choi state for $\mathcal{E}$.*

First we show that the two definitions are equivalent. Fix the channel of interest $\mathcal{E}$, and consider the set of feasible stabiliser states $\{\sigma\}$ such that $\Phi_{\mathcal{E}} \leq \lambda\sigma$ for some $\lambda$, with $\mathrm{Tr}_A[\sigma] = \frac{\mathbb{1}_n}{2^n}$. From Theorem 1.11 in Section 1.3, we know that $\sigma$ must be the Choi state for some stabiliser channel $\mathcal{E}_+ \in \mathrm{SP}_{n,n}$. Now consider the Hermitian operator $\rho_-$ defined

$$\rho_- = \frac{\lambda\sigma - \Phi_{\mathcal{E}}}{\lambda - 1} \tag{4.57}$$

By construction, $\mathrm{Tr}[\rho_-] = 1$, and since $\Phi_{\mathcal{E}} \leq \lambda\sigma$, we have $\rho_- \geq 0$. Therefore $\rho_-$ is a valid density operator, so corresponds to some CP map $\mathcal{E}_-$. Moreover by linearity we have $\mathrm{Tr}_A(\rho_-) = \mathbb{1}_n/2^n$, so $\mathcal{E}_-$ is also trace-preserving. So,

$$\frac{\lambda\mathcal{E}_+ - \mathcal{E}}{\lambda - 1} = \mathcal{E}_- \quad \Longrightarrow \quad \mathcal{E} = \lambda\mathcal{E}_+ - (\lambda - 1)\mathcal{E}_-, \tag{4.58}$$

where $\mathcal{E}_\pm$ are both CPTP, and $\mathcal{E}_+ \in \mathrm{SP}_{n,n}$. Minimising $\lambda$ in in this decomposition, subject to $\lambda \geq 1$, is therefore equivalent to the minimisation in equation (4.56). Next we will prove the standard channel monotone properties.

**(P1) Faithfulness.** Suppose that $\mathcal{E} \in \mathrm{SP}_{n,n}$. Then trivially, $\mathcal{E} = \lambda\mathcal{E}$, with $\lambda = 1$. Conversely, suppose that $\mathcal{E}' \notin \mathrm{SP}_{n,n}$. Then in any decomposition $\mathcal{E}' = \lambda\mathcal{E}_+ - (\lambda - 1)\mathcal{E}_-$, where $\mathcal{E}_+ \in \mathrm{SP}_{n,n}$, it cannot be the case that $\lambda = 1$, otherwise $\mathcal{E}' = \mathcal{E}_+$, which is a contradiction. Therefore all feasible solutions have $\lambda$ strictly larger than 1.

**(P2) Convexity.** Consider a CPTP map of the form $\mathcal{E} = \sum_k q_k\mathcal{E}_k$, where $q_k$ are real, and each $\mathcal{E}_k$ is itself a CPTP map. For each $\mathcal{E}_k$, there exists some $\lambda_k \geq 1$ and $\sigma_k \in \mathrm{SP}_{n,n}$, $\mathrm{Tr}_A[\sigma_k] = \mathbb{1}_n/2^n$, such that $\Phi_{\mathcal{E}_k} \leq \lambda_k\sigma_k$, where $\Lambda^+_*(\mathcal{E}_k) = \lambda_k$. Then,

$$\sum_k |q_k|\lambda_k\sigma_k \geq \sum_k |q_k|\Phi_{\mathcal{E}_k} \geq \sum_k q_k\Phi_{\mathcal{E}_k} = \Phi_{\mathcal{E}}. \tag{4.59}$$

Now since $\sigma_k$ are all normalised stabiliser states, the operator

$$\sigma' = \sum_k p_k\sigma_k, \quad \text{where} \quad p_k = \frac{|q_k|\lambda_k}{\sum_j |q_j|\lambda_j} \tag{4.60}$$

is also a normalised stabiliser state. Moreover, $\mathrm{Tr}_A[\sigma'] = \sum_k p_k \mathrm{Tr}_A[\sigma_k] = \mathbb{1}_n/2^n$. Then letting $\lambda' = \sum_k |q_k|\lambda_k$, we have:

$$\lambda'\sigma' = \lambda'\sum_k \frac{|q_k|\lambda_k}{\lambda'}\sigma_k = \sum_k |q_k|\lambda_k\sigma_k \geq \Phi_{\mathcal{E}}. \tag{4.61}$$

This is a feasible solution to the optimisation problem for the state $\Phi_{\mathcal{E}}$, so $\lambda' \geq \Lambda^+_*(\mathcal{E})$. But recall that $\lambda_k = \Lambda^+_*(\mathcal{E}_k)$, so we have:

$$\Lambda^+_*(\sum_k q_k\mathcal{E}_k) \leq \lambda' = \sum_k |q_k|\Lambda^+_*(\mathcal{E}_k). \tag{4.62}$$

**(P3) Invariant under extension.** To show $\Lambda^+_*(\mathcal{E}) \leq \Lambda^+_*(\mathcal{E} \otimes \mathbb{1}_m)$ for $n$-qubit $\mathcal{E}$, consider the solution optimal for $\mathcal{E}$, i.e. $\Phi_{\mathcal{E}}^{AB} \leq \lambda\sigma$ for $\lambda = \Lambda_*^+(\mathcal{E})$ and some $\sigma \in \overline{\mathrm{STAB}}_{2n}$, such that $\mathrm{Tr}_A[\sigma] = \mathbb{1}_n/2^n$. Then the Choi state for $\mathcal{E} \otimes \mathbb{1}_m$ is

$$\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} = \Phi_{\mathcal{E}}^{A|B} \otimes \Phi_{\mathbb{1}_m}^{A'|B'} \leq \lambda\sigma \otimes \Phi_{\mathbb{1}_m} = \lambda\sigma'. \tag{4.63}$$

where $\sigma' = \sigma \otimes \Phi_{\mathbb{1}_m} \in \overline{\mathrm{STAB}}_{2(n+m)}$. But,

$$\mathrm{Tr}_{AA'}[\sigma'] = \mathrm{Tr}_A[\sigma] \otimes \mathrm{Tr}_{A'}[\Phi_{\mathbb{1}_m}] = \frac{\mathbb{1}_{n+m}}{2^{n+m}}, \tag{4.64}$$

so that $\Lambda_*^+(\mathcal{E} \otimes \mathbb{1}_m) \leq \lambda = \Lambda_*^+(\mathcal{E})$.

Now to prove the other direction, suppose we have the solution optimal for $(\mathcal{E} \otimes \mathbb{1}_m)$, so that $\Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} \leq \lambda'\sigma$ for some $\sigma \in \overline{\mathrm{STAB}}_{2(n+m)}$, with $\mathrm{Tr}_{AA'}[\sigma''] = \mathbb{1}_{n+m}/2^{n+m}$ and $\lambda' = \Lambda_*^+(\mathcal{E} \otimes \mathbb{1}_m)$. So $\rho = \lambda'\sigma - \Phi_{\mathcal{E}\otimes\mathbb{1}_m}^{AA'|BB'} \geq 0$ is some unnormalised density operator $\rho$, with trace $\mathrm{Tr}[\rho] = \lambda' - 1$ with partial trace over $AA'$ proportional to $\mathbb{1}_{n+m}/2^{n+m}$. Tracing out the subsystem $A'B'$, we obtain:

$$\rho' = \mathrm{Tr}_{A'B'}[\rho] = \lambda'\sigma' - \Phi_{\mathcal{E}} \geq 0 \tag{4.65}$$

where $\sigma' = \mathrm{Tr}_{A'B'}[\sigma]$. But tracing out subsystem A, on the left-hand side:

$$\mathrm{Tr}_A[\rho'] = \mathrm{Tr}_{B'}(\mathrm{Tr}_{AA'}[\rho]) = (\lambda' - 1)\,\mathrm{Tr}_{B'}\left[\frac{\mathbb{1}_{n+m}}{2^{n+m}}\right] = (\lambda' - 1)\frac{\mathbb{1}_n}{2^n}. \tag{4.66}$$

Whereas, on the right-hand side, $\mathrm{Tr}_A[\lambda'\sigma' - \Phi_{\mathcal{E}}] = \lambda'\,\mathrm{Tr}_A[\sigma'] - \mathbb{1}_n/2^n$, so that $\mathrm{Tr}_A[\sigma'] = \mathbb{1}_n/2^n$. Since $\sigma'$ is a stabiliser state satisfying the trace criterion, we have a feasible solution $\Phi_{\mathcal{E}} \leq \lambda'\sigma'$, so that we must have $\Lambda_*^+(\mathcal{E}) \leq \lambda' = \Lambda_*^+(\mathcal{E} \otimes \mathbb{1}_m)$. Having proved the inequality in both directions, we have $\Lambda_*^+(\mathcal{E}) = \Lambda_*^+(\mathcal{E} \otimes \mathbb{1}_m)$ for any CPTP map $\mathcal{E}$ and any $m > 0$.

**(P5) Submultiplicativity under composition.** Consider $\mathcal{E} = \mathcal{E}_2 \circ \mathcal{E}_1$, where $\mathcal{E}_1$ and $\mathcal{E}_2$ are CPTP. Let $\lambda_1 = \Lambda_*^+(\mathcal{E}_1)$ and $\lambda_2 = \Lambda_*^+(\mathcal{E}_2)$. Then there exists some stabiliser state $\sigma_1$, such that:

$$\lambda_1\sigma_1 - \Phi_{\mathcal{E}_1} \geq 0, \quad \mathrm{Tr}_A[\sigma_1] = \frac{\mathbb{1}_n}{2^n}. \tag{4.67}$$

First we deal with the special case where $\mathcal{E}_2$ is a completely stabiliser-preserving channel, so that $\Lambda^+_*(\mathcal{E}_2) = 1$. In this case, since $\mathcal{E}_2$ is completely positive,

$$\lambda_1(\mathcal{E}_2 \otimes \mathbb{1}_n)(\sigma_1) - (\mathcal{E}_2 \otimes \mathbb{1}_n)(\Phi_{\mathcal{E}_1}) \geq 0 \quad \Longrightarrow \quad \lambda\sigma' - \Phi_{\mathcal{E}_2 \circ \mathcal{E}_1} \geq 0 \tag{4.68}$$

where $\sigma' = (\mathcal{E}_2 \otimes \mathbb{1}_n)(\sigma_1)$. This gives us a feasible solution for $\mathcal{E} = \mathcal{E}_2 \circ \mathcal{E}_1$, so $\Lambda^+_*(\mathcal{E}) \leq \Lambda^+_*(\mathcal{E}_1) = \Lambda^+_*(\mathcal{E}_2)\Lambda^+_*(\mathcal{E}_1)$.

Now we deal with the general non-stabiliser case, where $\lambda_2 = \Lambda^+_*(\mathcal{E}_2) > 1$. Then there exists some CPTP completely stabiliser-preserving channel $\mathcal{E}_{2+}$, and CPTP map $\mathcal{E}_{2-}$ such that $\mathcal{E}_2 = \lambda_2\mathcal{E}_{2+} - (\lambda_2 - 1)\mathcal{E}_{2-}$. Rearranging, we obtain a map,

$$\mathcal{T} = (\lambda_2 - 1)\mathcal{E}_{2-} = \lambda_2\mathcal{E}_{2+} - \mathcal{E}_2. \tag{4.69}$$

But $\lambda_2 - 1 > 0$, so $\mathcal{T}$ is a completely positive map. Now let:

$$A = (\mathcal{E}_2 \otimes \mathbb{1}_n)(\lambda_1\sigma_1 - \Phi_{\mathcal{E}_1}) \geq 0, \tag{4.70}$$

where the inequality follows because $\mathcal{E}_2$ is a completely positive map. Let:

$$B = (\mathcal{T} \otimes \mathbb{1}_n)(\lambda_1\sigma_1) = (\lambda_2\mathcal{E}_{2+} \otimes \mathbb{1}_n)(\lambda_1\sigma_1) - (\mathcal{E}_2 \otimes \mathbb{1}_n)(\lambda_1\sigma_1) \geq 0 \tag{4.71}$$

where the inequality again follows from the fact that $\mathcal{T}$ is completely positive. Combinining the inequalities (4.70) and (4.71), we obtain:

$$\begin{aligned} 0 &\leq A + B \qquad (4.72)\\ &= (\mathcal{E}_2 \otimes \mathbb{1}_n)(\lambda_1\sigma_1) - (\mathcal{E}_2 \otimes \mathbb{1}_n)(\Phi_{\mathcal{E}_1}) + (\lambda_2\mathcal{E}_{2+} \otimes \mathbb{1}_n)(\lambda_1\sigma_1) - (\mathcal{E}_2 \otimes \mathbb{1}_n)(\lambda_1\sigma_1)\\ &= \lambda_1\lambda_2(\mathcal{E}_{2+} \otimes \mathbb{1}_n)(\sigma_1) - (\mathcal{E}_2 \otimes \mathbb{1}_n)(\Phi_{\mathcal{E}_1}). \qquad (4.73) \end{aligned}$$

But $(\mathcal{E}_2 \otimes \mathbb{1}_n)(\Phi_{\mathcal{E}_1}) = \Phi_{\mathcal{E}_2 \otimes \mathcal{E}_1} = \Phi_{\mathcal{E}}$, and $\mathcal{E}_{2+}$ is a completely stabiliser-preserving CPTP map, so that $(\mathcal{E}_{2+} \otimes \mathbb{1}_n)(\sigma_1) = \sigma'$ is a normalised stabiliser state such that $Tr_A[\sigma'] = \mathbb{1}_n/2^n$. Then we have $\lambda_1\lambda_2\sigma' \geq \Phi_{\mathcal{E}}$, which gives us a feasible solution for the composed channel, so

$$\Lambda^+_*(\mathcal{E}_2 \circ \mathcal{E}_1) \leq \lambda_1\lambda_2 = \Lambda^+_*(\mathcal{E}_1)\Lambda^+_*(\mathcal{E}_2). \tag{4.74}$$

### 4.2.5 Dyadic channel negativity

Next we define a channel monotone analogous to the dyadic negativity defined for states. It can also be viewed as a dyadic variant of the channel robustness. Although stabiliser dyads do not represent physical states, they can be represented and updated efficiently using stabiliser techniques, and so are useful for classical simulation, as we will discuss in detail in Chapter 5. Unphysical maps such as the dyadic Clifford maps, defined below, can be similarly useful.

**Definition 4.21** (Dyadic Clifford maps)**.** *We say that an n-qubit linear map $\mathcal{U}$ is **dyadic Clifford** if it can be expressed $\mathcal{U}(\cdot) = U_L(\cdot)U_R^\dagger$ where $U_L, U_R \in \mathrm{Cl}_n$ are Clifford operators.*

Channel robustness $\mathcal{R}_*$ was defined with respect to maps that were (i) CPTP and (ii) completely stabiliser-preserving. Dyadic maps are in general neither positive, trace- or Hermiticity- preserving. Nevertheless, they can be efficiently simulable under certain conditions. To show this, we first define a notion of completeness for dyadic maps. We then specialise to those that also preserve stabiliser structure.

**Definition 4.22** (Complete dyadic maps)**.** *We say that a linear n-qubit map $\mathcal{T}$ is a **complete dyadic map** if we can write it in the form $\mathcal{T}(\cdot) = \sum_j L_j(\cdot)R_j^\dagger$, where $\{L_j\}_j$ and $\{R_j\}_j$ separately form complete sets of Kraus operators,*

$$\sum_j L_j^\dagger L_j = \mathbb{1}, \quad \sum_j R_j^\dagger R_j = \mathbb{1}. \tag{4.75}$$

We use $\mathrm{CD}_n$ to denote the set of maps satisfying this condition. We now define the subset of maps in $\mathrm{CD}_n$ that are completely stabiliser-preserving. We will call this set the dyadic stabiliser channels, $\mathrm{DSP}_{n,n}$.

**Definition 4.23** (Dyadic stabiliser channels)**.** *We say that the n-qubit map $\mathcal{T}(\cdot) = \sum_j L_j(\cdot)R_j^\dagger$ is a **dyadic stabiliser channel**, $\mathcal{T} \in \mathrm{DSP}_{n,n}$ if it is a complete dyadic map, and each $L_j$ and $R_j$ is individually a completely-stabiliser-preserving operator, i.e. $L_j \otimes \mathbb{1}_n |\phi\rangle \propto |\phi'\rangle$, where $|\phi'\rangle \in \mathrm{STAB}_{2n}$, $\forall\, |\phi\rangle \in \mathrm{STAB}_{2n}$, and likewise for all $R_j$.*

Definition 4.23 is chosen so as to ensure a well-behaved monotone, but also so that we can exploit the structure of these dyadic channels for classical simulation in Chapter 5. A key property in this respect is that such maps are contractive with respect to the Schatten 1-norm (see Appendix D.3). We now define our monotone.

**Definition 4.24** (Dyadic channel negativity, $\Lambda_*$)**.** *Let $\mathcal{E}$ be a CPTP map on n qubits. The* ***dyadic channel negativity*** $\Lambda_*$ *is defined as*

$$\Lambda_*(\mathcal{E}) = \min\{\|\alpha\|_1 : \mathcal{E} = \sum_k \alpha_k \mathcal{T}_k, \quad \mathcal{T}_k \in \mathrm{DSP}_{n,n}\}. \tag{4.76}$$

We can immediately see that this quantity is upper bounded by the channel robustness $\mathcal{R}_*$, since the CPTP maps in $\mathrm{SP}_{n,n}$ are a subset of $\mathrm{DSP}_{n,n}$. Next, we show that the dyadic negativity of the Choi state lower bounds $\Lambda_*$. Given any decomposition of the form above, we can write the Choi state $\Phi_{\mathcal{E}} = \sum_k \alpha_k \Phi_{\mathcal{T}_k}$. The Choi state for each sub-map $\mathcal{T}_k$ can be written

$$\Phi_{\mathcal{T}} = \sum_j (L_j \otimes \mathbb{1}_n) |\Phi\rangle\langle\Phi| (R_j^\dagger \otimes \mathbb{1}_n) = \sum_j v_j w_j |L_j\rangle\langle R_j|. \tag{4.77}$$

where, e.g.

$$|L_j\rangle = \frac{(L_j \otimes \mathbb{1}_n)|\Phi\rangle}{\|(L_j \otimes \mathbb{1}_n)|\Phi\rangle\|} = \frac{(L_j \otimes \mathbb{1}_n)|\Phi\rangle}{v_j}, \tag{4.78}$$

and similarly $w_j = \|(R_j \otimes \mathbb{1}_n)|\Phi\rangle\|$, so each $|L_j\rangle\langle R_j|$ is a normalised stabiliser dyad. By the Cauchy-Schwarz inequality, for each $\mathcal{T}_k$, we have $\sum_j v_j^{(k)} w_j^{(k)} \leq 1$ (see Appendix D.3). So we can write $\Phi_{\mathcal{E}} = \sum_{j,k} \alpha_k v_j^{(k)} w_j^{(k)} |L_{j,k}\rangle\langle R_{j,k}|$. The $\ell_1$-norm of this distribution is

$$\sum_{j,k} \left|\alpha_k v_j^{(k)} w_j^{(k)}\right| = \sum_k |\alpha_k| \mathbf{v}^{(k)} \cdot \mathbf{w}^{(k)} \leq \|\alpha\|_1. \tag{4.79}$$

Therefore, for any channel distribution of the above form, $\|\alpha\|_1 \geq \Lambda(\Phi_{\mathcal{E}})$, and in particular this is true of the optimal channel decomposition w.r.t the definition (4.76), so $\Lambda(\Phi_{\mathcal{E}}) \leq \Lambda_*(\mathcal{E})$.

**Faithfulness (P1)**. If $\mathcal{E}$ is a CPTP map in $\mathrm{SP}_{n,n}$, then it can be written $\mathcal{E}(\cdot) = \sum_j K_j(\cdot)K_j^\dagger$, where $K_j$ are all completely stabiliser-preserving. Then $\mathcal{E} \in \mathrm{DSP}_{n,n}$, so there exists a trivial decomposition with $\ell_1$-norm equal to 1, so $\Lambda_*(\mathcal{E}) \leq 1$. But the Choi state negativity lower bounds $\Lambda_*$, and $\Lambda(\Phi_{\mathcal{E}}) = 1$ for $\mathcal{E} \in \mathrm{SP}_{n,n}$, so $\Lambda_*(\mathcal{E}) = 1$.

Conversely suppose $\mathcal{E}$ is CPTP but not stabiliser-preserving. Then its Choi state is not a stabiliser state, so $\Lambda_*(\mathcal{E}) \geq \Lambda(\Phi_{\mathcal{E}}) > 1$.

**Convexity (P2)**. Suppose we have $\mathcal{E} = \sum_k \beta_k \mathcal{E}_k$. There exist optimal decompositions for the sub-channels $\mathcal{E}_k = \sum_j \alpha_{j,k} \mathcal{T}_{j,k}$, such that $\Lambda_*(\mathcal{E}_k) = \sum_{j,k} |\alpha_{j,k}|$. Then the channel $\mathcal{E}$

can be decomposed $\mathcal{E}\sum_{j,k}\beta_k\alpha_{j,k}\mathcal{T}_{j,k}$. The $\ell_1$-norm of this decomposition is

$$\sum_{j,k}|\beta_k\alpha_{j,k}| = \sum_k|\beta_k|\sum_j|\alpha_{j,k}| = \sum_k|\beta_k|\Lambda_*(\mathcal{E}_k). \tag{4.80}$$

So $\Lambda_*(\mathcal{E}) \leq \sum_k|\beta_k|\Lambda_*(\mathcal{E}_k)$.

**Invariance under extension (P3).** It is straightforward to show $\Lambda_*(\mathcal{E}\otimes\mathbb{1}_m) \leq \Lambda_*(\mathcal{E})$. Given a valid optimal decomposition $\mathcal{E} = \sum_k\alpha_k\mathcal{T}_k$, with $\mathcal{T}_k(\cdot) = \sum_j L_{j,k}(\cdot)R^\dagger_{j,k}$, and $\Lambda_*(\mathcal{E}) = \|\alpha\|_1$ we can always extend it with the $m$-qubit identity map:

$$\mathcal{E}\otimes\mathbb{1}_m = \sum_k\alpha_k\mathcal{T}_k\otimes\mathbb{1}_m(\cdot) = \sum_k\alpha_k\mathcal{T}'_k(\cdot), \tag{4.81}$$

So that the individual Kraus operators are extended as $L'_{j,k} = L_{j,k}\otimes\mathbb{1}_m$, and $R'_{j,k} = R_{j,k}\otimes\mathbb{1}_m$. Moreover, a completely stabiliser-preserving Kraus operator remains so when tensored with the identity. Therefore, the extended maps $\mathcal{T}'_k = \mathcal{T}_k\otimes\mathbb{1}_m$ satisfy all the necessary criteria, so that $\mathcal{T}'_k \in \mathrm{DSP}_{n+m,n+m}$. So the $\ell_1$-norm of the distribution is an upper bound for the dyadic channel negativity of $\mathcal{E}\otimes\mathbb{1}_m$,

$$\Lambda_*(\mathcal{E}\otimes\mathbb{1}_m) \leq \|\alpha\|_1 = \Lambda_*(\mathcal{E}). \tag{4.82}$$

Now to prove the other direction, suppose an optimal decomposition of the channel extended by the identity map can be written $\mathcal{E}^A\otimes\mathbb{1}^{A'}_m = \sum_k\alpha_k\mathcal{T}^{AA'}_k$, where the maps $\mathcal{T}_k = \sum_j L_{j,k}(\cdot)R^\dagger_{j,k}$ are in $\mathrm{DSP}_{n+m,n+m}$. We have Choi state decomposition $\Phi_{\mathcal{E}^A\otimes\mathbb{1}^{A'}_m} = \sum_k\alpha_k\Phi_k$, where

$$\Phi_k = (\mathcal{T}^{AA'}_k\otimes\mathbb{1}^{BB'}_{n+m})(|\Phi_{n+m}\rangle\langle\Phi_{n+m}|^{AA'|BB'}) = \sum_j|L_{j,k}\rangle\langle R_{j,k}| \tag{4.83}$$

where $|L_{j,k}\rangle = (L^{AA'}_{j,k}\otimes\mathbb{1}^{BB'}_{n+m})\,|\Phi_{n+m}\rangle^{AA'|BB'}$ are sub-normalised pure stabiliser states, and similarly for $|R_{j,k}\rangle$. Since the Choi state for a tensor product channel factorises, $\Phi_{\mathcal{E}\otimes\mathbb{1}_m} = \Phi_{\mathcal{E}}\otimes\Phi_{\mathbb{1}_m}$, to obtain $\Phi_{\mathcal{E}}$, we trace over subsystems $A'B'$,

$$\Phi_{\mathcal{E}} = \mathrm{Tr}_{A'B'}[\Phi_{\mathcal{E}^A\otimes\mathbb{1}^{A'}_m}] = \sum_k\alpha_k\,\mathrm{Tr}_{A'B'}\,\Phi_k. \tag{4.84}$$

Using stabiliser bases $\{|p\rangle\}$ for $A'$ and $\{|q\rangle\}$ for $B'$, and using the shorthand

$|p,q\rangle = |p\rangle \otimes |q\rangle$,

$$\mathrm{Tr}_{A'B'}\,\Phi_k = \sum_j \sum_{p,q} \langle p,q|L_{j,k}\rangle \langle R_{j,k}|p,q\rangle . \tag{4.85}$$

For a single bra-ket:

$$\langle p,q|L_{j,k}\rangle = \langle p|^{A'} \langle q|^{B'} L_{j,k}^{AA'} \otimes \mathbb{1}_{n+m}^{BB'} |\Phi_{n+m}\rangle^{AA'|BB'} \tag{4.86}$$

$$= \left( \langle p|^{A'} L_{j,k}^{AA'} \otimes \mathbb{1}_n^B \right) |\Phi_n\rangle^{AB} \otimes \langle q|^{B'} |\Phi_m\rangle^{A'B'} \tag{4.87}$$

$$= \frac{1}{2^{\frac{m}{2}}} \langle p|^{A'} L_{j,k}^{AA'} |q\rangle^{A'} \otimes \mathbb{1}_n^B |\Phi_n\rangle^{AB} . \tag{4.88}$$

Similarly, $\langle p,q|R_{j,k}\rangle = 2^{-\frac{m}{2}} \langle p| R_{j,k} |q\rangle \otimes \mathbb{1}_n |\Phi_n\rangle$. For each $k$, we use the form of the above decomposition to define new sets of $n$-qubit Kraus operators labelled by indices $(j,p,q)$,

$$L'_{(j,p,q),k} = \frac{1}{2^{\frac{m}{2}}} \langle p|^{A'} L_{j,k}^{AA'} |q\rangle^{A'} , \quad R'_{(j,p,q),k} = \frac{1}{2^{\frac{m}{2}}} \langle p|^{A'} R_{j,k}^{AA'} |q\rangle^{A'} . \tag{4.89}$$

Then the (pure), Choi state for each Kraus operator, e.g. $L'_{(j,p,q),k} \otimes \mathbb{1}_m |\Phi_n\rangle = \langle p,q|L_{j,k}\rangle$, is a sub-normalised stabiliser state, since $L_{j,k}$ was completely stabiliser-preserving, and $|p\rangle$, $|q\rangle$ were stabiliser states. Therefore $L'_{(j,p,q),k}$ are completely stabiliser-preserving, and likewise for $R'_{(j,p,q),k}$. Moreover, for each $k$,

$$\sum_{j,p,q} L'^{\dagger}_{(j,p,q),k} L'_{(j,p,q),k} = 2^{-m} \sum_{j,q} \langle q| L_{j,k}^{\dagger} \sum_p |p\rangle\langle p| L_{j,k} |q\rangle \tag{4.90}$$

$$= 2^{-m} \sum_q \langle q| \sum_j L_{j,k}^{\dagger} L_{j,k} |q\rangle \tag{4.91}$$

$$= 2^{-m} \sum_q \langle q| \mathbb{1}_{n+m} |q\rangle = 2^{-m} \sum_q \langle q|q\rangle \mathbb{1}_n = \mathbb{1}_n \tag{4.92}$$

where we used $\mathbb{1}_n \otimes \sum_p |p\rangle\langle p| = \mathbb{1}_{n+m}$ in line (4.90), and in the last line the fact that $|q\rangle$ are basis vectors for an $m$-qubit subsystem, so there are $2^m$ terms in the summation. Therefore, $\{L'_{(j,p,q),k}\}_{j,p,q}$ (and likewise $\{R'_{(j,p,q),k}\}_{j,p,q}$) form a complete set of stabiliser Kraus operators, so that we can define reduced maps:

$$\mathcal{T}'_k(\cdot) = \sum_{j,p,q} L'_{(j,p,q),k} (\cdot) R'^{\dagger}_{(j,p,q),k} \tag{4.93}$$

such that $\mathcal{T}'_k \in \mathrm{DSP}_{n,n}$, and comparing with equations (4.85) and (4.88), we see that

$\mathrm{Tr}_{A'B'}\Phi_k = \Phi'_k$, where $\Phi'_k$ is the Choi state for $\mathcal{T}'_k$. Combined with equation (4.84), this implies that there exists a feasible decomposition of the $n$-qubit channel, $\mathcal{E} = \sum_k \alpha_k \mathcal{T}'_k$. Then by the usual argument, $\Lambda_*(\mathcal{E}) \leq \|\alpha\|_1 = \Lambda_*(\mathcal{E} \otimes \mathbb{1}_m)$ for any $m > 0$. Having proved the inequality in both directions, we have $\Lambda_*(\mathcal{E}) = \Lambda_*(\mathcal{E} \otimes \mathbb{1}_m)$, for any CPTP map $\mathcal{E}$ and any $m > 0$.

**Submultiplicativity under composition (P5)** Suppose $\mathcal{T}_1, \mathcal{T}_2 \in \mathrm{DSP}_{n,n}$, having decompositions $\mathcal{T}_k(\cdot) = \sum_j L_{j,k}(\cdot)R^\dagger_{j,k}$. Then:

$$\mathcal{T}_2 \circ \mathcal{T}_1(\cdot) = \sum_{j,j'} L_{j,2}L_{j',1}(\cdot)R^\dagger_{j,1}R^\dagger_{j',2}. \tag{4.94}$$

If we define $L_{j,j'} = L_{j,2}L_{j',1}$, then we immediately see that

$$\sum_{j,j'} L^\dagger_{j,j'}L_{j,j'} = \sum_{j'} L^\dagger_{j',1} \sum_j L^\dagger_{j,2}L_{j,2}L_{j',1} = \sum_{j'} L^\dagger_{j',1}L_{j',1} = \mathbb{1}_n, \tag{4.95}$$

and similarly for $R_{j,j'}$. Each Kraus operator is also completely stabiliser-preserving, since it is the composition of two completely stabiliser-preserving Kraus operators. Therefore for any $\mathcal{T}_1, \mathcal{T}_2 \in \mathrm{DSP}_{n,n}$ the composed map $\mathcal{T}' = \mathcal{T}_2 \circ \mathcal{T}_1$ is also in $\mathrm{DSP}_{n,n}$.

Now consider that for any two CPTP maps $\mathcal{E}_1$ and $\mathcal{E}_2$, each will have an optimal decomposition $\mathcal{E}_t = \sum_k \alpha_{k,t}\mathcal{T}_{k,t}$, such that $\Lambda_*(\mathcal{E}_t) = \sum_k |\alpha_{k,t}|$. Then the composed CPTP map can be expressed as $\mathcal{E}_2 \circ \mathcal{E}_1 = \sum_{j,k} \alpha_{k,2}\alpha_{j,1}\mathcal{T}_{k,2} \circ \mathcal{T}_{j,1}$. This is a feasible decomposition with $\ell_1$-norm $\sum_{j,k} |\alpha_{j,1}||\alpha_{k,2}| = \Lambda_*(\mathcal{E}_1)\Lambda_*(\mathcal{E}_2)$. So we have $\Lambda_*(\mathcal{E}_1 \circ \mathcal{E}_2) \leq \Lambda_*(\mathcal{E}_1)\Lambda_*(\mathcal{E}_2)$. We have shown that the dyadic channel negativity is a well-behaved channel monotone, but it is not clear how to compute it exactly for all channels. In Section 4.2.2, we showed that the condition on the Choi state for trace-preserving maps $\mathrm{Tr}_A[\Phi_\mathcal{E}] = \mathbb{1}_n/2^n$ provided a practical means to compute channel robustness. One might ask whether analogous conditions can be found for complete dyadic channels as defined by equation (4.75). For the interested reader, in Appendix D.4 we present partial results towards better characterisation of the set of complete dyadic channels, based on a polar decomposition of the Choi matrix. We hope that in future work this will lead to improved methods for finding optimal dyadic decompositions. For the moment, we are in general only able to bound the dyadic channel negativity. The dyadic negativity of the Choi state provides a lower bound, and this is tractable for tensor products of single-qubit channels. We next discuss decompositions over discrete subsets of $\mathrm{DSP}_{n,n}$, which yield upper bounds.

#### 4.2.5.1 Upper bounds via dyadic Clifford maps

By definition, any unitary operation trivially satisfies $U^\dagger U = \mathbb{1}$, and the Clifford gates are manifestly completely stabiliser-preserving. Therefore, any map of the form $\mathcal{T} = L(\cdot)R$, where $L$ and $R$ are Clifford gates, is in $\mathrm{DSP}_{n,n}$. Provided a channel can be written as a complex combination of such maps, the following quantity upper bounds the dyadic channel negativity.

$$\Lambda_{\mathrm{Cl}}(\mathcal{E}) = \min\{\|\alpha\|_1 : \sum_k \alpha_k L_k(\cdot)R_k = \mathcal{E}(\cdot), \quad L_k, R_k \in \mathrm{Cl}_n\}. \tag{4.96}$$

We call $\Lambda_{\mathrm{Cl}}$ the dyadic Clifford negativity, but we note that it is not a well-behaved channel monotone. While it is convex and submultiplicative under composition, it is not faithful with respect to maps in $\mathrm{SP}_{n,n}$. We show this in Appendix D.2. However, it is clearly the case that $\Lambda_{\mathrm{Cl}}(\mathcal{E}) = 1$ for any channel $\mathcal{E}$ that can be written as a convex mixture of Clifford operations. In Chapter 6 we will see that $\Lambda_{\mathrm{Cl}}(\mathcal{E}) = \Lambda_*(\mathcal{E})$ for certain classes of noisy single-qubit operations, including noisy $T$-gates.

Before moving on to a more comprehensive set of dyadic stabiliser channels, we reflect on the size of the optimisation problem defined by equation (4.96). There are 24 single-qubit Clifford gates, so for the single-qubit channel problem there will be 576 vertices. However, for two qubits, there are 11,520 Clifford gates, and therefore over 132 million vertices, so the problem already looks formidable for two-qubit channels. Therefore if this technique is to be useful for multi-qubit channels, techniques for reducing the problem size using symmetry will be important.

#### 4.2.5.2 Upper bounds using dyadic projective channels

When proving submultiplicativity, we showed that the composition of any two maps in $\mathrm{DSP}_{n,n}$ is also in $\mathrm{DSP}_{n,n}$. Therefore we can form a larger subclass of dyadic stabiliser channels by composing physical stabiliser channels $\mathcal{O}$ with Clifford dyadic maps $\mathcal{U}_{j,k}(\cdot) = U_j(\cdot)U_k^\dagger$. Here we consider adaptive stabiliser channels of the form:

$$O(\rho) = \sum_j K_j \rho K_j^\dagger, \tag{4.97}$$

where we assume that the kraus operator has the form $K_j = V_j\Pi_j$ for some Clifford $V_j$ and projector $\Pi_j$, where $\{\Pi_j\}$ are a set of orthogonal projectors. Operationally, these are channels where a Clifford $V_j$ is carried out conditioned on obtaining the measurement outcome $j$ associated with the projector $\Pi_j$. We then form the dyadic stabiliser channel $\mathcal{T} = \mathcal{U}_{j,k} \circ \mathcal{O} \circ \mathcal{U}_{l,m}$, where $\mathcal{U}$ are dyadic Clifford maps. Explicitly, these have the form,

$$\mathcal{T}(\cdot) = \sum_s U_l K_s U_j (\cdot) U_k^\dagger K_s^\dagger U_m^\dagger. \tag{4.98}$$

We can compact this notation: consider the operators multiplying the input dyad on the left. If $K_s = V_s\Pi_s$ for some Clifford $V_s$ and projector $\Pi_s$, then

$$U_l V_s \Pi_s U_j = U_l V_s U_j U_j^\dagger \Pi_s U_j = V_s' \Pi_s' = K_s' \tag{4.99}$$

where $V_s' = U_l V_s U_j$ is some new Clifford gate, and $\Pi_s' = U_j^\dagger \Pi_s U_j$ is some new stabiliser-preserving projector. Since conjugation by a unitary operator preserves orthogonality, the new projectors $\{\Pi_s'\}_s$ are orthogonal, and $\{K_s'\}_s$ remains a complete set of Kraus operators. We can do the same for the operators multiplying on the right, so that we have $\mathcal{T}(\ldots) = \sum_s K_s'(\ldots)K_s''^\dagger$, where the Kraus operators on the right are related to those on the left by

$$K_s'' = U_m U_l^\dagger K_s' U_j^\dagger U_k = U K_s' V^\dagger. \tag{4.100}$$

In the last expression we absorb all the Clifford gates into a pair of Cliffords $U$ and $V$. We can therefore fully specify $\mathcal{T}$ by a complete set of stabiliser-preserving Kraus operators $\{K_s'\}_s$ and a pair of Clifford gates $U$ and $V$. We call the class of maps having the form above the *projective dyadic stabiliser channels*, PDSP. We can use this class to define a quantity that upper bounds dyadic channel negativity, which we call the *projective dyadic negativity*.

$$\Lambda_P(\mathcal{E}) = \min\{\|\alpha\|_1 : \sum_k \alpha_k \mathcal{T}_k = \mathcal{E}, \quad \mathcal{T} \in \text{PDSP}\}. \tag{4.101}$$

The class PDSP includes as base cases those maps where $U = V = \mathbb{1}$, i.e. physical stabiliser channels of the form (4.97). This ensures that important stabiliser-preserving channels such as state injection gadgets and Pauli-reset channels have $\Lambda_P = 1$. This also ensures that any CPTP map can be decomposed using only maps in PDSP, as it was shown by Bennink et al. that Clifford gates and Pauli-reset channels are sufficient for this purpose [45]. We show in Appendix D.5 that projective dyadic negativity is computable for single-qubit channels, but likely intractable in general for $n$-qubit channels with $n \geq 2$. Nevertheless, the scheme could be used, for example, for circuits decomposed into CNOTs and noisy single-qubit gates. Alternatively the circuit could be decomposed into single-qubit operations supplemented with a minimal number of multi-qubit operations, all of which may be non-stabiliser. Single-qubit operations would be decomposed using the dyadic channel scheme, while multi-qubit operations could be decomposed using standard channel robustness.

## 4.3 Magic-increasing capacity of quantum channels

We have so far considered channel monotones based on different types of stabiliser decomposition of a channel. We now introduce another family of channel monotones that quantify the ability of a channel to to increase the resourcefulness of an initial state. Each is defined with respect to a particular magic *state* monotone $\mathcal{M}$. It is natural to consider the largest possible increase in $\mathcal{M}$, over any initial state.

**Definition 4.25** (Magic capacity with respect to generic monotones)**.** *Given a generic magic monotone $\mathcal{M}$, we define the magic capacity $\mathcal{C}_{\mathcal{M}}$ with respect to $\mathcal{M}$,*

$$\mathcal{C}_{\mathcal{M}}(\mathcal{E}) = \max_{\rho \in \mathbb{D}_{2n}} \frac{\mathcal{M}[(\mathcal{E} \otimes \mathbb{1}_n)(\rho)]}{\mathcal{M}(\rho)}. \tag{4.102}$$

*where $\mathbb{D}_{2n}$ is the set of $2n$-qubit density operators.*

Analogous quantities have previously been defined for the resource theories of entanglement [160] and coherence [161]. Notice that the maximisation on the right-hand side must be normalised by the resourcefulness of the initial state, $\mathcal{M}(\rho)$. We will show that when defined with respect to several of the magic state monotones already encountered, the magic capacity possesses the five standard channel monotone properties (P1) to (P5), as well as another feature we call stabiliser-saturation (P6). The essence of this property is that we only need optimise over pure stabiliser states. For mixed states or even non-stabiliser states, the capacity still captures the largest possible increase in robustness of magic. We define this as follows.

**(P6) stabiliser-saturation:** We say that the magic capacity $\mathcal{C}_{\mathcal{M}}$ is stabiliser-saturated if for any $n$-qubit state $\rho$, which may be non-stabiliser, and for any CPTP quantum channel $\mathcal{E}$, the following holds:

$$\frac{\mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(\rho)]}{\mathcal{M}(\rho)} \leq \mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|)], \tag{4.103}$$

for some pure stabiliser state $|\phi_*\rangle \in \mathrm{STAB}_{2n}$. When this property is satisfied, $\mathcal{C}_{\mathcal{M}}(\mathcal{E}) = \mathcal{M}[\mathcal{E}\otimes\mathbb{1}_n(|\phi_*\rangle\langle\phi_*|)]$.

The magic capacity $\mathcal{C}_{\mathcal{M}}$ inherits properties (P1) and (P2) directly from the magic state monotone $\mathcal{M}$, as described in the following theorem.

**Theorem 4.26.** *Let $\mathcal{M}$ be a magic monotone (non-increasing under channels $\mathcal{E} \in \mathrm{SP}_{n,n}$), and let $\mathcal{C}_{\mathcal{M}}$ be the associated magic capacity as defined in equation* (4.102)*. Then the following statements hold.*

1. *Capacity $\mathcal{C}_{\mathcal{M}}$ is faithful for any faithful monotone $\mathcal{M}$. (P1)*
2. *If $\mathcal{M}$ is convex then $\mathcal{C}_{\mathcal{M}}$ is convex. (P2)*

Properties (P3 - P5) then hinge on proving stabiliser-saturation (P6), as follows.

**Theorem 4.27.** *Suppose $\mathcal{M}$ is a magic monotone that is convex, faithful, and submultiplicative under tensor product, and let $\mathcal{C}_{\mathcal{M}}$ be the associated magic capacity. Suppose that $\mathcal{C}_{\mathcal{M}}$ is stabiliser-saturated (P6); that is, we assume that for any* $2n$*-qubit state $\rho$ and $n$-qubit channel the associated magic monotone $\mathcal{M}$ satisfies*

$$\frac{\mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)\rho]}{\mathcal{M}(\rho)} \leq \mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(|\phi\rangle\langle\phi|)] = \mathcal{C}_{\mathcal{M}}(\mathcal{E}), \tag{4.104}$$

*for some pure stabiliser state $|\phi\rangle \in \mathrm{STAB}_{2n}$. Then $\mathcal{C}_{\mathcal{M}}$ is invariant under extension, and submultiplicative under tensor product (P4) and composition (P5):*

$$\text{(P3)} \quad \mathcal{C}_{\mathcal{M}}(\mathcal{E}\otimes\mathbb{1}_m) = \mathcal{C}_{\mathcal{M}}(\mathcal{E}), \quad \forall m > 0 \tag{4.105}$$

$$\text{(P4)} \quad \mathcal{C}_{\mathcal{M}}(\mathcal{E}_2\otimes\mathcal{E}_1) \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}_2)\mathcal{C}_{\mathcal{M}}(\mathcal{E}_2) \tag{4.106}$$

$$\text{(P5)} \quad \mathcal{C}_{\mathcal{M}}(\mathcal{E}_2\circ\mathcal{E}_1) \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}_2)\mathcal{C}_{\mathcal{M}}(\mathcal{E}_1). \tag{4.107}$$

We will next prove Theorems 4.26 and 4.27 for the general case. In the following subsections we show that $\mathcal{C}_{\mathcal{M}}$ is stabiliser-saturated when defined with respect to standard and generalised robustness of magic, and dyadic negativity.

*Proof of Theorem 4.26.* To prove statement (1), we note that if $\mathcal{M}$ is monotone under completely stabiliser-preserving channels, then $\mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(\rho)]/\mathcal{M}(\rho)\leq 1$ for any $\mathcal{E}\in\mathrm{SP}_{n,n}$. But if $\mathcal{M}$ is faithful then $\mathcal{M}(\sigma)=1$ for all $\sigma\in\mathrm{STAB}_{2n}$. It follows that the inequality is saturated by stabiliser initial state, $\mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(\sigma)]\mathcal{M}(\sigma)=1$, and $\mathcal{C}_{\mathcal{M}}(\mathcal{E})=1$ if $\mathcal{E}\in\mathrm{SP}_{n,n}$. Conversely, suppose instead that $\mathcal{E}$ is non-stabiliser-preserving, but still CPTP. Then there exists at least one stabiliser state $\sigma\in\mathrm{STAB}_{2n}$ such that $(\mathcal{E}\otimes\mathbb{1})(\sigma)$ is not a stabiliser state. Then by faithfulness of $\mathcal{M}$, $\mathcal{M}[(\mathcal{E}\otimes\mathbb{1})(\sigma)]>1$, and so $\mathcal{C}_{\mathcal{M}}(\mathcal{E})>1$.

Statement (2) follows straightforwardly from convexity of $\mathcal{M}$. Suppose a channel $\mathcal{E}$ has decomposition $\mathcal{E}=\sum_k q_k\mathcal{E}_k$, for some $n$-qubit CPTP maps $\mathcal{E}_k$. There exists some optimal state $\rho_*$ that achieves $\mathcal{C}_{\mathcal{M}}(\mathcal{E})=\mathcal{M}[\mathcal{E}\otimes\mathbb{1}(\rho_*)]$. Then

$$\frac{\mathcal{M}[(\mathcal{E}\otimes\mathbb{1}_n)(\rho_*)]}{\mathcal{M}(\rho_*)}=\frac{\mathcal{M}[\sum_k q_k(\mathcal{E}_k\otimes\mathbb{1}_n)(\rho_*)]}{\mathcal{M}(\rho_*)}\leq\sum_k|q_k|\frac{\mathcal{M}[(\mathcal{E}_k\otimes\mathbb{1}_n)(\rho_*)]}{\mathcal{M}(\rho_*)},\qquad(4.108)$$

where the last line follows by convexity of $\mathcal{M}$. But by definition, each ratio $\mathcal{M}[(\mathcal{E}_k\otimes\mathbb{1}_n)(\rho_*)]/\mathcal{M}(\rho_*)$ can be no larger than $\mathcal{C}_{\mathcal{M}}(\mathcal{E}_k)$. So we have,

$$\mathcal{C}_{\mathcal{M}}(\sum_k q_k\mathcal{E}_k)\leq\sum_k|q_k|\mathcal{C}_{\mathcal{M}}(\mathcal{E}_k).\qquad(4.109)$$

This proves the second statement. □

*Proof of Theorem 4.27.* By assumption, $\mathcal{M}$ is convex, faithful and submultiplicative under tensor product, and $\mathcal{C}_{\mathcal{M}}$ is stabiliser-saturated. We first prove invariance under extension (P3). We claim that if $\mathcal{M}$ is faithful and submultiplicative under tensor product, then for $m>0$, for any $|\phi\rangle\in\mathrm{STAB}_{2n+m}$, there exists some state $|\psi\rangle\in\mathrm{STAB}_{2n}$ such that

$$\mathcal{M}\left[\left(\mathcal{E}^A\otimes\mathbb{1}_{n+m}\right)(|\phi\rangle\langle\phi|^{AB})\right]=\mathcal{M}\left[\left(\mathcal{E}^A\otimes\mathbb{1}_n\right)(|\psi\rangle\langle\psi|^{AB'})\right].\qquad(4.110)$$

Consider a $(2n+m)$-qubit stabiliser state $|\phi\rangle$, with partition $A|B$ between the first $n$ and last $n+m$ qubits. Ref. [57] shows that the state $|\phi\rangle^{AB}$ is local Clifford-equivalent to $p$ independent Bell pairs entangled across the partition $A|B$ (here "local" means with respect to the bipartition rather than per qubit). Since there are $n$ qubits in partition $A$, $p$ is at most $n$. Let $B'|B''$ be a partition of $B$ into $n$ and $m$ qubits. Then by local permutation of qubits within B, we can take these $p\leq n$ Bell pairs to be entangled across $A|B'$. So

$|\phi\rangle^{AB} = \left(\mathbb{1}_n \otimes U^B\right)|\psi\rangle^{AB'}|\psi'\rangle^{B''}$ where $U^B$ is a Clifford operation, $|\psi\rangle^{AB'} \in \mathrm{STAB}_{2n}$ and $|\psi'\rangle^{B''} \in \mathrm{STAB}_m$. So writing the channel corresponding to $U^B$ as $\mathcal{U}^B$, for any $\mathcal{E}$ on $n$ qubits, we know that

$$\mathcal{M}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_{n+m}\right)(|\phi\rangle\langle\phi|^{AB})\right] = \mathcal{M}\left[(\mathbb{1}_n \otimes \mathcal{U}^B)\left(\left(\mathcal{E}^A \otimes \mathbb{1}_n\right)\left(|\psi\rangle\langle\psi|^{AB'}\right) \otimes |\psi'\rangle\langle\psi'|^{B''}\right)\right]. \tag{4.111}$$

Since $\mathbb{1}_n \otimes \mathcal{U}^B$ represents a (reversible) Clifford gate, by monotonicity of $\mathcal{M}$:

$$\begin{aligned}\mathcal{M}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_{n+m}\right)(|\phi\rangle\langle\phi|^{AB})\right] &= \mathcal{M}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_n\right)\left(|\psi\rangle\langle\psi|^{AB'}\right) \otimes |\psi'\rangle\langle\psi'|^{B''}\right] \\ &= \mathcal{M}\left[\left(\mathcal{E}^A \otimes \mathbb{1}_n\right)(|\psi\rangle\langle\psi|^{AB'})\right], \end{aligned} \tag{4.112}$$

where in the last line we used the fact that $|\psi'\rangle^{B''}$ is a stabiliser state, and hence does not contribute to the monotone value, since $\mathcal{M}$ is submultiplicative under tensor product, and by its faithfulness property, $\mathcal{M}(|\psi'\rangle\langle\psi'|) = 1$. The state $|\psi\rangle\langle\psi|^{AB'}$ is a $2n$-qubit state, so equation (4.110) holds.

So, the maximum $\mathcal{M}$ generated by $\mathcal{E}^A \otimes \mathbb{1}_{n+m}$ over all $(2n+m)$-qubit stabiliser states is no larger than that generated by $\mathcal{E}^A \otimes \mathbb{1}_n$, and $\mathcal{C}_{\mathcal{M}}(\mathcal{E}^A \otimes \mathbb{1}_m) \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}^A)$. The other direction can be easily proved by noting that for any $2n$-qubit state $\rho = (\mathcal{E} \otimes \mathbb{1}_n)(|\phi\rangle\langle\phi|)$, where $|\phi\rangle \in \mathrm{STAB}_{2n}$, the monotone value does not change if we form the tensor product with some $m$-qubit stabiliser state $|\phi'\rangle \in \mathrm{STAB}_m$. That is, $\mathcal{M}(\rho) = \mathcal{M}(\rho \otimes |\phi'\rangle\langle\phi'|)$, and consequently $\mathcal{C}_{\mathcal{M}}(\mathcal{E}^A \otimes \mathbb{1}_m) = \mathcal{C}_{\mathcal{M}}(\mathcal{E}^A)$.

Next we prove (P5), submultiplicativity under composition, $\mathcal{C}_{\mathcal{M}}(\mathcal{E}_1 \circ \mathcal{E}_2) \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}_1)\mathcal{C}_{\mathcal{M}}(\mathcal{E}_2)$. Consider the $n$-qubit channel $\mathcal{E} = \mathcal{E}_2 \circ \mathcal{E}_1$, where $\mathcal{E}_1$ and $\mathcal{E}_2$ are CPTP maps. Suppose $\sigma_* = |\phi_*\rangle\langle\phi_*|$ is an optimal choice of initial stabiliser state, such that $\mathcal{C}_{\mathcal{M}}(\mathcal{E}) = \mathcal{M}[(\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*)]$. Then,

$$\begin{aligned}\mathcal{C}_{\mathcal{M}}(\mathcal{E}) = \mathcal{M}[(\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*)] &= \mathcal{M}[(\mathcal{E}_2 \otimes \mathbb{1}_n)(\rho)] &(4.113)\\ &= \frac{\mathcal{M}[(\mathcal{E}_2 \otimes \mathbb{1}_n)(\rho)]\mathcal{M}(\rho)}{\mathcal{M}(\rho)}. &(4.114)\end{aligned}$$

where $\rho = (\mathcal{E}_1 \otimes \mathbb{1}_n)\,|\phi_*\rangle\langle\phi_*|$. Then, since the relation (4.104) holds by assumption,

$$\mathcal{C}_{\mathcal{M}}(\mathcal{E}) = \frac{\mathcal{M}[(\mathcal{E}_2 \otimes \mathbb{1}_n)(\rho)]\mathcal{M}(\rho)}{\mathcal{M}(\rho)} \leq \mathcal{M}(\rho)\mathcal{C}_{\mathcal{M}}(\mathcal{E}_2) \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}_1)\mathcal{C}_{\mathcal{M}}(\mathcal{E}_2) \qquad (4.115)$$

where the last inequality follows because by definition,

$$\mathcal{M}(\rho) = \mathcal{M}[(\mathcal{E}_1 \otimes \mathbb{1}_n)(|\sigma_*\rangle\langle\sigma_*|)] \leq \mathcal{C}_{\mathcal{M}}(\mathcal{E}_1). \qquad (4.116)$$

This proves $\mathcal{C}_{\mathcal{M}}$ is submultiplicative under composition.

Finally, we have already shown for previous channel monotones that invariance under extension (P3) and submultiplicativity under composition (P5) together imply submultiplicativity under tensor product (P4), and the argument does not depend on any other details of the monotone. This completes the proof. □

Taken together, Theorems 4.26 and 4.27 imply that given a magic state monotone $\mathcal{M}$ known to be convex, faithful and multiplicative under tensor product, we need only prove stabiliser-saturation (P6) to show that $\mathcal{C}_{\mathcal{M}}$ satisfies properties (P1)-(P5). In the next three subsections, we prove (P6) holds for capacities with respect to standard and generalised robustness of magic, and dyadic negativity.

### 4.3.1 Magic capacity with respect to robustness of magic

**Definition 4.28** (Magic capacity with respect to robustness of magic)**.** *For any n-qubit channel, the magic capacity is defined,*

$$\mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \max_{|\phi\rangle \in \mathrm{STAB}_{2n}} \mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi\rangle\langle\phi|)], \qquad (4.117)$$

*where $\mathcal{R}$ is the robustness of magic.*

We prove stabiliser-saturation below, so that this definition is equivalent to the definition for the generic case (Definition 4.25). Since $\mathcal{R}$ is convex, faithful and submultiplicative under tensor project, by Theorems 4.26 and 4.27, the capacity $\mathcal{C}_{\mathcal{R}}$ also has properties (P1-P5). Recall that to prove stabiliser-saturation, we need to show that for any $n$-qubit state $\rho$, including magic states, and for any CPTP quantum channel $\mathcal{E}$, we have

$$\frac{\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)(\rho))}{\mathcal{R}(\rho)} \leq \mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|)), \qquad (4.118)$$

for some pure stabiliser state $|\phi_*\rangle \in \mathrm{STAB}_{2n}$. We use similar arguments to those deployed in [160]. Consider an $n$-qubit channel $\mathcal{E}$. Any $2n$-qubit input state $\rho$ will have an optimal stabiliser decomposition $\rho = \sum_j q_j |\phi_j\rangle\langle\phi_j|$, where $\sum_j q_j = 1$, and such that $\mathcal{R}(\rho) = \sum_j |q_j|$. By linearity $(\mathcal{E}\otimes\mathbb{1}_n)(\rho) = \sum_j q_j(\mathcal{E}\otimes\mathbb{1}_n)(|\phi_j\rangle\langle\phi_j|)$. Then by convexity of RoM,

$$\mathcal{R}((\mathcal{E}\otimes\mathbb{1}_n)(\rho)) \leq \sum_j |q_j| \mathcal{R}\big((\mathcal{E}\otimes\mathbb{1}_n)(|\phi_j\rangle\langle\phi_j|)\big). \tag{4.119}$$

For any stabiliser state $|\phi_j\rangle$, the optimal pure stabiliser state $|\phi_*\rangle$, satisfies

$$\mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}((\mathcal{E}\otimes\mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|)) \geq \mathcal{R}\big((\mathcal{E}\otimes\mathbb{1}_n)(|\phi_j\rangle\langle\phi_j|)\big) \tag{4.120}$$

$$\implies \quad \mathcal{R}((\mathcal{E}\otimes\mathbb{1}_n)(\rho)) \leq \mathcal{R}((\mathcal{E}\otimes\mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|))\sum_j |q_j| = \mathcal{C}_{\mathcal{R}}(\mathcal{E})\mathcal{R}(\rho). \tag{4.121}$$

Rearranging we obtain inequality (4.118).

### 4.3.2 Magic capacity with respect to generalised robustness

Here we consider the magic capacity $\mathcal{C}_{\Lambda^+}$ with respect to $\Lambda^+$, the generalised robustness of magic (gRoM), which we introduced in Definition 4.4.

**Definition 4.29** (Magic capacity with respect to generalised robustness of magic)**.** *For any n-qubit channel, the magic capacity is defined,*

$$\mathcal{C}_{\Lambda^+}(\mathcal{E}) = \max_{|\phi\rangle\in\mathrm{STAB}_{2n}} \Lambda^+[(\mathcal{E}\otimes\mathbb{1}_n)(|\phi\rangle\langle\phi|)] \tag{4.122}$$

The gRoM is convex, faithful, submultiplicative under tensor product, and monotone under channels in $\mathrm{SP}_{n,n}$. So, once again we complete the proof that capacity with respect to gRoM, $\mathcal{C}_{\Lambda^+}$, satisifies all six properties by showing that it is stabiliser-saturated (P6).

Let $\rho$ be any $2n$-qubit density operator, and let $\mathcal{E}$ be any $n$-qubit CPTP map. Recall that the gRoM for $2n$-qubit states can be defined:

$$\Lambda^+(\rho) = \min\{\lambda : \rho \leq \lambda\sigma, \sigma \in \overline{\mathrm{STAB}}_{2n}\}. \tag{4.123}$$

Then there exists some $\lambda_* \geq 1$ and mixed stabiliser state $\sigma_*$ such that $\lambda_*\sigma_* - \rho \geq 0$, and

$\Lambda^+(\rho) = \lambda_*$. Since $\mathcal{E}$ is a CPTP map, it preserves the positivity of the operator $\lambda_*\sigma_* - \rho$:

$$(\mathcal{E} \otimes \mathbb{1}_n)[\lambda_*\sigma_* - \rho] \geq 0 \quad \Longrightarrow \quad \lambda_*(\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*) \geq (\mathcal{E} \otimes \mathbb{1}_n)(\rho), \tag{4.124}$$

where the second line follows by linearity of $\mathcal{E}$. Let $\rho' = (\mathcal{E} \otimes \mathbb{1}_n)(\rho)$ and $\rho'' = (\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*)$. In general both operators can be non-stabiliser states. Using the definition of $\Lambda_+$ again, there must exist $\sigma''_* \in \overline{\mathrm{STAB}}_{2n}$ such that $\rho'' \leq \lambda''_*\sigma''_*$ for $\lambda''_* = \Lambda^+(\rho'')$. It follows that $\rho' \leq \lambda_*\rho'' \leq \lambda_*\lambda''_*\sigma''_*$. This expression gives us a feasible solution for the optimisation problem in equation (4.123), with respect to $\rho'$ as the state of interest. Therefore we must have $\Lambda^+(\rho') \leq \lambda_*\lambda''_* = \Lambda^+(\rho)\Lambda^+(\rho'')$. We can rewrite this as

$$\Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(\rho)] \leq \Lambda^+(\rho)\Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*)]. \tag{4.125}$$

Now $\sigma_*$ is a mixed stabiliser state so has some ensemble decomposition $\sigma_* = \sum_j p_j \left|\phi_j\right\rangle\!\left\langle\phi_j\right|$, where $\left|\phi_j\right\rangle \in \mathrm{STAB}_{2n}$, $p_j \geq 0$ and $\sum_j p_j = 1$. By convexity of $\Lambda_+$ and linearity of $\mathcal{E}$,

$$\begin{aligned}
\Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(\sigma_*)] &\leq \sum_j p_j \Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(\left|\phi_j\right\rangle\!\left\langle\phi_j\right|)] && (4.126)\\
&\leq \sum_j p_j \Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\!\langle\phi_*|)] && (4.127)\\
&= \Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\!\langle\phi_*|)] && (4.128)
\end{aligned}$$

where in the second line $|\phi_*\rangle$ is any stabiliser state that maximises $\Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi\rangle\!\langle\phi|)]$ over $|\phi\rangle \in \mathrm{STAB}_{2n}$. Combining equation (4.128) with (4.125) and rearranging,

$$\frac{\Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(\rho)]}{\Lambda^+(\rho)} \leq \Lambda^+[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\!\langle\phi_*|)], \tag{4.129}$$

proving that $\mathcal{C}_{\Lambda^+}$ is stabiliser-saturated.

### 4.3.3 Magic capacity with respect to dyadic negativity

For dyadic negativity, it is convenient to define the capacity in terms of the maximum increase over stabiliser dyads. However, to do this, we must first extend the definition of dyadic negativity, previously defined for density matrices, to general operators on $n$-qubit

Hilbert spaces. For any operator $A$, we define this as:

$$\Lambda(A) = \min\{\|\alpha\|_1 : \sum_j \alpha_j \left|L_j\right\rangle\!\left\langle R_j\right| = A, \quad \left|L_j\right\rangle, \left|R_j\right\rangle \in \mathrm{STAB}_n\}. \tag{4.130}$$

This is always well-defined, since the stabiliser dyads form an overcomplete basis for the space of $n$-qubit operators. One can easily check that this quantity remains convex, so that for any $A = \sum_j \alpha_j B_j$, we have $\Lambda(A) \leq \sum_j |\alpha_j| \Lambda(B_j)$. We can also extend the notion of faithfulness, using normalisation appropriate for dyads. We treat an operator as free if $\Lambda(A)/\|A\|_1 = 1$. This is equivalent to saying that when the operator is normalised, it can be written as a convex combination of stabiliser dyads. Conversely, if $\Lambda(A)/\|A\|_1 > 1$, then the operator is non-stabiliser. The extended dyadic negativity is also submultiplicative under tensor product for any pair of operators. If we have optimal decompositions $A = \sum_j \alpha_j \sigma_j$ and $B = \sum_k \beta_k \sigma_k$, where $\sigma_j$ are stabiliser dyads, then $A \otimes B = \sum_{j,k} \alpha_j \beta_k \sigma_j \otimes \sigma_k$ gives a feasible solution since the tensor product of two stabiliser dyads remains a stabiliser dyad. Therefore $\Lambda(A \otimes B) \leq \sum_{j,k} |\alpha_j| \cdot |\beta_k| = \Lambda(A)\Lambda(B)$.

We now consider the increase in dyadic negativity under CPTP channels. Given any $n$-qubit operator $\rho$ with optimal decomposition $\rho = \sum_j \alpha_j \left|L_j\right\rangle\!\left\langle R_j\right|$,

$$(\mathcal{E} \otimes \mathbb{1}_n)(\rho) = \sum_j \alpha_j (\mathcal{E} \otimes \mathbb{1}_n)(\left|L_j\right\rangle\!\left\langle R_j\right|). \tag{4.131}$$

Then by convexity,

$$\Lambda[(\mathcal{E} \otimes \mathbb{1}_n)(\rho)] \leq \sum_j |\alpha_j| \Lambda[(\mathcal{E} \otimes \mathbb{1}_n)(\left|L_j\right\rangle\!\left\langle R_j\right|)] \tag{4.132}$$

$$\leq \Lambda(\rho)\Lambda[(\mathcal{E} \otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)] \tag{4.133}$$

where $|L_*\rangle\langle R_*|$ is the dyad that maximises $\Lambda[(\mathcal{E} \otimes \mathbb{1}_n)(|L\rangle\langle R|)]$ over all $|L\rangle, |R\rangle \in \mathrm{STAB}_{2n}$. Therefore we define the magic capacity with respect to $\Lambda$, which we call the dyadic magic capacity, as

$$\mathcal{C}_\Lambda(\mathcal{E}) = \max_{|L\rangle, |R\rangle \in \mathrm{STAB}_{2n}} \Lambda[(\mathcal{E} \otimes \mathbb{1}_n)(|L\rangle\langle R|)], \tag{4.134}$$

so the quantity is stabiliser-saturated in the sense that the maximum increase in negativity is achieved for a stabiliser dyad. The dyadic magic capacity inherits convexity and faithfulness from dyadic negativity as per the arguments in the proof of Theorem 4.26. We can confirm that $\mathcal{C}_\Lambda$ is invariant under extension by adapting the proof of Theorem 4.27. We simply need to notice that the dyadic negativity is invariant under Clifford dyadic maps of the form $U(\cdot)V^\dagger$, where $U$ and $V$ are Clifford gates, since these map stabiliser dyads to stabiliser dyads. Then the proof can be modified by replacing stabiliser projectors by stabiliser dyads, and conjugation with a single Clifford gate by application of a map $U(\cdot)V^\dagger$. Similarly we can substitute projectors for dyads in the rest of the proof to show that $\mathcal{C}_\Lambda$ is submultiplicative under composition and tensor product.

### 4.3.4 Sandwich theorems

The magic capacities defined in the previous sections can be harder to compute than some of the other channel monotones we have introduced, since without further simplification they involve evaluating magic state monotones once for each $2n$-qubit stabiliser state (or dyad). We will discuss methods for easing this calculation for diagonal channels in 4.4. For now, we establish some results that provide rigorous upper and lower bounds for the magic capacity. We first consider capacity of some channel $\mathcal{E}$ with respect to robustness. The lower and upper bounds are given by the robustness of the Choi state $\mathcal{R}(\Phi_\mathcal{E})$ and the channel robustness $\mathcal{R}_*(\mathcal{E})$, respectively.

**Theorem 4.30** (Sandwich theorem for $\mathcal{R}$-based monotones)**.** *For any CPTP map $\mathcal{E}$, the following inequalities hold,*

$$\mathcal{R}(\Phi_\mathcal{E}) \leq \mathcal{C}_\mathcal{R}(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E}). \tag{4.135}$$

*Moreover, if the unitary operation $\mathcal{U}$ is in the third level of the Clifford hierarchy, then we have equality, $\mathcal{R}(\Phi_\mathcal{U}) = \mathcal{C}_\mathcal{R}(\mathcal{U}) = \mathcal{R}_*(\mathcal{U})$.*

*Proof.* By definition $\Phi_\mathcal{E} = (\mathcal{E} \otimes \mathbb{1}_n)(|\Phi_n\rangle\langle\Phi_n|)$. But $|\Phi_n\rangle$ is a stabiliser state, so $\mathcal{R}(\Phi_\mathcal{E})$ can be no larger than $\mathcal{R}((\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|)) = \mathcal{C}_\mathcal{R}(\mathcal{E})$, where $|\phi_*\rangle$ is the stabiliser state that achieves the capacity, and so $\mathcal{R}(\Phi_\mathcal{E}) \leq \mathcal{C}_\mathcal{R}(\mathcal{E})$.

Now suppose $\mathcal{E} = (1+p)\mathcal{E}_+ - p\mathcal{E}_-$ is the optimal decomposition of $\mathcal{E}$ into CPTP stabiliser-preserving maps, $\mathcal{E}_\pm \in \mathrm{SP}_{n,n}$, so that $\mathcal{R}_*(\mathcal{E}) = 1+2p$. Then for any *input* stabiliser state $\sigma \in \mathrm{STAB}_{2n}$, we can write down a valid stabiliser decomposition of the *output* state,

$$(\mathcal{E} \otimes \mathbb{1}_n)(\sigma) = (1+p)(\mathcal{E}_+ \otimes \mathbb{1}_n)(\sigma) - p(\mathcal{E}_- \otimes \mathbb{1}_n)(\sigma). \tag{4.136}$$

In particular this is true for the stabiliser state $\sigma_* = |\phi_*\rangle\langle\phi_*|$ that is optimal with respect to the capacity. But equation (4.136) could be a non-optimal decomposition, so its $\ell_1$-norm $1+2p$ is at least as large as $\mathcal{R}((\mathcal{E}\otimes\mathbb{1}_n)(\sigma_*))$. So,

$$\mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}[(\mathcal{E}\otimes\mathbb{1}_n)(\sigma_*)] \leq 1+2p = \mathcal{R}_*(\mathcal{E}), \tag{4.137}$$

completing the proof of the first statement. Having done so, to prove the second statement it suffices to show that $\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{R}_*(\mathcal{U})$.

For any $n$-qubit unitary channel $\mathcal{U} = U(\cdot)U^\dagger$ from the third level of the Clifford hierarchy, deterministic state injection is possible [35, 89]. That is, given a Hilbert space $\mathcal{H} = \mathcal{H}_A\otimes\mathcal{H}_B\otimes\mathcal{H}_C\otimes\mathcal{H}_D$, where each subspace is comprised of $n$ qubits, there exists a completely stabiliser-preserving circuit $\mathcal{E}_{\mathrm{SI}}$ such that, for any $2n$-qubit input state $\rho$,

$$\mathrm{Tr}_{BC}\left[\mathcal{E}_{\mathrm{SI}}(\Phi^{AB}_{\mathcal{U}}\otimes\rho^{CD})\right] = \mathcal{U}\otimes\mathbb{1}_n(\rho^{AD}) \tag{4.138}$$

Where $\Phi_{\mathcal{U}}$ is the Choi state for $\mathcal{U}$. The circuit $\mathcal{E}_{\mathrm{SI}}$ comprises a complete Bell measurement on $BC$, followed by a Clifford correction on subspace $A$ conditioned on the outcome of the Bell measurement. It can be represented by Kraus operators,

$$K_j = (C^A_j\otimes\mathbb{1}^{BCD}_{3n})M_j, \tag{4.139}$$

where $M_j = \mathbb{1}^A\otimes\left|\Phi_j\right\rangle\!\left\langle\Phi_j\right|^{BC}\otimes\mathbb{1}^D$ are the Kraus operators corresponding to elements of the Bell basis $\left|\Phi_j\right\rangle$, and $C_j$ is some unitary Clifford correction. Now consider an optimal decomposition of the Choi state, $\Phi_{\mathcal{U}} = (1+p)\rho_+ - p\rho_-$, such that $\mathcal{R}(\Phi_{\mathcal{U}}) = 1+2p$. We now show that by substitution into equation (4.138) we can obtain a decomposition of the channel that satisfies the trace-preservation condition required for channel robustness. We have

$$\Phi^{AD}_{\mathcal{U}} = \mathcal{U}\otimes\mathbb{1}_n(|\Phi\rangle\langle\Phi|^{AD}) = \mathrm{Tr}_{BC}\left[\mathcal{E}_{\mathrm{SI}}(\Phi^{AB}_{\mathcal{U}}\otimes|\Phi\rangle\langle\Phi|^{CD})\right] \tag{4.140}$$

$$= (1+p)\widetilde{\rho}^{AD}_+ - p\widetilde{\rho}^{AD}_-, \tag{4.141}$$

where $\widetilde{\rho}^{AD}_\pm = \mathrm{Tr}_{BC}[\mathcal{E}_{\mathrm{SI}}(\rho^{AB}_\pm\otimes|\Phi\rangle\langle\Phi|^{CD})] \in \mathrm{STAB}_{2n}$, since $\mathcal{E}_{\mathrm{SI}} \in \mathrm{SP}_{n,n}$. We want to show that $\mathrm{Tr}_A(\widetilde{\rho}_\pm) = \mathbb{1}_n/2^n$. First, note that $\mathrm{Tr}_A(\widetilde{\rho}_\pm)$ is independent of the Cliffords $C_j$, since the partial trace depends only on the Bell measurement outcome probabilities, $p_j = \mathrm{Tr}\left[M_j(\rho_{\pm}^{AB} \otimes |\Phi\rangle\langle\Phi|^{CD})M_j^{\dagger}\right]$. Therefore $\mathrm{Tr}_A(\widetilde{\rho}_{\pm}) = \mathrm{Tr}_{ABC}(\rho'_{\pm})$, where $\rho'_{\pm} = \sum_j M_j(\rho_{\pm}^{AB} \otimes |\Phi\rangle\langle\Phi|^{CD})M_j^{\dagger}$ is the post-measurement state. Then,

$$\begin{aligned}
\mathrm{Tr}_A(\widetilde{\rho}_{\pm}) &= \sum_j \mathrm{Tr}_{ABC}\left[M_j\left(\rho_{\pm}^{AB} \otimes |\Phi\rangle\langle\Phi|^{CD}\right)M_j^{\dagger}\right] && (4.142)\\
&= \mathrm{Tr}_{ABC}\left[\sum_j M_j^{\dagger}M_j\left(\rho_{\pm}^{AB} \otimes |\Phi\rangle\langle\Phi|^{CD}\right)\right] && (4.143)\\
&= \mathrm{Tr}_{ABC}\left[\rho_{\pm}^{AB} \otimes |\Phi\rangle\langle\Phi|^{CD}\right] && (4.144)\\
&= \mathrm{Tr}_C\left[|\Phi\rangle\langle\Phi|^{CD}\right] = \frac{\mathbb{1}_n}{2^n}. && (4.145)
\end{aligned}$$

In going to the second line, we used the fact that the partial trace over $BC$ is cyclic with respect to operators that act non-trivially only on $\mathcal{H}_B \otimes \mathcal{H}_C$. In going from the second to the third line, we used the fact that $\{M_j\}$ is a complete set of Kraus operators, so $\sum_j M_j^{\dagger}M_j = \mathbb{1}$. We have shown that the decomposition (4.141) satisfies the trace preservation criterion. Since the decomposition may not be optimal, we have that $\mathcal{R}_*(\mathcal{U}) \leq 1+2p = \mathcal{R}(\Phi_{\mathcal{U}})$. But from the proof of the first statement $\mathcal{R}(\Phi_U) \leq \mathcal{C}_{\mathcal{R}}(\mathcal{U}) \leq \mathcal{R}_*(\mathcal{U})$, so it must be that equality holds. □

We note that the result that $\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{R}_*(\mathcal{U})$ for third-level gates carries over to the case of decompositions of $\mathcal{U}(|+\rangle\langle+|^{\otimes n})$ for diagonal third-level gates. That is, there always exists a decomposition satisfying the constraints of equation (4.34) that is optimal with respect to $\mathcal{R}(\Phi_{\mathcal{U}}) = \mathcal{R}(\mathcal{U}(|+\rangle\langle+|^{\otimes n}))$. This can be seen by following the argument of Theorem 4.30, but replacing the full $4n$-qubit teleportation circuit with a $2n$-qubit state injection circuit, as discussed in Section 1.4.

**Theorem 4.31** (Sandwich theorem for $\Lambda^+$-based channel monotones)**.** *For any CPTP map $\mathcal{E}$, the following inequalities hold,*

$$\Lambda^+(\Phi_{\mathcal{E}}) \leq \mathcal{C}_{\Lambda^+}(\mathcal{E}) \leq \Lambda^+_*(\mathcal{E}). \tag{4.146}$$

*Moreover, if the unitary operation $\mathcal{U}$ is in the third level of the Clifford hierarchy, then we have equality, $\Lambda^+(\Phi_{\mathcal{U}}) = \mathcal{C}_{\Lambda^+}(\mathcal{U}) = \Lambda^+_*(\mathcal{U})$.*

We omit the full proof as it is almost identical to that given for Theorem 4.30, with

$\mathcal{R}$, $\mathcal{C}_{\mathcal{R}}$ and $\mathcal{R}_*$ replaced with $\Lambda^+$, $\mathcal{C}_{\Lambda^+}$ and $\Lambda^+_*$ respectively. We can also prove a sandwich theorem for the channel monotones based on dyadic negativity.

**Theorem 4.32** (Sandwich theorem for $\Lambda$-based channel monotones)**.** *For any CPTP map $\mathcal{E}$, the following inequalities hold,*

$$\Lambda(\Phi_{\mathcal{E}}) \leq \mathcal{C}_{\Lambda}(\mathcal{E}) \leq \Lambda_*(\mathcal{E}). \tag{4.147}$$

*Proof.* Assume $\mathcal{E}$ is an $n$-qubit channel. The first inequality, $\Lambda(\Phi_{\mathcal{E}}) \leq \mathcal{C}_{\Lambda}(\mathcal{E})$, is proved using the same steps as for Theorem 4.30, so we omit the details. To prove the second inequality, recall that there exists a decomposition over stabiliser dyadic channels $\mathcal{T}_j$ optimal with respect to dyadic negativity $\Lambda_*$, that is,

$$\mathcal{E} = \sum_j \alpha_j \mathcal{T}_j \quad \text{s.t.} \quad \Lambda_*(\mathcal{E}) = \sum_j |\alpha_j|. \tag{4.148}$$

Meanwhile, there exists a normalised stabiliser dyad $|L_*\rangle\langle R_*|$ that achieves the dyadic capacity, $\mathcal{C}_{\Lambda}(\mathcal{E}) = \Lambda[(\mathcal{E}\otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)]$ (recall that we are using the extended definition of dyadic negativity given in Section 4.3.3). Then using the dyadic channel decomposition (4.148) and the convexity of dyadic negativity,

$$\mathcal{C}_{\Lambda}(\mathcal{E}) = \Lambda\left[\sum_j \alpha_j (\mathcal{T}_j \otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)\right] \leq \sum_j |\alpha_j| \Lambda[(\mathcal{T}_j \otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)]. \tag{4.149}$$

Now recall that stabiliser dyadic channels take the form $\mathcal{T}(\cdot) = \sum_k L_k(\cdot)R_k^\dagger$, where $\{L_k\}_k$ and $\{R_k\}_k$ are both complete sets of completely stabiliser-preserving Kraus operators. Let $l_k = \| L_k \otimes \mathbb{1}_n |L_*\rangle \|$ and let $|L_k\rangle = L_k \otimes \mathbb{1}_n |L_*\rangle / l_k$ and likewise for the Kraus operators acting on the right, i.e. $r_k = \|R_k \otimes \mathbb{1}_n |R_*\rangle\|$. Then using convexity again, for any stabiliser dyadic channel $\mathcal{T}$ we obtain

$$\Lambda[(\mathcal{T}\otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)] \leq \sum_k l_k r_k \Lambda[|L_k\rangle\langle R_k|] = \sum_k l_k r_k, \tag{4.150}$$

where $|L_k\rangle\langle R_k|$ is a normalised stabiliser dyad. But $\sum_k l_k^2 = \sum_k r_k^2 = 1$ by the completeness of the Kraus operator sets $\{L_k\}_k$ and $\{R_k\}_k$. Then using the Cauchy-Schwarz inequality, $\sum_k l_k r_k \leq \sqrt{(\sum_k l_k^2)\cdot(\sum_{k'} r_{k'}^2)} = 1$, and so $\Lambda[(\mathcal{T}\otimes \mathbb{1}_n)(|L_*\rangle\langle R_*|)] \leq 1$ for all

| $n$ | $N_S$ |
|---|---|
| 1 | 6 |
| 2 | 60 |
| 3 | 1,080 |
| 4 | 36,720 |
| 5 | 2,423,520 |
| 6 | 315,057,600 |

**Table 4.1:** Number of pure stabiliser states $N_S$ for number of qubits $n$.

stabiliser dyadic channels $\mathcal{T}$. Using this fact with the inequality (4.149) we have $\mathcal{C}_\Lambda(\mathcal{E}) \leq \sum_j |\alpha_j| = \Lambda_*(\mathcal{E})$, completing the proof. □

## 4.4 Computing monotones for diagonal channels

Direct computation of the robustness of magic $\mathcal{R}$ for an $n$-qubit state involves solving a linear program where the number of vertices is equal to the number of $n$-qubit pure stabiliser states, $N_S$ [103, 133]. The size of this problem quickly becomes prohibitively large, since $N_S$ increases super-exponentially with $n$ (Table 4.1), and in practice we are limited to five qubits. Similarly, computing generalised robustness and dyadic negativity for entangled states is tractable up to three qubits. Typically computing channel monotones is even harder. An $n$-qubit optimisation problem over channels is equivalent to a $2n$-qubit optimisation over states, so naively, direct computation appears impractical for $n > 2$. Meanwhile computing projective dyadic negativity is tractable only for single-qubit operations. This difficulty is aggravated when calculating magic capacities as in principle we have to repeat the optimisation for every state $(\mathcal{E} \otimes \mathbb{1}_n)(|\phi\rangle\langle\phi|)$ such that $|\phi\rangle \in \mathrm{STAB}_{2n}$. The inability to compute monotones exactly in the most general case does not prohibit the use of sub-optimal decompositions for classical simulation. For example, given channel decompositions $\mathcal{E}_1 = \sum_k p_k \mathcal{T}_k$ and $\mathcal{E}_2 = \sum_k q_k \mathcal{T}_k$, we can always construct decompositions for $\mathcal{E}_1 \otimes \mathcal{E}_2$ or $\mathcal{E}_2 \circ \mathcal{E}_1$ in the obvious way, such that the $\ell_1$-norm will be the product of the two. However, ideally we would like to increase the number of qubits for which the problem is tractable, so as to optimise simulation costs.

In Ref. [133], Heinrich and Gross used permutation and Clifford symmetries of the Hadamard and face states states to simplify the problem for multiple copies, e.g. states of the form $|\psi^n\rangle = |\psi\rangle^{\otimes n}$. This allowed them to compute $\mathcal{R}(|\psi\rangle)$ exactly for up to 10 copies, and to solve the problem approximately for up to 26 qubits, including determining an associated sub-optimal decomposition. However, their method depends on the specific symmetries of these states, and cannot be applied to more general states. In the channel picture, in some cases we can ameliorate the difficulty of computation by looking for Clifford gates that commute with the channel of interest. In the rest of this section, we present techniques that ease computation of our monotones for diagonal channels.

### 4.4.1 Reducing the problem size

Consider the case where $\mathcal{E}$ is a diagonal channel, meaning it has a Kraus representation where each Kraus operator is diagonal in the computational basis. This includes diagonal unitaries as a special case. It is straightforward to see how the problem can be simplified for standard and generalised channel robustness, $\mathcal{R}_*$ and $\Lambda^+_*$, as well as magic state monotones for the Choi state. If $\mathcal{E}$ is diagonal, the operation $\mathcal{E} \otimes \mathbb{1}_n$ commutes with any sequence of CNOTs targeted on the last $n$ qubits. But the maximally entangled state $|\Phi_n\rangle$ can be written $|\Phi_n\rangle = U_C(|+^n\rangle \otimes |0^n\rangle)$. Here $U_C = \otimes_{j=1}^n U_j$, where $U_j$ is the CNOT controlled on qubit $j$ and targeted on qubit $n+j$. Given any magic state monotone $\mathcal{M}$, and letting $\rho_{\mathcal{E}} = \mathcal{E}(|+^n\rangle\langle +^n|)$,

$$\mathcal{M}[\Phi_{\mathcal{E}}] = \mathcal{M}[(\mathcal{E} \otimes \mathbb{1}_n)(|\Phi_n\rangle\langle\Phi_n|)] = \mathcal{M}[\rho_{\mathcal{E}} \otimes |0^n\rangle\langle 0^n|] = \mathcal{M}[\rho_{\mathcal{E}}]. \tag{4.151}$$

For the channel robustness we aim to decompose $\rho_{\mathcal{E}}$ in terms of states $\rho_\pm \in \overline{\mathrm{STAB}}_n$, but need to take care that the trace condition $\mathrm{Tr}_A(\rho'_\pm) = \mathbb{1}_n/2^n$ is satisfied for the equivalent $2n$-qubit Choi states $\rho'_\pm$. In Section 4.4.1.1 we show that this holds provided all diagonal elements of $\rho_\pm$ are equal to $1/2^n$. So we can write,

$$\mathcal{R}_*(\mathcal{E}) = \min_{\rho_\pm \in \mathbb{T}_n} \left\{1 + 2p : (1+p)\rho_+ - p\rho_- = \mathcal{E}(|+\rangle\langle +|^{\otimes n}), p \geq 0, \right\}, \tag{4.152}$$

when $\mathcal{E}$ is diagonal and where $\mathbb{T}_n = \{\rho : \rho \in \overline{\mathrm{STAB}}_n \text{and } \langle x| \rho_\pm |x\rangle = 1/2^n, \forall x\}$. The same argument applies for generalised channel robustness, except that we remove the constraint that $\rho_- \in \overline{\mathrm{STAB}}_n$. So provided $\mathcal{E}$ is diagonal, calculation is tractable up to $n = 5$ for $R_*(\mathcal{E})$ and $\mathcal{R}(\Phi_{\mathcal{E}})$ , and up to $n = 3$ for $\Lambda^+_*(\mathcal{E})$ and $\Lambda^+(\Phi_{\mathcal{E}})$. We show in Section 4.4.2 that computing magic capacities can also be greatly simplified.

#### 4.4.1.1 Trace condition for diagonal channels

Consider that the Choi state for a diagonal channel has a decomposition

$$\Phi_{\mathcal{E}} = U_C\big(\rho_{\mathcal{E}} \otimes |0\rangle\langle 0|^{\otimes n}\big)U_C^\dagger = (1+p)\rho_+ - p\rho_-, \tag{4.153}$$

where $\rho_{\mathcal{E}} = \mathcal{E}(|+^n\rangle\langle +^n|)$ and $U_C = \otimes_{j=1}^n U_j$ is the tensor product of CNOTs $U_j$ that are controlled on the $j$-th qubit and targeted on the $(n+j)$-th. Then

$$\rho_{\mathcal{E}} \otimes |0\rangle\langle 0|^{\otimes n} = (1+p)\rho'_+ - p\rho'_-, \tag{4.154}$$

where $\rho'_\pm \in \overline{\mathrm{STAB}}_n$ since $U_C$ is Clifford. Now applying the stabiliser-preserving channel $\mathbb{1}_n \otimes \mathcal{E}_0$ that resets the last $n$ qubits to $|0\rangle\langle 0|^{\otimes n}$ to both sides we obtain

$$\rho_{\mathcal{E}} \otimes |0\rangle\langle 0|^{\otimes n} = (1+p)\rho''_+ \otimes |0\rangle\langle 0|^{\otimes n} - p\rho''_- \otimes |0\rangle\langle 0|^{\otimes n}. \tag{4.155}$$

So from equation (4.153), we obtain $\rho_\pm = U_C\big(\rho''_\pm \otimes |0\rangle\langle 0|^{\otimes n}\big)U_C^\dagger$, and the TP condition becomes

$$\frac{\mathbb{1}_n}{2^n} = \mathrm{Tr}_A(\rho_\pm) = \sum_x \langle x|^A U_C\big(\rho''_\pm \otimes |0\rangle\langle 0|^{\otimes n}\big)U_C^\dagger |x\rangle^A, \tag{4.156}$$

where $|x\rangle$ are the computational basis states on subsystem $A$. Recall that $U_C$ can be written as a tensor product of CNOTs $U_C = \otimes_{j=1}^n U_j$. The CNOT gate controlled on the j-th qubit and targeted on the $(n+j)$-th qubit can be expressed:

$$U_j = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes X = \sum_{x_j=0}^{1} |x_j\rangle\langle x_j| \otimes X_{n+j}^{x_j}. \tag{4.157}$$

Then the tensor product of all $n$ CNOT gates can be written:

$$U_C = \otimes_{j=1}^n \sum_{x_j=0}^{1} |x_j\rangle\langle x_j| \otimes X_j^{x_j} = \sum_x |x\rangle\langle x| \otimes X(x) \tag{4.158}$$

where $x$ is an $n$-bit string, and $X(x) = X_{n+1}^{x_1}X_{n+2}^{x_2}\ldots X_{2n}^{x_n}$. But $X(x)^B|\psi\rangle^A \otimes |0^n\rangle^B = |\psi\rangle^A \otimes |x\rangle^B$ for any $|\psi\rangle$. So we have:

$$U_C\big(\rho''_\pm \otimes |0\rangle\langle 0|^{\otimes n}\big)U_C^\dagger = \sum_x \langle x|\rho''_\pm|x\rangle\, |x\rangle\langle x|^A \otimes |x\rangle\langle x|^B. \tag{4.159}$$

Combined with equation (4.156), the trace-preservation criterion can be expressed as $\frac{\mathbb{1}_n}{2^n} = \sum_x \langle x| \rho''_\pm |x\rangle |x\rangle\langle x|$. Therefore, $\mathcal{E}''_\pm$ are trace-preserving provided that all the diagonal elements of $\rho''_\pm$ are equal to $1/2^n$.

For a given diagonal channel, there always exists a decomposition that satisfies these conditions and has $\ell_1$-norm equal to the channel robustness as defined for the full Choi state. We do not give the full proof here, but sketch the argument. Given any decomposition of the full Choi state $\Phi_{\mathcal{E}} = (1+p)\rho_+ - p\rho_-$ for diagonal $\mathcal{E}$, satisfying the trace condition, one can always find a new decomposition $\Phi_{\mathcal{E}} = (1+p)\mathcal{E}_{\text{DIAG}}(\rho_+) - p\mathcal{E}_{\text{DIAG}}(\rho_-)$ where $\mathcal{E}_{\text{DIAG}}(\rho_\pm)$ still satisfy $\text{Tr}_A(\mathcal{E}_{\text{DIAG}}(\rho_\pm))$, but are now the Choi states for diagonal channels. Here $\mathcal{E}_{\text{DIAG}}$ is a stabiliser circuit that maps any stabiliser Choi state $\Phi_{\mathcal{T}}$ to some other stabiliser Choi state $\Phi_{\mathcal{T}'}$, where $\mathcal{T}'$ is diagonal. Notice that the Choi states for diagonal maps $\mathcal{T}$ have the general form:

$$\Phi_{\mathcal{T}} = \frac{1}{2^n}\sum_{j,k} c_{j,k} |j\rangle^A |j\rangle^B \langle k|^A \langle k|^B . \tag{4.160}$$

In general $c_{j,k}$ can be complex or zero, but terms on the diagonal are constrained. In particular, trace-preserving diagonal channels cannot change the weight of particular computational basis states, so the probability distribution for measurements in the standard basis is the same as for $|\Phi_n\rangle$, that is, $\langle p,q| \Phi_{\mathcal{T}} |p,q\rangle = \frac{1}{2^n}\delta_{p,q}$. The circuit $\mathcal{E}_{\text{DIAG}}$ is defined by the following steps. For each $j$ from 1 to $n$:

1. Perform a parity measurement $(Z\otimes Z)$ between qubits $j$ and $n+j$.
2. If even parity (+1 outcome), do nothing. If odd parity (-1 outcome), perform an $X$ gate on qubit $j$.

This stabiliser-preserving channel leaves Choi states for diagonal maps (and crucially, the target Choi state $\Phi_{\mathcal{E}}$) invariant, but updates general Choi states to have the form (4.160). One can check that the circuit preserves the property $\text{Tr}_A(\rho_\pm) = \mathbb{1}_n/2^n$. We then obtain a decomposition in the desired form:

$$\mathcal{E}_{\text{DIAG}}(\Phi_{\mathcal{E}}) = \Phi_{\mathcal{E}} = (1+p)\mathcal{E}_{\text{DIAG}}(\rho_+) - p\mathcal{E}_{\text{DIAG}}(\rho_-) \tag{4.161}$$

Where $\mathcal{E}_{\text{DIAG}}(\rho_\pm) = (\mathcal{E}_\pm \otimes \mathbb{1})(|\Phi\rangle\langle\Phi|)$ are Choi states for $n$-qubit diagonal channels $\mathcal{E}_\pm$. But as described above, the CNOT sequence $U_C$ commutes with diagonal channels acting on the first $n$ qubits, so we can obtain $n$-qubit representatives of these channels:

$$\mathcal{E}_{\pm}(|+^n\rangle\langle+^n|)\otimes|0\rangle\langle 0|^{\otimes n}=U_C[(\mathcal{E}_{\pm}\otimes\mathbb{1})(|\Phi\rangle\langle\Phi|)]U_C^{\dagger}. \tag{4.162}$$

Discarding the last $n$ qubits we obtain the desired $n$-qubit decomposition:

$$\mathcal{E}(|+^n\rangle\langle+^n|)=(1+p)\mathcal{E}_{+}(|+^n\rangle\langle+^n|)-p\mathcal{E}_{-}(|+^n\rangle\langle+^n|). \tag{4.163}$$

Once again, the same arguments apply for generalised robustness, since we only need discard the constraint that $\mathcal{E}_{-}$ is stabiliser-preserving.

### 4.4.2 Magic capacity in the affine space picture

In this section we will make use of the formalism due to Dehaene and De Moor (Section 1.1.5.3), in which stabiliser states are cast in terms of affine spaces and quadratic forms over binary vectors [54, 55], to prove the following theorem.

**Theorem 4.33** (Magic capacity for diagonal operations)**.** *Suppose the n-qubit channel $\mathcal{E}_D$ is diagonal. Let*

$$|\mathcal{K}\rangle=\frac{1}{|\mathcal{K}|^{1/2}}\sum_{\mathbf{x}\in\mathcal{K}}|\mathbf{x}\rangle\,, \tag{4.164}$$

*where $\mathbf{x}\in\mathbb{F}_2^n$ are binary vectors and $\mathcal{K}\subseteq\mathbb{F}_2^n$ is an affine space. Then, for any well-behaved magic state monotone $\mathcal{M}$, we have $\mathcal{C}_{\mathcal{M}}(\mathcal{E}_D)=\max_{\mathcal{K}}\mathcal{M}[\mathcal{E}_D(|\mathcal{K}\rangle\langle\mathcal{K}|)]$.*

That is, given an $n$-qubit channel $\mathcal{E}$, provided the channel is diagonal, the capacity $\mathcal{C}_{\mathcal{M}}(\mathcal{E})$ may be calculated by optimisation over only the $n$-qubit states $|\mathcal{K}\rangle$ as defined in equation (4.164), rather than over all $2n$-qubit stabiliser states.

Recall that any pure $n$-qubit stabiliser state can be expressed in the form

$$|\mathcal{K},q,\mathbf{d}\rangle=\sqrt{|\mathcal{K}|}\sum_{\mathbf{x}\in\mathcal{K}}i^{\mathbf{d}^T\mathbf{x}}(-1)^{q(\mathbf{x})}|\mathbf{x}\rangle\,, \tag{4.165}$$

where $\mathcal{K}\subseteq\mathbb{F}_2^n$ is an affine space, $\mathbf{d}$ is some fixed binary vector, and $q(\mathbf{x})$ has the form $q(\mathbf{x})=\mathbf{x}^TQ\mathbf{x}+\lambda^T\mathbf{x}$. Here $Q$ is a binary, strictly upper triangular matrix, $\lambda$ is a vector, and addition is modulo 2. Conversely, any state that can be written in this way is a stabiliser state. Also recall that every affine space is related to exactly one linear subspace, by $\mathcal{K}=\mathcal{L}+\mathbf{h}$ for some shift vector $\mathbf{h}$, and the dimension $k=\dim(\mathcal{K})$ of an affine space means the dimension of the corresponding subspace. Instead of enumerating all elements of an affine space, we can specify it by a shift vector $\mathbf{h}$ and an $n \times k$ matrix where each column is one of the generators of the corresponding linear space,

$$G = \begin{pmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_k \end{pmatrix} = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,k} \\ \vdots & \vdots & & \vdots \\ g_{j,1} & g_{j,2} & \cdots & g_{j,k} \\ \vdots & \vdots & & \vdots \\ g_{n,1} & g_{n,2} & \cdots & g_{n,k} \end{pmatrix}. \tag{4.166}$$

We have freedom in our choice of $k$ independent generators, and we can transform between equivalent generating sets by adding any two columns of $G$. We are also free to swap any two columns. A general transform between generating sets can therefore be represented by an invertible matrix $S$ of dimension $k \times k$, multiplying on the right $G \longrightarrow GS$.

Any linear transformation of the affine space can be fully specified by the transformation of the generators and the shift vector. In particular, we can represent the action of a single CNOT by multiplication on the left by a matrix $C$ [54, 61]. If the CNOT has control qubit $j$ and target qubit $k$, then $C$ has 1s on the diagonal, a 1 in the $j$th element of the $k$th row, and zeroes everywhere else. A sequence for a $2n$-qubit system, in which CNOTs are always controlled on the first $n$ qubits, and targeted on the last $n$ qubits can be represented in block form as

$$C = \begin{pmatrix} \mathbb{1} & 0 \\ M & \mathbb{1} \end{pmatrix}, \tag{4.167}$$

where each block has dimension $n \times n$, and $M$ can be any binary matrix. We use this formalism to prove the following lemma, which leads directly to Theorem 4.33.

**Lemma 4.34** (Equivalences for diagonal channels)**.** *Suppose $\mathcal{E}_D$ is a diagonal CPTP channel, and let $\mathcal{M}$ be a well-behaved magic state monotone (faithful, monotone under completely stabiliser-preserving channels and submultiplicative under tensor product). Then the following statements hold.*

1. *All initial stabiliser states with the same affine space $\mathcal{K}$ result in the same value of*

$\mathcal{M}$. *That is, for any valid* $q$, $q'$, $\mathbf{d}$ *and* $\mathbf{d}'$,

$$\mathcal{M}[(\mathcal{E}_D \otimes \mathbb{1})(|\mathcal{K}, q, \mathbf{d}\rangle\langle\mathcal{K}, q, \mathbf{d}|)] = \mathcal{M}[(\mathcal{E}_D \otimes \mathbb{1})(|\mathcal{K}, q', \mathbf{d}'\rangle\langle\mathcal{K}, q', \mathbf{d}'|)]. \quad (4.168)$$

2. *Given a* $2n$*-qubit state* $|\phi\rangle \in \mathrm{STAB}_{2n}$ *, there exists some* $n$*-qubit* $|\phi'\rangle \in \mathrm{STAB}_n$ *such that* $\mathcal{M}[(\mathcal{E}_D \otimes \mathbb{1}_n)(|\phi\rangle\langle\phi|)] = \mathcal{M}[\mathcal{E}_D(|\phi'\rangle\langle\phi'|)]$.

*Proof.* We first prove statement 1. Since any magic monotone must be invariant under Clifford operations, we need to show that there exists a Clifford unitary $U$ that converts $(\mathcal{E}_D \otimes \mathbb{1})(|\mathcal{K}, q, \mathbf{d}\rangle\langle\mathcal{K}, q, \mathbf{d}|)$ to $(\mathcal{E}_D \otimes \mathbb{1})(|\mathcal{K}, q', \mathbf{d}'\rangle\langle\mathcal{K}, q', \mathbf{d}'|)$. A suitable choice for $U$ is one such that $U\left|\phi_{\mathcal{K},q,\mathbf{d}}\right\rangle = \left|\phi_{\mathcal{K},q',\mathbf{d}'}\right\rangle$, and, crucially, that commutes with the channel $\mathcal{E}_D$. Since $\mathcal{E}_D$ is given to be diagonal, any diagonal Clifford $U$ will suffice. The affine space $\mathcal{K}$ remains unchanged, so we only need show there is always a diagonal Clifford that maps $q \to q'$ and $\mathbf{d} \to \mathbf{d}'$ for any $q$, $q'$, $\mathbf{d}$ and $\mathbf{d}'$. That this is always possible is perhaps already evident from Ref. [54], but for completeness we give the argument here.

We can convert $\mathbf{d}$ to $\mathbf{d}'$ using appropriately chosen $S_j$ gates, meaning the gate $\mathrm{diag}(1, i)$ acting on the $j$th qubit. Consider the action of $S_j$ on a basis vector:

$$S_j|\mathbf{x}\rangle = \begin{cases} |\mathbf{x}\rangle & \text{if } x_j = 0 \\ i|\mathbf{x}\rangle & \text{if } x_j = 1 \end{cases} \quad (4.169)$$

If we define basis vector $\mathbf{e}_j$ so that it has 1 in the $j$th position and zeroes elsewhere, we can write the action of $S_j$ as $S_j|\mathbf{x}\rangle = i^{\mathbf{e}_j^T\mathbf{x}}|\mathbf{x}\rangle$. Note that the form of this equation is independent of the value of $\mathbf{x}$, so we can write:

$$\begin{aligned} S_j\left|\phi_{\mathcal{K},q,\mathbf{d}}\right\rangle &= \frac{1}{|\mathcal{K}|^{1/2}} \sum_{\mathbf{x}\in\mathcal{K}} i^{\mathbf{d}^T\mathbf{x}}(-1)^{q(\mathbf{x})} S_j|\mathbf{x}\rangle & (4.170) \\ &= \sum_{\mathbf{x}\in\mathcal{K}} i^{(\mathbf{d}^T+\mathbf{e}_j^T)\mathbf{x}}(-1)^{q(\mathbf{x})} S_j|\mathbf{x}\rangle. & (4.171) \end{aligned}$$

So, we can flip any bit of $\mathbf{d}$ by applying the correct $S$ gate, leaving $q(\mathbf{x})$ unchanged.

Now consider $q(\mathbf{x}) = \mathbf{x}^T Q\mathbf{x} + \lambda^T\mathbf{x}$, which we must convert to some other $q'(\mathbf{x}) = \mathbf{x}^T Q'\mathbf{x} + \lambda'^T\mathbf{x}$. We can use the same trick as above to convert any $\lambda$ to any other $\lambda'$, by replacing $S_j$ with the $Z_j$ gate, i.e. $\mathrm{diag}(1, -1)$ acting on the $j$th qubit. For $Q$ we can use the controlled-$Z$ gate between the $j$th and $k$th qubit, which we denote $CZ_{jk}$.

This acts on basis states as $CZ_{jk}\left|\mathbf{x}\right\rangle = (-1)^{\mathbf{x}^T M_{jk}\mathbf{x}}\left|\mathbf{x}\right\rangle$, where $M_{jk}$ is the $n \times n$ matrix with a 1 in position $(j,k)$ and zeroes everywhere else. The set of all $\{M_{jk}\}$ form a basis for $n \times n$ binary matrices, hence we can convert any $Q$ to any other $Q'$ by an appropriately chosen sequence of $CZ$ gates, leaving $\mathbf{d}$ and $\lambda$ untouched. This completes the proof of statement 1.

Now to prove statement 2. From statement 1 any stabiliser state $\left|\phi\right\rangle$ is equivalent to $\left|\mathcal{K}\right\rangle = |\mathcal{K}|^{-1/2}\sum_{\mathbf{x}\in\mathcal{K}}\left|\mathbf{x}\right\rangle$, up to some diagonal Clifford, for some $\mathcal{K}$. The strategy is to find a Clifford unitary $U$ that commutes with $\mathcal{E}_D$, and converts the $2n$-qubit stabiliser state $\left|\mathcal{K}\right\rangle$ to some product of two $n$-qubit states $\left|\mathcal{K}'\right\rangle = \left|\mathcal{K}'_A\right\rangle \otimes \left|\mathcal{K}'_B\right\rangle$. If such a Clifford gate exists, we would have

$$\begin{aligned}\mathcal{M}[(\mathcal{E}_D \otimes \mathbb{1}_n)(\left|\mathcal{K}\right\rangle\!\left\langle\mathcal{K}\right|)] &= \mathcal{M}\left[(\mathcal{E}_D \otimes \mathbb{1}_n)(\left|\mathcal{K}'_A\right\rangle\!\left\langle\mathcal{K}'_A\right| \otimes \left|\mathcal{K}'_B\right\rangle\!\left\langle\mathcal{K}'_B\right|)\right] \qquad (4.172)\\ &= \mathcal{M}\left[\mathcal{E}_D(\left|\mathcal{K}'_A\right\rangle\!\left\langle\mathcal{K}'_A\right|) \otimes \left|\mathcal{K}'_B\right\rangle\!\left\langle\mathcal{K}'_B\right|\right] = \mathcal{M}\left[\mathcal{E}_D(\left|\mathcal{K}'_A\right\rangle\!\left\langle\mathcal{K}'_A\right|)\right],\end{aligned}$$

where the last step follows as $\left|\mathcal{K}'_B\right\rangle$ is a stabiliser state so makes no contribution to the robustness. The state $\left|\mathcal{K}'\right\rangle = U\left|\mathcal{K}\right\rangle$ can be factored as $\left|\mathcal{K}'_A\right\rangle \otimes \left|\mathcal{K}'_B\right\rangle$ provided its generator $G'$ can be written in diagonal matrix form,

$$G' = \begin{pmatrix} G'_A & 0 \\ 0 & G'_B \end{pmatrix}, \qquad (4.173)$$

where $G'_A$ and $G'_B$ have $n$ rows, representing generators for affine spaces $\mathcal{K}'_A$ and $\mathcal{K}'_B$.

We now show that we can always reach this form by a Clifford $U_C$ comprised of a sequence of CNOTs targeted on the last $n$ qubits. Such a sequence always commutes with $\mathcal{E}_D \otimes \mathbb{1}_n$. Suppose we have some $2n \times k$ generator $G$ for an affine space $\mathcal{K}$ with $k = \dim(\mathcal{K})$:

$$G = \begin{pmatrix} G_A \\ G_B \end{pmatrix}, \qquad (4.174)$$

where $G_A$ and $G_B$ are each $n \times k$ submatrices. The full matrix $G$ will have rank $k$, and $G_A$ will have some rank $m \leq k$. Either $G_A$ is already full rank ($m = k$), or it can be reduced to the following form by elementary column operations, which is equivalent to

| $n$ | Stabiliser states | Total affine spaces | Non-trivial affine spaces |
|---|---|---|---|
| 2 | 60 | 11 | 7 |
| 3 | 1,080 | 51 | 43 |
| 4 | 36,720 | 307 | 291 |
| 5 | 2,423,520 | 2451 | 2419 |

**Table 4.2:** Number of $n$-qubit stabiliser states compared with number of affine spaces. By trivial affine spaces we mean those comprised of a single element, which correspond to computational basis states. Diagonal CPTP channels act as the identity on such states.

multiplication on the right by a $k \times k$ matrix $S$:

$$G_A \longrightarrow G_A S = \begin{pmatrix} G'_A & 0 \end{pmatrix}, \tag{4.175}$$

where $G'_A$ is $n \times m$ (and hence full column rank), and 0 is $n \times (k-m)$. Multiplying $G$ on the right by $S$, we interpret as a change in the choice of generating set:

$$G \longrightarrow GS = \begin{pmatrix} G_A S \\ G_B S \end{pmatrix} = \begin{pmatrix} G'_A & 0 \\ G''_B & G'_B \end{pmatrix}. \tag{4.176}$$

Now, apply the Clifford $U_C$ described by the matrix $C$ in equation (4.167). This transforms the generator to:

$$G' = CGS = \begin{pmatrix} \mathbb{1} & 0 \\ M & \mathbb{1} \end{pmatrix} \begin{pmatrix} G'_A & 0 \\ G''_B & G'_B \end{pmatrix} = \begin{pmatrix} G'_A & 0 \\ MG'_A + G''_B & G'_B \end{pmatrix}. \tag{4.177}$$

Note that if $G_A$ was already full rank, the change of generating set is not necessary. If we can set the bottom-left submatrix to zero, then $U_C |\mathcal{K}\rangle$ can be factored as described above. This is possible if there exists a binary matrix $M$ such that $MG'_A = G''_B$. But $G'_A$ has full column rank $m$, so there exists an $m \times n$ left-inverse $G'^{-1}_{A,\text{left}}$ such that $G'^{-1}_{A,\text{left}} G'_A = \mathbb{1}$, where $\mathbb{1}$ is $m \times m$. Then we can set $M = G''_B G'^{-1}_{A,\text{left}}$, so that:

$$MG'_A = G''_B G'^{-1}_{A,\text{left}} G'_A = G'_B \mathbb{1} = G''_B. \tag{4.178}$$

Then $G' = CGS$ is in the form (4.173), so $U_C |\mathcal{K}\rangle = |\mathcal{K}'_A\rangle \otimes |\mathcal{K}'_B\rangle$, as required. □

Lemma 4.34 shows that if $\mathcal{E}_D$ is diagonal then for any $2n$-qubit stabiliser state $|\phi\rangle$ we have that $\mathcal{M}[(\mathcal{E}_D \otimes \mathbb{1}_n)(|\phi\rangle\langle\phi|)] = \mathcal{M}[\mathcal{E}_D(|\mathcal{K}\rangle\langle\mathcal{K}|)]$ for some $n$-qubit affine space $\mathcal{K}$. So, the capacity with respect to any well-behaved magic state monotone can be calculated by maximising over just the representative states $|\mathcal{K}\rangle$, proving Theorem 4.33. Table 4.2 illustrates the reduction in problem size. For example, whereas naively for a two-qubit channel we would need to compute $\mathcal{M}$ for all $36,720$ four-qubit stabiliser states, using the result above we only need check one stabiliser state for each of the 7 non-trivial two-qubit affine spaces. The problem therefore becomes tractable up to five qubits for $\mathcal{C}_{\mathcal{R}}$ and up to three for $\mathcal{C}_{\Lambda^+}$.

### 4.4.3 Dimension of affine space

We next make further observations that will later help interpret numerical results in Chapter 6.

**Observation 4.35** (Dimension of initial affine space limits generation of magic)**.** *Suppose $U$ is a diagonal unitary acting on $n$ qubits, and suppose $|\mathcal{K}\rangle$ is a stabiliser state associated with some affine space $\mathcal{K}$, $k = \dim(\mathcal{K})$. Then for any well-behaved magic state monotone $\mathcal{M}$, we find that $\mathcal{M}(U|\mathcal{K}\rangle) = \mathcal{M}(U'|\phi'\rangle)$ where $U'|\phi'\rangle$ is a state on only $k$ qubits, and $U'$ is some $k$-qubit unitary. Therefore $\mathcal{M}(U|\mathcal{K}\rangle)$ is upper-bounded by the maximum $\mathcal{M}(\rho)$ achievable for an $k$-qubit state $\rho$.*

*Proof.* We prove the result by showing that there is a sequence of Clifford gates that takes $U|\mathcal{K}\rangle$ to the product of a $k$-qubit state and an $(n-k)$-qubit stabiliser state. We know from Lemma 4.34 that for diagonal unitaries, all states with same affine space have the same magic, so it is enough to consider the state $|\mathcal{K}\rangle = \frac{1}{\sqrt{|\mathcal{K}|}}\sum_{\mathbf{x}\in\mathcal{K}}|\mathbf{x}\rangle$. A diagonal unitary will map this to some state $U|\mathcal{K}\rangle = \frac{1}{\sqrt{|\mathcal{K}|}}\sum_{\mathbf{x}\in\mathcal{K}} e^{i\theta_{\mathbf{x}}}|\mathbf{x}\rangle$, where $\left\{e^{i\theta_{\mathbf{x}}}\right\}$ will be a subset of the diagonal elements of $U$. The affine space $\mathcal{K}$ has a generator matrix of rank $k$. Lemma 4.34 showed that a sequence of elementary row operations on the generator matrix can be realised by a sequence of CNOT gates. So we can use Clifford gates to transform any rank $k$ generator matrix as

$$G \longrightarrow G' = AG = \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix}, \tag{4.179}$$

where $\mathbb{1}$ is the $k\times k$ identity. Each element of $\mathcal{K}$ can be written $\mathbf{x} = \sum_j \mathbf{g}_j + \mathbf{h}$, where $\sum_j \mathbf{g}_j$ is some combination of columns of $G$, and $\mathbf{h}$ is a fixed shift vector. The matrix $A$ represents a sequence of CNOTs collected in a single Clifford unitary $U_A$, that acts on $n$-qubit computational basis states $|\mathbf{x}\rangle$, where $\mathbf{x}\in\mathcal{K}$, as $U_A|\mathbf{x}\rangle = |\mathbf{y}(\mathbf{x})\rangle\otimes|\mathbf{h}'\rangle$, where $\mathbf{h}'$

is an $(n-k)$-length vector, and $\mathbf{y}(\mathbf{x})$ is a $k$-length vector given by,

$$\begin{pmatrix} \mathbf{y}(\mathbf{x}) \\ \mathbf{h}' \end{pmatrix} = A\mathbf{x} = \sum_j A\mathbf{g}_j + A\mathbf{h}. \tag{4.180}$$

Note that $\mathbf{y}(\mathbf{x})$ is only defined for $\mathbf{x} \in \mathcal{K}$, and that $\mathbf{h}'$ is independent of $\mathbf{x}$. Elements $\mathbf{x} \in \mathbb{F}_2^n$ that are not in $\mathcal{K}$ could be mapped to a vector where the last $n-k$ bits are not $\mathbf{h}'$, but these never appear as terms of $U\left|\mathcal{K}\right\rangle$. Since $U_A$ must preserve orthogonality, each $\left|\mathbf{x}\right\rangle$, where $\mathbf{x} \in \mathcal{K}$, maps to a distinct element of the $k$-qubit basis set $\{\left|\mathbf{y}\right\rangle\}$. Since $\mathbf{y}$ are length $k$ and there are $2^k$ distinct elements, they must form the $k$-bit linear space $\mathcal{L}' = \mathbb{F}_2^k$. So we can write

$$U_A U \left|\mathcal{K}\right\rangle = \frac{1}{\sqrt{|\mathcal{L}'|}} \sum_{\mathbf{y} \in \mathcal{L}'} e^{i\theta'_{\mathbf{y}}} \left|\mathbf{y}\right\rangle \otimes \left|\mathbf{h}'\right\rangle = (U' \left|\mathcal{L}'\right\rangle) \otimes \left|\mathbf{h}'\right\rangle, \tag{4.181}$$

where $\left|\mathcal{L}'\right\rangle$ is a $k$-qubit stabiliser state, and $U'$ is the $k$-qubit diagonal unitary with $e^{i\theta'_{y(\mathbf{x})}} = e^{i\theta_{\mathbf{x}}}$ as the non-zero elements. The state $\left|\mathbf{h}'\right\rangle$ is a stabiliser state, so does not increase the monotone value of $U_A U\left|\mathcal{K}\right\rangle$, and therefore $\mathcal{M}(U\left|\mathcal{L}(\mathcal{K})\right\rangle) = \mathcal{M}(U_A U \left|\mathcal{L}(\mathcal{K})\right\rangle) = \mathcal{M}(U'\left|\mathcal{L}'\right\rangle)$, where $U'\left|\mathcal{L}'\right\rangle$ is a $k$-qubit state. □

Finally, consider the special case of multi-control phase gates $M_{t,n}$, defined

$$M_{t,n} = \mathrm{diag}(\exp\left(i\pi/2^t\right), 1, 1, \ldots, 1), \quad t \in \mathbb{Z}. \tag{4.182}$$

By convention, controlled-phase gates typically apply the phase to the all-one state $\left|1^n\right\rangle$, but the form given above is Clifford-equivalent to the conventional version, and is more convenient for the arguments below. Notice that $M_{t,n}$ acts as the identity on states $\left|\mathcal{K}\right\rangle$ unless $\mathcal{K}$ contains the zero vector $0^n = (0,\ldots,0)^T$, so if $0^n \notin \mathcal{K}$, we get $\mathcal{R}(M_{t,n}\left|\mathcal{K}\right\rangle) = 1$. But if $0^n \in \mathcal{K}$, then $\mathcal{K}$ is a linear subspace. So for this type of gate, to find all possible values of $\mathcal{R}(M_{t,n}\left|\mathcal{K}\right\rangle) > 1$ we need only consider linear subspaces. The following theorem implies that we only need solve one optimisation for each possible $k \leq n$ rather than one for every linear subspace.

**Theorem 4.36.** *Consider the n-qubit gate $M_{t,n}$ defined by equation* (4.182)*, and let $\mathcal{L}_A$ and $\mathcal{L}_B$ be linear subspaces such that* $\dim(\mathcal{L}_A) = \dim(\mathcal{L}_B) = k$*. Then for any well-*

*behaved magic monotone $\mathcal{M}$,*

$$\mathcal{M}[M_{t,n}\,|\mathcal{L}_A\rangle] = \mathcal{M}[M_{t,n}\,|\mathcal{L}_B\rangle]. \tag{4.183}$$

*Proof.* We largely repeat the arguments of Observation 4.35, for the special case where the phases are given by

$$\theta_{\mathbf{x}} = \begin{cases} \pi/2^t & \text{if} \quad \mathbf{x} = 0^n \\ 0 & \text{otherwise.} \end{cases} \tag{4.184}$$

Since $\dim(\mathcal{L}_A) = \dim(\mathcal{L}_B)$, their generator matrices $G_A$ and $G_B$ have the same rank. It follows from the arguments of Observation 4.35 that there exists an invertible $C$, corresponding to a sequence of CNOT gates, such that $G_B = CG_A$, and $|\mathcal{L}_B\rangle = U_C\,|\mathcal{L}_A\rangle$, where $U_C$ is a unitary Clifford operation. The state $M_{t,n}\,|\mathcal{L}_A\rangle$ can be expressed with precisely the same basis vectors as $|\mathcal{L}_A\rangle$, but with updated phase. Clearly, since any CNOT acts as the identity on $|0^n\rangle$, we obtain

$$U_C M_{t,n}\,|\mathcal{L}_A\rangle = \frac{1}{2^{k/2}} \sum_{\mathbf{x}\in\mathcal{L}_B} \exp(i\theta_{\mathbf{x}})\,|\mathbf{x}\rangle = M_{t,n}\,|\mathcal{L}_B\rangle \tag{4.185}$$

Since $U_C$ is a reversible Clifford, equation (4.183) follows by monotonicity. □

From Theorem 4.36, then, to find $\mathcal{C}_{\mathcal{M}}(M_{t,n})$, we only need calculate $\mathcal{M}(M_{t,n}\,|\mathcal{L}\rangle)$ for a single representative subspace for each possible value of $\dim(\mathcal{L})$. Recall that for $n$-qubit stabiliser states $|\mathcal{L}\rangle$, $k = \dim\mathcal{L}$ can take integer values from 0 to $n$. The states with $k = 0$ correspond to single computational basis states without superposition, so are unaffected by phase gates. That is, for $n$-qubit multicontrol phase gates we only have to evaluate $\mathcal{M}$ for $n$ distinct states. Compare this to the number of optimisation problems we would need to solve without using the above observations (Table 4.2).

We can go further. From Observation 4.35, for a subspace with $\dim(\mathcal{L}) = k < n$, it must be the case that $M_{t_n}\,|\mathcal{L}\rangle$ is Clifford-equivalent to $(U'\,|\mathcal{L}'\rangle) \otimes |h'\rangle$ for the $k$-qubit state $|\mathcal{L}'\rangle$ and $(n-k)$-qubit computational basis state $|h'\rangle$, and some diagonal $k$-qubit unitary $U'$. By inspection of the phases given by equation (4.184), $U'$ can only be the $k$-qubit multicontrol gate $M_{t,k}$. This leads to the following.

**Observation 4.37** ($n$-qubit multicontrol gates)**.** *For any fixed $t$ and $n$-qubit state $|\mathcal{L}\rangle$*

*where* $\dim(\mathcal{L}) = k < n$*, we have:*

$$\mathcal{M}(M_{t,n}|\mathcal{L}\rangle) = \mathcal{M}(M_{t,k}|\mathcal{L}'\rangle) \tag{4.186}$$

*where* $|\mathcal{L}'\rangle$ *is the k-qubit state with* $\mathcal{L}' = \mathbb{F}_2^k$*, and* $\mathcal{M}$ *is any well-behaved monotone.*

We have seen in this section that we can significantly ease the computation of magic capacities for all diagonal channels, and we can simplify the problem still further for highly structured gates like the multicontrol gates considered above. Our results showing Clifford-equivalences between certain magic states of the form $\mathcal{E}(|\mathcal{K},q,\mathbf{d}\rangle\langle\mathcal{K},q,\mathbf{d}|)$ may also be of interest for the study of state conversion.

## 4.5 Summary and conclusions

In this chapter we presented several contributions to the resource theory of magic for states and operations on $n$-qubit systems. We summarise the measures of magic presented in the preceding sections in Table 4.3. Of the quantities listed in this table, all are novel contributions except for robustness of magic $\mathcal{R}$, which was introduced by Howard and Campbell [103]. All are well-behaved magic monotones in the sense we described earlier in the chapter, with the exception of channel extent, whose domain restricted to channels in the convex hull of the unitary operations. The general pattern has been that we started with state monotones related to decompositions of a particular type, then defined channel monotones corresponding to analogous decompositions of channels. Finally we defined magic capacities, based on the maximum increase in magic with respect to one of the magic state monotones. In the last part of the chapter we showed that computation of the monotones can be significantly simplified for the case of diagonal channels. We note that the family of extent-type monotones is incomplete compared to the other

| Type | State monotone | Channel monotone | Capacity | Sandwich theorem |
|---|---|---|---|---|
| Robustness | $\mathcal{R}$ (Ref. [103]) | $\mathcal{R}_*$ | $\mathcal{C}_\mathcal{R}$ | Y |
| Gen. robustness | $\Lambda^+$ | $\Lambda^+_*$ | $\mathcal{C}_{\Lambda^+}$ | Y |
| Dyadic | $\Lambda$ | $\Lambda_*$ | $\mathcal{C}_\Lambda$ | Y |
| Extent | $\Xi$ | $\Xi_*$ (restricted) | - | N |

**Table 4.3:** Families of magic monotones presented in this chapter. “Sandwich theorem” column indicates whether we were able to prove the chain of inequalities $\mathcal{M}(\Phi_\mathcal{E}) \leq \mathcal{M}_*(\mathcal{E}) \leq \mathcal{C}_\mathcal{M}(\mathcal{E})$ for each family. Note that the channel extent $\Xi_*$ is defined only for channels that can be expressed as a convex combination of unitary operations.

monotones we studied. In addition to the fact that it is not clear how to define an appropriate extent monotone for general CPTP maps, we did not discuss capacity with respect to density-operator extent $\Xi$. In part this is because computing this capacity would likely be intractable. Although computing $\Xi$ is a solved problem for tensor products of single-qubit states due to the results of Section 4.1, it remains hard in general for multi-qubit states. Consequently we do not expect the associated magic capacity to be computable in the general case. Nevertheless, if a generalised channel extent could be defined, it may be possible to bound the capacity using a sandwich theorem. We leave these matters as open questions for future work.

In the next chapter we will show that for every one of the magic monotones we have presented, there exists a classical simulation algorithm where the monotone quantifies the performance in some respect. We will see that the different types of decomposition used in the definition of each monotone lead to simulators that are useful for distinct simulation tasks.

# Chapter 5

# Classical simulation algorithms for non-stabiliser circuits

We next introduce several improved classical algorithms, addressing previously unresolved questions on the classical simulation of stabiliser circuits supplemented by magic resources. We first introduce an efficient subroutine for updating a stabiliser state subject to a completely stabiliser-preserving CPTP map. Surprisingly, while the Gottesman-Knill theorem is well known, and various extensions have appeared in the literature, to our knowledge the classical simulability of general completely stabiliser-preserving CPTP maps $\mathrm{SP}_{n,n}$ has not previously been studied.

Having shown in Chapter 4 that the connection between stabiliser extent and robustness of magic can be understood formally via three new magic monotones for states, in Section 5.2 we show that each monotone is associated with a different classical simulator for stabiliser circuits with magic state inputs. We demonstrate how stabiliser rank methods, previously limited to pure state evolution under unitary gates and projective Pauli measurements, can be extended to the density operator picture, admitting mixed magic state inputs and noisy stabiliser channels. In the process we obtain a significant improvement in performance. We also present a generalisation of the quasiprobability sampling method discussed in Chapter 2, improving the exponent in the exponential scaling of the runtime with magic resources.

We then present a suite of simulation algorithms for circuits that involve general magic channels as well as magic states. Each simulator is linked with one of the channel monotones defined in Chapter 4. These can broadly be classed into *static* simulators, based on precomputed channel decompositions, and *dynamic* simulators, which call optimisation subroutines during the simulation. In this sense the Oak Ridge simulator [45] discussed in Section 2.5.1 was an example of a static simulator, though its performance was not related to a true magic monotone, and its efficiency was limited by a restricted set of free operations. We show how it can be generalised so that is efficient for all CPTP completely stabiliser-preserving maps on bounded number of qubits $\mathrm{SP}_{n,n}$, with runtime measured by the channel robustness. This yields reduced runtime for certain non-stabiliser channels. We go on to present static simulators with performance quantified by generalised channel robustness and dyadic channel negativity. We also introduce a bit-string sampling algorithm for noisy non-stabiliser circuits restricted to unital channels, with the simulation cost of each circuit element given by its channel extent.

Next we present a family of dynamic channel simulators associated with magic capacity monotones. These can lead to reduced sample complexity, at the cost of having to solve few-qubit optimisation problems on the fly during the simulation. In Chapter 6, we will combine the insights from the resource theory of magic with the understanding of classical simulation overhead developed in this chapter to estimate simulation costs for specific classes of channel, so as to compare the performance of our various simulators in different contexts. For now, we introduce an efficient stabiliser-Kraus subroutine that functions as a key part of several of our algorithms.

## 5.1 Stabiliser-Kraus subroutine

The original Gottesman-Knill algorithm showed that the update of stabiliser states under Clifford operations, and the computation of both the outcome probabilities and post-measurement states after a Pauli measurement can be done efficiently. Efficient update rules for certain composite stabiliser operations can be devised straightforwardly. In particular one can efficiently simulate the channel $\mathcal{E} = \sum_j p_j \mathcal{E}_j$ provided the update rule for each $\mathcal{E}_j$ is known and $\{p_j\}_j$ can be efficiently sampled from. Indeed, the Oak Ridge simulator [45] can efficiently simulate convex combinations of elements in CPR, the set of Clifford operations supplemented by Pauli reset channels. However, there exist completely stabiliser-preserving channels outside the convex hull of CPR, and for these the algorithm as stated in Ref. [45] is not efficient. Here we present a subroutine for probabilistically performing the update of an $n$-qubit stabiliser state for any CPTP map in $\mathrm{SP}_{m,m}$, $m \leq n$. This allows efficient simulation of any completely stabiliser-preserving

CPTP map, subject to modest assumptions about the channel decomposition. We present two versions of the subroutine, one of which first appeared in our article Ref. [1], the other being a variant introduced in our pre-print Ref. [2]. To our knowledge, a result for the efficient simulation of general stabiliser-preserving CPTP maps had not previously appeared in the literature prior to Ref. [1].

Both variants use the fact that the Choi state for any completely stabiliser-preserving CPTP map $\mathcal{E} \in \mathrm{SP}_{m,m}$ is a mixed stabiliser state, so can be decomposed as a convex combination $\Phi_{\mathcal{E}} = \sum_j p_j |\phi_j\rangle\langle\phi_j|$, where $|\phi_j\rangle \in \mathrm{STAB}_{2m}$ are pure stabiliser states. In turn, each pure stabiliser state corresponds to a completely stabiliser-preserving Kraus operator in the decomposition of the channel, $|\phi_j\rangle = K_j \otimes \mathbb{1}_m |\Phi_m\rangle$. In general $K_j$ may not preserve trace, so that transition probabilities for each operator depend on the initial state $\rho$. However, the total channel $\mathcal{E}$ is trace-preserving, which means that the final state $\mathcal{E}(\rho)$ will be *some* probability distribution over pure stabiliser states. Given initial state $|\phi\rangle$, the probability of obtaining the outcome corresponding to the $j$-th Kraus operator is given by $P_j = p_j \big\| K_j |\phi\rangle \big\|^2$, where $|\Omega_m\rangle = 2^{-m/2} \sum_j |j\rangle \otimes |j\rangle$. This probability can always be computed efficiently and exactly (up to machine precision in the coefficient $p_j$). The full probability distribution can then be computed efficiently provided the number of Kraus operators is not too large. We will make this precise shortly. The first version of the subroutine is given in Algorithm 4. We assume that the channel is supplied as an $N_K$-length list of couples $\mathbb{L} = \{(p_1, \phi_1), \ldots, (p_{N_K}, \phi_{N_K})\}$, where each couple corresponds to the coefficient $p_j$ and pure stabiliser state $\phi_j$ from the $j$-th term of $\Phi_{\mathcal{E}} = \sum_j p_j |\phi_j\rangle\langle\phi_j|$. For concreteness, we assume that the description of the initial state $\psi$ and each $\phi_j$ is given in stabiliser tableau format [31], requiring $\mathcal{O}(m^2)$ classical bits to describe each Kraus operator, and $\mathcal{O}(n^2)$ for the initial state. Without loss of generality we assume that $\mathcal{E}$ is applied to the first $m$ qubits.

**Algorithm 4** Stabiliser Kraus update subroutine (Teleportation scheme)

```
Input: Initial n-qubit stabiliser tableau for state $|\psi\rangle$, length $N_K$ list $\mathbb{L}$ with entries $(p_j, \phi_j)$
    describing the Choi state decomposition for m-qubit channel $\mathcal{E}$.
Output: Updated n-qubit tableau for state $|\psi'\rangle$.
for $r \leftarrow 1$ to $N_K$ do
    Prepare tableau $T_r$ for $(2m+n)$-qubit state $|\phi_r\rangle^{AB} \otimes |\psi\rangle^{CD}$.
    Use $T_r$ to compute $P_r \leftarrow p_r \left\| (\mathbb{1}_n^{AD} \otimes |\Phi_m\rangle\langle\Phi_m|^{BC}) \, |\phi_r\rangle^{AB} \otimes |\psi\rangle^{CD} \right\|^2$.
end for
Sample index $s$ from distribution $\{P_s\}$.
Prepare tableau $T_s$ for $|\phi_s\rangle^{AB} \otimes |\psi\rangle^{CD}$.
Update tableau $T_s'$ after projection $(\mathbb{1}_n^{AD} \otimes |\Phi_m\rangle\langle\Phi_m|^{BC})$.
Find reduced tableau $T_s''$ after discarding subsystem $BC$.
```

We call this the teleportation scheme, since the Choi state is employed as a resource state for gate teleportation [89]. Recall that for an $m$-qubit linear map $\mathcal{T}$,

$$\mathcal{T}(\rho^A) = \langle\Phi_m|^{BC} \Phi_{\mathcal{T}}^{AB} \otimes \rho^C \, |\Phi_m\rangle^{BC} . \tag{5.1}$$

Here, the Choi state $\Phi_{K_r} = |\phi_r\rangle\langle\phi_r|$ is a pure state, and the map acts on the partition $A$ which comprises the first $m$ qubits, so that partition $D$ contains the last $(m-n)$ qubits. We can then write the state update as

$$|\psi'\rangle = K^A \otimes \mathbb{1}^D \, |\psi\rangle^{AD} = (\mathbb{1}_m^A \otimes \langle\Phi_m|^{BC} \otimes \mathbb{1}_{n-m}^D) \, |\phi_r\rangle^{AB} \otimes |\psi\rangle^{CD} . \tag{5.2}$$

Thus the input state $|\psi\rangle$ is first prepared on the partition $CD$, and the resource state $|\phi_r\rangle$ is prepared on partition $AB$. In our algorithm, this corresponds to preparing the $(n+2m)$-qubit tensor product tableaux $T_r$ and $T_s$ in steps 2 and 6, which takes $\mathcal{O}\big((n+2m)^2\big)$ steps (Section 1.2). We then perform a post-selected Bell measurement on the $2m$-qubit partition $BC$. This corresponds to a projector,

$$|\Phi_m\rangle\langle\Phi_m| = \prod_{j=1}^{m} \frac{(\mathbb{1} + Z_j Z_{m+j})(\mathbb{1} + X_j X_{m+j})}{4} . \tag{5.3}$$

So, evaluating each probability $P_r$ in step 3 and performing the update in step 7 amounts to simulating a sequence of $2m$ weight-2 Pauli measurements using a tableau of size $(n+2m)$, which takes time $\mathcal{O}\big(m(n+2m)^2\big)$, (see Section 1.2). In step 8 we need to discard the unwanted partition $BC$. It follows from the Gottesman-Knill theorem and

Ref. [57] that if we know that an $(n+2m)$-qubit stabiliser state is in product form, we can prepare the tableau for one of the tensor factors in time $\mathcal{O}\big((n+2m)^2\big)$. In all, since we need to evaluate $N_K$ probabilities $P_r$ and then perform the final update, the algorithm divides into $2N_K+3$ computations, each taking time $\mathcal{O}\big(m(n+2m)^2\big)$. Since $m \leq n$, the whole subroutine is completed in time $\mathcal{O}\big(N_K mn^2\big)$, so the runtime is efficient in $n$ provided $N_K \leq \mathrm{poly}(n)$.

We call the second variant of the subroutine the polar scheme (Algorithm 5), as it relies on the fact that the Kraus operator corresponding to a normalised $2m$-qubit pure stabiliser Choi state $|\phi\rangle$ can be written in a canonical polar form $K = 2^{h/2}U\Pi$, where $U$ is a Clifford gate, and $\Pi$ is a stabiliser projector of rank $2^{m-h}$ [63]. This scheme may reduce the runtime in some cases, at the cost of requiring polar decomposition of each Kraus operator. The channel description is supplied as a list of triples $(p_j, U_j, \Pi_j, h_j)$, where $\left|\phi_j\right\rangle = 2^{h_j/2}U_j\Pi_j \otimes \mathbb{1}_m \left|\Phi_m\right\rangle$. For fixed initial state, the probability $2^{h_j} p_j \left\|\Pi_j \left|\psi\right\rangle\right\|^2$ depends only on the projector $\Pi_j$ and coefficient $2^{h_j}p_j$, since $U$ leaves the norm invariant. $K_j$ corresponds to a unitary Clifford operation when $\Pi_j = \mathbb{1}_m$. In this case, $h_j = 0$ and $\left\|K_j \left|\psi\right\rangle\right\|^2 = 1$, so the probability for sampling this Kraus operator is simply $p_j$. One can therefore reduce the overhead by splitting $\mathbb{L}$ into a unitary part $\mathbb{L}_U = \{(p_j, U_j)\}_j$ and non-unitary part $\mathbb{L}_{NU} = \{(p'_j, U'_j, \Pi_j, h_j)\}_j$, with $N_U$ and $N_{NU}$ entries, respectively, so that $N_K = N_U + N_{NU}$. Assuming one can efficiently compute $P_U = \sum_{j=1}^{N_U} p_j$, where the sum is over all coefficients for Clifford operations, one can then flip a biased coin to randomly choose the unitary or non-unitary branch of the channel. If we select the unitary branch, we simply read off the probability for selecting the $j$-th Clifford, and transition probabilities for non-unitary terms need not be computed.

To analyse the runtime, consider that the rank $2^{m-h_r}$ stabiliser projector may be written as the product of $h_r$ Pauli projections, so computing $\|\Pi_r \left|\psi\right\rangle\|^2$ takes time $\mathcal{O}\big(h_r n^2\big)$. Since $h_r \leq m$, the for-loop starting in step 4 is completed in runtime $\mathcal{O}\big(N_{NU}mn^2\big)$. Meanwhile any $m$-qubit Clifford gate can be decomposed as a sequence of $\mathcal{O}\big(m^2/\log m\big)$ elementary Clifford operations from the set $\{H, S, CNOT\}$, each simulable in time $\mathcal{O}\big(n^2\big)$, so step 14 takes time $\mathcal{O}\big(N_{NU}mn^2\big)$. If we suppose that $N_K \leq \mathrm{poly}(n)$, then $N_U, N_{NU} \leq \mathrm{poly}(n)$, and step 1 completes in time $\mathrm{poly}(n)$. The worst-case runtime is then $\mathrm{poly}(n)$ as for the teleportation scheme. However, for certain channels the polar scheme may perform better on average, since computing transition probabilities is only

**Algorithm 5** Stabiliser Kraus update subroutine (Polar scheme)

```
Input: Initial $n$-qubit stabiliser tableau $T$ for state $|\psi\rangle$, description of the channel $\mathcal{E}$ in
    the form of a length-$N_U$ list $\mathbb{L}_U$ list of Clifford operations $(p'_j, U'_j)$, and length-$N_{NU}$
    list of non-unitary operations $(p_j, U_j, \Pi_j, h_j)$.
Output: Updated $n$-qubit tableau for state $|\psi'\rangle$.
Compute $P_U = \sum_j p'_j$.
Set PATH ← “U” with probability $P_U$, else PATH ← “NU”
if PATH = “NU” then
    for $r \leftarrow 1$ to $N_{NU}$ do
        Compute $P_r \leftarrow 2^{h_r} p_r \|\Pi_r |\psi\rangle\|^2 / (1 - P_U)$
    end for
    Sample index $s$ from distribution $\{P_s\}$.
    $U \leftarrow U_s$; $\Pi \leftarrow \Pi_s$.
    Compute tableau $T'$ for $|\psi'\rangle = \Pi |\psi\rangle / \|\Pi |\psi\rangle\|$
else
    Sample index $s$ from distribution $\{p_s/P_U\}_{s=1}^{N_U}$.
    $U \leftarrow U'_s$; $T' \leftarrow T$;
end if
Update tableau $T''$ for state following Clifford, $|\psi''\rangle = U |\psi'\rangle$.
```

necessary if the non-unitary path is sampled.

The restriction to poly$(n)$ Clifford terms can be relaxed if we modify Algorithm 5 to allow that, rather than an explicit list, we have access to a function that efficiently returns the description of a Clifford gate $U_j$ with probability $p_j$. For example, consider a random sequence of $t$ Clifford gates, each chosen from a library of at most $b$ elements. This is a product distribution, so sampling a gate sequence is efficient in $t$ despite the fact that the total number of Clifford terms $N_U$ in the channel decomposition is $\mathcal{O}(b^t)$. This motivates the following definition.

**Definition 5.1** (Simulable channel decompositions)**.** *Suppose an $n$-qubit completely stabiliser-preserving CPTP map $\mathcal{E} \in \mathrm{SP}_{n,n}$ has Kraus decomposition,*

$$\mathcal{E}(\cdot) = \sum_{j=1}^{N_U} p_j U_j (\cdot) U_j^\dagger + \sum_{k=1}^{N_{NU}} q_k K_k (\cdot) K_k^\dagger \tag{5.4}$$

*where $p_j, q_k \geq 0$, $U_j$ are Clifford operators and $K_k$ is a non-unitary completely stabiliser-preserving Kraus operator. We call this a simulable decomposition if:*

1. *The number of non-unitary operators is bounded by $N_{NU} \leq \mathrm{poly}(n)$;*

2. *There exists a procedure that can compute $P_U = \sum_j p_j$, sample from $\{p_j/P_U\}_j$ and*

*compute a description of the unitary $U_j$, all in* poly$(n)$ *time.*

By ensemble ambiguity, this definition of simulability refers to a specific decomposition (e.g. described by a data structure $\mathbb{L}$), rather than being a property of the map $\mathcal{E}$ itself, though there can exist channels that do not admit a simulable decomposition. The class of simulable channel decompositions encompasses a wide range of practically important stabiliser operations, including the random Clifford circuits already mentioned. Simulable decompositions can also be found for any channel of the form $\mathcal{E} = \mathcal{E}' \otimes \mathbb{1}_{n-b}$, where $\mathcal{E}' \in \mathrm{SP}_{b,b}$ and $b$ is a small constant [1]. This follows because by Carathéodory's theorem, the Choi state $\Phi_{\mathcal{E}'}$ can be expressed as a convex combination of at most $4^{2b}$ pure stabiliser states. For circuit families where $b$ has a fixed upper bound, $N_K$ does not grow with $n$. This restriction is not too onerous, since practical quantum algorithms are typically synthesised in terms of one-, two- and three-qubit gates, and noise channels are often assumed to act locally. Moreover, sparse stabiliser decompositions are often known for important channels. $T$-gate injection gadgets and the single-qubit depolarizing channel can be decomposed with only two and four Kraus operators respectively. Analogously, a decomposition of a stabiliser state is called simulable if we can sample a pure state from the ensemble in poly$(n)$ time. Having defined simulable decompositions, we can state the following.

**Theorem 5.2** (Gottesman-Knill for CPTP maps)**.** *Given a sequence of $n$-qubit CPTP maps $\mathcal{E} = \mathcal{E}_T \circ \ldots \circ \mathcal{E}_1$ of length $T$, where each $\mathcal{E}_t \in \mathrm{SP}_{n,n}$ has a known simulable decomposition, and initial stabiliser state with simulable decomposition $\rho = \sum_j p_j |\phi_j\rangle\langle\phi_j|$, then the following procedures are possible.*

1. *Sampling of a pure stabiliser state $|\phi'\rangle$ from an ensemble $\rho' = \sum_j p'_j |\phi'_j\rangle\langle\phi'_j|$, such that $\rho' = \mathcal{E}(\rho)$, in worst-case runtime $T \cdot \mathrm{poly}(n)$.*

2. *Sampling of a bit-string $\mathbf{x}$ of length $w$ from a distribution $\Pr(\mathbf{x}) = \mathrm{Tr}[\Pi_{\mathbf{x}}\mathcal{E}(\rho)]$, where $\Pi_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w}$ in worst-case runtime $(T+w) \cdot \mathrm{poly}(n)$.*

3. *Estimation of the expected value of a stabiliser observable $O$ (meaning a generalised Pauli operator or the Born rule probability for a stabiliser measurement), such that the estimate $\mu$ satisfies $\|\mu - \mathrm{Tr}[O\mathcal{E}(\rho)]\| \leq \varepsilon$ with probability at least $(1-p)$. The estimate is computed in worst-case runtime $T \cdot \mathrm{poly}(n) \cdot \frac{2}{\varepsilon^2} \log\left(\frac{2}{p}\right)$.*

*Proof.* The validity of this theorem follows by application of the stabiliser-Kraus update subroutine. Procedures 2 and 3 both take procedure 1 as a starting point. Let $\mathcal{E}_t = \sum_j q_j^{(t)} \mathcal{K}_j^{(t)}$ be the decomposition of the $t$-th circuit element into maps corresponding to conjugation with a single Kraus operator, $\mathcal{K}_j^{(t)} = K_j^{(t)}(\cdot)K_j^{(t)\dagger}$. Then the final state $\rho'$ can be decomposed as an ensemble:

$$\rho = \sum_{\mathbf{j}} p_{j_0} q_{\mathbf{j}} K_{\mathbf{j}} \left|\phi_{j_0}\right\rangle\!\left\langle\phi_{j_0}\right| K_{\mathbf{j}}^{\dagger} \tag{5.5}$$

where $\mathbf{j}$ is a vector of length $T+1$ representing a trajectory through the circuit and choice of initial state, so that $q_{\mathbf{j}} = p_{j_0} \prod_{t=1}^{T} q_{j_t}^{(T)}$ and $K_{\mathbf{j}} = \prod_{t=1}^{t} K_{j_t}^{(t)}$. It is possible that $K_{\mathbf{j}} \left|\phi_{j_0}\right\rangle = 0$ for some $\mathbf{j}$. Let $\mathbb{P}$ denote the set of trajectories such that $K_{\mathbf{j}} \left|\phi_{j_0}\right\rangle \neq 0$. Then let $|\phi_{\mathbf{j},j_0}\rangle$ be the normalised final state for a trajectory $\mathbf{j} \in \mathbb{P}$,

$$|\phi_{\mathbf{j}}\rangle = \frac{\sqrt{q_{\mathbf{j}}} K_{\mathbf{j}} \left|\phi_{j_0}\right\rangle}{\sqrt{P_{\mathbf{j}}}}, \quad \text{where} \quad P_{\mathbf{j}} = q_{\mathbf{j}} \left\| K_{\mathbf{j}} \left|\phi_{j_0}\right\rangle \right\|^2. \tag{5.6}$$

Then the final state $\rho'$ can be written as an ensemble over pure stabiliser states,

$$\rho' = \mathcal{E}(\rho) = \sum_{\mathbf{j} \in \mathbb{P}} P_{\mathbf{j}} \left|\phi_{\mathbf{j}}\right\rangle\!\left\langle\phi_{\mathbf{j}}\right|. \tag{5.7}$$

Let $\mathbf{j}_t$ denote the trajectory up to the $t$-th circuit element. Let $|\phi_{\mathbf{j}_{t-1}}\rangle$ be the normalised pure state at step $t-1$. Then let $|\phi_{\mathbf{j}_t}\rangle$ be the pure state renormalised after applying the $t$-th Kraus operator,

$$|\phi_{\mathbf{j}_t}\rangle = \frac{K_{j_t}^{(t)} |\phi_{\mathbf{j}_{t-1}}\rangle}{\|K_{j_t}^{(t)} |\phi_{\mathbf{j}_{t-1}}\rangle\|}. \tag{5.8}$$

But we can also write $|\phi_{\mathbf{j}_t}\rangle = K_{\mathbf{j}_t} |\phi_{j_0}\rangle / \|K_{\mathbf{j}_t} |\phi_{j_0}\rangle\|$, and $K_{\mathbf{j}_t} = K_{j_t}^{(t)} K_{\mathbf{j}_{t-1}}$. Combining this with equation (5.8), we obtain the recurrence relation,

$$\|K_{\mathbf{j}_t} |\phi_{j_0}\rangle\| = \|K_{j_t}^{(t)} |\phi_{\mathbf{j}_{t-1}}\rangle\| \cdot \|K_{\mathbf{j}_{t-1}} |\phi_{j_0}\rangle\|. \tag{5.9}$$

Then by induction, using equation (5.6) we can factorise the probabilities $P_{\mathbf{j}}$ as,

$$P_{\mathbf{j}} = p_{j_0} \prod_{t=1}^{T} q_{j_t}^{(t)} \|K_{j_t}^{(t)} |\phi_{\mathbf{j}_{t-1}}\rangle\|^2. \tag{5.10}$$

But $q^{(t)}_{j_t} \| K^{(t)}_{j_t} \left|\phi_{\mathbf{j}_{t-1}}\right\rangle \|^2$ is precisely the probability of sampling the pure state $\left|\phi_{\mathbf{j}_t}\right\rangle$ using the stabiliser-Kraus subroutine described above. So by sampling an initial state $\left|\phi_{j_0}\right\rangle$ with probability $p_{j_0}$, and sampling a new pure state update for each circuit element $\mathcal{E}_t$ using the stabiliser-Kraus subroutine, we reproduce the statistics for the ensemble in equation (5.7). Each of the T subroutine calls takes $\text{poly}(n)$ time, so the total runtime for procedure 1 is $T \cdot \text{poly}(n)$.

The statements on the other two simulation procedures follow almost immediately. To perform procedure 2, having sampled a tableau describing a pure state $|\phi'\rangle$ from the ensemble $\rho' = \mathcal{E}(\rho)$ using procedure 1, we can sample a bit string of length $w$ by simulating a $Z$-basis measurement on each of the first $w$ qubits in turn [60], each taking time $\text{poly}(n)$. There are $w$ of these measurements so the total runtime is $(T+w)\cdot\text{poly}(n)$. Then the probability of sampling the string $\mathbf{x}$ given state $|\phi'\rangle$ is $\Pr(\mathbf{x}|\phi') = \text{Tr}[\Pi_{\mathbf{x}} |\phi'\rangle\langle\phi'|]$. The probability obtaining state $|\phi_{\mathbf{j}}\rangle$ from procedure 1 is $\Pr(\phi_{\mathbf{j}}) = P_{\mathbf{j}}$, so the marginal probability $\sum_{\mathbf{j}} \Pr(\mathbf{x}|\phi_{\mathbf{j}}) \Pr(\phi_{\mathbf{j}})$ for sampling $\mathbf{x}$ is given by $\text{Tr}\left[\Pi_{\mathbf{x}} \sum_{\mathbf{j}} P_{\mathbf{j}} \left|\phi_{\mathbf{j}}\right\rangle\!\left\langle\phi_{\mathbf{j}}\right|\right] = \text{Tr}[\Pi_{\mathbf{x}}\mathcal{E}(\rho)]$, as required.

Finally to perform procedure 3, assume for simplicity that we want to estimate some Pauli observable $O$; a near-identical argument holds for Born rule probabilities. First we sample a pure stabiliser state $|\phi_{\mathbf{j}}\rangle$ using procedure 1, then evaluate $\mu = \left\langle\phi_{\mathbf{j}}\middle|O\middle|\phi_{\mathbf{j}}\right\rangle$ using the standard tableau method. Repeating this sampling procedure $M$ times and taking the mean yields an unbiased estimator for $\langle O\rangle$, as $\mathbb{E}(\mu) = \sum_{\mathbf{j}} P_{\mathbf{j}} \text{Tr}\left[O\left|\phi_{\mathbf{j}}\right\rangle\!\left\langle\phi_{\mathbf{j}}\right|\right] = \text{Tr}[O\mathcal{E}(\rho)]$. Then using Hoeffding's inequalities, to obtain additive precision $\varepsilon$ with probability at least $(1-p)$ we require $M = \left\lceil \frac{2}{\varepsilon^2}\log\left(\frac{2}{p}\right)\right\rceil$ samples. Sampling each pure state from the channel and evaluating $\mu$ takes time $T\cdot\text{poly}(n)$, so this gives us the stated runtime. □

We end this section by defining the notion of a simulable circuit decomposition.

**Definition 5.3** (Simulable circuit decomposition)**.** *Let $\mathcal{E} \in \text{SP}_{n,n}$ be a CPTP, completely stabiliser-preserving map, with a decomposition into a sequence of T circuit elements $\mathcal{E} = \mathcal{E}_T \circ \ldots \mathcal{E}_1$. Let $\mathbb{D}$ be a data structure describing $\mathcal{E}$, constructed as follows. Let $\mathbb{D} = \{\mathbb{L}_1, \ldots, \mathbb{L}_T\}$, such that for each j, $\mathbb{L}_j$ is a list describing a circuit element $\mathcal{E}_j \in \text{SP}_{n,n}$, in the format described earlier in this section. Since $\mathbb{D}$ describes a sequence of T circuit elements, we say that it is a decomposition of length T. We say that $\mathbb{D}$ is a* ***simulable circuit decomposition*** *if $T \leq \text{poly}(n)$ and each $\mathbb{L}_j$ describes a simulable decomposition* *(as per Definition 5.1) of the channel $\mathcal{E}_j$.*

The efficient simulability of such decompositions is the starting point for the main results of this chapter, which deal with the classical overhead involved in simulating resourceful states and operations. Before considering general magic-generating channels, we first consider the case where all operations are stabiliser-preserving, but the circuit has mixed magic state inputs.

## 5.2 Algorithms for stabiliser circuits with magic state inputs

We now present three classical simulation algorithms associated with the new magic monotones for states introduced in Section 4.1, which first appeared in our article Ref. [2]. Recall that the generalised robustness (gRoM) $\Lambda^+$ and dyadic negativity $\Lambda$ can be considered relaxations of robustness of magic (RoM) $\mathcal{R}$ [103], while the density-operator stabiliser extent $\Xi$ is closely related to pure-state extent [3] and stabiliser rank [59, 60]. All three monotones coincide for pure states, and for $n$-fold tensor products of $n$-qubit mixed states. We will shortly introduce three very different simulation algorithms, whose classical simulation overhead is respectively quantified by one of the three monotones. The *dyadic frame* and *constrained path* simulators are descendants of quasiprobability methods [103, 124], whereas the *density operator stabiliser rank* simulator generalises stabiliser rank methods [3, 59, 60] which previously only admitted pure state evolution. In this way the magic monotone framework presented in Chapter 4 provides the formal link between previously unconnected simulation methods.

We first introduce two quasiprobability-like simulators. The first is the *constrained path* simulator, associated with gRoM. In standard quasiprobability methods, the runtime needed to achieve any fixed precision with high probability is a function of the $\ell_1$-norm of the distribution. The constrained path technique trades the ability to estimate to arbitrary precision for efficient runtime; the maximum precision is a function of the monotone $\Lambda^+$, while the runtime is not. The second quasiprobability simulator we call the *dyadic frame* simulator. This improves on previous quasiprobability simulators by using a novel choice of frame, namely the set of stabiliser dyads, so that “free” operators are those in the convex hull of $e^{i\phi}\left|L\right\rangle\!\left\langle R\right|$, where $\left|L\right\rangle, \left|R\right\rangle \in \mathrm{STAB}_n$. In contrast with a conventional choice of frame elements, dyads are not guaranteed to be Hermitian or positive semidefinite, and can be traceless. We show that these technical issues can be circumvented, and in doing so we can reduce the sampling overhead.

The third simulator will be the *density operator stabiliser rank* simulator. We extend the method of Ref. [3] to deal with general density operators (and therefore mixed states). This enables powerful stabiliser rank methods to be applied to the simulation of noisy circuits with imperfect magic state inputs. In Ref. [3], it was shown that using sparsification, the runtime for their simulator can connected to the pure-state extent. We present a refinement to the sparsification procedure that allows us to sidestep some technical issues with the results of Ref. [3] (see Section 3.4), and obtain a significant improvement in performance.

### 5.2.1 Constrained path simulator

Recall that in standard quasiprobability method, a target state $\rho$ is expressed as an affine combination of elements of some frame $\mathcal{F}$. The frame is chosen so that free operations can be efficiently simulated when applied to elements in the frame. In Ref. [103] Howard and Campbell took the frame to be the set of pure stabiliser state projectors, so that the target state is expressed $\rho = \sum_j q_j \left|\phi_j\right\rangle\!\left\langle\phi_j\right|$, where $\left|\phi_j\right\rangle \in \mathrm{STAB}_n$. We can always combine all the positive and negative terms of the decomposition into convex combinations $\sigma_+$ and $\sigma_-$ respectively, to write $\rho = \lambda\sigma_+ - (\lambda - 1)\sigma_-$ , where $\lambda = \sum_{q_j \geq 0} q_j \geq 1$, so that $\|q\|_1 = 2\lambda - 1$. Recall from Chapter 2 that $\mathrm{Tr}[E\mathcal{E}(\rho)]$ for simulable $\mathcal{E} \in \mathrm{SP}_{n,n}$ and stabiliser observable $E$ can be estimated to arbitrary additive precision $\varepsilon > 0$ in time $\mathcal{O}\big(\|q\|_1^2\varepsilon^{-2}\big)$ using the standard quasiprobability sampling procedure. The procedure for each sample can be split into two steps: (i) randomly sample the positive or negative path with probability $\lambda/\|q\|_1$ or $(\lambda - 1)/\|q\|_1$ ; (ii) sample an individual frame element $|\phi\rangle\langle\phi| \in \mathcal{F}$ from the selected convex combination $\sigma_\pm$, and then compute $\mathrm{Tr}[E\mathcal{E}(|\phi\rangle\langle\phi|)]$. Since step (ii) is efficient, any increased runtime for simulating magic states arises in step (i) rather than step (ii). In other words, sampling a frame element from the convex combination $\sigma_\pm$ does not incur additional overhead, once we have taken into account the renormalisation of the distribution in step (i).

The alternative strategy we present here is to constrain sampling to the positive path so that step (i) is avoided. This is equivalent to making the approximation $\rho \approx \lambda\sigma_+$, and comes at the cost of an unavoidable systematic error of size $|(\lambda - 1)\,\mathrm{Tr}[E\mathcal{E}(\sigma_-)]|$. The

| | $\lambda < 2/(1+c)$ | $\lambda \geq 2/(1+c)$ |
|---|---|---|
| $\lvert P_\sigma \rvert \leq \lvert \lambda(1+c) - 2 \rvert$ | $\Delta = \lambda(1+c) - 1$ | “FAIL” |
| $\lvert P_\sigma \rvert > \lvert \lambda(1+c) - 2 \rvert$ | $\Delta = \frac{\lambda(1+c) - \lvert P_\sigma \rvert}{2}$ | |

**Table 5.1:** Error bounds for constrained path simulator (Algorithm 6) in different parameter regimes, for input state such that $\rho \leq \lambda\sigma$, where $\sigma \in \overline{\mathrm{STAB}}_n$. The output from the algorithm depends on the parameter $\lambda$, and a variable $P_\sigma \in [-\lambda, \lambda]$ which is computed in the course of the algorithm, and satisfies $|P_\sigma - \lambda \operatorname{Tr}[P\sigma]| \leq \delta$ with high probability. When $|P_\sigma|$ is large (bottom row), the algorithm always outputs non-trivial bounds. If $|P_\sigma|$ is small (top row), the algorithm returns a meaningful estimate if $\lambda < 2/(1+c)$, but returns the “FAIL” flag otherwise.

advantage to this approach is that since $\operatorname{Tr}[E\mathcal{E}(\sigma_-)]$ is no longer evaluated explicitly, $\sigma_-$ need not be an efficiently simulable state. It is natural to connect this strategy with primal solutions for the gRoM problem,

$$\rho = \Lambda^+(\rho)\sigma - (\Lambda^+(\rho) - 1)\rho_-, \tag{5.11}$$

where $\sigma$ is a mixed stabiliser state, but $\rho_-$ can be any density operator. Since systematic error is unavoidable, the first term need not be estimated to high precision. We will prove the following theorem.

**Theorem 5.4** (Constrained path simulator)**.** *Let $\rho$ be an n-qubit density matrix s.t. $\rho \leq \lambda\sigma$, where $\sigma \in \mathrm{STAB}_n$, let $\mathcal{E} \in \mathrm{SP}_{n,n}$ be a completely stabiliser-preserving CPTP map with a simulable circuit decomposition $\mathbb{D}$, and let $P$ be a Pauli observable. Then for any $c, p_{\text{fail}}$ there exists an efficient classical algorithm, terminating in runtime $\tau$, that either outputs an estimate $\widehat{P}$ and error bound $\Delta$, such that $\left|\widehat{P} - \langle P \rangle\right| \leq \Delta$ with probability at least $1 - p_{\text{fail}}$, or returns a flag* “FAIL” *indicating that with high probability it is not able to do so given the provided inputs. The runtime $\tau$ for the algorithm is bounded by*

$$\tau \leq 2\frac{1}{c^2}\log\left(\frac{2}{p_{\text{fail}}}\right) \cdot \mathrm{poly}(n). \tag{5.12}$$

*The size of $\Delta$, and whether the algorithm returns* “FAIL” *depends on $\lambda$ and a variable $P_\sigma \in [-\lambda, \lambda]$ computed by the algorithm (see Table 5.1). When $\lambda \leq 2/(1+c)$ the algorithm outputs a non-trivial estimate with probability at least $1 - p_{\text{fail}}$.*

*Proof.* To prove the theorem we give the algorithm, show that the error bounds are satisfied for each case, and then analyse the runtime. We rely on the stabiliser-Kraus subroutine (see Section 5.1). Let $\textsc{stabiliserCircuit}(A, \mathbb{D}, P, \delta, p_{\text{fail}})$ be a function that takes as input a quasiprobability decomposition of operator $A$, a simulable circuit decomposition $\mathbb{D}$ of stabiliser channel $\mathcal{E}$, Pauli observable $P$, and parameters $\delta, p_{\text{fail}}$, and, using the procedure described in the proof of Theorem 5.2, outputs an estimate $P_\sigma$ such that $|P_\sigma - \text{Tr}[P\mathcal{E}(A)]| \leq \delta$ with probability at least $1 - p_{\text{fail}}$. Pseudocode for our simulator is given in Algorithm 6.

**Algorithm 6** Constrained Path Simulator

```
Input: Target state $\rho$; real numbers $\lambda, c, p_{\text{fail}} > 0$, and stabiliser state $\sigma \in \overline{\text{STAB}}_n$ s.t.
    $\rho \leq \lambda\sigma$ and $c, p_{\text{fail}} \ll 1$; Pauli observable $P$ and channel $\mathcal{E} \in \text{SP}_{n,n}$ with known
    simulable decomposition $\mathbb{D}$.
Output: Estimate $\widehat{P}$ and error bound $\Delta$, s.t. $|\widehat{P} - \text{Tr}(P\mathcal{E}[\rho])| \leq \Delta$ with probability $1 -$
    $p_{\text{fail}}$.
 1: $\delta \leftarrow c\lambda$
 2: $P_\sigma \leftarrow \textsc{stabiliserCircuit}(\lambda\sigma, \mathbb{D}, P, \delta, p_{\text{fail}})$.
 3: $P_{\max} \leftarrow \min\{1, P_\sigma + \delta + \lambda - 1\}$
 4: $P_{\min} \leftarrow \max\{-1, P_\sigma - \delta - \lambda + 1\}$
 5: if $P_{\max} - P_{\min} < 2$ then
 6:     $\widehat{P} \leftarrow (P_{\max} + P_{\min})/2$
 7:     $\Delta \leftarrow (P_{\max} - P_{\min})/2$
 8:     return $\widehat{P}, \Delta$
 9: else
10:     return "FAIL"
11: end if
```

Choosing $P_{\max}$ and $P_{\min}$ to be given by the expressions in steps 3 and 4 ensures that for all $\lambda$ and $P_\sigma$, $|\widehat{P} - \text{Tr}(P\mathcal{E}[\rho])| \leq \Delta$ holds with probability $1 - p_{\text{fail}}$. The major caveat is that there are certain regimes (for large $\lambda$ and small $P_\sigma$, as shown in Table 5.1), where the algorithm fails by determining bounds that are trivially true, $-1 \leq \text{Tr}[P\mathcal{E}(\rho)] \leq 1$. Nevertheless, in some regimes we efficiently obtain a biased but non-trivial estimate. We first briefly explain the rationale for the min and max expressions given in steps 3 and 4, before analysing the error bound and runtime.

For target state $\rho$, when there exists $\lambda > 0$ and $\sigma \in \overline{\text{STAB}}_n$ such that $\rho \leq \lambda\sigma$, then there exists some density matrix $\rho_-$ such that $\rho = \lambda\sigma - (\lambda - 1)\rho_-$. Step 2 estimates $P_\sigma$ such that $|P_\sigma - \lambda\,\text{Tr}(P\mathcal{E}[\sigma])| \leq \delta$ with probability $1 - p_{\text{fail}}$. We use this to place bounds on $\mathrm{Tr}(P\mathcal{E}[\rho])$,

$$\begin{aligned}\mathrm{Tr}(P\mathcal{E}[\rho]) &= \lambda\,\mathrm{Tr}(P\mathcal{E}[\sigma]) - (\lambda-1)\,\mathrm{Tr}(P\mathcal{E}[\rho_-]) && (5.13)\\ &\leq P_\sigma + \delta + (\lambda-1) && (5.14)\end{aligned}$$

Here we have used the fact that $\rho_-$ is a density operator, $\mathcal{E}$ is a CPTP map, and $P$ has eigenvalues ±1, so $|\mathrm{Tr}(P\mathcal{E}[\rho_-])| \leq 1$. Similarly one obtains $\mathrm{Tr}(P\mathcal{E}[\rho]) \geq P_\sigma - \delta - (\lambda - 1)$. Trivially, $|\mathrm{Tr}(P\mathcal{E}[\rho])| \leq 1$, so in case either expression exceeds this (for example if $P_\sigma$ is close to $\pm 1$) we simply take either $P_{\max} = 1$ or $P_{\min} = -1$ as necessary. If *both* $P_{\max} = 1$ or $P_{\min} = -1$, then the procedure gives the trivial result that $-1 \leq \mathrm{Tr}(P\mathcal{E}[\rho]) \leq 1$, so in this case we have the algorithm return "FAIL". We now show how to obtain the non-trivial bounds in Table 5.1.

**Case 1:** $|P_\sigma| > |\lambda(1+c)-2|$.

If $|P_\sigma| > |\lambda(1+c)-2|$ then either $P_\sigma > 0$ or $P_\sigma < 0$. Assume first that $P_\sigma > 0$. Then $P_\sigma > |\lambda(1+c)-2| > 2-\lambda(1+c)$ and $P_\sigma > \lambda(1+c)-2$. Recalling that $\delta = c\lambda$, we have $P_\sigma + \delta + \lambda - 1 > 1$ and $P_\sigma - \delta - \lambda + 1 > -1$. Therefore step 3 sets $P_{\max} = 1$ and step 4 sets $P_{\min} = P_\sigma - \delta - \lambda + 1$. Then with probability at least $1 - p_{\text{fail}}$ we know that,

$$-1 < P_\sigma - \delta - \lambda + 1 \leq \mathrm{Tr}[P\mathcal{E}(\rho)] \leq 1. \tag{5.15}$$

So in steps 6 and 7, we obtain

$$\widehat{P} = \frac{P_\sigma - \lambda(1+c)+2}{2}, \quad \Delta = \frac{\lambda(1+c) - |P_\sigma|}{2}. \tag{5.16}$$

In the case when $|P_\sigma| > |\lambda(1+c)-2|$ and $P_\sigma < 0$, by the same argument,

$$-1 \leq \mathrm{Tr}[P\mathcal{E}(\rho)] \leq P_\sigma + \delta + \lambda - 1 < 1. \tag{5.17}$$

It follows that $\widehat{P} = (P_\sigma + \delta + \lambda - 2)/2$ and $\Delta$ has the same value as in equation (5.16). So in Case 1 the estimation error is at most $\lambda(1+c)/2$, and tends to $\delta/2 = c\lambda/2$ as $|P_\sigma| \to \lambda$.

**Case 2:** $|P_\sigma| \leq |\lambda(1+c)-2|$ **and** $\lambda < 2/(1+c)$.

Combining these conditions we have $\pm P_\sigma \leq 2-\lambda-\delta$, so that

$$P_\sigma + \delta + \lambda - 1 \leq 1, \quad \text{and} \quad P_\sigma - \delta - \lambda + 1 \geq -1. \tag{5.18}$$

Then the algorithm sets $P_{\max} = P_\sigma + \delta + \lambda - 1$ and $P_{\min} = P_\sigma - \delta - \lambda + 1$. We have $P_{\min} \leq \mathrm{Tr}[P\mathcal{E}(\rho)] \leq P_{\max}$ with probability at least $1 - p_{\mathrm{fail}}$, and we obtain:

$$\widehat{P} = P_\sigma, \quad \Delta = \lambda(1+c) - 1. \tag{5.19}$$

So if $|P_\sigma|$ is small, we obtain less information about $\mathrm{Tr}[P\mathcal{E}(\rho)]$, but we still have a non-trivial estimate as long as $\lambda < 2/(1+c)$.

**Case 3:** $|P_\sigma| \leq |\lambda(1+c)-2|$ **and** $\lambda \geq 2/(1+c)$.

Here we have that $|\lambda(1+c)-2| = \lambda(1+c)-2$, so $|P_\sigma| \leq |\lambda(1+c)-2|$. Following the same argument as for Case 2 but with inequalities reversed, we find:

$$P_\sigma + \delta + (\lambda - 1) \geq 1, \tag{5.20}$$

$$P_\sigma - \delta - (\lambda - 1) \leq -1. \tag{5.21}$$

But this means that the algorithm must set $P_{\max} = 1$ and $P_{\max} = -1$. In this case estimating $\lambda\,\mathrm{Tr}[P\mathcal{E}(\sigma)]$ yields no useful information, so we return the "FAIL" flag.

**Runtime analysis:** The runtime is dominated by the estimation of $\lambda\mathrm{Tr}(P\mathcal{E}[\sigma])$ using the stabiliser-Kraus subroutine, as the other steps are trivial to evaluate, and can be computed in constant time. Since $\sigma \in \overline{\mathrm{STAB}}_n$, and by assumption $\mathbb{D}$ provides a simulable circuit decomposition for $\mathcal{E} \in \mathrm{SP}_{n,n}$, there is no additional sampling overhead due to negativity. The prefactor $\lambda$ increases the variance of the estimator, but we compensate by setting the precision to $\delta = c\lambda$, where $0 < c \ll 1$, so that $\delta \ll \lambda$. The rationale for this is that the systematic error due to our ignorance of $\rho_-$ is unavoidable, and this error is of size $\lambda - 1$. Therefore there is a limit to the precision we can achieve by increasing the runtime of the sampling step, and we should set the precision commensurate with the size of $\lambda$. By the Hoeffding inequality the smallest number of samples $M$ sufficient to achieve this precision is

$$M = \lceil 2\lambda^2\delta^{-2}\log(2p_{\mathrm{fail}}^{-1})\rceil = \lceil 2c^{-2}\log(2p_{\mathrm{fail}}^{-1})\rceil. \tag{5.22}$$

Each sample is obtained in poly$(n)$ time (Theorem 5.2), yielding (5.12). □

Our explicit algorithm for estimating Pauli expectation values easily adapts to estimate Born rule probabilities for stabiliser projectors $\Pi$ by replacing the assumption $|\mathrm{Tr}(P\mathcal{E}(\rho))| \leq 1$ for any $\rho$ with $0 \leq \mathrm{Tr}(\Pi\mathcal{E}(\rho)) \leq 1$. We emphasise that the runtime for the constrained path simulator is constant with respect to $\lambda$ (i.e. the generalised robustness $\Lambda^{+}(\rho)$ when the decomposition is optimal), depending only on the parameters $c$ and $p_{\mathrm{fail}}$. In this sense, we achieve efficient runtime by trading off against precision in the estimate; it is the error $\Delta$ which scales with the magic monotone rather than the runtime. We envisage the constrained path algorithm having application as a fast observable estimation method for circuits with a relatively small amount of magic, or where high precision is not required. For example, it may be of use for the simulation of noisy near-term devices. The fact that the simulator can often return a "FAIL" outcome when $\lambda > 2/(1+c)$ limits its application to large-scale circuits with significant amounts of magic. However, a surprising feature is that even when $\lambda$ is large, the simulator can still efficiently return non-trivial estimates provided $\mathrm{Tr}[P\mathcal{E}(\rho)]$ is sufficiently close to either $+1$ or $-1$. This suggests it may be useful for the study of highly structured circuits where a final measurement is expected to take the value either $+1$ or $-1$ with near certainty.

### 5.2.2 Dyadic frame simulator

Next we introduce a new simulator for estimation of stabiliser observables up to arbitrary additive precision, with runtime that scales with a generalised notion of negativity, namely the dyadic negativity defined in Chapter 4. By generalising the notion of frame as defined by Pashayan et al. [124], the dyadic frame simulator can outperform all previously known $n$-qubit quasiprobability simulators.

Recall that the RoM simulator was based on decompositions $\rho = \sum_j q_j |\phi_j\rangle\langle\phi_j|$ where $q_j \in \mathbb{R}$ and $\sum_j q_j = 1$. When $\rho$ is a magic state, some of the $q_j$ must be negative, so that $||q||_1 > 1$. The runtime of the associated Monte Carlo simulator scales with $||q||_1^2$, so that the optimal runtime for the simulator is quantified by RoM, $\mathcal{R}(\rho)^2 = (\min ||q||_1^2)$. The simulator can be related to the formalism of Pashayan et al. [124] by allowing the specification of frame and dual frame to be state-dependent, so that making the optimal choice recovers the robustness of magic (See Appendix C.2). In Ref. [124], it is implicitly assumed that frame elements are Hermitian. We show that this requirement can be relaxed, so that we can supplement stabiliser frames with non-Hermitian stabiliser dyads. The runtime of the resultant simulator scales with the square of the dyadic negativity $\Lambda(\rho)^2$. Dyadic negativity is often significantly smaller than RoM, dramatically improving the performance in some cases. Operators are now considered free if they are in the convex hull of the dyads $e^{i\phi}\left|L\right\rangle\!\left\langle R\right|$, where $\left|L\right\rangle, \left|R\right\rangle \in \mathrm{STAB}_n$ and $\phi \in \mathbb{R}$. In particular, because dyadic negativity is faithful, a density operator $\rho$ is free if and only if $\rho \in \overline{\mathrm{STAB}}_n$.

There are two technical issues to be overcome in designing our dyadic frame simulator. The first relates to the fact when a stabiliser projector is updated by conjugation with a stabiliser Kraus operator, any global phase is cancelled, so the sign of terms in the quasiprobability distribution remains invariant under these updates. This is not the case for dyads since, for example, the same Clifford operator can result in a different phase depending on the input state. The second difficulty relates to the computation of intermediate transition probabilities. For the stabiliser-Kraus updates described in Section 5.1, the transition probability for a Kraus operator $K$ was computed as $\Pr(K|\phi) = \mathrm{Tr}\left[K\left|\phi\right\rangle\!\left\langle\phi\right|K^\dagger\right]$. This strategy fails for non-Hermitian dyads, since they are not normalised by trace, but we will see that the subroutine can be amended to recover an unbiased estimator. We first illustrate the basic algorithm with a simplified version, where the stabiliser circuit elements are restricted to be convex mixtures of Clifford gates. We subsequently generalise to cover all completely stabiliser-preserving circuits with magic state inputs, in particular showing how the stabiliser-Kraus subroutine can be updated to appropriately compute transition probabilities for Kraus operators applied to general stabiliser dyads.

For the simplified algorithm we assume the following restricted simulation setting. The input to the algorithm will consist of (i) a known dyadic decomposition of a mixed magic state $\rho = \sum_j \alpha_j \left|L_j\right\rangle\!\left\langle R_j\right|$; (ii) a circuit description comprising a list of $T$ quantum operations $\{O^{(1)}, \ldots, O^{(T)}\}$; and (iii) a stabiliser observable $P$. We stipulate that each $O^{(t)}$ must be a convex mixture of Clifford channels, $O^{(t)} = \sum_k p_k^{(t)} U_k(\cdot)U_k^\dagger$ , and we assume this decomposition is known and simulable as per Definition 5.1. We use the term *mixed Clifford circuit* to describe the class of circuits satisfying these conditions. The output of the algorithm is again an estimate for a stabiliser observable $\langle P\rangle = \mathrm{Tr}[P\mathcal{E}(\rho)]$. For simplicity we assume $P$ is a Pauli operator, but the procedure can be easily adapted for stabiliser projectors. Here $\mathcal{E} = O^{(T)} \circ \ldots \circ O^{(1)}$. The restriction on $O^{(t)}$ means that the whole circuit can be expressed as an ensemble over unitary Clifford gates $\mathcal{E}(\cdot) = \sum_{\mathbf{k}} p_{\mathbf{k}} U_{\mathbf{k}}(\cdot) U_{\mathbf{k}}^\dagger$, where $\mathbf{k} = (k_1, k_2, \ldots, k_T)$ is a vector that represents a Clifford trajectory through the circuit $U_{\mathbf{k}} = U_{k_T} \ldots U_{k_2} U_{k_1}$, and $p_{\mathbf{k}}$ is a product distribution and so can be efficiently sampled from. The simulator (Algorithm 7) proceeds by sampling elements from the initial distribution, computing an estimate, and averaging over many samples.

**Algorithm 7** Dyadic Frame Simulator - Mixed Clifford circuits

```
Input: Initial state with known decomposition $\rho = \sum_j \alpha_j |L_j\rangle\langle R_j|$, where $|L_j\rangle, |R_j\rangle \in$
    STAB$_n$; simulable decomposition $\mathbb{D}$ for mixed Clifford circuit $\mathcal{E}(\cdot) = \sum_{\mathbf{k}} p_{\mathbf{k}} U_{\mathbf{k}}(\cdot) U_{\mathbf{k}}^\dagger$
    of length $T$; target Pauli $P$; number of samples $M$.
Output: Estimate $\widehat{P}$ for Tr$[P\mathcal{E}(\rho)]$.
1: for $m \leftarrow 1$ to $M$ do
2:     Randomly sample index $j$ with probability $|\alpha_j|/\|\alpha\|_1$.
3:     Randomly sample $\mathbf{k}$ with probability $p_{\mathbf{k}}$.
4:     $e^{i\theta'_{j,\mathbf{k}}} |L'_{j,\mathbf{k}}\rangle\langle R'_{j,\mathbf{k}}| \leftarrow e^{i\theta_j} U_{\mathbf{k}} |L_j\rangle \langle R_j| U_{\mathbf{k}}^\dagger$
5:     $\widehat{P}_m \leftarrow \|\alpha\|_1 \mathrm{Re}\{e^{i\theta'_{j,\mathbf{k}}} \langle R'_{j,\mathbf{k}}|P|L'_{j,\mathbf{k}}\rangle\}$
6: end for
7: $\widehat{P} \leftarrow \sum_m \widehat{P}_m / M$
8: return $\widehat{P}$
```

In step 4, $e^{i\theta'}$ is a final global phase taking into account the initial phase $e^{i\theta_j} = \alpha_j/|\alpha_j|$ and the action of the sampled unitary circuit on $|L_j\rangle$ and $|R_j\rangle$ respectively. Whereas the RoM simulator dealt with projectors $|\phi\rangle\langle\phi|$, so that any global phase on $|\phi\rangle$ is unimportant, here $|L_j\rangle$ and $|R_j\rangle$ can represent different stabiliser states and the combined phase can affect both the magnitude and sign of the real-valued sample $\widehat{P}$. The standard stabiliser tableau method [30] (Section 1.2) cannot track this global phase, but subsequent extensions can [3, 59, 60]. For concreteness, we assume that we use the CH-simulator of Ref. [3] (Section 1.2.1). For step 5, recall that any Pauli operator is also a Clifford gate, so to compute any $\langle R'|P|L'\rangle$ for $|L'\rangle, |R'\rangle \in$ STAB$_n$ we can perform the update $|L''\rangle = P|L'\rangle$, then compute the the complex inner product $\langle L'|R''\rangle$. Both these steps are efficient using the CH-simulator [3, 59, 60]. Thus steps 3 and 4 are efficient. Note that the two parts of the dyad $U_{\mathbf{k}} |L_j\rangle$ and $\langle R_j| U_{\mathbf{k}}^\dagger = (U_{\mathbf{k}} |R_j\rangle)^\dagger$ are updated independently.

We can check that the method gives an unbiased estimator,

$$\mathbb{E}(\widehat{P}) = \sum_{j,\mathbf{k}} \frac{|\alpha_j|}{\|\alpha\|_1} p_{\mathbf{k}} \left( \|\alpha\|_1 \mathrm{Re}\{e^{i\theta'_{j,\mathbf{k}}} \langle R'_{j,\mathbf{k}}|P|L'_{j,\mathbf{k}}\rangle\} \right) \tag{5.23}$$

$$= \mathrm{Re}\{\sum_{j,\mathbf{k}} e^{i\theta_j} |\alpha_j| p_{\mathbf{k}} \mathrm{Tr}\left[PU_{\mathbf{k}} \left|L_j\right\rangle \left\langle R_j\right| U_{\mathbf{k}}^{\dagger}\right]\} \tag{5.24}$$

$$= \mathrm{Re}\{\mathrm{Tr}\left[P \sum_{\mathbf{k}} p_{\mathbf{k}} U_{\mathbf{k}} (\sum_j \alpha_j \left|L_j\right\rangle \left\langle R_j\right|) U_{\mathbf{k}}^{\dagger})\right]\} = \mathrm{Tr}[P\mathcal{E}(\rho)]. \tag{5.25}$$

We can apply Hoeffding's inequality in the same way as for the standard quasiprobability technique. Using the fact that each $\widehat{P}$ is in the range $[-\|\alpha\|_1, +\|\alpha\|_1]$, we find that the total number of samples needed to achieve additive error $\varepsilon$ and success probability $1 - p_{\text{pfail}}$ is $M = \lceil 2\|\alpha\|_1^2 \varepsilon^{-2} \log(2p_{\text{fail}}^{-1}) \rceil$. When the decomposition of $\rho$ is optimal with respect to dyadic negativity as per Definition 4.2, we have that $\|\alpha\|_1 = \Lambda(\rho)$. In this case, the worst-case runtime will be $\mathcal{O}(\Lambda(\rho)^2)$.

This simplified algorithm can be used only in the case where the stabiliser circuit is a convex mixture of unitary Clifford operations, so channels are restricted to be unital. Our main goal, however, is to admit more general stabiliser channels. In particular, extending to adaptive Clifford circuits with mixed magic state inputs allows for universal quantum computation [35]. We will shortly show how appropriate transition probabilities can be computed efficiently even when the input operator is not a state but a dyad, provided that we restrict to simulable channel decompositions as set out in Definition 5.1. This leads to the following theorem.

**Theorem 5.5.** *Let $\rho = \sum_j \alpha_j \left|L_j\right\rangle \left\langle R_j\right|$, be a known stabiliser dyadic decomposition of an initial n-qubit state, where $\alpha_j \in \mathbb{C}$ and the probability distribution $\{|\alpha_j|/\|\alpha\|_1\}$ can be efficiently sampled. Let $\mathcal{E} = O^{(T)} \circ \ldots \circ O^{(1)}$ be a sequence of completely stabiliser-preserving channels with simulable circuit decomposition. Then, given a stabiliser observable E, one can estimate $\langle E \rangle = \mathrm{Tr}(E\mathcal{E}[\rho])$ within additive error $\varepsilon$, with success probability at least $1 - p_{\text{fail}}$ and worst-case runtime*

$$\tau = \frac{\|\alpha\|_1^2}{\varepsilon^2} \log\left(\frac{2}{p_{\text{fail}}}\right) T \mathrm{poly}(n). \tag{5.26}$$

*Furthermore, if the dyadic decomposition of $\rho$ is optimal then $\|\alpha\|_1 = \Lambda(\rho)$.*

To prove Theorem 5.5, we give pseudocode for the dyadic frame simulator and prove its validity and runtime. For brevity the pseudocode assumes that each channel $O^{(t)}$ is specified by a list of at most poly$(n)$ Kraus operators. Once we have proved the validity of this algorithm, we will argue it can be extended to general simulable circuit decompositions. The algorithm has two subroutines: (i) Algorithm 8, which is a variant of the stabiliser-Kraus subroutines, adapted to probabilistically update stabiliser dyads; and (ii) Algorithm 9, which is an outer quasiprobability sampling routine that samples an initial dyad from the initial non-stabiliser state and propagates the dyad through the circuit.

**Algorithm 8** Stabiliser Kraus update subroutine (dyadic scheme)

**Input:** Initial stabiliser dyad $\sigma = |L\rangle\langle R|$; length-$N_K$ list of pairs $\mathbb{L}$ with entries $(q_j, U_j, \Pi_j, h_j)$ representing Kraus operators $\sqrt{q_j}K_j = \sqrt{q_j}2^{h_j/2}U_j\Pi_j$, where $q_j > 0$ are weights, $U_j$ are Clifford gates, and $\Pi_j$ are stabiliser projectors of rank $2^{n-h_j}$.

**Output:** Updated dyad $\sigma' = |L'\rangle\langle R'|$.

```
function DYADUPDATE($\sigma$,$\mathbb{L}$)
    for $r \leftarrow 1$ to $N_K$ do
        $P_r \leftarrow \|q_r\Pi_r\sigma\Pi_r^\dagger\|_1$
    end for
    $P_0 \leftarrow 1 - \sum_{r=1}^{N_K} P_r$
    Sample $s$ from $\{0,\dots,N_K\}$ with probability $P_s$
    if s = 0 then
        $\sigma' \leftarrow 0$
    else
        $\sigma' \leftarrow |L'\rangle\langle R'| = q_s U_s \Pi_s |L\rangle\langle R| \Pi_s^\dagger U_s^\dagger$    ▷ Computed with CH-simulator.
    end if
    return $\sigma'$
end function
```

In Algorithm 8, we use the trace norm, $\|A\|_1 = \mathrm{Tr}\left[\sqrt{A^\dagger A}\right]$, rather than the trace used in the usual Born rule to calculate the transition probabilities for propagating with a particular Kraus operator. While $\mathrm{Tr}(\Pi\rho) = \|\Pi\rho\Pi^\dagger\|_1$ for physical states $\rho$, this does not hold for all dyads $|L\rangle\langle R|$. We illustrate that the trace norm is the appropriate choice with a toy example. Consider the scenario where the penultimate dyad is $\sigma^{(T-1)} = |{+}0\rangle\langle{-}0|$, the final stabiliser channel $O^{(T)}$ is defined by Kraus operators $K_1 = \mathbb{1} \otimes |0\rangle\langle 0|$ and $K_2 = U \otimes |1\rangle\langle 1|$ for some Clifford $U$, and the observable $E$ to be evaluated is the stabiliser projector $E = |1\rangle\langle 1| \otimes \mathbb{1}$. Now, the channel $O^{(T)}$ leaves $\sigma^{(T-1)}$ unchanged, $O^{(T)}(|{+}0\rangle\langle{-}0|) = |{+}0\rangle\langle{-}0|$. It is therefore clear that the correct contribution

**Algorithm 9** Dyadic frame simulator

```
Input: Initial state $\rho$ with known dyadic stabiliser decomposition $\rho = \sum_j \alpha_j |L_j\rangle \langle R_j|$;
    number of samples $M$; set of channel decompositions $\{\mathbb{L}^{(1)}, \ldots, \mathbb{L}^{(T)}\}$ describing the
    completely stabiliser-preserving circuit $\mathcal{E} = O^{(T)} \circ \ldots \circ O^{(1)}$; stabiliser observable $E$.
Output: Estimate $\widehat{E}$ for $\mathrm{Tr}[E\mathcal{E}(\rho)]$, where $\mathcal{E} = O^{(T)} \circ \ldots \circ O^{(1)}$.

 1: Let $P_j = |\alpha_j|/\|\alpha\|_1$ define a probability distribution.
 2: for $m \leftarrow 1$ to $M$ do
 3:     Sample $r$ with probability $P_r$.
 4:     $\sigma^{(0)} \leftarrow |L^{(0)}\rangle \langle R^{(0)}| = |L_r\rangle \langle R_r|$,    $e^{i\theta_r} \leftarrow \alpha_r/|\alpha_r|$.
 5:     for $t \leftarrow 1$ to $T$ do
 6:         $\sigma^{(t)} \leftarrow$ DYADUPDATE($\sigma^{(t-1)}$, $\mathbb{L}^{(t)}$)
 7:         if $\sigma^{(t)} = 0$ then                ▷ Terminate trajectory if “zero” selected.
 8:             $\sigma^{(T)} \leftarrow 0$
 9:             break
10:         end if
11:     end for
12:     $E_m \leftarrow \mathrm{Re}\left\{\|\alpha\|_1 e^{i\theta_r}\,\mathrm{Tr}\left[E\sigma^{(T)}\right]\right\}$                ▷ Computed with CH-simulator.
13: end for
14: return $\widehat{E} \leftarrow \sum_m E_m/M$
```

to the expectation value estimate (line 12 in Alg. 9) should be

$$E_m = \|\alpha\|_1 \mathrm{Re}\{\mathrm{Tr}\left[\Pi\sigma^{(T)}\right]\} \tag{5.27}$$

$$= \|\alpha\|_1 \mathrm{Re}\{\mathrm{Tr}[(|1\rangle\langle 1| \otimes \mathbb{1})\,|{+}0\rangle\langle{-}0|]\} \tag{5.28}$$

$$= \|\alpha\|_1 \mathrm{Re}\{\langle -|1\rangle \langle 1|+\rangle\} = -\|\alpha\|_1/2, \tag{5.29}$$

where we used cyclicity of the trace, and neglect the phase $e^{i\theta_r}$ for brevity. We must ensure that the transition probabilities computed in line 3 of Algorithm 8 produce statistics that converge to this contribution. Supposing we were to naively use the trace to compute transition probabilities, $P_{\mathrm{Tr},j} = \mathrm{Tr}\left[K_j \sigma^{(T-1)} K_j^\dagger\right]$, we would obtain

$$P_{\mathrm{Tr},1} = \mathrm{Tr}[(\mathbb{1} \otimes |0\rangle\langle 0|)\,|{+}0\rangle\langle{-}0|\,(\mathbb{1} \otimes |0\rangle\langle 0|)] = \langle -|+\rangle = 0, \tag{5.30}$$

$$P_{\mathrm{Tr},2} = \mathrm{Tr}\left[(U \otimes |1\rangle\langle 1|)\,|{+}0\rangle\langle{-}0|\,(U^\dagger \otimes |1\rangle\langle 1|)\right] = \langle -|\,U^\dagger U\,|+\rangle\,|\,\langle 1|0\rangle\,|^2 = 0.$$

Here we have a problem, because both paths evaluate to zero, preventing $E_m$ from making any non-zero contribution to our estimate. By contrast, in our algorithm we use the

Schatten 1-norm to compute transition probabilities,

$$P_1 = \|(\mathbb{1} \otimes |0\rangle\langle 0|)\, |{+}0\rangle\langle{-}0|\, (\mathbb{1} \otimes |0\rangle\langle 0|)\|_1 = \|\, |{+}0\rangle\langle{-}0|\, \|_1 = 1 \tag{5.31}$$

$$P_2 = \|(U \otimes |1\rangle\langle 1|)\, |{+}0\rangle\langle{-}0|\, (U^\dagger \otimes |1\rangle\langle 1|)\|_1 \tag{5.32}$$

$$= |\langle 1|0\rangle|^2 \|(U\,|+\rangle)\, |1\rangle\langle 1|\, (\langle -|)U^\dagger\|_1 = 0. \tag{5.33}$$

This method correctly tells us that we should select Kraus operator $K_1$ with certainty, resulting in the correct contribution $E_m = -\|\alpha\|_1/2$.

Below we prove that this strategy leads to an unbiased estimator for $\langle E\rangle$, where each individual sample is bounded as $|E_m| \leq \|\alpha\|_1$. As per standard quasiprobability simulators (see Chapter 2), to estimate an observable within additive error of $\varepsilon$ with success probability $p_{\text{suc}} \geq 1 - p_{\text{fail}}$, we require $M \geq 2\frac{\|\alpha\|_1^2}{\varepsilon^2} \log\left(\frac{2}{p_{\text{fail}}}\right)$ samples from our algorithm [103, 124]. The total runtime given in Theorem 5.5 is the product of $M$ and the runtime to compute each sample. To prove the validity of our algorithm we must: (i) explain how the probabilistic stabiliser dyad update $\sigma \to q_r K_r \sigma K_r^\dagger$ can be carried out efficiently, (ii) show that the values $P_r$ in steps 2-5 of Algorithm 8 form a proper probability distribution, and (iii) show that $\widehat{E}$ returned by Algorithm 9 is an unbiased estimator for $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(\rho)]$.

**(i) Efficient stabiliser update with Kraus operators.** In Algorithm 8 we must compute the trace norm $\|q_r K_r |L\rangle\langle R| K_r^\dagger\|_1$, where $K_r = 2^{h_r/2} U_r \Pi_r$ for all $N_K$ entries $(q_r, U_r, \Pi_r, h_r) \in \mathcal{L}$, and then perform the update $|L\rangle\langle R| \to |L'\rangle\langle R'|$. Any accumulated phase is tracked throughout, but here we absorb this factor in $|L'\rangle\langle R'|$ for brevity. Unlike the trace, the trace norm does not depend on the overlap between $K_r|L\rangle$ and $K_r|R\rangle$, and their vector norms can be calculated separately. Since $U_r$ leaves the norm invariant, the transition probability depends only on the initial dyad, projector and normalisation,

$$\|q_r K_r |L\rangle\langle R| K_r^\dagger\|_1 = q_r \|K_r|L\rangle\| \cdot \|K_r|R\rangle\| = q_r 2^h \|\Pi_r|L\rangle\| \cdot \|\Pi_r|R\rangle\|. \tag{5.34}$$

The projection of each pure state onto a stabiliser subspace can be computed using standard stabiliser simulation techniques in time $\mathcal{O}(hn^2)$ [31], as we did for the stabiliser-Kraus subroutine discussed in 5.1. We must compute the norm for $2N_K$ projected stabiliser states, so the total runtime for computing all transition probabilities for a single step $t$ is $\mathcal{O}(hN_K n^2)$.

Once all $N$ transition probabilities are computed, the algorithm randomly selects some $K_s \propto U_s\Pi_s$ , and in step 10 computes $|L'\rangle\langle R'| \propto U_s\Pi_s\,|L\rangle\langle R|\,\Pi_s U_s^{\dagger}$. Assume we use the phase-sensitive CH-simulator [3] (Section 1.2.1). It was shown in Ref. [3] that the update corresponding to the projection $(\mathbb{1}+Q)/2$, where $Q$ is a Pauli operator, can be carried out in $\mathcal{O}(n^2)$ steps. A rank $2^{n-h}$ stabiliser projector can be decomposed as a product of $h$ Pauli projections, so the projective part of the update takes time $\mathcal{O}(hn^2)$. Meanwhile, any $n$-qubit Clifford operation can be written in canonical form comprising $\mathcal{O}(n^2/\log(n))$ gates from the standard gate set $\{CNOT, H, S\}$ [31]. $CNOT$ and $S$ updates can be performed in time $\mathcal{O}(n)$, and $H$ in time $\mathcal{O}(n^2)$ [3], so the Clifford update for $U_r$ can be completed in time $\mathcal{O}(n^4/\log(n))$. Since $h \leq n$ and we have assumed that $N_K \leq \text{poly}(n)$, the time taken is $\text{poly}(n)$.

Combining all steps, the total time for a single call to DYADUPDATE will be $\mathcal{O}(h(N_K+1)n^2)+\mathcal{O}(n^4/\log(n))$. Since $h \leq n$ and for simulable decompositions $N_K \leq \text{poly}(n)$, the call is completed in $\text{poly}(n)$ time in the general case. In the special case where we restrict each $O^{(t)}$ to act on at most $b$ qubits, for some fixed $b$, the runtime for a single call can be improved considerably. In that case, $N_K \leq 4^{2b}$ and the worst-case runtime will be $\mathcal{O}(b(4^{2b}+1)n^2)+\mathcal{O}(b^2n^2/\log(b))$.

**(ii) Valid probability distribution.** From the definition of $P_0$ in step 5 of Algorithm 8, it is clear that $\sum_{r=0}^{N} P_r = 1$ and $P_1, \ldots, P_N \geq 0$. Hence, to show that $\{P_r\}$ is a probability distribution, it suffices to show that $P_0 \geq 0$. It is given that the channel $O^{(t)}$ is a CPTP map, so its Kraus representation $O^{(t)} = \sum_r q_r K_r(\cdot)K_r^{\dagger}$ must be complete, $\sum_{r=1}^{N_K}(\sqrt{q_r}K_r)^{\dagger}(\sqrt{q_r}K_r) = \mathbb{1}$. Then for any pure state $|\psi\rangle$,

$$1 = \text{Tr}[O(|\psi\rangle\langle\psi|)] = \sum_{r=1}^{N_K} \text{Tr}\left[\sqrt{q_r'}K_r\,|\psi\rangle\langle\psi|\,K_r^{\dagger}\sqrt{q_r'}\right] = \sum_{r=1}^{N_K} \left\|\sqrt{q_r'}K_r\,|\psi\rangle\right\|^2. \quad (5.35)$$

Let $\boldsymbol{Q}^{(\psi)}$ be the $N_K$-element real vector where the $r$-th entry is $Q_r^{(\psi)} = \left\|\sqrt{q_r'}K_r\,|\psi\rangle\right\|$. From equation (5.35), we have that $\left\|\boldsymbol{Q}^{(\psi)}\right\| \leq 1$. Then for any normalised dyad $|L\rangle\langle R|$

we can express $\sum_{r=1} P_r$ as a scalar product between $\boldsymbol{Q}^{(L)}$ and $\boldsymbol{Q}^{(R)}$,

$$\sum_{r=1}^{N_K} P_r = \sum_{r=1} \| \sqrt{q_r} K_r |L\rangle \| \cdot \| \sqrt{q_r} K_r |R\rangle \| \tag{5.36}$$

$$= \sum_{r=1} Q_r^{(L)} Q_r^{(R)} = \boldsymbol{Q}^{(L)} \cdot \boldsymbol{Q}^{(R)} \leq \left\| \boldsymbol{Q}^{(L)} \right\| \cdot \left\| \boldsymbol{Q}^{(R)} \right\| \leq 1, \tag{5.37}$$

where in the last line we used the Cauchy-Schwarz inequality to show that $\sum_{r \geq 1} P_r \leq 1$, as promised. We note that the strategy of using an 'abort' outcome $P_0$ was deployed in the appendix of Ref. [142] to simulate post-selective channels, that is, maps that do not have a complete Kraus representation. In our case each map does have a complete set of Kraus operators; the fact that $P_r$ for $r \geq 1$ can sum to less than 1 instead arises from the non-Hermiticity of the initial dyad $\sigma$.

**(iii) Unbiased estimator.** Finally we show that the expected value of $\widehat{E}$ in Algorithm 9 is $\mathrm{Tr}[E\mathcal{E}(\rho)]$. Let the $(T+1)$-element vector $\mathbf{r} = (r_0, r_1, \ldots, r_T)$ label a particular trajectory through the circuit, in the following sense. The first entry $r_0$ labels the initial dyad $\sigma_{\mathbf{r}}^{(0)} = |L_{r_0}\rangle\langle R_{r_0}|$ selected in steps 3-4. For $t \geq 1$, the entry $r_t$ gives the index of the Kraus operator chosen at the $t$-th circuit element and we write $\mathcal{K}_{\mathbf{r}}^{(t)}(\cdot) = K_{r_t}^{(t)}(\cdot) K_{r_t}^{(t)\dagger}$, and use $q_{\mathbf{r}}^{(t)}$ to denote the corresponding prefactor. Let $\sigma_{\mathbf{r}}^{(t)}$ denote the current dyad updated up to the $t$-th Kraus operator along the trajectory $\mathbf{r}$, so that $\sigma_{\mathbf{r}}^{(t)} = q_{\mathbf{r}}^{(t)} \mathcal{K}_{\mathbf{r}}^{(t)}(\sigma_{\mathbf{r}}^{(t-1)})/P_{\mathbf{r}}^{(t)}$, where $P_{\mathbf{r}}^{(t)}$ is the probability of obtaining the outcome corresponding to the map $\mathcal{K}_{\mathbf{r}}^{(t)}$. The probability $P_{\mathbf{r}}$ of choosing the trajectory $\mathbf{r}$ is given by $P_{\mathbf{r}} = \prod_{t=0}^{T} P_{\mathbf{r}}^{(t)}$, where $P_{\mathbf{r}}^{(0)} = |\alpha_{r_0}|/\|\alpha\|_1$ is the probability of sampling the initial dyad $\sigma_{\mathbf{r}}^{(0)}$. For $t \geq 1$, $P_{\mathbf{r}}^{(t)} = \|q_{\mathbf{r}}^{(t)} \mathcal{K}_{\mathbf{r}}^{(t)}(\sigma_{\mathbf{r}}^{(t-1)})\|_1$ is calculated in the $t$-th call to Algorithm 8. Then the final dyad $\sigma_{\mathbf{r}}^{(T)}$ for trajectory $\mathbf{r}$ is

$$\sigma_{\mathbf{r}}^{(T)} = \frac{q_{\mathbf{r}} \mathcal{K}_{\mathbf{r}}(\sigma_{\mathbf{r}}^{(0)})}{P_{\mathbf{r}}/P_{\mathbf{r}}^{(0)}}, \tag{5.38}$$

where $\mathcal{K}_{\mathbf{r}}(\cdot) = \mathcal{K}_{\mathbf{r}}^{(T)} \circ \ldots \circ \mathcal{K}_{\mathbf{r}}^{(1)}(\cdot)$ and $q_{\mathbf{r}} = \prod_{t=1}^{T} q_{\mathbf{r}}^{(t)}$. This dyad satisfies $\|\sigma_{\mathbf{r}}^{(T)}\|_1 = 1$, but is only defined for those trajectories with $P_{\mathbf{r}}^{(t)} > 0$ for all $t$. We write $\mathbb{P}$ to denote the set of all such non-zero probability trajectories.

Now, there are two mutually exclusive possibilities for a given iteration of Algorithm 9: either we pick $r_t > 0$ at each circuit element, so that $\mathbf{r} \in \mathbb{P}$, and we obtain normalised $\sigma_{\mathbf{r}}^{(T)}$, or $r_t = 0$ for some step $t$, and the iteration terminates with $\sigma_{\mathbf{r}}^{(T)} = 0$. Since these are the only possible outcomes, the total probability of terminating must be $P_{\text{term}} = 1 - \sum_{\mathbf{r}\in\mathbb{P}} P_{\mathbf{r}}$. The expectation value of the random variable $E_m$ in step 12 can now be explicitly written as

$$\mathbb{E}(E_m) = P_{\text{term}} \cdot 0 + \sum_{\mathbf{r}\in\mathbb{P}} P_{\mathbf{r}} \text{Re}\{\|\alpha\|_1 e^{i\theta_{r_0}} \operatorname{Tr}\left[E\sigma_{\mathbf{r}}^{(T)}\right]\} \tag{5.39}$$

$$= \sum_{\mathbf{r}\in\mathbb{P}} P_{\mathbf{r}}^{(0)} \text{Re}\{\|\alpha\|_1 e^{i\theta_{r_0}} \operatorname{Tr}\left[E q_{\mathbf{r}} \mathcal{K}_{\mathbf{r}}(\sigma_{\mathbf{r}}^{(0)})\right]\}, \tag{5.40}$$

where in the second line we have cancelled the factors $P_{\mathbf{r}}^{(t)}$ for $t \geq 1$ with those in the denominator of equation (5.38). The real vectors $\mathbf{r} \notin \mathbb{P}$ are never chosen when running the algorithm, since they correspond to paths where $P_{\mathbf{r}}^{(t)} = 0$ for some $t$, and hence $\mathcal{K}_{\mathbf{r}}^{(t)}(\sigma_{\mathbf{r}}^{(t-1)}) = 0$. Since $\mathcal{K}_{\mathbf{r}}(\sigma_{\mathbf{r}}^{(0)}) = 0$ for all $\mathbf{r} \notin \mathbb{P}$, we can freely add these zero-probability trajectories to the summation (5.40). Thus

$$\mathbb{E}(E_m) = \sum_{\mathbf{r}} P_{\mathbf{r}}^{(0)} \text{Re}\{\|\alpha\|_1 e^{i\theta_{r_0}} \operatorname{Tr}\left[E q_{\mathbf{r}} \mathcal{K}_{\mathbf{r}}(\sigma_{\mathbf{r}}^{(0)})\right]\} \tag{5.41}$$

$$= \sum_{r_0} P_{\mathbf{r}}^{(0)} \text{Re}\{\|\alpha\|_1 e^{i\theta_{r_0}} \operatorname{Tr}\left[E \sum_{r_1,\ldots,r_T} q_{\mathbf{r}} \mathcal{K}_{\mathbf{r}}(|L_{r_0}\rangle\langle R_{r_0}|)\right]\},$$

where in the second line $P_{\mathbf{r}}^{(0)}$ is taken outside the inner sum since it is independent of $r_t$ for $t \geq 1$. The inner summation is over all Kraus trajectories, and by linearity

$$\sum_{r_1,\ldots,r_T} q_{\mathbf{r}} \mathcal{K}_{\mathbf{r}} = \sum_{r_T} q_{\mathbf{r}}^{(T)} \mathcal{K}_{\mathbf{r}}^{(T)} \circ \ldots \circ \sum_{r_1} q_{\mathbf{r}}^{(1)} \mathcal{K}_{\mathbf{r}}^{(1)} = \mathcal{O}^{(T)} \circ \ldots \circ \mathcal{O}^{(1)} = \mathcal{E}. \tag{5.42}$$

Hence, using $P_{\mathbf{r}}^{(0)} e^{i\theta_{r_0}} = \alpha_{r_0}/\|\alpha\|_1$,

$$\mathbb{E}(E_m) = \sum_{r_0} P_{\mathbf{r}}^{(0)} \text{Re}\{\|\alpha\|_1 e^{i\theta_{r_0}} \operatorname{Tr}[E\mathcal{E}(|L_{r_0}\rangle\langle R_{r_0}|)]\} \tag{5.43}$$

$$= \text{Re}\{\operatorname{Tr}\left[E\mathcal{E}(\sum_{r_0} \alpha_{r_0} |L_{r_0}\rangle\langle R_{r_0}|)\right]\} = \operatorname{Tr}[E\mathcal{E}(\rho)], \tag{5.44}$$

We have proved that $\mathbb{E}(\widehat{E}) = \operatorname{Tr}[\Pi\mathcal{E}(\rho)]$, so $\widehat{E}$ is an unbiased estimator, with each sample satisfying $|E_m| \leq \|\alpha\|_1$. This implies we need $\lceil 2\|\alpha\|_1^2 \varepsilon^{-2} \log(2p_{\text{fail}}^{-1}) \rceil$ samples. To compute each sample, we need to make $T$ calls to StabiliserUpdate, and we showed in part (i) that each call takes poly$(n)$ time. Therefore the total runtime is

$\|\alpha\|_1^2\varepsilon^{-2}\log\left(p_{\text{fail}}^{-1}\right)T\text{poly}(n)$, as stated in Theorem 5.5.

We have proved the statement of the theorem for circuit decompositions $\{\mathcal{E} = O^{(T)} \circ \ldots \circ O^{(t)}\}$, where each $O^{(t)}$ is specified by $\text{poly}(n)$ Kraus operators. Recall that simulable channels (Definition 5.1) are those with at most $\text{poly}(n)$ non-unitary Kraus operators, but where the number of unitary Kraus operators could be larger than $\text{poly}(n)$ provided they could be efficiently sampled from. Since unitary operations $U_j$ leave the Schatten 1-norm invariant, for any normalised dyad $\sigma$ we always have $P_j = \|q_j U_j \sigma U_j^\dagger\|_1 = q_j$ in step 3 of Algorithm 8, so the transition probability can be read off from the coefficient. Therefore, we can modify Algorithm 8 to include a unitary path where transition probabilities do not need to be computed in a dyad-dependent fashion (see Algorithm 5 in Section 5.1). Consequently the result also holds for completely stabiliser-preserving channels with simulable decomposition (Definition 5.1). This completes the proof of Theorem 5.5.

### 5.2.3 A stabiliser rank simulator for density operators

In Chapter 3 we reviewed the development of stabiliser rank simulators [3, 59, 60]. For circuits to which they are applicable, namely those involving pure initial states and standard stabiliser operations such as Clifford gates and Pauli measurements, they typically perform better than quasiprobability simulators. Indeed, the sparsification and fast norm estimation procedures presented in Ref. [3] allow computation of Born rule probabilities up to multiplicative error in time linear in the pure-state stabiliser extent of the initial magic state. This enables simulated sampling of the output from small-to-intermediate size quantum computers. However, the method has previously only been applied to circuits with pure state inputs, preventing a direct comparison with simulators for more general noisy states and operations such as the RoM simulator [103], or the Oak Ridge simulator [45] which can simulate any few-qubit CPTP map, subject to additional runtime dependent on the negativity.

In this section we extend the method of Ref. [3] (which we call the BBCCGH simulator) from the state vector picture to general density operators, allowing the application of stabiliser rank techniques to mixed states and noisy circuits. Through the lens of the density operator picture, we are able to recast the sparsification technique as a procedure which samples from an ensemble. This leads to a refinement of the sparsification lemma upon which the BBCCGH simulator rests, avoiding some technical obstables and significantly improving performance. We first present and prove our new sparsification results, before introducing our bit-string sampling simulator for noisy quantum circuits.

### 5.2.3.1 Sparsification lemma for density operators

Stabiliser rank simulators exploit the fact that any state vector $|\psi\rangle$ can be expressed as a linear combination of stabiliser states, $|\psi\rangle = \sum_{j=1}^{k} c_j |\phi_j\rangle$, where $c_j$ are complex. Recall that the stabiliser rank $\chi(\psi)$ is the smallest number of terms $k$ needed for a given state $|\psi\rangle$ [3, 59, 60]. The runtime for computing stabiliser observables using an exact decomposition is lower bounded by $\chi(\psi)$. Computing the exact stabiliser rank is intractable for many-qubit states, and known upper bounds can be large. Instead, the strategy of BBCCGH is to approximate $|\psi\rangle$ with a sparsified $k$-term vector $|\Omega\rangle$ of smaller stabiliser rank, using the subroutine SPARSIFY (see Section 3.4). To recap, BBCCGH [3, Lem. 6]) showed that for any pure state $|\psi\rangle$ and any integer $k > 0$, one can use SPARSIFY to generate random (un-normalised) states $|\Omega\rangle$ with $k$ stabiliser terms such that $\mathbb{E}(\| |\psi\rangle - |\Omega\rangle \|^2) \leq \|\mathbf{c}\|_1^2/k$. In Appendix E.1 we present a simple corollary of [3, Lem. 6]), which implies that

$$\mathbb{E}(\| |\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega| \|_1) \leq 2\frac{\|\mathbf{c}\|_1}{\sqrt{k}} + \frac{\|\mathbf{c}\|_1^2}{k} \approx 2\frac{\|\mathbf{c}\|_1}{\sqrt{k}}. \tag{5.45}$$

For any target precision $\delta_S > 0$, choosing $k \geq 4\|c\|_1^2/\delta_S^2$,

$$\mathbb{E}(\| |\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega| \|_1) \leq \delta_S + \mathcal{O}(\delta_S^2). \tag{5.46}$$

We call this the BBCCGH sparsification lemma [3]. With high probability and subject to some technical caveats (see Section 3.4), by combining sparsification with fast norm estimation, BBCCGH simulates sampling from the quantum distribution $P(\mathbf{x}) = |\langle\mathbf{x}|\psi\rangle|^2$ up to trace-norm error $\delta_S$ in runtime $\|c\|_1^2\delta_S^{-4}\text{poly}(n,w)$. So, assuming an optimal decomposition ($\xi(\psi) = \|c\|_1^2$), the runtime scales linearly with extent $\xi$. Below we improve on this algorithm in three main respects: (i) we extend the simulator from pure to mixed magic state inputs, so that the average-case runtime is proportional to the density-operator extent $\Xi$ (Definition 4.1); (ii) we show that important cases admit decompositions such that $\Xi$ yields the worst-case runtime; and (iii) we derive a new sparsification lemma that, with minor caveats, improves the runtime over the BBCCGH simulator by a factor of $1/\delta_S$. Our new sparsification lemma also removes some lim-

**Algorithm 10** BBCCGH sparsification procedure. (Bravyi et al. [3])

```
Input: n-qubit state decomposition $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, where $|\phi_j\rangle \in \mathrm{STAB}_n$; $k \geq 0$.
Output: Sparsified vector $|\Omega\rangle$ with $k$ terms.
function SPARSIFY($|\psi\rangle$,$k$)
    for $\alpha \leftarrow 1$ to $k$ do
        Sample $|\omega_\alpha\rangle \leftarrow (c_{j_\alpha}/|c_{j_\alpha}|)\,|\phi_{j_\alpha}\rangle$ with probability $|c_{j_\alpha}|/\|\mathbf{c}\|_1$.
    end for
    $|\Omega\rangle \leftarrow (\|\mathbf{c}\|_1/k)\sum_{\alpha=1}^{k}|\omega_\alpha\rangle$
    return $|\Omega\rangle$
end function
```

itations that made certain parameter regimes problematic for the original simulator(see Section 3.4). In Section 5.2.3.2 we present the algorithm in detail. First we state and prove our improved lemma.

For convenience we restate the BBCCGH sparsification procedure in Algorithm 10. Given an integer $k$ and state vector with known decomposition $|\psi\rangle = \sum_{j=1}^{k} c_j |\phi_j\rangle$, the SPARSIFY subroutine outputs a random $k$-term vector,

$$|\Omega\rangle = \frac{\|\mathbf{c}\|_1}{k}\sum_{\alpha=1}^{k}|\omega_\alpha\rangle, \quad \text{where} \quad \Pr\left(|\omega_\alpha\rangle = \frac{c_j}{|c_j|}|\phi_j\rangle\right) = \frac{|c_j|}{\|\mathbf{c}\|_1}, \tag{5.47}$$

so that each term is i.i.d sampled stabiliser state. It follows that $\mathbb{E}(|\omega_\alpha\rangle) = |\psi\rangle/\|\mathbf{c}\|_1$, and in turn $\mathbb{E}(|\Omega\rangle) = |\psi\rangle$.

Since the sparsification $|\Omega\rangle$ is a random superposition of non-orthogonal terms, it need not have unit norm. In Ref. [3], after obtaining a state $|\Omega\rangle$ from SPARSIFY, one estimates its Euclidean norm, and discards the state if its norm is not close to 1. A state post-selected in this way will be close to the target state with high probability, given some assumptions (recall Section 3.4). In our extended simulator, we instead use a sampling strategy that avoids the post-selection step. After SPARSIFY returns a random $|\Omega\rangle$, we estimate $\|\Omega\|$, and normalise the vector. Then, instead of bounding the error between an individual sample and the target state $|\psi\rangle$, we bound the error between $|\psi\rangle$ and the whole ensemble as captured by the density matrix

$$\rho_1 := \mathbb{E}\left[\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}\right] = \sum_{\Omega}\Pr(\Omega)\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}. \tag{5.48}$$

Intuitively, this is advantageous because coherent errors in each sample smooth out to a less harmful stochastic error. Similarly, randomising coherent errors improves error bounds in the setting of circuit compilation [162–165]. Our refinement to the BBCCGH sparsification lemma is summarised in the following theorem.

**Theorem 5.6.** *Let $\rho_1$ be the mixed state in equation* (5.48). *Let $|\psi\rangle$ be an input state with known decomposition $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, where $|\phi_j\rangle$ are stabiliser states, let* $\mathbf{c}$ *be the vector whose elements are the coefficients $c_j$, and let*

$$C_{\psi,\mathbf{c}} = \|\mathbf{c}\|_1 \sum_j |c_j| |\langle \psi|\phi_j\rangle|^2. \tag{5.49}$$

*Then there is a critical precision $\delta_c = 8(C_{\psi,\mathbf{c}} - 1)/\|\mathbf{c}\|_1^2$ such that for every target precision $\delta_S$ for which $\delta_S \geq \delta_c$, we can sample pure states from an ensemble $\rho_1$, where every pure state drawn from $\rho_1$ has stabiliser rank at most $\lceil 4\|\mathbf{c}\|_1^2/\delta_S \rceil$ and*

$$\|\rho_1 - |\psi\rangle\langle\psi|\,\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2). \tag{5.50}$$

*When $|\psi\rangle$ is a Clifford magic state (see below) [3], the critical precision is $\delta_c = 0$, and sampled pure states in $\rho_1$ have stabiliser rank at most $\lceil (2+\sqrt{2})\|\mathbf{c}\|_1^2/\delta_S \rceil$.*

Notice that the theorem sets a critical precision $\delta_c$ above which we can achieve the promised $1/\delta_S$ improvement in the runtime over BBCCGH [3]. For $\delta_S \ll \delta_c$, our runtime has the same leading order $\delta_S$-scaling as BBCCGH but with a much smaller constant prefactor, yielding improved performance. Clifford magic states were defined in Ref. [3] as those states $|\psi\rangle$ that are stabilised by a group $\mathcal{Q}$ of Clifford unitary operators with generators $UX_jU^\dagger$, where $X_j$ is the Pauli $X$ operator that acts on the $j$-th qubit and U is unitary. This includes the important case of $T$ states $|T\rangle = T|+\rangle$. For Clifford magic states the improvement holds for arbitrarily small $\delta_S$.

We first argue that Theorem 5.6 follows from two lemmata. We will then prove the lemmata. First, Lemma 5.7 captures the idea that the ensemble (5.48) can be made close in the trace norm to the target state $|\psi\rangle\langle\psi|$ by choosing sufficiently large $k$, up to a term that depends on the variance of $\langle\Omega|\Omega\rangle$. The second lemma then bounds this variance in terms of $C_{\psi,c}$, $\|\mathbf{c}\|_1$ and $k$.

**Lemma 5.7 (Ensemble sampling lemma).** *Given a state $|\psi\rangle = \sum_j c_j |\phi_j\rangle$ where $\phi_j$ are stabiliser states, we can sample from an ensemble of pure states $\rho_1$ such that every sam-*

*pled pure state has stabiliser rank no larger than k and*

$$\|\rho_1 - |\psi\rangle\langle\psi|\,\|_1 \leq \frac{2\|\mathbf{c}\|_1^2}{k} + \sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]} \tag{5.51}$$

*where* $|\Omega\rangle$ *is the random sparsified vector defined in equation* (5.47).

We will show that this can be proved by splitting $\|\rho_1 - |\psi\rangle\langle\psi|\,\|_1$ into two terms via the triangle inequality, then showing that they are upper bounded by $\sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]}$ and $2\|\mathbf{c}\|_1^2/k$ respectively. With this established, it remains to bound the variance.

**Lemma 5.8 (Sparsification variance bound).** *Using the notation of Lemma 5.7 the variance of* $\langle\Omega|\Omega\rangle$ *satisfies the bound*

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq \frac{4(C-1)}{k} + \frac{2\|\mathbf{c}\|_1^4}{k^2} + \mathcal{O}\left(\frac{C}{k^3}\right), \tag{5.52}$$

*where* $C = C_{\psi,c}$ *is as given in equation* (5.49). *When* $|\psi\rangle$ *is a Clifford magic state as defined in Ref. [3],*

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq \frac{2\|\mathbf{c}\|_1^4}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right). \tag{5.53}$$

We will prove Lemma 5.8 by expanding $\mathrm{Var}[\langle\Omega|\Omega\rangle]$ as a series of terms of the form $\mathbb{E}(\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\lambda|\omega_\mu\rangle)$, treating the cases where the indices $\alpha$, $\beta$, $\lambda$ and $\mu$ are all distinct, then where $\alpha = \beta$, but $(\alpha,\lambda,\mu)$ are all distinct and so on. Assuming these lemmata to hold, we can prove Theorem 5.6.

*Proof of Theorem 5.6.* Substituting $k = 4\|\mathbf{c}\|_1^2/\delta_S$ and $\delta_S \geq 8(C-1)/\|\mathbf{c}\|_1^2$ into equation (5.52) from Lemma 5.8, we obtain

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq \frac{\delta_S^2}{4}\left(1 + \mathcal{O}\left(\frac{\delta_S}{\|\mathbf{c}\|_1^4}\right)\right), \tag{5.54}$$

and hence, using $\sqrt{1+x} \leq 1 + |x|$,

$$\sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]} \leq \frac{\delta_S}{2} + \mathcal{O}(\delta_S^2). \tag{5.55}$$

Using Lemma 5.7 with $k = 4\|\mathbf{c}\|_1^2/\delta_S$ and (5.55), we obtain the main result of Theorem 5.6, that $\|\rho_1 - |\psi\rangle\langle\psi|\,\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2)$. When $|\psi\rangle$ is a Clifford magic state, equa-

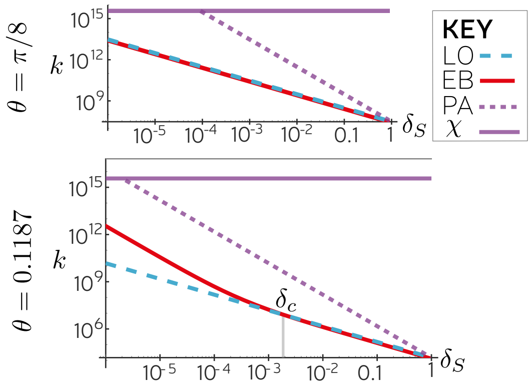


**Figure 5.1:** For the target state $|\psi\rangle = (\cos(\theta)|0\rangle + \sin(\theta)|1\rangle)^{\otimes 100}$ with two choices of $\theta$, we plot the trace norm error $\delta_S$ when using a $k$-term sparsification. EB (Exact Bound) refers to equation (5.51) and is valid for all $\delta_S$, with the variance exactly bounded by equation (5.93). LO (Leading Order) refers to our Theorem 5.6 expression $k = 4\|\mathbf{c}\|_1^2/\delta_S$, and is valid provided $\delta_S \geq \delta_c$ with $\delta_c$ highlighted by a vertical line. Note $\theta = \pi/8$ corresponds to the Clifford magic state $|H\rangle$, for which $\delta_c = 0$. PA (Prior Art) shows the cost of Ref. [3]. The exact stabiliser rank is $\chi$ (see Theorem 2 of Ref. [3]) and this is upper bounds PA. When $C \neq 1$ and $\delta_S < \delta_c$, then EB shows that there is still a large saving even though LO is not valid in this regime. To better understand the deviation of the leading order expression from the exact bound, we refer the reader to the proof of Lemma 5.8 and in particular Fig. 5.2, and to the discussion below following equation (5.56). Plot produced by Earl Campbell, reproduced from Ref. [2] under CC-BY 4.0 license.

tion (5.53) combined with Lemma 5.7 gives

$$\|\rho_1 - |\psi\rangle\langle\psi|\,\|_1 \leq \frac{(2+\sqrt{2})\|\mathbf{c}\|_1^2}{k} + \mathcal{O}\left(\frac{1}{k^2}\right). \tag{5.56}$$

We then obtain the second statement of the theorem by setting $k = \lceil (2+\sqrt{2})\|c\|_1^2/\delta_S \rceil$.

□

Theorem 5.6 tells us that provided the target error $\delta_S$ is no smaller than a critical precision $\delta_c$, one can sample from an ensemble of sparsified states $\rho_1$ that is $\delta_S$-close in the trace norm to $|\psi\rangle$, where the number of stabiliser terms is $k = \lceil 4\|\mathbf{c}\|_1^2/\delta_S \rceil$. This gives a factor $1/\delta_S$ improvement over the BBCCGH [3] sparsification lemma, where $k = \lceil 4\|\mathbf{c}\|_1^2/\delta_S^2 \rceil$. When $\delta_S < \delta_c$, one can compute $C$ and obtain a sharp bound on the trace-norm error by using Lemmas 5.7 and 5.8 directly. In this case, the $\delta_S^{-2}$ scaling of $k$ is recovered, but with a prefactor often much smaller than in the original BBCCGH sparsification lemma. This is because one typically finds that $(C-1)/\|\mathbf{c}\|_1^2 \ll 1$ for many-qubit magic states. We illustrate this in Fig. 5.1, where we compare our sharpened trace-norm bound with that of Ref. [3] for states of the form $|\psi_N\rangle = |\psi\rangle^{\otimes N}$, where

$|\psi\rangle$ are single-qubit magic states, and $N = 100$. While $\delta_S \geq 8(C-1)/\|\mathbf{c}\|_1^2$ we have a quadratic improvement over equation (5.45), but even in the high-precision regime, we find a significant reduction in $k$. We now prove Lemmata 5.7 and 5.8, upon which the above proof of Theorem 5.6 rests.

*Proof of Lemma 5.7.* Given target state $|\psi\rangle = \sum_j c_j |\phi_j\rangle$, we must bound the error $\delta_S = \|\rho_1 - |\psi\rangle\langle\psi|\,\|_1$ where $\rho_1$ is defined

$$\rho_1 = \mathbb{E}\left[\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}\right] = \sum_{\Omega} \Pr(\Omega)\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}. \tag{5.57}$$

Here $|\Omega\rangle = (\|\mathbf{c}\|_1/k)\sum_\alpha |\omega_\alpha\rangle$ are the random sparse vectors returned by SPARSIFY (Algorithm 10) with probability $\Pr(\Omega)$. First we introduce the operator $\rho_2 = \frac{1}{\mu}\mathbb{E}\left[|\Omega\rangle\langle\Omega|\right]$, where $\mu = \mathbb{E}[\langle\Omega|\Omega\rangle]$. Then using the triangle inequality,

$$\delta_S = \|\rho_1 + \rho_2 - \rho_2 - |\psi\rangle\langle\psi|\,\|_1 \leq \|\rho_1 - \rho_2\|_1 + \|\rho_2 - |\psi\rangle\langle\psi|\,\|_1. \tag{5.58}$$

For the first term,

$$\|\rho_1 - \rho_2\|_1 = \|\mathbb{E}\left[|\Omega\rangle\langle\Omega|\left(\frac{1}{\langle\Omega|\Omega\rangle} - \frac{1}{\mu}\right)\right]\|_1. \tag{5.59}$$

Using Jensen's inequality, that is, $\mathbb{E}(f(X)) \leq f(\mathbb{E}(X)))$ for random variable $X$ and convex function $f$, we can bring the expectation value outside the norm,

$$\|\rho_1 - \rho_2\|_1 \leq \mathbb{E}\left|\langle\Omega|\Omega\rangle\left(\frac{1}{\langle\Omega|\Omega\rangle} - \frac{1}{\mu}\right)\right| = \frac{1}{\mu}\mathbb{E}|\mu - \langle\Omega|\Omega\rangle\,|.$$

That $\mu = \mathbb{E}[\langle\Omega|\Omega\rangle] = 1 + (\|\mathbf{c}\|_1^2 - 1)/k$ was shown in Ref. [3], and explained in Section 3.4. Loosening (5.60) with $\mu^{-1} \leq 1$ gives $\|\rho_1 - \rho_2\|_1 \leq \mathbb{E}|\mu - \langle\Omega|\Omega\rangle\,|$, which is simply the average deviation from the mean. Using Jensen's inequality,

$$\mathbb{E}|\mu - \langle\Omega|\Omega\rangle\,| \leq \sqrt{\mathbb{E}|\mu - \langle\Omega|\Omega\rangle\,|^2} = \sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]}, \tag{5.60}$$

$$\implies \quad \|\rho_1 - \rho_2\|_1 \leq \sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]}. \tag{5.61}$$

Next, we bound $\|\rho_2 - |\psi\rangle\langle\psi|\,\|_1$, by first finding an explicit form for $\rho_2$. Observe that $|\Omega\rangle\langle\Omega| = \|c\|_1^2 k^{-2}\sum_{\alpha,\beta}|\omega_\alpha\rangle\langle\omega_\beta|$. Recall from Algorithm 10 that $|\omega_\alpha\rangle = c_j|\phi_j\rangle/|c_j|$

with probability $|c_j|/\|\mathbf{c}\|_1$, so that $\mathbb{E}(|\omega_\alpha\rangle) = |\psi\rangle / \|\mathbf{c}\|_1$ [3]. Then,

$$\mathbb{E}(|\Omega\rangle\langle\Omega|) = \mu\rho_2 = \frac{\|c\|_1^2}{k^2} \sum_{\alpha,\beta} \mathbb{E}[|\omega_\alpha\rangle\langle\omega_\beta|]. \tag{5.62}$$

Let $\sigma := \mathbb{E}[\,|\omega_\alpha\rangle\langle\omega_\alpha|]$. We split (5.62) into two summations as follows:

$$\begin{aligned}\mu\rho_2 &= \frac{\|c\|_1^2}{k^2}\left[\left(\sum_{\alpha\neq\beta}\mathbb{E}[|\omega_\alpha\rangle\langle\omega_\beta|]\right) + \left(\sum_\alpha \mathbb{E}[\,|\omega_\alpha\rangle\langle\omega_\alpha|]\right)\right], && (5.63)\\ &= \frac{\|c\|_1^2}{k^2}\left[\left(\sum_{\alpha\neq\beta}\frac{|\psi\rangle\langle\psi|}{\|\mathbf{c}\|_1^2}]\right) + \left(\sum_\alpha \sigma\right)\right], && (5.64)\\ &= \frac{1}{k^2}\left(\sum_{\alpha\neq\beta}|\psi\rangle\langle\psi|\right) + \frac{\|c\|_1^2}{k^2}\sum_\alpha \sigma. \end{aligned}$$

In the first contribution, we used the independence of $\omega_\alpha$ and $\omega_\beta$ when $\alpha \neq \beta$, so that $\mathbb{E}[\,|\omega_\alpha\rangle\langle\omega_\beta|] = \mathbb{E}(\,|\omega_\alpha\rangle)\mathbb{E}(\,\langle\omega_\beta|)$, and that $\mathbb{E}[|\omega_\alpha\rangle] = |\psi\rangle / \|\mathbf{c}\|_1$. Next, there are $k(k-1)$ terms and $k$ terms in the first and second summations respectively, so

$$\mu\rho_2 = \left(1 - k^{-1}\right)|\psi\rangle\langle\psi| + \frac{\|c\|_1^2}{k}\sigma. \tag{5.65}$$

Using this form for $\rho_2$, we have that

$$\begin{aligned}\|\rho_2 - |\psi\rangle\langle\psi|\,\|_1 &= \mu^{-1}\|\mu\rho_2 - \mu\,|\psi\rangle\langle\psi|\,\|_1, && (5.66)\\ &= \mu^{-1}\|(1-k^{-1}-\mu)\,|\psi\rangle\langle\psi| + \|c\|_1^2 k^{-1}\sigma\|_1.\end{aligned}$$

Substituting in the value of $\mu$ we find $1 - k^{-1} - \mu = -\|\mathbf{c}\|_1^2/k$ and so

$$\|\rho_2 - |\psi\rangle\langle\psi|\,\|_1 = \frac{\|c\|_1^2}{k\mu}\|\sigma - |\psi\rangle\langle\psi|\,\|_1 \leq 2\frac{\|c\|_1^2}{k}, \tag{5.67}$$

where we have used the triangle inequality, $\|\sigma\|_1$ and $\mu^{-1} \leq 1$. Substituting equation (5.61) and equation (5.67) into equation (5.58), we obtain

$$\delta_S = \|\rho_1 - |\psi\rangle\langle\psi|\,\|_1 \leq \frac{2\|\mathbf{c}\|_1^2}{k} + \sqrt{\mathrm{Var}[\langle\Omega|\Omega\rangle]}, \tag{5.68}$$

completing the proof. □

*Proof of Lemma 5.8.* In Ref. [3] it was shown that

$$\mu = \mathbb{E}[\langle\Omega|\Omega\rangle] = \frac{||\mathbf{c}||_1^2}{k} + \frac{||\mathbf{c}||_1^2}{k^2}\mathbb{E}\left(B\right), \tag{5.69}$$

where $B = \sum_\alpha \sum_{\beta\neq\alpha}\langle\omega_\alpha|\omega_\beta\rangle$. Since $|\omega_\alpha\rangle$ and $\left|\omega_\beta\right\rangle$ are independently sampled for distinct $\alpha$ and $\beta$, we get

$$\mathbb{E}\left(\langle\omega_\alpha|\omega_\beta\rangle\right) = \mathbb{E}(\langle\omega_\alpha|)\mathbb{E}(\left|\omega_\beta\right\rangle) = \frac{\langle\psi|\psi\rangle}{||\mathbf{c}||_1^2}. \tag{5.70}$$

We use similar proof techniques to bound $\mathbb{E}[\langle\Omega|\Omega\rangle^2]$, and in turn bound the variance. We begin with

$$\begin{aligned}\langle\Omega|\Omega\rangle^2 &= \frac{||\mathbf{c}||_1^4}{k^4}\left(\sum_{\alpha,\beta}\left\langle\omega_\alpha\middle|\omega_\beta\right\rangle\right)^2, &(5.71)\\ &= \frac{||\mathbf{c}||_1^4}{k^4}\left(\sum_\alpha\left(\langle\omega_\alpha|\omega_\alpha\rangle + \sum_{\beta\neq\alpha}\left\langle\omega_\alpha\middle|\omega_\beta\right\rangle\right)\right)^2, &(5.72)\\ &= \frac{||\mathbf{c}||_1^4}{k^4}(k^2+2kB+B^2), &(5.73)\end{aligned}$$

since there are $k$ terms in the sum over $\alpha$. Whereas from equation (5.69) we have

$$\mathbb{E}[\langle\Omega|\Omega\rangle]^2 = \frac{||\mathbf{c}||_1^4}{k^4}(k^2+2k\mathbb{E}(B)+\mathbb{E}(B)^2). \tag{5.74}$$

Comparing these expressions, for the variance we obtain

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] = \mathbb{E}[\langle\Omega|\Omega\rangle^2] - \mathbb{E}[\langle\Omega|\Omega\rangle]^2 = \frac{||\mathbf{c}||_1^4}{k^4}(\mathbb{E}(B^2)-\mathbb{E}(B)^2). \tag{5.75}$$

By counting terms in the summation $B$, and using the relation (5.70), we find

$$\mathbb{E}(B)^2 = \left(k(k-1)\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle]\right)^2 = \frac{k^2(k-1)^2}{||\mathbf{c}||_1^4}. \tag{5.76}$$

Expanding $B^2$, we get

$$B^2 = \left(\sum_{\alpha}\sum_{\beta\neq\alpha} \langle \omega_\alpha | \omega_\beta \rangle\right)\left(\sum_{\lambda}\sum_{\mu\neq\lambda} \langle \omega_\lambda | \omega_\mu \rangle\right) \tag{5.77}$$

$$= \sum_{(\alpha,\beta,\lambda,\mu)\in\mathcal{A}} \langle \omega_\alpha | \omega_\beta \rangle \langle \omega_\lambda | \omega_\mu \rangle + B' \tag{5.78}$$

where $\mathcal{A}$ is the set of all possible combinations $(\alpha,\beta,\lambda,\mu)$ where all four indices are distinct, and $B'$ denotes the remaining terms where at least two of the indices are equal. Now, if $(\alpha,\beta,\lambda,\mu)$ are all distinct, then $\langle \omega_\alpha | \omega_\beta \rangle$ and $\langle \omega_\lambda | \omega_\mu \rangle$ are i.i.d. variables, so $\mathbb{E}(\langle \omega_\alpha | \omega_\beta \rangle \langle \omega_\lambda | \omega_\mu \rangle) = \mathbb{E}(\langle \omega_\alpha | \omega_\beta \rangle)\mathbb{E}(\langle \omega_\lambda | \omega_\mu \rangle)$. This yields

$$\mathbb{E}(B^2) = \frac{k(k-1)(k-2)(k-3)}{\|\mathbf{c}\|_1^4} + \mathbb{E}(B'). \tag{5.79}$$

Substituting the expressions (5.76) and (5.79) back into (5.75), we obtain

$$\mathrm{Var}[\langle \Omega | \Omega \rangle] = \frac{\|\mathbf{c}\|_1^4}{k^4}\mathbb{E}(B') - \frac{k(k-1)(4k-6)}{k^4}. \tag{5.80}$$

We must now consider terms $\langle \omega_\alpha | \omega_\beta \rangle \langle \omega_\lambda | \omega_\mu \rangle$ in the expansion of $B^2$ where $(\alpha,\beta,\lambda,\mu)$ are not all distinct. We use the notation $B_{j=k}$ to indicate the sum of all terms where indices $j$ and $k$ are equal but all others are distinct, e.g. $B_{\lambda=\alpha} = \sum_{\alpha,\beta,\mu} \langle \omega_\alpha | \omega_\beta \rangle \langle \omega_\alpha | \omega_\mu \rangle$, where the summation is over terms such that $\alpha$, $\beta$ and $\mu$ are all distinct, and so on. There are $k(k-1)(k-2)$ terms in each summation of this type. Similarly for the terms sharing two pairs of indices, we use the notation $B_{\lambda=\alpha;\mu=\beta} = \sum_{\alpha\neq\beta} \langle \omega_\alpha | \omega_\beta \rangle \langle \omega_\alpha | \omega_\beta \rangle$, and these have $k(k-1)$ terms. From equation (5.77), we never have terms where $\alpha = \beta$ or $\lambda = \mu$. We then have

$$B' = B_{\lambda=\alpha} + B_{\mu=\alpha} + B_{\lambda=\beta} + B_{\mu\beta} + B_{\lambda=\alpha;\mu=\beta} + B_{\mu=\alpha;\lambda=\beta}. \tag{5.81}$$

One can check that $\mathbb{E}[B^*_{\lambda=\alpha}] = \mathbb{E}[B_{\mu=\beta}]$ and $\mathbb{E}[B^*_{\mu=\alpha}] = \mathbb{E}[B_{\lambda=\beta}]$. Therefore

$$\mathbb{E}[B'] = 2\mathrm{Re}\{\mathbb{E}[B_{\lambda=\beta}] + \mathbb{E}[B_{\mu=\beta}]\} + \mathbb{E}[B_{\lambda=\alpha;\mu=\beta} + B_{\mu=\alpha;\lambda=\beta}]. \tag{5.82}$$

Next we note that

$$\mathbb{E}[B_{\lambda=\beta}] = \sum_{\alpha}\sum_{\beta\neq\alpha}\sum_{\alpha\neq\mu\neq\beta} \mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\mu\rangle] \tag{5.83}$$

$$= k(k-1)(k-2)\mathbb{E}[\langle\omega_\alpha|]\mathbb{E}[|\omega_\beta\rangle\langle\omega_\beta|]\mathbb{E}[|\omega_\mu\rangle]$$

$$= \frac{k(k-1)(k-2)}{\|\mathbf{c}\|_1^2}\langle\psi|\sigma|\psi\rangle, \tag{5.84}$$

where $\sigma = \mathbb{E}[|\omega_\beta\rangle\langle\omega_\beta|] = \sum_j(|c_j|/\|\mathbf{c}\|_1)|\phi_j\rangle\langle\phi_j|$, since the probability of sampling $|\omega_\beta\rangle\langle\omega_\beta| = |\phi_j\rangle\langle\phi_j|$ is defined as $p_j = |c_j|/\|\mathbf{c}\|_1$. Next we consider $\mathbb{E}[B_{\mu=\beta}]$. Taking the modulus and using the triangle inequality we obtain

$$|\mathbb{E}[B_{\mu=\beta}]| \leq \sum_{\alpha}\sum_{\beta\neq\alpha}\sum_{\alpha\neq\lambda\neq\beta} \mathbb{E}[|\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\lambda|\omega_\beta\rangle|] \tag{5.85}$$

$$= k(k-1)(k-2)\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\lambda\rangle] \tag{5.86}$$

$$= \frac{k(k-1)(k-2)}{\|\mathbf{c}\|_1^2}\langle\psi|\sigma|\psi\rangle. \tag{5.87}$$

Similarly, for the last two terms $B'' = B_{\lambda=\alpha;\mu=\beta} + B_{\mu=\alpha;\lambda=\beta}$, we obtain

$$|\mathbb{E}[B'']| \leq \sum_{\alpha}\sum_{\beta\neq\alpha}\mathbb{E}[|\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\alpha|\omega_\beta\rangle|] + \sum_{\alpha}\sum_{\beta\neq\alpha}\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\alpha\rangle]$$

$$= 2\sum_{\alpha}\sum_{\beta\neq\alpha}\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\alpha\rangle]. \tag{5.88}$$

By cyclicity of the trace, $\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\alpha\rangle] = \mathbb{E}[\mathrm{Tr}[|\omega_\alpha\rangle\langle\omega_\alpha|\,|\omega_\beta\rangle\langle\omega_\beta|]]$, so

$$\mathbb{E}[\langle\omega_\alpha|\omega_\beta\rangle\langle\omega_\beta|\omega_\alpha\rangle] = \mathrm{Tr}[\mathbb{E}[|\omega_\alpha\rangle\langle\omega_\alpha|]\mathbb{E}[|\omega_\beta\rangle\langle\omega_\beta|]] = \mathrm{Tr}[\sigma^2], \tag{5.89}$$

$$\implies \quad |\mathbb{E}[B'']| \leq 2k(k-1)\,\mathrm{Tr}[\sigma^2] \leq 2k(k-1). \tag{5.90}$$

Combining the results (5.82), (5.84), (5.87) and (5.90) gives us

$$\mathbb{E}[B'] \leq 4\frac{k(k-1)(k-2)}{\|\mathbf{c}\|_1^2}\langle\psi|\sigma|\psi\rangle + 2k(k-1). \tag{5.91}$$

Writing

$$C = \|\mathbf{c}\|_1^2\langle\psi|\sigma|\psi\rangle = \|\mathbf{c}\|_1\sum_j|c_j||\langle\psi|\phi_j\rangle|^2 \tag{5.92}$$

and substituting the expression for $\mathbb{E}(B')$ into equation (5.80) we obtain

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq 4\frac{k^3-3k^2+2k}{k^4}C+2\frac{\|\mathbf{c}\|_1^4}{k^2}\left(1-\frac{1}{k}\right)-\frac{4k^3-10k^2+6k}{k^4}, \tag{5.93}$$

which to leading order in $1/k$ gives us the general bound appearing in Lemma 5.8,

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq \frac{4(C-1)}{k}+2\left(\frac{\|\mathbf{c}\|_1^2}{k}\right)^2+\mathcal{O}\left(\frac{C}{k^3}\right). \tag{5.94}$$

For Clifford magic states stabilised by a group $\mathcal{Q}$ of Clifford operators, there exists [3] an optimal decomposition

$$|\psi\rangle = \sum_{q\in\mathcal{Q}} c_q\,|\phi_q\rangle = \frac{1}{|\mathcal{Q}|\,\langle\psi|\phi_0\rangle}\sum_{q\in\mathcal{Q}} q\,|\phi_0\rangle, \tag{5.95}$$

where $|\phi_0\rangle$ is some stabiliser state that maximises $|\langle\psi|\phi_0\rangle|$. Then

$$\|\mathbf{c}\|_1 = |\mathcal{Q}|\cdot(|\mathcal{Q}||\langle\psi|\phi_0\rangle|)^{-1} = |\langle\psi|\phi_0\rangle|^{-1} \tag{5.96}$$

and $\sigma = \sum_{q\in\mathcal{Q}} p_q q\,|\phi_0\rangle\langle\phi_0|\,q^\dagger$, where $p_q = |\mathcal{Q}|^{-1}$. This yields

$$\langle\psi|\,\sigma\,|\psi\rangle = \sum_{q\in\mathcal{Q}} p_q\,\langle\psi|\,q\,|\phi_0\rangle\langle\phi_0|\,q^\dagger\,|\psi\rangle \tag{5.97}$$

$$= \sum_{q\in\mathcal{Q}} p_q\,\langle\psi|\phi_0\rangle\,\langle\phi_0|\psi\rangle = |\langle\psi|\phi_0\rangle|^2 = \frac{1}{\|\mathbf{c}\|_1^2}, \tag{5.98}$$

where we used the Hermiticity of $q$ and $q\,|\psi\rangle = |\psi\rangle$. This shows that for optimal decompositions of Clifford magic states, $C=1$, and leads to the simplified bound

$$\mathrm{Var}[\langle\Omega|\Omega\rangle] \leq 2\left(\frac{\|\mathbf{c}\|_1^2}{k}\right)^2+\frac{2}{k^3}, \tag{5.99}$$

completing the proof. □

Finally, we comment on the effect of the constant $C$ when $|\psi\rangle$ is not a Clifford magic state. Recall that $C$ can be written in terms of the expected overlap, $C = \|\mathbf{c}\|_1^2\mathbb{E}\left[|\langle\psi|\omega\rangle|^2\right]$, and enters into Theorem 5.6 via the critical precision $\delta_c = 8(C-1)/\|\mathbf{c}\|_1^2$. Consider $|\psi\rangle = |\psi'\rangle^{\otimes N}$ where $|\psi'\rangle$ are pure states. When $|\psi\rangle$ is a product of

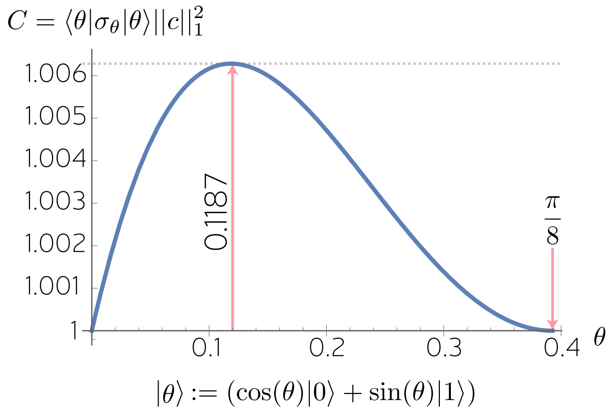


**Figure 5.2:** The variable $C$ as introduced in equation (5.92), plotted as a function of the angle $\theta$ for a class of single-qubit states. This is the $C$ value for one copy of the state, for $n$ copies we must raise to the $n^{\text{th}}$ power. The prefactor $C-1$ appearing in equation (5.93) is key to the behaviour of the bound, as when $C=1$, the variance scales asymptotically as $\mathcal{O}(1/k^2)$. We highlight two specific angles $\theta = \{\pi/8, 0.1187\}$ that correspond to angles used in Fig. 5.1. For $\theta = \pi/8$, we have $C-1=0$ and so the $\mathcal{O}(1/k^2)$ is exact as can be seen in Fig. 5.1. For $\theta = 0.1187$, we have the maximal possible value of $C$ and Fig. 5.1 shows the maximal deviation from $\mathcal{O}(1/k^2)$ scaling. Plot generated by Earl Campbell, reproduced from Ref. [2] under CC-BY 4.0 license.

$N$ pure states, we can write each randomly sampled stabiliser state as $|\omega\rangle = \bigotimes_{\alpha=1}^{N} |\omega_\alpha\rangle$, where $|\omega_\alpha\rangle$ are i.i.d. random vectors. It follows that $\mathbb{E}\left[|\langle\psi|\omega\rangle|^2\right] = (\mathbb{E}\left[|\langle\psi'|\omega_\alpha\rangle|^2\right])^N$. Since $|\omega_\alpha\rangle$ are always stabiliser states, when $|\psi'\rangle$ are non-stabiliser states, we have $|\langle\psi'|\omega_\alpha\rangle|^2 < 1$. Therefore the threshold precision $\delta_c < 8C/\|\mathbf{c}\|_1^2 = 8(\mathbb{E}\left[|\langle\psi'|\omega_\alpha\rangle|^2\right])^N$ vanishes for large $N$ when $|\psi\rangle$ is a tensor product of $N$ pure states. Moreover, in Figure 5.2 we plot values of $C$ for a class of single-qubit states, showing that $C-1$ is close to zero even when $N$ is not large.

### 5.2.3.2 Bit-string sampling algorithm

We next show how to simulate sampling from the output distribution of a circuit with noisy magic state inputs. Assume we measure the first $w$ qubits. Let $\Pi_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w}$ be the projector representing the outcome where we obtain bit-string $\mathbf{x}$. The probability of obtaining the string $\mathbf{x}$ is given by the Born rule, $P(\mathbf{x}) = \mathrm{Tr}[\Pi_{\mathbf{x}}\rho]$. We call $P$ the *quantum* probability distribution. Here we deal with the simulation task of classically sampling from a probability distribution $P_{\text{sim}}(\mathbf{x})$ over $w$-bit strings $\mathbf{x}$ such that $P_{\text{sim}}$ is $\delta$-close in $\ell_1$-norm to $P$, with high probability. We saw in Section 5.1 that this can be done efficiently for poly-sized stabiliser circuits. The BBCCGH algorithm introduced in Ref. [3] performs this task for pure magic states $|\psi\rangle$, with runtime that scales linearly with the

**Algorithm 11** Bit-string sampling algorithm (simplified)

```
Input: $n$-qubit density operator $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$, with known decompositions $\left|\psi_j\right\rangle = \sum_j c_r^{(j)} \left|\phi_r\right\rangle$; number of bits $w$; target precision $\delta > \delta_{\psi_j,c}, \forall j$, where $\delta_{\psi_j,c}$ is the critical precision for state $\psi_j$ (see Theorem 5.6).
Output: $w$-bit string $\mathbf{x}$.
Randomly draw $\psi_j$ with probability $p_j$.
$k \leftarrow \lceil 12\|\mathbf{c}^{(j)}\|_1^2/\delta \rceil$
$|\Omega\rangle \leftarrow$ SPARSIFY$(\psi_j, k)$
Estimate $\Pr(x_1 = 0) = \frac{\|\Pi_0|\Omega\rangle\|^2}{\||\Omega\rangle\|^2}$ using FASTNORM [3]. $\mathbf{x} \leftarrow (x_1)$
$x_1 \leftarrow 0$ with probability $\Pr(x_1 = 0)$, $x_1 \leftarrow 1$ otherwise.
for $b \leftarrow 2$ to $w$ do
    Estimate $\Pr(x_b = y|\mathbf{x}) = \frac{\Pr[(\mathbf{x},y)]}{\Pr[\mathbf{x}]} = \frac{\left\|\Pi_{(\mathbf{x},y)}|\Omega\rangle\right\|^2}{\|\Pi_{\mathbf{x}}|\Omega\rangle\|^2}$ using FASTNORM.
    Sample $x_b \leftarrow 0$ or $x_b \leftarrow 1$ accordingly and concatenate, $\mathbf{x} \leftarrow (\mathbf{x}, x_b)$.
end for
```

pure state extent $\xi(\psi)$.

The algorithm presented in this section allows sampling from mixed magic states $\rho$ in average runtime linear in density-operator extent $\Xi(\rho)$. The simulator is closely related to the BBCCGH simulator [3], differing in two key respects: (i) whereas BBCCGH deals only with pure states, our variant admits general mixed states; and (ii) we employ our improved sparsification lemma to reduce runtime. We also avoid a post-selection step needed for the BBCCGH algorithm (recall Section 3.4). We give simplified pseudocode showing the key steps in our procedure (Algorithm 11), for the case where the target error $\delta$ is greater than the critical value $\delta_c$ introduced in Theorem 5.6. Full pseudocode for the case of arbitrary precision is given in Appendix E.2, where we also prove the validity and runtime for the general case. The main steps in the algorithm are (1) the sampling of a random pure state $\left|\psi_j\right\rangle$ from the ensemble $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$, (2) a call to SPARSIFY to generate the $k$-term approximation $|\Omega\rangle$, and (3) computation of a chain of conditional probabilities using at most $2w+1$ calls to the subroutine FASTNORM which estimates the norm of a given vector (see Section 3.5 and Refs. [3, 60]). Here we first sketch the proof before discussing the runtime improvement over BBCCGH [3]. We initially assume that $\delta > \delta_c$, returning to the $\delta \leq \delta_c$ case at the end of the section.

We want to show that the classical probability distribution $P_{\text{sim}}$ satisfies the inequality $\|P_{\text{sim}} - P\|_1 \leq \delta + \mathcal{O}(\delta^2)$, where $P$ is the quantum distribution. We split the proof into two parts. First, we consider an idealised algorithm EXACT where the calls to FAST-

NORM are replaced by an oracle that can compute $\|\Pi_{\mathbf{x}}|\Omega\rangle\|$ exactly given $k$-term sparsification $|\Omega\rangle$. Let $P_{\text{ex}}(\mathbf{x})$ be the probability of obtaining the string $\mathbf{x}$ as the output of EXACT. We will first show that $P_{\text{ex}}$ is $\delta_S$-close to the quantum distribution $P$ in $\ell_1$-norm, and then show that $P_{\text{sim}}$ is $\varepsilon$-close to $P_{\text{ex}}$. We then split the error budget so that $\delta = \delta_S + \varepsilon$. In Appendix E.2 we show that the optimal strategy is to set $\delta_S = \delta/3$ and $\varepsilon = 2\delta/3$.

Let $\mathbf{x}_m = (x_1, \ldots, x_m)$ be the bit string comprising the first $m$ bits of $\mathbf{x}$, and let $|\Omega_m\rangle = \Pi_{\mathbf{x}_m}|\Omega\rangle$ be the projection of the first $m$ qubits of $|\Omega\rangle$. We can obtain the probability $\Pr(\mathbf{x}|\Omega) = \Pr(x_1)\Pr(x_2|\mathbf{x}_1)\ldots\Pr(x_w|\mathbf{x}_{w-1})$ of sampling $\mathbf{x}$ from EXACT given fixed sparsification $|\Omega\rangle$ by multiplying the conditional probabilities $\Pr(x_b|\mathbf{x}_{b-1})$ computed in step 7 of Algorithm 11,

$$\Pr(\mathbf{x}|\Omega) = \frac{\||\Omega_1\rangle\|^2}{\||\Omega\rangle\|^2}\frac{\||\Omega_2\rangle\|^2}{\||\Omega_1\rangle\|^2}\cdots\frac{\|\Pi_{\mathbf{x}}|\Omega\rangle\|^2}{\||\Omega_{w-1}\rangle\|^2} = \frac{\|\Pi_{\mathbf{x}}|\Omega\rangle\|^2}{\||\Omega\rangle\|^2} = \mathrm{Tr}\left[\Pi_{\mathbf{x}}\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}\right]. \quad (5.100)$$

Thus EXACT simulates sampling from the quantum state $|\Omega\rangle/\||\Omega\rangle\|$ exactly; any error arises solely from the sparsification procedure. Now consider that randomly choosing a pure state $\psi_j$ from $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, generating a random approximation $|\Omega\rangle$ using SPARSIFY and then normalising is equivalent to sampling a pure state from the ensemble,

$$\sigma = \sum_j p_j \sum_\Omega \Pr\big(\Omega|\psi_j\big)\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle} = \sum_j p_j \rho_1^{(j)} \quad (5.101)$$

where $\Pr\big(\Omega|\psi_j\big)$ is the probability of SPARSIFY outputting the vector $|\Omega\rangle$ , and $\rho_1^{(j)}$ is the expected projector $\mathbb{E}(|\Omega\rangle\langle\Omega| / \langle\Omega|\Omega\rangle)$ as defined in equation (5.48), both conditioned on the input to SPARSIFY being $|\psi_j\rangle$. From our argument above it follows that $P_{\text{ex}}(\mathbf{x}) = \mathrm{Tr}[\Pi_{\mathbf{x}}\sigma]$. A key conceptual difference between our method and that of Ref. [3] is that while the BBCCGH sparsification results are concerned with the distance $\|\psi - \Omega\|^2$ for a particular random sparsification $\Omega$, here we compare the target state $\rho$ with the full ensemble over sparsifications $\sigma$. From our sparsification lemma (Theorem 5.6), for each pure state $|\psi_j\rangle$, we have that $\|\rho_1^{(j)} - |\psi_j\rangle\langle\psi_j|\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2)$. It follows that $\|\sigma - \rho\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2)$, and so,

$$\|P_{\text{ex}} - P\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2). \quad (5.102)$$

Next we argue that $P_{\text{ex}}$ is $\varepsilon$-close to $P_{\text{sim}}$, the distribution arising from our full classical algorithm, including the calls FastNorm. Recall from Section 3.5 that FastNorm is able to output an estimate $\eta$ for $||\Omega_m||^2$ up to some multiplicative error $\varepsilon_{\text{FN}}$ with probability $(1-p_{\text{FN}})$, which we can set arbitrarily small, such that

$$(1-\varepsilon_{\text{FN}})||\Omega_m||^2 \leq \eta \leq (1+\varepsilon_{\text{FN}})||\Omega_m||^2, \tag{5.103}$$

and that each call takes time $\mathcal{O}(kn^3\varepsilon_{\text{FN}}^{-2}\log(p_{\text{FN}}^{-1}))$ when the vector $|\Omega_m\rangle$ has $k$ stabiliser terms. One can show (see Appendix E.2) that estimating the chain of $w$ conditional probabilities (5.100) using FastNorm leads to a total multiplicative error $3w\varepsilon_{\text{FN}}$ in the distribution sampled from, i.e.

$$(1-3w\varepsilon_{\text{FN}})P_{\text{ex}}(\mathbf{x}) \leq P_{\text{sim}}(\mathbf{x}) \leq (1+3w\varepsilon_{\text{FN}})P_{\text{ex}}(\mathbf{x}), \tag{5.104}$$

so to achieve multiplicative error $\varepsilon$ we must set $\varepsilon_{\text{FN}} = \varepsilon/3w$. One can similarly show that to achieve overall success probability at least $1-p_{\text{fail}}$, it is necessary to set the parameters for FastNorm so that $p_{\text{FN}} = p_{\text{fail}}/(2w)$. This governs the runtime of each call to FastNorm. By combining this result with equation (5.102) we have $||P_{\text{sim}} - P||_1 \leq \delta + \mathcal{O}(\delta^2)$.

To analyse the runtime of our simulator, we define $\widetilde{\Xi} = \sum_j p_j ||\mathbf{c}^{(j)}||_1^2$, where $\mathbf{c}^{(j)}$ is the vector of coefficients in the decomposition $|\psi_j\rangle = \sum_r c_r^{(j)} |\phi_j\rangle$. Recall that for an $n$-qubit state vector with $k$ terms, the runtime of FastNorm is $\mathcal{O}(kn^3\varepsilon_{\text{FN}}^{-2})$. From the previous discussion, if we selected the $j$-th pure state in the decomposition of $\rho$, we will have set $k \propto ||\mathbf{c}^{(j)}||_1^2\delta^{-1}$ and $\varepsilon_{\text{FN}} \propto \delta w^{-1}$. In a single run of the full algorithm, FastNorm is called $\mathcal{O}(w)$ times. Therefore the runtime to generate a single $w$-length bit string is $T = \mathcal{O}(||\mathbf{c}^{(j)}||_1^2 w^3 n^3 \delta^{-3})$ with probability $p_j$. So the *average-case* runtime is $\mathcal{O}(\widetilde{\Xi} w^3 n^3 \delta^{-3})$. Through $\widetilde{\Xi}$, this average-case runtime is sensitive to the particular decomposition of $\rho$ supplied to the simulator. In the case where the decomposition is optimal with respect to the density-operator extent $\Xi$ (Definition 4.2), we have $\widetilde{\Xi} = \Xi(\rho)$, so that the average-case runtime is linear in $\Xi(\rho)$. Recall from Theorem 4.8 in Section 4.1.4 that all single-qubit states admit an equimagical decomposition that naturally extends to all tensor products of single-qubit states. In that case $||\mathbf{c}^{(j)}||_1^2 = \Xi(\rho)$ for all $j$, so that we can give the *worst-case* runtime as $\mathcal{O}(\Xi(\rho))$.

The runtime scaling of $\mathcal{O}(\delta^{-3})$ holds provided that the sparsification error $\delta_S$ is not smaller than the critical threshold $\delta_c = 8(C-1)/\|\mathbf{c}\|_1^2$, where $C$ is defined in equation (5.49). However, the algorithm is still valid for the case of arbitrary precision, $\delta_S < \delta_c$. In this case we recover the same leading order scaling as Bravyi *et al.*, namely $\mathcal{O}(\delta^{-4})$ [3], but typically with a prefactor improved by several orders of magnitude (see Figure 5.1). A detailed technical analysis is provided in Appendix E.2, including explicit proof of the following theorem, which captures the results discussed above.

**Theorem 5.9.** *Let $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$ be an n-qubit state where every pure state has a known stabiliser decomposition $\left|\psi_j\right\rangle = \sum_r c_r^{(j)} \left|\phi_r\right\rangle$. For every $\left|\psi_j\right\rangle$, let $C_j = \|\mathbf{c}^{(j)}\|_1 \sum_r |c_r^{(j)}| \left|\langle\psi|\phi_r\rangle\right|^2$. Let $\widetilde{\Xi} = \sum_j p_j \|\mathbf{c}^{(j)}\|_1^2$, and let $D = \max\{(C_j - 1)/\|\mathbf{c}^{(j)}\|_1^2\}$. Then for any $p_{\text{fail}} > 0$, and $\delta \geq 24D$ there exists a classical algorithm that, with success probability $(1-p_{\text{fail}})$, samples a bit-string $\mathbf{x}$ of length w with probability $P_{\text{sim}}(\mathbf{x})$ such that:*

$$\|P_{\text{sim}} - P\|_1 \leq \delta + \mathcal{O}(\delta^2), \tag{5.105}$$

*where $P(\mathbf{x}) = \mathrm{Tr}(\Pi_{\mathbf{x}}\rho)$, and $\Pi_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w}$ is a projector. The algorithm returns $\mathbf{x}$ with random runtime T where the **average** runtime is*

$$\mathbb{E}(T) = \mathcal{O}(w^3 n^3 \widetilde{\Xi} \delta^{-3} \log(w/p_{\text{fail}})). \tag{5.106}$$

*If the decomposition of $\rho$ is optimal with respect to the definition* (4.2)*, then the expected runtime is $\mathcal{O}(\Xi(\rho))$. Moreover, if the state decomposition is equimagical, then the right side of* (5.106) *also bounds the **worst-case** runtime.*

*If arbitrary precision $\delta \leq 24D$ is required, this can be achieved at the cost of an increased runtime:*

$$\mathbb{E}(T) = \mathcal{O}(w^3 n^3 \widetilde{\Xi}(\delta^{-3} + 3D\delta^{-4}) \log(w/p_{\text{fail}})). \tag{5.107}$$

Finally we note that the method can be adapted to classically estimate Born rule probabilities. In this case, rather than drawing a single sparsified state $|\Omega\rangle\langle\Omega| / \langle\Omega|\Omega\rangle$ from the ensemble and computing a chain of conditional probabilities, we generate a large number of sparsifications $|\Omega\rangle$, compute $\|\Pi|\Omega\rangle\|^2/\||\Omega\rangle\|^2$, and take the mean, using the Hoeffding bound [128] to determine the number of samples needed to achieve given precision. For technical details see Appendix E.3. This procedure outputs an estimate $\eta$

satisfying

$$|\eta - \mathrm{Tr}[\Pi\rho]| \leq 3\varepsilon_{\mathrm{FN}}\mathrm{Tr}[\Pi\rho] + \delta + \mathcal{O}\big(\delta^2\big) \tag{5.108}$$

with probability $(1-p)$ in average runtime

$$\mathbb{E}(\tau) = \mathcal{O}\left(\Xi n^3 \varepsilon_{\mathrm{FN}}^{-2} \delta^{-3} \log\left(\frac{2}{p}\right)\right). \tag{5.109}$$

Note that the error in the estimate includes both a multiplicative error contribution due to fast norm estimation, and an additive error contribution arising from sparsification and Hoeffding sampling. While the linear scaling with $\Xi$ suggests the method might perform better than quasiprobability methods, whose performance scales with the square of the associated monotone, the scaling with $\delta$ is worse for the present method, so care must be taken in choosing the optimal strategy. We discuss this further in Section 6.3 of Chapter 6.

## 5.3 Algorithms for magic-generating channels

In this section we present classical simulation algorithms for non-stabiliser circuits, where the runtime is quantified by an associated channel magic monotone, so that the simulators always deal efficiently with completely stabiliser-preserving circuit elements. In all simulators considered here, we assume a circuit composed from a sequence of channels $\{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_L\}$ acting on an initial stabiliser state, which without loss of generality we take to be $|0^n\rangle$. The circuit ends with some final state $\rho = \mathcal{E}_L \ldots \circ \mathcal{E}_2 \circ \mathcal{E}_1(|0^n\rangle\langle 0^n|)$, and we simulate measurements performed on this final system state. We typically assume that each channel acts non-trivially on a bounded number of qubits (e.g. 2 or 3). This setting allows the (inefficient) classical simulation of universal quantum circuits without the need for cumbersome gadgetisation. Our algorithms are not restricted to discrete gate sets such as Clifford + T, and therefore admit more direct simulation of arbitrary non-Clifford gates, leading to reduced overhead in many cases.

We first introduce a class of simulators we call *static* simulators, as they make use of precomputed, and therefore static, quasiprobability distributions for each element of the circuit. Our first static simulator works by sampling a trajectory of CPTP maps in $\mathrm{SP}_{n,n}$ from a quasiprobability decomposition of a non-stabiliser circuit, and its runtime is directly related to the channel robustness of each circuit element. The constrained path channel simulator, dyadic channel simulator and stabiliser rank simulator for noisy rotations that follow have performance quantified by one of our measures of magic for channels, and employ related channel decompositions. In contrast, the class of simulation algorithms we call *dynamic* channel simulators recompute magic state decompositions at each step in the circuit, leading to subtle trade-offs in the runtime complexities. For example, the dynamic simulator associated with magic capacity (w.r.t. robustness of magic) can reduce sample complexity compared to the static channel simulator associated with channel robustness, at the cost of increased runtime per sample.

We will make repeated use of the stabiliser-Kraus subroutines described earlier (Section 5.1), which guarantee that simulation of circuits built from few-qubit stabiliser channels is efficient, as well as helping to reduce overhead for the simulation of non-stabiliser operations. For example, we will show that our quasiprobability simulators always perform at least as well as the Oak Ridge algorithm [45] in terms of sample complexity, and give examples of specific cases where we obtain a speedup. For most of the simulators we discuss, the precise variant of the stabiliser-Kraus subroutine used is unimportant; we simply assume that we have access to a procedure STABILISERKRAUS defined as follows.

**Definition 5.10** (Generic stabiliser-Kraus function)**.** *Let $\mathcal{E}(\cdot) = \sum_j K_j(\cdot)K_j^\dagger$ be a completely stabiliser-preserving CPTP map, specified by a simulable channel decomposition $\mathbb{L}$, where the $j$-th element represents the Kraus operator $K_j$. Then let* STABILISERKRAUS *be a procedure that takes as input a representation of a stabiliser state $|\phi\rangle$, and a channel decomposition $\mathbb{L}$ and probabilistically outputs a representation of an updated pure stabiliser state,*

$$|\phi'\rangle \leftarrow \textsc{stabiliserKraus}(|\phi\rangle, \mathbb{L}), \tag{5.110}$$

*such that with probability $p_j = \left\|K_j|\phi\rangle\right\|^2$ the updated state is chosen to be $|\phi'\rangle = K_j|\phi\rangle/\sqrt{p_j}$. The update is completed in* poly$(n)$ *time.*

### 5.3.1 Static channel simulator

Here we present the static channel simulator. This can be viewed as a generalisation of the Oak Ridge simulator [45] (Section 2.5.1), differing in two important ways; firstly in the set of operations treated as "free", and secondly in the way sampling over free operations is performed. Whereas the Oak Ridge algorithm employed the Cliffords and Pauli-reset channels (CPR) as free operations, in our simulator we optimise channel decompositions with respect to $\mathrm{SP}_{m,m}$, so that all stabiliser channels are represented non-negatively. This includes CPR as a strict subset. While the Oak Ridge simulator sampled over a discrete set of operations where each was itself a CPTP stabiliser channel, we first sample either the positive or negative path of the decomposition, before using the stabiliser-Kraus subroutine from Section 5.1 to sample individual Kraus operators.

We assume that each circuit element $\mathcal{E}_j$ in the circuit decomposition $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$ acts non-trivially on at most $m$ qubits, and is decomposed,

$$\mathcal{E}^{(j)} = (1 + p_j)\mathcal{E}_0^{(j)} - p_j\mathcal{E}_1^{(j)}, \quad \text{where} \quad \mathcal{E}_0^{(j)}, \mathcal{E}_1^{(j)} \in \mathrm{SP}_{m,m} \text{ and CPTP.} \tag{5.111}$$

We quantify the negativity of the decomposition by defining $Q_j = 1 + 2p_j$. Recall that the channel robustness $\mathcal{R}_*$ (Definition 4.18) is the optimal value of $Q_j$, that is, $\mathcal{R}_*(\mathcal{E}^{(j)}) = \min Q_j$. The runtime to precompute decompositions is bounded since we restrict $\mathcal{E}^{(j)}$ to be $m$-qubit circuit elements. We then assume that a description of each $\mathcal{E}_{0,1}^{(j)}$ is provided to the main simulator as a list $\mathbb{L}_{0,1}^{(j)}$, in one of the formats described in Section 5.1. One can define a composite quasiprobability distribution,

$$q_{\mathbf{k}} = \prod_{j:k_j=0} (1 + p_j) \prod_{j:k_j=1} (-p_j), \tag{5.112}$$

where $\mathbf{k} \in \mathbb{F}_2^L$ is a vector representing a choice of either $\mathcal{E}_0^{(j)}$ or $\mathcal{E}_1^{(j)}$ at each circuit element $\mathcal{E}_j$. We can renormalise this to obtain a product probability distribution.

$$p_{\mathbf{k}} = \prod_{j:k_j=0} \frac{(1 + p_j)}{Q_j} \prod_{j':k_{j'}=1} \frac{(p_{j'})}{Q_{j'}}. \tag{5.113}$$

Hence we can write the density operator for the final state of the circuit as follows,

$$\mathcal{E}(|\phi_0\rangle\langle\phi_0|) = Q \sum_{\mathbf{k}} p_{\mathbf{k}} \lambda_{\mathbf{k}} \mathcal{E}_{\mathbf{k}}(|\phi_0\rangle\langle\phi_0|) \tag{5.114}$$

where each $\mathcal{E}_{\mathbf{k}} = \mathcal{E}_{k_L}^{(L)} \circ \ldots \circ \mathcal{E}_{k_1}^{(1)}$ gives a trajectory of $\mathrm{SP}_{m,m}$ channels through the circuit, $\lambda_{\mathbf{k}} = \mathrm{sign}(q_{\mathbf{k}})$, and $Q = \prod_{j=1}^{T} Q_j$. As usual, we give pseudocode for the static channel simulator (Algorithm 12), before arguing for its validity and analysing runtime. The

**Algorithm 12** Static channel simulator

```
Input: Circuit description $\left\{\mathcal{E}^{(1)},\mathcal{E}^{(2)},\ldots,\mathcal{E}^{(T)}\right\}$, where each $\mathcal{E}^{(j)}$ has a known decomposition with $\ell_1$-norm $Q_j$ as per equations (5.111), and each stabiliser sub-channel $\mathcal{E}^{(j)}_{0,1} \in \mathrm{SP}_{m,m}$ is specified by a simulable channel decomposition $\mathbb{L}^{(j)}_{0,1}$; stabiliser observable $E$; initial stabiliser state $|\phi_0\rangle$; number of samples $M$.
Output: Estimate $\widehat{E}$ for expectation value $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle)]$.
Set $\widetilde{E} \leftarrow 0$ and $Q \leftarrow \prod_{j=1}^{T} Q_j$
for $i = 1$ to $M$ do
    Sample vector $\mathbf{k}$ according to the distribution $\{p_{\mathbf{k}}\}$.
    Prepare representation of initial state $|\phi_0\rangle$.
    for $j = 1$ to $T$ do
        $|\phi_j\rangle \leftarrow \textsc{StabiliserKraus}(|\phi_{j-1}\rangle, \mathbb{L}^{(j)}_{k_j})$
    end for
    $\widetilde{E}_i \leftarrow \mathrm{sign}(q_{\mathbf{k}})Q\langle\phi_T|E|\phi_T\rangle$
    $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_i$
end for
return $\widehat{E} \leftarrow \widetilde{E}/M$
```

expectation value $\langle E\rangle$ at the end of the circuit is given by

$$\mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)] = Q\sum_{\mathbf{k}} p_{\mathbf{k}}\lambda_{\mathbf{k}}\,\mathrm{Tr}[E\mathcal{E}_{\mathbf{k}}(|\phi_0\rangle\langle\phi_0|)]. \tag{5.115}$$

This decomposition yields a quasi-probability distribution with $\ell_1$-norm

$$||\mathbf{q}||_1 = \sum_{\mathbf{k}} |p_{\mathbf{k}}\lambda_{\mathbf{k}}Q| = \sum_{\mathbf{k}} |p_{\mathbf{k}}|Q = \prod_j Q_j. \tag{5.116}$$

where $\mathbf{q}$ is a $2^T$-element vector with entries $q_{\mathbf{k}}$. By sampling a vector $\mathbf{k}$ in step 3, we randomly choose a stabiliser-preserving trajectory through the circuit. Crucially, (5.113) gives a product distribution, so this first sampling step is efficient. If we could efficiently compute the mean value $\langle E\rangle_{\mathbf{k}} = \lambda_{\mathbf{k}}Q\mathrm{Tr}[E\mathcal{E}_{\mathbf{k}}(|\phi_0\rangle\langle\phi_0|)]$ for any given stabiliser sub-channel, we would have an unbiased estimator for $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$. The variance is increased by $Q > 1$, but by the standard arguments [45, 103, 124]any fixed additive error $\varepsilon > 0$ in the estimate could be achieved with high probability by repeating the sampling procedure $M = \mathcal{O}(\varepsilon^{-2}||\mathbf{q}||_1^2)$ times. However, computing each $\langle E\rangle_{\mathbf{k}}$ exactly cannot always be done efficiently in the depth of the circuit, since if each channel $\mathcal{E}^{(j)}_{k_j}$ has $N_K$ Kraus operators, then (without further knowledge of its structure), the Kraus decomposition of the trajectory $\mathcal{E}_{\mathbf{k}}$ has $N_K^T$ terms. So even when $N_K$ is bounded, the time to compute all contributions $\langle E\rangle_{\mathbf{k}}$ would be exponential in $T$.

Instead, the simulator tracks the evolution of a probabilistically updated pure stabiliser state through the sampled trajectory $\mathcal{E}_{\mathbf{k}}$. In each iteration of step 6, the algorithm selects the channel decomposition $\mathbb{L}_{k_j}^{(j)}$ for either the positive ($k_j = 0$) or negative ($k_j = 1$) part of the decomposition of non-stabiliser channel $\mathcal{E}^{(j)}$. This is passed to the STABILISERKRAUS subroutine, which then probabilistically selects a Kraus operator with which to update the stabiliser state. Because $\mathbb{L}_{k_j}^{(j)}$ represents a CPTP completely stabiliser-preserving map, it in effect has $\ell_1$-norm 1, so the internal sampling carried out by STABILISERKRAUS does not increase the variance, and the sample complexity of the simulator depends only on $Q = \prod_j Q_j$. The end result is that by combining the outer quasiprobability sampling routine, with the inner sampling routine, we have an unbiased estimator for $\langle E\rangle$. To generate a single sample, there are $T$ calls to STABILISERKRAUS. Since we stipulate that each $\mathcal{E}^{(j)}$ acts non-trivially on no more than $m$ qubits, the number of Kraus operators is bounded, so each call to STABILISERKRAUS completes in poly$(n)$ time. Therefore to generate $M$ samples the total runtime $\tau$ is

$$\tau = M\cdot T\cdot \text{poly}(n) = \left\lceil \frac{2}{\varepsilon}\|q\|_1^2 \ln\left(\frac{2}{p_{\text{fail}}}\right)\right\rceil \cdot T\cdot \text{poly}(n), \tag{5.117}$$

where $M$ is chosen to achieve additive error $\varepsilon$ with probability at least $1 - p_{\text{fail}}$, using a Hoeffding inequality in the usual way [124, 128]. Recalling that $\|q\|_1^2 = \prod_{j=1}^{T} Q_j$, and that $\mathcal{R}_*(\mathcal{E}^{(j)})$ optimises $Q_j$, we see that when we have access to optimal decompositions for each circuit element $\mathcal{E}^{(j)}$ the runtime is quantified by the channel robustness. Our arguments above lead to the following result.

**Theorem 5.11** (Static channel simulator)**.** *Suppose $\mathcal{E}$ is an n-qubit non-stabiliser circuit that can be decomposed as a sequence of circuit elements, $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$, where each $\mathcal{E}^{(j)}$ acts non-trivially on at most m qubits, so that the channel robustness can be computed. Then for any stabiliser state $|\phi_0\rangle$ and stabiliser observable E, and parameters $\delta, p_{\text{fail}} > 0$, there is a classical algorithm that outputs an estimate $\widehat{E}$ for the expectation value $\langle E\rangle = \text{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$ in time:*

$$\tau = \left\lceil \frac{2}{\varepsilon} \ln\left(\frac{2}{p_{\text{fail}}}\right)\right\rceil \cdot T\cdot \text{poly}(n)\cdot \prod_{j=1}^{T} [\mathcal{R}_*(\mathcal{E}^{(j)})]^2, \tag{5.118}$$

*where the estimate satisfies:*

$$\left|\widehat{E} - \langle E \rangle\right| \leq \varepsilon \tag{5.119}$$

*with probability at least* $1 - p_{\mathrm{fail}}$.

To compare our runtime with that of the Oak Ridge simulator [45], recall that their algorithm has cost function

$$\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}) = \min_{\Lambda_j \in \mathrm{CPR}} \left\{ ||p||_1 : \sum p_j \Lambda_j = \mathcal{E} \right\}, \tag{5.120}$$

so that simulating a given circuit element $\mathcal{E}^{(j)}$ contributes a factor $\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}^{(j)})^2$ to the runtime. Since $\mathrm{CPR} \subseteq \mathrm{SP}_{n,n}$, it must be the case that $\mathcal{R}_* \leq \mathcal{R}_{\mathrm{CPR}}$, potentially leading to lower simulation sample complexity if there exist channels with $\mathcal{R}_* < \mathcal{R}_{\mathrm{CPR}}$. We now give a toy example demonstrating a significant advantage to our simulator. Consider the single-qubit CPTP map $\mathcal{E}_H$ defined by a $Z$-measurement followed by a Hadamard gate conditioned on the "-1" outcome. This has Kraus representation $K_1 = |0\rangle\langle 0|, \quad K_2 = |-\rangle\langle 1|$. This is clearly a completely stabiliser-preserving map, so has channel robustness $\mathcal{R}_*(\mathcal{E}_H) = 1$. For a single qubit, CPR consists of the 24 Clifford gates, and 6 Pauli-reset channels. We find that $\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}_H) = 2$. Since $\Lambda_H \in \mathrm{SP}_{1,1}$, this confirms that CPR is a strict subset of the completely stabiliser-preserving channels, and indicates that $\mathcal{R}_{\mathrm{CPR}}$ is not a monotone under stabiliser operations. We also note that the calculated value is larger than the robustness of magic for any single-qubit state, despite $\Lambda_H$ being a stabiliser operation. For a circuit containing $T$ uses of the channel $\Lambda_H$, the samples required for a CPR simulator would be proportional to $\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}_H)^{2T} = 4^T$. Whereas, for our simulator $\mathcal{R}_*(\Lambda_H)^{2T} = 1$, so simulation of this circuit element is efficient.

While the above example is quite artificial, a reduction in sample complexity is also achieved for channels where $\mathcal{R}_{\mathrm{CPR}}(\mathcal{E}) > \mathcal{R}_*(\mathcal{E}) > 1$. Given a circuit with $L$ non-stabiliser elements $\mathcal{E}^{(j)}$ , the sample complexity for the CPR simulator would be proportional to $\prod_j^L \mathcal{R}_{\mathrm{CPR}}(\mathcal{E}^{(j)})^2$. Since $\mathcal{R}_*(\mathcal{E}^{(j)}) \leq \mathcal{R}_{\mathrm{CPR}}(\mathcal{E}^{(j)})$ for any channel, the sample complexity for our simulator will never be greater. While our simulator sometimes incurs a modest increase in the runtime per sample, this must be weighed against a reduction in runtime by a factor exponential in the number of circuit elements where $\mathcal{R}_*(\mathcal{E}^{(j)}) < \mathcal{R}_{\mathrm{CPR}}(\mathcal{E}^{(j)})$. The obvious next question is whether there are any natural non-trivial examples where this happens. In Chapter 6 we will show that gate sequences subject to amplitude-damping noise provide one such case.

Finding optimal CPR decompositions is only tractable for one- and two-qubit circuit elements, as the three-qubit case already involves a linear program with nearly 93 million variables [45]. For general channels, our simulator suffers from a similar problem, as three-qubit channel decompositions are optimised over six-qubit stabiliser states. However, the problem can be greatly simplified for diagonal channels (Section 4.4), and remains tractable for five-qubit operations. This allows our algorithm to take advantage of the submultiplicativity of channel robustness. When $\mathcal{R}(\mathcal{E}^{\otimes n}) < \mathcal{R}(\mathcal{E})^n$, it is advantageous to decompose the composite channel $\mathcal{E}^{\otimes n}$, provided the linear program is tractable. In Chapter 6 we consider this strategy for sequences of single-qubit $Z$-rotations.

### 5.3.2 Static constrained path channel simulator

We have seen that the gRoM $\Lambda^{+}(\rho)$ for states lower bounds the additive error sustained by a fast but noisy simulator (Section 5.2.1), which admits stabiliser circuits with magic state inputs. We now show how this constrained path simulator can be extended to admit magic channels, so that the performance is quantified by generalised *channel* robustness $\Lambda^{+}_{*}$. Recall that for a channel $\mathcal{E}$, this quantity is the minimum $\lambda \geq 1$ such that $\mathcal{E} = \lambda\mathcal{E}_{+} - (\lambda-1)\mathcal{E}_{-}$, where $\mathcal{E}_{+}$ is a CPTP, completely stabiliser-preserving map, whereas $\mathcal{E}_{-}$ can be non-stabiliser. Equivalently, in the Choi state picture, $\Lambda^{+}_{*}(\mathcal{E}) = \min\lambda : \Phi_{\mathcal{E}} \leq \lambda\sigma$, where $\sigma$ is a normalised stabiliser state such that $\mathrm{Tr}_A[\sigma] = \mathbb{1}_n/2^n$. We showed in Section 4.2.4 that given two channels $\mathcal{E}_1$ and $\mathcal{E}_2$, if feasible solutions for the above optimisation problem are given by $\Phi_{\mathcal{E}_1} \leq \lambda_1\sigma_1$ and $\Phi_{\mathcal{E}_2} \leq \lambda_2\sigma_2$, then a feasible solution for the composed channel $\mathcal{E} = \mathcal{E}_2 \circ \mathcal{E}_1$ is given by $\Phi_{\mathcal{E}} \leq \lambda_1\lambda_2\sigma'$ for some stabiliser state $\sigma'$ satisfying $\mathrm{Tr}_A[\sigma'] = \mathbb{1}_n/2^n$. This can be extended to a sequence of circuit elements of length $T$. That is, suppose we are given a circuit $\mathcal{E} = \mathcal{E}^{(j)} \circ \ldots \circ \mathcal{E}^{(1)}$ where for each $\mathcal{E}^{(j)}$ we have a known solution $\Phi_{\mathcal{E}^{(j)}} \leq \lambda_j\sigma_j$, such that $\sigma_j = (\mathcal{E}^{(j)}_{+} \otimes \mathbb{1}_n)(|\Omega_n\rangle\langle\Omega_n|)$ is the Choi state for some completely stabiliser-preserving channel $\mathcal{E}^{(j)}_{+}$. We called $\mathcal{E}^{(j)}_{+}$ the constrained sub-channel of $\mathcal{E}^{(j)}$. It follows that there exists a decomposition:

$$\mathcal{E} = \lambda\mathcal{E}_{+} - (\lambda-1)\mathcal{E}_{-} \tag{5.121}$$

$$\text{where} \quad \lambda = \prod_{j=1}^{T}\lambda_j, \quad \mathcal{E}_{+} = \mathcal{E}^{(T)}_{+} \circ \ldots \circ \mathcal{E}^{(1)}_{+}. \tag{5.122}$$

**Algorithm 13** Constrained path channel simulator

```
Input: Simulable constrained path decomposition $\mathbb{D} = \{(\lambda_j, \mathbb{L}^{(j)})\}_{j=1}^{T}$ for a circuit $\mathcal{E}$;
    initial stabiliser state $\sigma$; stabiliser observable $E$; number of samples $M$; parameters
    $c, p_{\text{fail}} > 0$.
Output: Estimate $\widehat{E}$ and error bound $\Delta$, s.t. $|\widehat{E} - \mathrm{Tr}(E\,|\phi_0\rangle\langle\phi_0|)| \leq \Delta$ with probability
    $1 - p_{\text{fail}}$.
 1: $\lambda \leftarrow \prod_{j=1}^{T} \lambda_j$
 2: $\delta \leftarrow c\lambda$
 3: $\mathbb{D}' \leftarrow \{\mathbb{L}^{(j)}\}_{j=1}^{T}$                ▷ Extract sequence of stabiliser channels.
 4: $E_\sigma \leftarrow$ STABILISERCIRCUIT$(\lambda\sigma, \mathbb{D}', E, \delta, p_{\text{fail}})$.
 5: $E_{\max} \leftarrow \min\{1, E_\sigma + \delta + \lambda - 1\}$
 6: $E_{\min} \leftarrow \max\{-1, E_\sigma - \delta - \lambda + 1\}$
 7: if $E_{\max} - E_{\min} < 2$ then
 8:     $\widehat{E} \leftarrow (E_{\max} + E_{\min})/2$
 9:     $\Delta \leftarrow (E_{\max} - E_{\min})/2$
10:     return $\widehat{E}, \Delta$
11: else
12:     return "FAIL"
13: end if
```

We can use this to extend the constrained path simulator for magic states to non-stabiliser channels. We say that the circuit $\mathcal{E} = \mathcal{E}_T \circ \ldots \mathcal{E}_1$ has a simulable constrained path decomposition $\mathbb{D}$ if there is a feasible solution for each $\mathcal{E}^{(j)}$, such that the constrained stabiliser sub-channel $\mathcal{E}_{+}^{(j)}$ has a simulable decomposition $\mathbb{L}_j$ as per Definition 5.1. The simulable constrained path decomposition can be specified as a list of coefficients paired with channel decompositions, $\mathbb{D} = \{(\lambda_j, \mathbb{L}_j)\}_{j=1}^{T}$. As usual, a simulable decomposition can always be found when each circuit element acts on a small bounded number of qubits. Pseudocode is given in Algorithm 13. In principle the initial state can be a magic state as in Algorithm 6, but for brevity we assume that the initial state is a stabiliser state. As in the original constrained path simulator, the runtime is dominated by the procedure STABILISERCIRCUIT in step 4. This subroutine uses the techniques of Section 5.1 to simulate the evolution of the stabiliser state $\sigma$ through the stabiliser circuit $\mathbb{D}'$, outputting an additive error estimate for the mean value of the observable $E$ with high probability. Since $\mathbb{D}'$ comprises a sequence of $T$ simulable decompositions of stabiliser channels, this subroutine completes in time $\tau = T \cdot 2\frac{1}{c^2}\log\left(\frac{2}{p_{\text{fail}}}\right) \cdot \mathrm{poly}(n)$.

Once again, the algorithm is efficient with respect to system size, and it is the error bound $\Delta$ that increases with magic. The analysis of the error bound is identical to that summarised in Table 5.1 for the constrained path simulator, except that here the variable $\lambda$ is the product of $\lambda_j$ for the composed channels (step 1), rather than arising from decomposition of the initial state. In the case where we are able to find an optimal decomposition for each circuit element, then $\lambda = \prod_{j=1}^{T} \Lambda_*^+(\mathcal{E}^{(j)})$, so that the accuracy of the estimator is governed by the generalised channel robustness. As mentioned in Section 4.2.4, where we defined $\Lambda_*^+$, late in the preparation of this thesis we became aware of complementary work on the generalisation of channel robustness by Saxena and Gour [159]. They also propose a constrained path channel simulator which differs in some details. In particular, we note that their algorithm may have wider application than the one presented here, as it allows precision to be improved at the cost of greater runtime.

### 5.3.3 Static dyadic channel simulator

The channel robustness monotone associated with the static channel simulator [1] introduced in Section 5.3.1 is closely related to RoM, so that the static simulator can be seen as the counterpart in the channel picture to the RoM simulator for magic states [103] (Section 2.3). By extending the frame to include non-Hermitian stabiliser dyads, the dyadic frame simulator of Section 5.2.2 was able to achieve the same simulation tasks as the RoM simulator with significantly reduced runtime [2], since by Theorem 4.16, we have $\mathcal{R}(\rho^{\otimes n})/\Lambda(\rho^{\otimes n}) \geq 2^{\gamma n}$ for some $\gamma > 0$ any single-qubit magic state $\rho$. We can improve on the static channel simulator in an analogous manner by expanding the set of free operations from the completely stabiliser-preserving channels $\mathrm{SP}_{n,n}$ to the dyadic stabiliser channels $\mathrm{DSP}_{n,n}$ introduced in the previous chapter (see Definition 4.23 in Section 4.2.5). In practice, this strategy may be limited by the fact that it is not clear that computing the dyadic channel negativity exactly is tractable. Nevertheless, we showed in Section 4.2.5 that it is at least practical to optimise over the projective dyadic stabiliser channels PDSP, a subset of DSP, for the single-qubit case. Thus by supplementing CPTP stabiliser channels with single-qubit PDSP maps, there is potential to reduce simulation costs for some channels.

We must first make a further extension to the stabiliser-Kraus subroutine (Algorithm 14). We assume that the description of the dyadic map is provided analogously to the simulable channel decompositions we have discussed previously, except that instead of a single list of Kraus operators, the map is represented by a pair $(\mathbb{L}, \mathbb{R})$, where $\mathbb{L}$ specifies the Kraus operators that act on the left and $\mathbb{R}$ describes those that act on the right. The entries of each list have the format $(\alpha_j, U_j, \Pi_j, h_j)$, where $2^{h_j/2} U_j \Pi_j$ is the polar decomposition of a stabiliser Kraus operator (see Section 5.1), and $\alpha_j$ is a complex coefficient. We call $(\mathbb{L},\mathbb{R})$ a simulable decomposition if the list has at most poly$(n)$ entries. As in

**Algorithm 14** Dyadic stabiliser channel update subroutine

**Input:** Initial stabiliser dyad $\sigma = |\phi_L\rangle\langle\phi_R|$; pair of lists $(\mathbb{L},\mathbb{R})$, representing a map in DSP$_{n,n}$, where $\mathbb{L}$ is of length $N_K$ and has entries $(\alpha_j^{(L)}, U_j^{(L)}, \Pi_j^{(L)}, h_j^{(L)})$ representing polar-decomposed stabiliser Kraus operators $\alpha_j^{(L)}K_j^{(L)} = \alpha_j^{(L)}2^{h_j^{(L)}/2}U_j^{(L)}\Pi_j^{(L)}$, and similarly for $\mathbb{R}$.

**Output:** Updated dyad $\sigma' = |\phi'_L\rangle\langle\phi'_R|$.

```
1:  function DYADICMAPUPDATE($\sigma$,$(\mathbb{L},\mathbb{R})$)
2:      for $j \leftarrow 1$ to $N_K$ do
3:          $P_j \leftarrow 2^{(h_j^{(L)}+h_j^{(R)})/2}\|\alpha_j^{(L)}\alpha_j^{(R)}\Pi_j^{(L)}\sigma\Pi_j^{(R)\dagger}\|_1$
4:      end for
5:      $P_0 \leftarrow 1 - \sum_{r=1}^{N_K} P_r$
6:      Sample $s$ from $\{0,\dots,N_K\}$ with probability $P_s$
7:      if s = 0 then
8:          $\sigma' \leftarrow 0$
9:      else
10:         $\sigma' \leftarrow |\phi'_L\rangle\langle\phi'_R| = 2^{(h_j^{(L)}+h_j^{(R)})/2}\alpha_j^{(L)}\alpha_j^{(R)*}U_s^L\Pi_s^{(L)}|\phi_L\rangle\langle\phi_R|\Pi_s^{(R)\dagger}U_s^{(R)\dagger}$
11:     end if
12:     return $\sigma'$
13: end function
```

Section 4.2.5, the transition probabilities in step 3 can be evaluated by computing norms for each side of the dyad separately,

$$\|\alpha_j^{(L)}\alpha_j^R\Pi_j^{(L)}\sigma\Pi_j^{(R)\dagger}\|_1 = |\alpha_j^{(L)}\alpha_j^{(R)}|\cdot\left\|\Pi_j^{(L)}|\phi_L\rangle\right\|\cdot\left\|\Pi_j^{(R)}|\phi_R\rangle\right\|. \tag{5.123}$$

This is done in polynomial time using standard stabiliser techniques. Meanwhile the phase-sensitive state update in step 10 can be carried out efficiently using the CH-simulator (Section 1.2.1) [3]. The argument to check that the computed values $\{P_r\}_{r=0}^{N_K}$ form a proper probability distribution is virtually identical to that given in the proof of Theorem 5.5. We can define real-valued vectors $\mathbf{Q}^{(X)}$, where $X = L$ or $X = R$, with entries $Q_r^{(X)} = 2^{h_r^{(X)}/2}\left\|\alpha_r^{(X)}\Pi_r^{(X)}|\phi_X\rangle\right\|$. Then, by virtue of the definition of complete dyadic stabiliser channels (Definition 4.23 in Section 4.2.5), it is always the case that $\sum_{r=1}^{N_K} 2^{h_r^{(X)}}\left\|\alpha_r^{(X)}\Pi_r^{(X)}|\phi_X\rangle\right\|^2 = 1$. Then by the same chain of reasoning as in equations

(5.35) to (5.37), we have that $\left\|\mathbf{Q}^{(X)}\right\| = 1$ and,

$$\sum_{r=1}^{N_K} P_r \leq \left\|\mathbf{Q}^{(L)}\right\| \cdot \left\|\mathbf{Q}^{(R)}\right\| \leq 1. \tag{5.124}$$

Then since $P_0 = 1 - \sum_{r=1}^{N_K} P_r$ by definition, $\{P_r\}_{r=0}^{N_K}$ is a valid probability distribution.

We next modify the static channel simulator of Ref. [1] to admit decompositions over dyadic channels. Suppose we have a circuit $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \mathcal{E}^{(1)}$. For the static channel simulator, we split the quasiprobability decomposition for each circuit element $\mathcal{E}^{(t)}$ into a positive and negative part, each of which constituted a CPTP stabiliser channel (equation (5.111)). Here the coefficients in the decomposition are complex, so there may be more than two complete dyadic stabiliser channels in the decomposition,

$$\mathcal{E}^{(t)} = \sum_k \beta_k^{(t)} \mathcal{T}_k^{(t)} \quad \text{where} \quad \mathcal{T}_k^{(t)} \in \mathrm{DSP}_{n,n}, \quad B^{(t)} = \|\beta^{(t)}\|_1. \tag{5.125}$$

Recall that the dyadic channel negativity $\Lambda_*(\mathcal{E}^{(t)})$ is the minimal $\ell_1$-norm $\|\beta^{(t)}\|_1$ over all valid decompositions (Definition 4.24). We present the pseudocode in Algorithm 15. This algorithm yields an unbiased estimator for $\langle E \rangle$. Since the proof repeats arguments made for previous simulators, and involves some lengthy algebra, we omit it here, giving full technical details in Appendix F. The performance of the simulator can be stated in the following theorem, proved in the appendix.

**Theorem 5.12** (Dyadic channel simulator). *Suppose an n-qubit non-stabiliser circuit $\mathcal{E}$ has decomposition $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$, where each circuit element $\mathcal{E}^{(t)}$ has a known decomposition into simulable dyadic stabiliser channels as per Definition 4.23, with $\ell_1$-norm $B^{(t)}$. Then for any initial stabiliser state $|\phi^{(0)}\rangle\langle\phi^{(0)}|$, stabiliser observable E and any constants $\varepsilon, p_{\text{fail}} > 0$, an estimate $\widehat{E}$ for the mean value $\langle E \rangle = \mathrm{Tr}\left[E\mathcal{E}(|\phi^{(0)}\rangle\langle\phi^{(0)}|)\right]$ can be computed in time*

$$\tau = \frac{2}{\varepsilon} \ln\left(\frac{2}{p_{\text{fail}}}\right) \cdot T \cdot \mathrm{poly}(n) \cdot \prod_{t=1}^{T} [B^{(t)}]^2, \tag{5.126}$$

*such that with probability at least $1 - p_{\text{fail}}$, we obtain $\left|\widehat{E} - \langle E \rangle\right| \leq \varepsilon$. When the decomposition for each $\mathcal{E}^{(t)}$ is optimal with respect to the dyadic channel negativity, the runtime is $\mathcal{O}\left(\prod_t [\Lambda_*(\mathcal{E}^{(t)})]^2\right)$.*

**Algorithm 15** Static dyadic channel simulator

**Input:** Circuit description $\{\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \ldots, \mathcal{E}^{(T)}\}$, where each $\mathcal{E}^{(t)}$ has known decomposition with $\ell_1$-norm $B^{(t)}$ as per equation (5.125), with each dyadic stabiliser subchannel $\mathcal{T}_k^{(j)} \in DSP$ is specified by simulable decomposition $(\mathbb{L}_k^{(t)}, \mathbb{R}_k^{(t)})$; stabiliser observable $E$; initial stabiliser state $\sigma^{(0)} = |\phi^{(0)}\rangle\langle\phi^{(0)}|$; number of samples $M$.

**Output:** Estimate $\widehat{E}$ for expectation value $\langle E\rangle = \mathrm{Tr}\Big[E\mathcal{E}(|\phi^{(0)}\rangle\langle\phi^{(0)}|)\Big]$.

```
Set $\widetilde{E} \leftarrow 0$, $\theta \leftarrow 0$, $B \leftarrow \prod_{t=1}^{T} B^{(t)}$.        ▷ Initialise phase angle $\theta$.
for $j = 1$ to $M$ do
    Prepare representation of initial state $\sigma^{(0)}$.
    for $t = 1$ to $T$ do
        Sample $k_t$ with probability $|\beta_{k_t}^{(t)}|/B^{(t)}$.
        $\theta \leftarrow \theta + \arg\beta_{k_t}^{(t)}$        ▷ Update phase angle.
        $\sigma^{(t)} \leftarrow \textsc{dyadicMapUpdate}(\sigma^{(t-1)}, (\mathbb{L}_{k_t}^{(t)}, \mathbb{R}_{k_t}^{(t)}))$
        if $\sigma^{(t)} = 0$ then        ▷ Terminate trajectory if "zero" selected.
            $\sigma^{(T)} \leftarrow 0$
            break
        end if
    end for
    $\widetilde{E}_j \leftarrow \mathrm{Re}\Big\{Be^{i\theta}\,\mathrm{Tr}\Big[E\sigma^{(T)}\Big]\Big\}$        ▷ Computed with CH-simulator.
    $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_j$
end for
return $\widehat{E} \leftarrow \widetilde{E}/M$
```

### 5.3.4 Sampling simulator for non-stabiliser channels

We next sketch how the mixed-state stabiliser rank simulation techniques developed in Section 5.2.3 can be extended to simulate noisy unital channels, such as non-Clifford gates subject to dephasing or depolarising noise. Since the bit-string simulator for magic states was described in detail in Section 5.2.3 and Appendix E.2, we do not give full pseudocode. Instead, we modify the sparsification method to deal with this simulation setting, and then outline how the simulator proceeds.

Suppose that the system is initialised with a stabiliser state $|\phi_0\rangle$, and the circuit comprises $T$ channels $\mathcal{E} = \mathcal{E}^{(T)} \circ \mathcal{E}^{(T-1)} \circ \ldots \circ \mathcal{E}^{(1)}$, so that the final state is $\rho = \mathcal{E}(|\phi_0\rangle\langle\phi_0|)$. We assume that for each circuit element $\mathcal{E}^{(t)}$ we have a known decomposition:

$$\mathcal{E}^{(t)} = \sum_j p_{t,j}\mathcal{U}_{t,j}, \quad \mathcal{U}_{t,j}(\cdot) = U_{t,j}(\cdot)U_{t,j}^{\dagger}, \tag{5.127}$$

where each $U_{t,j}$ is a non-Clifford gate, with known sum-over-Clifford decomposition $U_{t,j} = \sum_k c_{t,j,k}V_{t,j,k}$ for Clifford gates $V_{t,j,k}$. We will assume that that the Clifford decomposition of $U_{t,j}$ is optimal with respect to unitary extent, so that $\xi(U_{t,j}) = \|\mathbf{c}_{t,j}\|_1^2$, where $\mathbf{c}_{t,j}$ is a vector with elements $c_{t,j,k}$. Then the final state of the circuit can be written as an ensemble over pure states, $\rho = \sum_{\mathbf{j}} p_{\mathbf{j}} |\psi_{\mathbf{j}}\rangle\langle\psi_{\mathbf{j}}|$, where $\mathbf{j}$ is a vector of indices $j_t$, such that $p_{\mathbf{j}} = p_{T,j_T} p_{T-1,j_{T-1}} \cdots p_{1,j_1}$ and

$$\left|\psi_{\mathbf{j}}\right\rangle = U_{T,j_T} U_{T-1,j_{T-1}} \ldots U_{1,j_1} |\phi_0\rangle. \tag{5.128}$$

In effect, the noisy circuit $\mathcal{E}$ is decomposed as a probabilistic mixture of ideal non-Clifford unitary circuits. In turn, using the Clifford decomposition of each $U_{t,j_t}$, we can express each pure state as a superposition of stabiliser states, $\left|\psi_{\mathbf{j}}\right\rangle = \sum_{\mathbf{k}} c_{\mathbf{j},\mathbf{k}} |\phi_{\mathbf{k}}\rangle$, where $|\phi_{\mathbf{k}}\rangle = V_{T,j_T,k_T} \ldots V_{1,j_1,k_1} |\phi_0\rangle$, and $c_{\mathbf{j},\mathbf{k}} = c_{T,j_T,k_T} \ldots c_{1,j_1,k_1}$. By Theorem 5.6, each pure state $\left|\psi_{\mathbf{j}}\right\rangle\!\left\langle\psi_{\mathbf{j}}\right|$ can be approximated up to $\mathcal{O}(\delta_S)$ error in the trace-norm by an ensemble $\rho_{\mathbf{j}}$ of random states $|\Omega\rangle\langle\Omega| \,/\, \langle\Omega|\Omega\rangle$, subject to the caveats on the size of $\delta_S$ explained in Section 5.2.3.1. The random vectors $|\Omega\rangle$ are defined by

$$|\Omega\rangle = \frac{\|\mathbf{c}_{\mathbf{j}}\|_1}{M} \sum_{\alpha=1}^{M} |\omega_\alpha\rangle, \quad \text{where} \quad \Pr\left\{|\omega_\alpha\rangle = c_{\mathbf{j},\mathbf{k}} |\phi_{\mathbf{k}}\rangle / |c_{\mathbf{j},\mathbf{k}}|\right\} = \frac{|c_{\mathbf{j},\mathbf{k}}|}{\|\mathbf{c}_{\mathbf{j}}\|_1}. \tag{5.129}$$

Here $\mathbf{c}$ is the vector with elements $c_{\mathbf{j},\mathbf{k}}$, and $M = \lceil 4\|\mathbf{c}_{\mathbf{j}}\|_1^2/\delta_S \rceil$. But each coefficient factorises as $c_{\mathbf{j},\mathbf{k}} = c_{T,j_t,k_T} \cdots c_{1,j_1,k_1}$, so we can factorise each $\ell_1$-norm as

$$\|\mathbf{c}_{\mathbf{j}}\|_1 = \sum_{\mathbf{k}} |c_{\mathbf{j},\mathbf{k}}| = \sum_{k_T,\ldots,k_1} \prod_{t=1}^{T} |c_{t,t_t,k_t}| = \prod_{t=1}^{T} \sqrt{\xi(U_t)}. \tag{5.130}$$

Moreover, the probability distribution used to sample the sparsified vectors $|\Omega\rangle$ is a product distribution,

$$\Pr\left\{|\omega_\alpha\rangle = c_{\mathbf{j},\mathbf{k}} |\phi_{\mathbf{k}}\rangle / |c_{\mathbf{j},\mathbf{k}}|\right\} = \frac{|c_{\mathbf{j},\mathbf{k}}|}{\|\mathbf{c}_{\mathbf{j}}\|_1} = \prod_{t=1}^{T} \frac{\left|c_{t,j_t,k_t}\right|}{\sqrt{\xi(U_{t,j})}}. \tag{5.131}$$

We can then approximate the final state of the circuit $\rho$ by an ensemble $\rho'$ defined

$$\rho' = \sum_{\mathbf{j}} p_{\mathbf{j}} \rho_{\mathbf{j}}, \quad \rho_{\mathbf{j}} = \sum_{\Omega} \Pr\left(\Omega|\psi_{\mathbf{j}}\right) \frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle} \tag{5.132}$$

satisfying $\|\rho - \rho'\|_1 \leq \delta_S + \mathcal{O}\left(\delta_S^2\right)$. Here $\Pr\left(\Omega|\psi_{\mathbf{j}}\right)$ is the probability of obtaining $|\Omega\rangle$ from the SPARSIFY subroutine given target pure state $\left|\psi_{\mathbf{j}}\right\rangle$. We simulate sampling a bit string from the output distribution of the circuit as follows.

1. Sample a unitary non-Clifford trajectory $\mathbf{j}$ from the product distribution.

2. Randomly compute a sparsified, unnormalised state vector $|\Omega\rangle$.

3. Use repeated calls to FASTNORM to sample one bit at a time as per Section 5.2.3 and Appendix E.2.

The rest of the analysis closely follows that for the mixed-state stabiliser rank simulator, so that given an $n$-qubit circuit, to sample a bit string of length $w$ with failure probability at most $p_{\text{fail}}$, the average runtime $\mathbb{E}(\tau)$ is given by

$$\mathbb{E}(\tau) = \mathcal{O}(w^3 n^3 \widetilde{\Xi}(\delta^{-3}) \log(w/p_{\text{fail}})), \tag{5.133}$$

$$\text{where} \quad \widetilde{\Xi} = \sum_{\mathbf{j}} p_{\mathbf{j}} \|\mathbf{c}_{\mathbf{j}}\|_1^2 = \sum_{\mathbf{j}} p_{\mathbf{j}} \prod_{t=1}^{T} \xi(U_{t,j_t}). \tag{5.134}$$

But $\{p_{\mathbf{j}}\}$ is a product distribution, and factorising each $p_{\mathbf{j}}$ yields $\widetilde{\Xi} = \prod_{t=1}^{T} \widetilde{\Xi}_t$, where $\widetilde{\Xi}_t = \sum_{j_t} p_{t,j_t} \xi(U_{t,j_t})$. So $\widetilde{\Xi}_t$ represents the multiplicative cost factor for simulating each non-stabiliser channel $\mathcal{E}^{(t)}$ in the sequence. This simulation technique depends on the existence and computability of decompositions $\mathcal{E}^{(t)} = \sum_j p_{t,j} \mathcal{U}_{t,j}$. Clearly such decompositions do not exist for non-unital channels, for example. However in Chapter 6 we show that computing optimal decompositions is practical for sequences of single-qubit rotations subject to commonly used unital noise models. For such optimal decompositions, $\widetilde{\Xi}_t$ is equal to the channel extent defined in Section 4.2.3.

### 5.3.5 Dynamic channel simulator

We now introduce a class of algorithmss we call *dynamic* simulators. In contrast to the static channel simulator, where all decompositions are assumed to have been pre-computed, dynamic simulators compute convex optimisations on-the-fly for few-qubit subsystems as the algorithm steps through the circuit. The additional overhead required to solve these intermediate optimisations can sometimes be traded with runtime savings elsewhere. We illustrate this idea with a simulator related to the magic capacity with respect to RoM, $\mathcal{C}_{\mathcal{R}}$. Consider the setting where we have an $n$-qubit circuit where each non-stabiliser circuit element acts on at most $m$ qubits, where $m \ll n$. Suppose we partition the circuit into an $m$-qubit subsystem $A$ on which a channel $\mathcal{E}$ acts non-trivially, and subsystem $B$ containing the other $n-m$ qubits. Naively it is not tractable to update the quantum state using matrix multiplication, since qubits in $A$ could be entangled with those $B$. So, without prior knowledge of the entanglement structure, we have to apply the channel to the whole $n$-qubit system, which takes time exponential in $n$. To overcome this difficulty, we use the fact that any bipartite stabiliser state is always local Clifford-equivalent to some number $0 \leq b \leq m$ of independent Bell pairs, tensor product with a separable state [57], where $m$ is the number of qubits in the smaller partition. Given an $m$-qubit circuit element, an $n$-qubit initial stabiliser state can always be efficiently manipulated so that the state update can be computed on a subsystem of at most $2m$ qubits. After computing a decomposition for the new $2m$-qubit magic state, we can sample a new pure stabiliser state and move to the next circuit element. In Section 6.1 of Chapter 6, we will show that there exist few-qubit channels $\mathcal{E}$ such that $\mathcal{C}_{\mathcal{R}}(\mathcal{E}) < \mathcal{R}_*(\mathcal{E})$. For such channels, and for any stabiliser state $\rho$, the robustness of the output state $\mathcal{E}(\rho)$ will always be less than the $\ell_1$-norm of the decomposition of $\mathcal{E}$ into stabiliser-preserving CPTP channels. Our dynamic simulator takes advantage of this.

We will shortly present pseudocode, but first make several definitions. First, let $\mathbb{Q} = \{(q_r, \phi_r)\}_r$ denote a stabiliser decomposition of a magic state $\rho = \sum_r q_r |\phi_r\rangle\langle\phi_r|$. Then, let $||\mathbb{Q}||_1$ denote the $\ell_1$-norm $||\mathbf{q}||_1$ for the vector of coefficients $q_r$. Let OPTIMISE be a function that takes as input a density matrix $\rho$ and outputs an optimal stabiliser decomposition $\mathbb{Q}$ and its corresponding norm $||\mathbb{Q}||_1$,

$$(\mathbb{Q}, ||\mathbb{Q}||_1) = \text{OPTIMISE}(\rho), \quad \text{where} \quad ||\mathbb{Q}||_1 = \mathcal{R}(\rho). \tag{5.135}$$

Then let QUASISAMPLE be a function that randomly selects a pure stabiliser $|\psi\rangle$ from $\mathbb{Q}$ with the appropriate weighting, as well as tracking the sign:

$$(|\psi\rangle, \lambda) \leftarrow \text{QUASISAMPLE}(\mathbb{Q}) \quad \text{where} \quad \Pr(|\psi\rangle \leftarrow |\phi_r\rangle) = \frac{|q_r|}{||Q||_1}, \quad \lambda = \text{sign}(q_r) \tag{5.136}$$

Fattal et al. showed in Ref. [57] that any bipartite stabiliser state $|\phi\rangle^{A|B}$ can always be transformed by local Clifford gates $V^A \otimes \mathbb{1}^B$ and $\mathbb{1}^A \otimes W^B$ into $b$ independent Bell pairs, entangled across the partition $A|B$. If the partition $A$ comprises $m$ qubits, then $b$ can be at most $m$. It follows that given a tripartite state $|\phi\rangle^{ABC}$, we can always find some Clifford local to subsystem $BC$ such that subsystem $AB$ is disentangled from $C$, i.e. there exists a Clifford gate $\mathbb{1}^A \otimes U^{BC}$ such that,

$$\mathbb{1}^A \otimes U^{BC} |\phi\rangle^{ABC} = |\phi'\rangle^{AB} \otimes |\phi''\rangle^{C} \tag{5.137}$$

Since the unitary acts as the identity on subsystem $A$, it commutes with any channel of the form $\mathcal{E}^A \otimes \mathbb{1}^{BC}$. So, applying the Clifford $U$, followed by the channel $\mathcal{E}^A \otimes \mathbb{1}^{BC}$ and finally applying the inverse $U^\dagger$ is equivalent to applying the channel alone,

$$(\mathbb{1}^A \otimes U^{\dagger BC}) \circ (\mathcal{E}^A \otimes \mathbb{1}^{BC}) \circ (\mathbb{1}^A \otimes U^{BC}) = \mathcal{E}^A \otimes \mathbb{1}^{BC}. \tag{5.138}$$

We can take advantage of this to apply the non-stabiliser channel to the disentangled $2m$-qubit tensor factor instead of the full $n$-qubit state. We can then decompose the updated state and sample a new pure stabiliser state from the quasiprobability distribution, before re-entangling with the global system.

We need to be able to appropriately partition the system prior to applying this disentangling Clifford. We specify each subsystem $S$ by $n$-bit string $\mathbf{s}$, where $a_k = 1$ if the $k$-th qubit is in $A$, and $a_k = 0$ otherwise. Suppose $\mathbf{a}$ enumerates the $m$ qubits on which the channel $\mathcal{E}$ acts non-trivially. Let $B$ be the subsystem containing the first $m$ qubits not in $A$, and let $b(\mathbf{a})$ be the string describing this subsystem. Then let $\mathbf{o} = (1, \ldots, 1)$ be the string where every entry is 1, and let $c(\mathbf{a}) = \mathbf{a} \oplus b(\mathbf{a}) \oplus \mathbf{o}$, so that $c(\mathbf{a})$ specifies the remaining $n-2m$ qubits. The results of Fattal et al. [57] show that an appropriate Clifford can be found in poly$(n)$ time. Define DISENTANGLE to be the classical procedure that performs this task, taking as input the stabiliser state $|\phi\rangle^{ABC}$ and vectors specifying the choice of partition $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and returning the disentangling Clifford and the the tensor factors of the disentangled state,

$$(|\phi'\rangle^{AB}, |\phi''\rangle^{C}, U^{BC}) \leftarrow \text{DISENTANGLE}(|\phi\rangle^{ABC}, \mathbf{a}, \mathbf{b}, \mathbf{c}). \tag{5.139}$$

In Algorithm 16 we present pseudocode for the dynamic channel simulator. As for the static channel simulator, we represent a sampled trajectory through the circuit by a vector $\mathbf{k}$, such that the output of the true quantum circuit would be $\rho = \sum_{\mathbf{k}} q_{\mathbf{k}} \sigma_{\mathbf{k}}$. The major difference is that $q_{\mathbf{k}}$ cannot be decomposed as a product distribution, as the quasiprobabilities for each intermediate decomposition will depend on the stabiliser state

**Algorithm 16** Dynamic channel simulator

```
Input: Circuit description $\left\{\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \ldots, \mathcal{E}^{(T)}\right\}$, where each $\mathcal{E}^{(j)}$ acts non-trivially on
    a partition comprising at most $m$ qubits, specified by a binary vector $\mathbf{a}^{(j)}$ ; stabiliser
    observable $E$; initial stabiliser state $|\phi_0\rangle$; number of samples $M$.
Output: Estimate $\widehat{E}$ for expectation value $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$.
 1: Set $\widetilde{E} \leftarrow 0$.
 2: for $i = 1$ to $M$ do
 3:     Prepare representation of initial state $|\phi_0\rangle$.
 4:     $Q \leftarrow 1$                    ▷ This variable tracks the sign and overhead due to sampling
 5:     for $j = 1$ to $T$ do
 6:         $\mathbf{a}, \mathbf{b}, \mathbf{c} \leftarrow \mathbf{a}^{(j)}, b(\mathbf{a}^{(j)}), c(\mathbf{a}^{(j)})$.                    ▷ Specify qubit partitions
 7:         $(|\phi'_{j-1}\rangle^{AB}, |\phi''_{j-1}\rangle^{C}, U^{BC}) \leftarrow$ DISENTANGLE$(\left|\phi_{j-1}\right\rangle^{ABC}, \mathbf{a}, \mathbf{b}, \mathbf{c})$.
 8:         Compute $\rho^{AB}_{j-1} \leftarrow (\mathcal{E}^{(j)} \otimes \mathbb{1}_m)(|\phi'_{j-1}\rangle\langle\phi'_{j-1}|^{AB})$
 9:         $(\mathbb{Q}_j, \|\mathbb{Q}_j\|_1) \leftarrow$ OPTIMISE$(\rho^{AB}_{j-1})$
10:         $(|\phi'_j\rangle^{AB}, \lambda_j) \leftarrow$ QUASISAMPLE$(\mathbb{Q}_j)$
11:         $\left|\phi_j\right\rangle^{ABC} = (\mathbb{1}_m \otimes U^{\dagger BC}) |\phi'_j\rangle^{AB} \otimes |\phi''_{j-1}\rangle^{C}$                    ▷ Undo the Clifford gate
12:         $Q \leftarrow Q \times \|\mathbb{Q}_j\|_1 \times \lambda_j$
13:     end for
14:     $\widetilde{E}_i \leftarrow Q\langle\phi_T|E|\phi_T\rangle$
15:     $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_i$
16: end for
17: return $\widehat{E} \leftarrow \widetilde{E}/M$
```

sampled in the previous step. Consider the $j$-th circuit element. After step 7 of the algorithm, we have some randomly selected stabiliser state $\sigma_{\mathbf{k}_{j-1}}$, where $\mathbf{k}_{j-1}$ labels the trajectory prior to the $j$-th channel $\mathcal{E}^{(j)}$. We do not assume $\mathbf{k}_j$ is a binary vector; instead its elements label the pure stabiliser states sampled in each step. After steps 8 and 9 we have a magic state $\rho_{\mathbf{k}_{j-1}} = (\mathcal{E}^A_j \otimes \mathbb{1}^{BC})(\sigma_{\mathbf{k}_{j-1}})$, up to a Clifford rotation on subsystem $BC$. This state is decomposed as,

$$\rho_{\mathbf{k}_{j-1}} = (\mathcal{E}^A_j \otimes \mathbb{1}^{BC})(\sigma_{\mathbf{k}_{j-1}}) = \sum_{k_j} q_{\mathbf{k}_j} \sigma_{\mathbf{k}_j}. \tag{5.140}$$

We need to take care here which trajectories are included in the summation. The initial state that is input to the channel is fixed by the vector $\mathbf{k}_{j-1}$. So by by summing over the final index $k_j$, we sum over all $j$-step trajectories $\mathbf{k}_j = (\mathbf{k}_{j-1}, k_j)$ consistent with the previous $(j-1)$-step trajectory labelled by $\mathbf{k}_{j-1}$. Let $\|\mathbb{Q}_{\mathbf{k}_{j-1}}\|_1 = \sum_{k_j} \left|q_{\mathbf{k}_j}\right|$ be the $\ell_1$-norm for the computed decomposition of $\rho_{\mathbf{k}_{j-1}}$. We assume that each circuit element acts on sufficiently few qubits that computing the optimal decomposition is tractable, so that $||\mathbb{Q}_{\mathbf{k}_{j-1}}||_1 = \mathcal{R}(\rho_{\mathbf{k}_{j-1}})$. Then in steps 10 and 12 we sample a stabiliser state $\sigma_{\mathbf{k}_j}$ with probability $|q_{\mathbf{k}_j}|/||\mathbb{Q}_{\mathbf{k}_{j-1}}||_1$, and the variable $Q$ picks up a factor commensurate with norm $||\mathbb{Q}_{\mathbf{k}_{j-1}}||_1$ and the sign of the sampled element.

Now, compare this with the true evolution of the quantum circuit. The final state of system prior to measurement may be expressed

$$\rho = \mathcal{E}^{(T)} \circ \mathcal{E}^{(T-1)} \circ \ldots \circ \mathcal{E}^{(1)}(|\phi_0\rangle\langle\phi_0|), \tag{5.141}$$

where for clarity we omit the the tensor product and partition notation (i.e. $\mathcal{E}^{(j)}$ is shorthand for $\mathcal{E}^{(j)} \otimes \mathbb{1}^{BC}_{n-m}$). Using the vector notation defined above, we can write $\sigma_{\mathbf{k}_0} = |\phi_0\rangle\langle\phi_0|$, since all trajectories must start with the specified initial state. Then from the relation (5.140), we can rewrite the state after the first circuit element as a quasiprobability distribution optimal with respect to robustness of magic, so that.

$$\rho = \sum_{k_1} q_{\mathbf{k}_1} \mathcal{E}^{(T)} \circ \mathcal{E}^{(T-1)} \circ \ldots \circ \mathcal{E}^{(2)} \sigma_{\mathbf{k}_1}. \tag{5.142}$$

Note that the decomposition $(\mathcal{E}^{(1)} \otimes \mathbb{1})(\sigma_{\mathbf{k}_0}) = \sum_{k_1} q_{\mathbf{k}_1} \sigma_{\mathbf{k}_1}$ is simply a rewriting of the quantum state after the first channel in the circuit has been applied, it is not an artefact of our algorithm. Applying this rewrite recursively to all circuit elements,

$$\rho = \sum_{k_1} \cdots \sum_{k_T} (\prod_{j=1}^{T} q_{\mathbf{k}_j}) \sigma_{\mathbf{k}_T} = \sum_{\mathbf{k}} q_{\mathbf{k}} \sigma_{\mathbf{k}}, \tag{5.143}$$

where $q_{\mathbf{k}} = (\prod_{j=1}^{T} q_{\mathbf{k}_j})$ and $\sigma_{\mathbf{k}} = \sigma_{\mathbf{k}_T}$. Normalising each coefficient $q_{\mathbf{k}_j}$ with the appropriate norm $||\mathbb{Q}_{\mathbf{k}_{j-1}}||_1 = \mathcal{R}(\rho_{\mathbf{k}_{j-1}})$, the final state is written,

$$\rho = \sum_{\mathbf{k}} p_{\mathbf{k}} Q_{\mathbf{k}} \sigma_{\mathbf{k}}, \quad \text{where} \quad p_{\mathbf{k}} = \prod_{j=1}^{T} \frac{|q_{\mathbf{k}_j}|}{\mathcal{R}(\rho_{\mathbf{k}_{j-1}})}, \quad Q_{\mathbf{k}} = \prod_{j=1}^{T} \operatorname{sign}(q_{\mathbf{k}_j}) \mathcal{R}(\rho_{\mathbf{k}_{j-1}}). \tag{5.144}$$

The true mean value for the observable $E$ given final state $\rho$ can be decomposed:

$$\mathrm{Tr}[E\rho] = \sum_{\mathbf{k}} p_{\mathbf{k}} Q_{\mathbf{k}} E_{\mathbf{k}}, \quad \text{where} \quad E_{\mathbf{k}} = \mathrm{Tr}[E\sigma_{\mathbf{k}}]. \tag{5.145}$$

Unlike the static simulator, the full distribution $\{p_{\mathbf{k}}\}$ is never explicitly computed. Indeed, it is not a product distribution, because each set of quasiprobabilities $\{q_{\mathbf{k}_j}\}_{k_j}$ and the corresponding $\ell_1$-norm depends on the trajectory $\mathbf{k}_{j-1}$ sampled in earlier steps. Nevertheless, on each iteration, the simulator randomly samples an output $Q_{\mathbf{k}}E_{\mathbf{k}}$ with probability $p_{\mathbf{k}}$. These probabilities exactly match the weightings in equation (5.145), which was derived from the quantum evolution of the circuit, so the simulator is an unbiased estimator for the mean value of observable $E$. The number of samples needed can therefore be computed in the usual way based on $\max\{|\widetilde{E}_i|\}$. For any trajectory $\mathbf{k}$, the value is bounded as $|Q_{\mathbf{k}}E_{\mathbf{k}}| \leq |Q_{\mathbf{k}}| \leq \prod_t \mathcal{R}(\rho_{\mathbf{k}_{t-1}}) \leq \prod_t \mathcal{C}_{\mathcal{R}}(\mathcal{E}^{(t)})$, where $\mathcal{C}_{\mathcal{R}}$ is the magic capacity with respect to RoM. This leads to the following result.

**Theorem 5.13** (Dynamic channel simulator)**.** *Suppose $\mathcal{E}$ is an $n$-qubit non-stabiliser circuit that can be decomposed as a sequence of circuit elements, $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$, where each $\mathcal{E}^{(j)}$ acts non-trivially on at most $m$ qubits, where $m$ is sufficiently small that the robustness of magic for a $2m$-qubit stabiliser state can be computed in time $\tau_{\mathcal{R}}$. Then for any stabiliser state $|\phi_0\rangle$ and stabiliser observable $E$, and parameters $\delta, p_{\text{fail}} > 0$, there exists a classical algorithm that outputs an estimate $\widehat{E}$ for the expectation value $\langle E \rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$ in time*

$$\tau = \frac{2}{\varepsilon}\ln\left(\frac{2}{p_{\text{fail}}}\right) \cdot T \cdot (\mathrm{poly}(n) + \tau_{\mathcal{R}}) \cdot \prod_{t=1}^{T}[\mathcal{C}(\mathcal{E}^{(t)})]^2, \tag{5.146}$$

*where the estimate satisfies $\left|\widehat{E} - \langle E \rangle\right| \leq \varepsilon$ with probability at least $1 - p_{\text{fail}}$.*

Notice that for every sample, $T$ optimisations are performed, as well as $T$ calls to DISENTANGLE, which has poly$(n)$ runtime. If $\mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ then we would simply not use this method, so that the only optimisations are in the preprocessing. On the other hand, if $\prod_t \mathcal{C}_{\mathcal{R}}(\mathcal{E}^{(t)}) \ll \prod_t \mathcal{R}_*(\mathcal{E}^{(t)})$ then the dynamic simulator can be faster than the static simulator; here we have a trade-off of increase in per-sample runtime, versus a multiplicative reduction in sample complexity. Indeed, the sampling cost is likely to be the bottleneck for highly non-stabiliser circuits. In Chapter 6 we discuss circuits where the dynamic simulator may have an advantage.

In theorem 5.13, we require that $m$ is small enough that finding optimal decompositions for $2m$-qubit states takes time $\tau_{\mathcal{R}}$. Since this time grows super-exponentially with $2m$, in practice we need $m \leq 2$ in the general case. However, for the case of diagonal channels, we can again use the results of Section 4.4 to increase the number of qubits. Lemma 4.34 showed that for diagonal $m$-qubit channel $\mathcal{E}$ and any $2m$-qubit state $|\phi\rangle$, there exists a stabiliser state $|\phi'\rangle$ on only $m$ qubits such that $\mathcal{R}[(\mathcal{E}\otimes\mathbb{1})(|\phi\rangle\langle\phi|)] = \mathcal{R}[\mathcal{E}(|\phi'\rangle\langle\phi'|)]$. The proof of the lemma hinged on the stronger result that $U|\phi\rangle = |\phi'\rangle\otimes|\phi''\rangle$ for some Clifford operation $U$ that commutes with $\mathcal{E}\otimes\mathbb{1}$. Therefore we can use this Clifford in the same way as that employed in the DISENTANGLE subroutine in Algorithm 5.3.5. For diagonal channels a modified DISENTANGLE procedure can be used so that the system can be put in the form $|\phi'_{j-1}\rangle^A \otimes |\phi''_{j-1}\rangle^B$ where subsystem $A$ comprises only $m$ qubits. This in principle allows the simulator to admit diagonal circuit elements on on up to five qubits.

### 5.3.6 Dynamic dyadic channel simulator

Here we show how the dynamic channel simulator of the previous subsection can be adapted to take advantage of the reduction in $\ell_1$-norm typically obtained by moving to a dyadic frame. The obvious change is that the optimisation routine used in step 9 of Algorithm 16 will be over dyads rather than stabiliser projectors, but several other modifications are needed. First, if the current stabiliser dyad is $|L\rangle\langle R|$, the Clifford that disentangles subsystems $AB$ from $C$ in step 7 is typically not the same for $|L\rangle$ as it is for $|R\rangle$. Therefore we must apply a dyadic Clifford map to the stabiliser dyad, rather than a unitary Clifford operation. Second, although each circuit element $\mathcal{E}^{(j)}$ is a CPTP map, the operator it is applied to in step 9 is a stabiliser dyad rather than a density operator. Therefore the output matrix $v_j = (\mathcal{E}^{(j)}\otimes\mathbb{1}n)(|L\rangle\langle R|)$ can be non-Hermitian and non-positive. This means we need to make use of the extended definition of dyadic negativity introduced in Section 4.3.3.

Let OPTIMISEDYAD be a procedure that finds a complex linear decomposition $\mathbb{Q}$ over stabiliser dyads for an input operator $v$, minimising $\ell_1$-norm $||\mathbb{Q}||_1$,

$$(\mathbb{Q}, ||\mathbb{Q}||_1) \leftarrow \textsc{optimiseDyad}(v) \quad \text{s.t.} \quad ||\mathbb{Q}||_1 = \Lambda(v). \tag{5.147}$$

Here $\mathbb{Q} = \{(\alpha_k, |L_k\rangle\langle R_k|)\}_k$ is returned as a list of complex coefficients $\alpha_k$ and stabiliser dyads $|L_k\rangle\langle R_k|$, such that $v = \sum_k \alpha_k |L_k\rangle\langle R_k|$. Let DYADSAMPLE be a subroutine that takes as input a distribution over stabiliser dyads $\mathbb{Q}$, and returns a randomly selected stabiliser dyad $|L_k\rangle\langle R_k|$ and the corresponding phase,

$$(|L'\rangle\langle R'|, \theta) \leftarrow \textsc{dyadSample}(\mathbb{Q}), \tag{5.148}$$

**Algorithm 17** Dynamic dyadic channel simulator

```
Input: Circuit description $\{\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \ldots, \mathcal{E}^{(T)}\}$, where each $\mathcal{E}^{(j)}$ acts non-trivially on a
    partition comprising at most $m$ qubits, specified by a binary vector $\mathbf{a}^{(j)}$ ; stabiliser
    observable $E$; initial stabiliser state $|\phi_0\rangle$; number of samples $M$.
Output: Estimate $\widehat{E}$ for expectation value $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle)]$.
Set $\widetilde{E} \leftarrow 0$.
for $i = 1$ to $M$ do
    Prepare representation of initial dyad $|L_0\rangle\langle R_0| = |\phi_0\rangle\langle\phi_0|$.
    $Q \leftarrow 1$, $\phi \leftarrow 0$          ▷ $Q$ and $\phi$ track sampling overhead and phase respectively.
    for $t = 1$ to $T$ do
        $\mathbf{a}, \mathbf{b}, \mathbf{c} \leftarrow \mathbf{a}^{(t)}, b(\mathbf{a}^{(t)}), c(\mathbf{a}^{(t)})$.          ▷ Specify qubit partitions
        $(\,|L'_{t-1}\rangle^{AB}, |L''_{t-1}\rangle^{C}, U^{BC}) \leftarrow$ DISENTANGLE$(\,|L_{t-1}\rangle^{ABC}, \mathbf{a}, \mathbf{b}, \mathbf{c})$.
        $(\,|R'_{t-1}\rangle^{AB}, |R''_{t-1}\rangle^{C}, V^{BC}) \leftarrow$ DISENTANGLE$(\,|R_{t-1}\rangle^{ABC}, \mathbf{a}, \mathbf{b}, \mathbf{c})$.
        Compute $v^{AB}_{t-1} \leftarrow (\mathcal{E}^{(t)} \otimes \mathbb{1}_m)(\,|L'_{t-1}\rangle\langle R'_{t-1}|^{AB})$
        $(\mathbb{Q}_t, \|\mathbb{Q}_t\|_1) \leftarrow$ OPTIMISEDYAD$(v^{AB}_{t-1})$
        $(|L'_t\rangle\langle R'_t|^{AB}, \theta_t) \leftarrow$ DYADSAMPLE$(\mathbb{Q}_t)$
        $|L_t\rangle^{ABC} = (\mathbb{1}_m \otimes U^{\dagger BC})\,|L'_t\rangle^{AB} \otimes |L''_{t-1}\rangle^{C}$          ▷ Undo left Clifford gate.
        $|R_t\rangle^{ABC} = (\mathbb{1}_m \otimes V^{\dagger BC})\,|R'_t\rangle^{AB} \otimes |R''_{t-1}\rangle^{C}$          ▷ Undo right Clifford gate.
        $Q \leftarrow Q \times \|\mathbb{Q}_t\|_1$, $\theta \leftarrow \theta + \theta_t$
    end for
    $\widetilde{E}_i \leftarrow \mathrm{Re}\{Qe^{i\theta}\,\langle R_T|E|L_T\rangle\}$
    $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_i$
end for
return $\widehat{E} \leftarrow \widetilde{E}/M$
```

so that $|L'\rangle\langle R'| = |L_k\rangle\langle R_k|$ with probability $|\alpha_k|/\|\mathbb{Q}\|_1$, and then $\theta = \arg(\alpha_j)$. We present pseudocode for the simulator in Algorithm 17. We make use of the same DISENTANGLE subroutine and partitioning scheme as Algorithm 16. The proof that the random variable $\widetilde{E}_i$ in Algorithm 17 is an unbiased estimator for $\langle E\rangle$ closely follows that for Algorithm 16. The key difference is that we express the final state of the quantum circuit as a dyadic decomposition, rather than in terms of stabiliser projectors (cf. equation (5.144)).

Let $\mathbf{k}$ be a vector labelling the full trajectory of stabiliser dyads chosen in the inner for-loop, and let $\mathbf{k}_t$ label the trajectory up to and including step $t$. At the beginning of loop $t$ the current dyad is $\left|L_{\mathbf{k}_{t-1}}\right\rangle\!\left\langle R_{\mathbf{k}_{t-1}}\right|$, up to a Clifford dyadic map that commutes with $\mathcal{E}^{(t)}$. Applying circuit element $t$ we obtain a non-stabiliser dyad,

$$v_{\mathbf{k}_{t-1}} = (\mathcal{E}^{(t)} \otimes \mathbb{1})(\left|L_{\mathbf{k}_{t-1}}\right\rangle\!\left\langle R_{\mathbf{k}_{t-1}}\right|) = \sum_{k_t} \alpha_{\mathbf{k}_j} \left|L_{\mathbf{k}_j}\right\rangle\!\left\langle R_{\mathbf{k}_j}\right|. \qquad (5.149)$$

Following the same reasoning as in Section 5.3.5, the final state can be expressed $\rho =$

$\sum_{\mathbf{k}} p_{\mathbf{k}} e^{i\phi_{\mathbf{k}}} Q_{\mathbf{k}} \left|L_{\mathbf{k}_{t-1}}\right\rangle\!\left\langle R_{\mathbf{k}_{t-1}}\right|$, where

$$p_{\mathbf{k}} = \prod_{t=1}^{T} \frac{|\alpha_{\mathbf{k}_t}|}{\Lambda(\nu_{\mathbf{k}_{t-1}})}, \quad Q_{\mathbf{k}} = \prod_{t=1}^{T} \Lambda(\nu_{\mathbf{k}_{t-1}}), \quad \phi_{\mathbf{k}} = \sum_{t=1}^{T} \phi_{\mathbf{k}_t}, \quad \phi_{\mathbf{k}_t} = \arg(\alpha_{\mathbf{k}_t}). \tag{5.150}$$

Again, we note that the above quasiprobability distribution is derived by considering the evolution of the full quantum state through the circuit, but each coefficient $p_{\mathbf{k}}$ is precisely the probability of the simulator selecting the trajectory $\mathbf{k}$, and $e^{i\phi_{\mathbf{k}}} Q_{\mathbf{k}}$ is the complex prefactor computed by the algorithm for that trajectory. Therefore the classical simulator gives an unbiased estimator for the quantum mean value, with each random variable $\widetilde{E}_i$ bounded by the product of dyadic magic capacities,

$$\left|\widetilde{E}_i\right| \leq Q_{\mathbf{k}} = \prod_{t=1}^{T} \Lambda(\nu_{\mathbf{k}_{t-1}}) \leq \prod_{t=1}^{T} \mathcal{C}_{\Lambda}(\mathcal{E}^{(t)}). \tag{5.151}$$

So, by the standard arguments, the number of samples $M$ required will be $\mathcal{O}(\prod_{t=1}^{T} \mathcal{C}_{\Lambda}(\mathcal{E}^{(t)})^2)$. We therefore have the following theorem.

**Theorem 5.14** (Dynamic dyadic channel simulator)**.** *Suppose $\mathcal{E}$ is an n-qubit non-stabiliser circuit that can be decomposed as a sequence of circuit elements, $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$, where each $\mathcal{E}^{(j)}$ acts non-trivially on at most m qubits, where m is small enough that the dyadic negativity for a 2m-qubit stabiliser state can be computed in time $\tau_{\Lambda}$. Then for any stabiliser state $|\phi_0\rangle$, stabiliser observable E, and parameters $\delta, p_{\text{fail}} > 0$, there exists a classical algorithm that outputs an estimate $\widehat{E}$ for the expectation value $\langle E\rangle = \mathrm{Tr}[E\mathcal{E}(|\phi_0\rangle\langle\phi_0|)]$ in time:*

$$\tau = \frac{2}{\varepsilon} \ln\left(\frac{2}{p_{\text{fail}}}\right) \cdot T \cdot (\mathrm{poly}(n) + \tau_{\Lambda}) \cdot \prod_{t=1}^{T} [\mathcal{C}_{\Lambda}(\mathcal{E}^{(t)})]^2, \tag{5.152}$$

*where the estimate satisfies $\left|\widehat{E} - \langle E\rangle\right| \leq \varepsilon$ with probability at least $1 - p_{\text{fail}}$.*

Since the dyadic negativity is typically significantly smaller than the robustness of magic, for many circuits the dyadic variant of the dynamic channel simulator may offer a significant reduction in sample complexity. As for the dynamic channel simulator of Section 5.3.5, the restriction to few-qubit operations can be partially reduced for diagonal channels by the addition of a subroutine to find a commuting Clifford gate that can disentangle each part of the stabiliser dyad. However, we should note that the optimisation time for dyadic negativity, $\tau_\Lambda$, can be significantly larger than that for robustness of magic, $\tau_\mathcal{R}$, as the number of qubits increases. Therefore the improvement in sample complexity must be weighed against the increased time to optimise decompositions for each sample in deciding which simulator to use for a given circuit.

## 5.4 Summary and conclusions

In this chapter we introduced a suite of classical simulators for magic states and operations, each of which is associated with one of the magic monotones defined in Chapter 4. We first introduced a triplet of simulators for stabiliser circuits with magic state inputs, corresponding to the three magic state monotones discussed in section 4.1. This completes the strand of work establishing the connection between quasiprobability and stabiliser rank methods, as we have shown that one can construct simulators of both types where the performance is quantified by magic monotones which coincide for important classes of states, namely pure states and tensor products of single-qubit states. Our stabiliser rank simulator allows bit-string sampling from circuits with noisy magic state inputs, with runtime comparable to pure state stabiliser rank simulators. Indeed, we have shown improved scaling with respect to precision compared to the BBCCGH sparsification technique [3]. Meanwhile our dyadic frame simulator enables observable estimation with runtime significantly reduced compared to prior qubit quasiprobability simulators.

Next we introduced the family of static simulators which make the conceptual move from decompositions of states to decompositions of channels, and have performance that scales with some channel monotone. While this move introduces some technical complications, it has the potential to yield faster simulation of certain circuits, as we can avoid overhead associated with gadgetisation. Moreover, it opens up the possibility of simulating non-stabiliser noise channels, such as amplitude damping, in a more efficient manner.

In the last part of the chapter we introduced the class of dynamic channel simulators. The sandwich theorems proven in section 4.3.4 of the previous chapter established that magic capacity monotones $\mathcal{C}_\mathcal{M}$ lower bound their associated decomposition-based monotones $\mathcal{M}_*$. Although the two types of channel monotone can coincide in some cases, in Chapter 6 we give examples where there is a gap. Our dynamic simulators exploit this gap to reduce the number of samples needed to achieve a given precision, trading off against the need to solve linear programs on-the-fly. A detailed analysis of this trade-off is left for future work.

Thus far we have focused on establishing the validity of our classical simulators, and their relationship to well-behaved magic monotones. The performance of our algorithms in practice will depend on the computed monotone value for the particular circuit decomposition to be simulated. In the next chapter we illustrate how the monotones can be employed in more realistic settings.

## Chapter 6

# Resource costs for elements of quantum circuits

In this final research chapter we investigate simulation overhead for practically relevant scenarios by numerically or analytically evaluating the magic monotones introduced earlier, or bounding their value when direct computation is not practical. We hope that the illustrative examples we provide in this chapter suggest regimes in which the algorithms presented in Chapter 5 may be useful, and point the way toward finer-grained circuit analyses, and indeed practical classical simulations, that could be carried out in future work. It is beyond the scope of this thesis to conduct a review of the literature on quantum algorithms. Instead we briefly review some basic principles of quantum simulation [10], in order to motivate our consideration of particular types of operation in this chapter.

The starting point for any quantum simulation using qubits is to fix how the physical system of interest is to be represented. Systems of spins can be directly identified with qubit systems, and their Hamiltonians are naturally expressed as a sum of Pauli terms. However, many important problems in materials science and quantum chemistry are concerned with systems of electrons [11, 166]. For electronic systems it is vital to ensure that the encoding respects the correct fermionic statistics. The standard approach is to work in the second quantised picture, and a typical electronic Hamiltonian has the form

$$H = \sum_{i,j} h_{i,j} a_i^\dagger a_j + \frac{1}{2} \sum_{i,j,k,l} h_{i,j,k,l} a_i^\dagger a_j^\dagger a_k a_l, \tag{6.1}$$

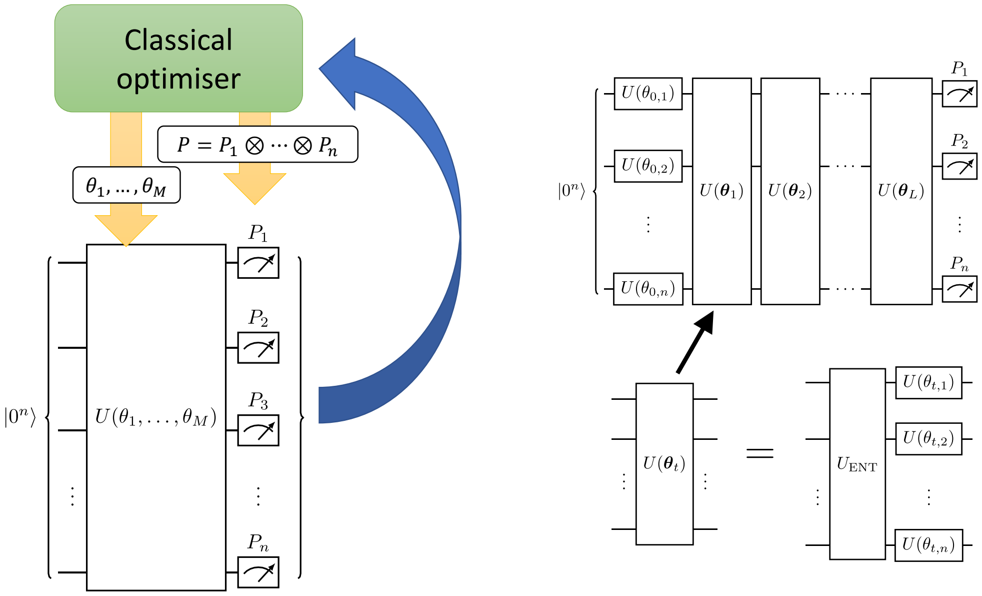
where $a_j^\dagger$ ($a_j$) is the creation (annihilation) operator for the $j$-th fermionic mode, and $h_{i,j}$ and $h_{i,j,k,l}$ are coefficients. The Jordan-Wigner [167, 168], Bravyi-Kitaev [169] and other [170–172] encodings provide a mapping from fermionic to qubit operators while preserving all fermionic anticommutation relations. In this way the second quantised Hamiltonian can be re-expressed as a sum of Pauli terms.

Having fixed a representation, a key task is to simulate time evolution under the Hamiltonian, $U(t) = \exp(-iHt)$ for some time $t$. However, the problem of synthesising an $n$-qubit unitary operator using a standard library of gates is a non-trivial task, and cannot be done efficiently for arbitrary $H$ (recall Section 1.4). One approach is to approximate the time evolution using a Trotter-Suzuki decomposition [39, 173–177]. For a Hamiltonian written as a sum of $k$-local Pauli terms $H = \sum_j h_j H_j$, the evolution for the full time $t$ is written as a product of short time-steps, $\Delta t$, $U(t) = (\exp(-iH\Delta t))^{t/\Delta t}$. One can then use the first-order Trotter formula [168],

$$U(\Delta t) = \prod_j \exp\big[-iH_j\Delta t\big] + \mathcal{O}\big((\Delta t)^2\big), \tag{6.2}$$

so that for small $\Delta t$, the first term is a good approximation for $U(\Delta t)$. Then by repeating the sequence $t/\Delta t$ we can recover an approximation for the longer time evolution, $U(t) \approx \big(\prod_j \exp\big[-iH_j\Delta t\big]\big)^{t/\Delta t}$. The shorter the time-step $\Delta t$, the more accurate the approximation, but the greater the number of few-qubit gates $\exp\big(-iH_j\Delta t\big)$ that must be implemented. More sophisticated techniques have been proposed to reduce the number of gates, such as the use of higher-order Trotter product formulae [175, 176] or random compilation [165, 178]. The long sequences of small-angle Pauli rotations needed for this type of algorithm motivates our study of these gates later in the chapter. However, for such schemes, each logical gate must be executed with extremely high fidelity, as otherwise errors will accumulate and overwhelm the intended computation. High accuracy digital quantum simulation of the type described above therefore falls within the domain of error-corrected, fault-tolerant quantum computation, and remains out of reach for the present.

Recently, approaches better suited to NISQ devices have been proposed, including classical-quantum hybrid algorithms such as the variational quantum eigensolver (VQE) [78], illustrated in Figure 6.1a. The method is based on the variational principle that the energy expectation value for any state $|\psi\rangle$ can be no smaller than that of the ground state of the Hamiltonian $|E_0\rangle$, $\langle\psi|H|\psi\rangle \geq \langle E_0|H|E_0\rangle = E_0$. A quantum subroutine is


Classical optimiser
$P = P_1 \otimes \cdots \otimes P_n$
$\theta_1, \ldots, \theta_M$
$|0^n\rangle$
$U(\theta_1, \ldots, \theta_M)$
$P_1$
$P_2$
$P_3$
$P_n$
$|0^n\rangle$
$U(\theta_{0,1})$
$U(\theta_{0,2})$
$U(\theta_{0,n})$
$U(\boldsymbol{\theta}_1)$
$U(\boldsymbol{\theta}_2)$
$U(\boldsymbol{\theta}_L)$
$P_1$
$P_2$
$P_n$
$U(\boldsymbol{\theta}_t)$
=
$U_{\text{ENT}}$
$U(\theta_{t,1})$
$U(\theta_{t,2})$
$U(\theta_{t,n})$

**(a)** VQE optimisation loop **(b)** Hardware-efficient VQE circuit

**Figure 6.1:** (a) Variational quantum eigensolver. A quantum subroutine is embedded in a classical optimisation loop. The quantum circuit is parameterised by a set of angles $\{\theta_1, \ldots, \theta_M\}$ which specify an ansatz state, and a set of Pauli observables to be measured. A classical control computer attempts to minimise $H(\theta_1, \ldots, \theta_M) = \sum_P h_P \langle P \rangle$, by varying the parameters $\{\theta_1, \ldots, \theta_M\}$ sent to the quantum processor based on the statistics obtained from each iteration. (b) Illustration of a hardware-efficient layout. Layers of hardware-native entangling gates $U_{\text{ENT}}$ are interleaved with layers comprising single-qubit gates $U(\theta_{t,j})$. The entangling layers remain fixed, while the single-qubit layers are parameterised by a set of vectors $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L\}$.

parameterised by a vector of classical parameters $\boldsymbol{\theta}$, so that the circuit prepares an ansatz state $|\boldsymbol{\theta}\rangle$. The average energy $\langle H \rangle$ is estimated by measuring each term in the Pauli decomposition $H$. This quantum subroutine is nested in a classical optimisation loop, which iteratively sends updated $\boldsymbol{\theta}$ to the quantum processor using a standard method such as gradient descent.

The ansatz circuit can either be physically motivated [11, 78, 179], or based on a hardware-efficient construction drawn from the gates that can be easily implemented on a given quantum device [180]. In the former case, a time evolution operation can be Trotterised, but with far fewer gates than would be needed for conventional quantum simulation [11, 179]. Meanwhile, a typical strategy in hardware-efficient schemes is to interleave layers of fixed entangling gates and parameterised single-qubit gates [180] (Figure 6.1b), chosen from a hardware-native gate set. In contrast to conventional digital quantum simulation, VQE circuits are designed to have low depth, and there is no restriction to small-angle rotations. It has been argued that VQE is robust against certain types of noise [11, 181, 182], and techniques have been developed for further mitigating against errors [49, 79, 181], making it attractive for implementation on NISQ hardware.

In this chapter, the question we are interested in is how far quantum subroutines of the types described above can themselves be classically simulated using stabiliser techniques. This is relevant both for the verification and benchmarking of prototype devices, and as an indicator of how large circuits must be to exceed the reach of classical simulation. In addition to long sequences of small-angle rotations motivated by Trotter methods, and shorter sequences of larger angle noisy rotations motivated by VQE, we will also consider multi-qubit controlled phase gates, which may be useful in more efficient synthesis of universal circuits beyond the Clifford + T library [183]. Code to compute the numerical results presented in this chapter was written in MATLAB [184], using the package CVX [132] to solve convex optimisation problems, and has been made available at the public repository Ref. [185]. We also made use of MATLAB code written by Toby Cubitt, for permuting subsystems and generating Haar-random unitaries [186]. We first consider resource costs for estimating observables using our robustness-based channel simulators (Section 6.1), comparing our methods with the Oak Ridge algorithm [45]. We will then show how certain types of noisy gate can be decomposed for use with the bit-string sampling simulator introduced in Section 5.3.4. Finally in Section 5.3.4 we look at the cost of observable estimation using the dyadic frame simulator and compare this with the sparsification method over a range of parameters.

## 6.1 Robustness-based channel simulators

### 6.1.1 Single-qubit rotations with amplitude damping

Consider a sequence of noisy operations, where a noise channel $\mathcal{E}_p$, acts between each unitary gate $\mathcal{U}_j$, and where $p$ is a parameter denoting the error rate,

$$\mathcal{E} = \mathcal{E}_p \circ \mathcal{U}_L \circ \ldots \circ \mathcal{E}_p \circ \mathcal{U}_2 \circ \mathcal{E}_p \circ \mathcal{U}_1. \tag{6.3}$$

This could, for example, model a layer of parameterised single-qubit rotations in a VQE circuit (Figure 6.1b). Below, we compute the simulation cost for a single step in such a computation, comprising a single-qubit rotation and an amplitude-damping channel. We compute the cost with respect to our static and dynamic channel simulators (Algorithms

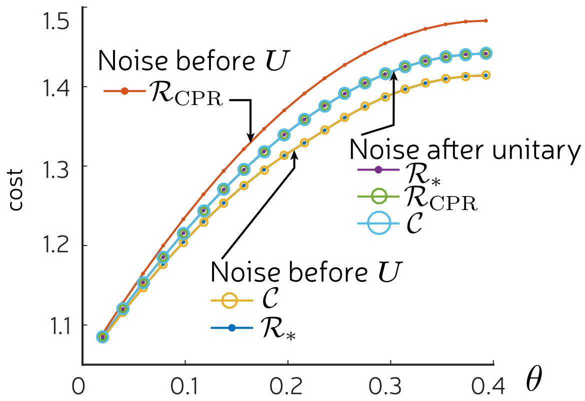


**Figure 6.2:** Comparison of $\mathcal{C}(\mathcal{E})$, $\mathcal{R}_*(\mathcal{E})$ and $\mathcal{R}_{CPR}(\mathcal{E})$, where $\mathcal{E}$ is a single-qubit $X$-rotation $\mathcal{U}(\theta)$ composed with an amplitude damping channel $\mathcal{E}_p$ for noise $p = 0.1$. We consider both possible orderings: noise after unitary ($\mathcal{E}_p \circ \mathcal{U}(\theta)$), and noise before unitary ($\mathcal{U}(\theta) \circ \mathcal{E}_p$). Figure adapted from our article Ref. [1]

12 and 16). Recall that the worst-case overhead factors for these algorithms are given by the square of the channel robustness $\mathcal{E}_*$ and magic capacity $\mathcal{C}$ respectively. Since the static simulator generalises the Oak Ridge simulator[45] (see Section 2.5.1), we compare with the analogous cost function for that algorithm, $\mathcal{R}_{\mathrm{CPR}}$. We note that amplitude-damping noise is not stabiliser-preserving, so in general can increase the robustness. This is in contrast to dephasing and depolarising noise, which we will study in Section 6.2, and which tend to reduce overhead in our simulators.

In modelling the effect of noise for intermediate steps in the sequence (6.3), we have a choice of ordering. We can take the circuit elements to be either $\mathcal{E}_p \circ \mathcal{U}_j$ or $\mathcal{U}_j \circ \mathcal{E}_p$. These choices are equivalent in the sense that

$$(\mathcal{E}_p \circ \mathcal{U}_2) \circ (\mathcal{E}_p \circ \mathcal{U}_1) \circ \mathcal{E}_p = \mathcal{E}_p \circ (\mathcal{U}_2 \circ \mathcal{E}_p) \circ (\mathcal{U}_1 \circ \mathcal{E}_p), \tag{6.4}$$

but could lead to different sample complexity when $\mathcal{E}_p$ and $\mathcal{U}_j$ do not commute. We studied circuit elements made up of a single-qubit Pauli $X$-rotation $U(\theta) = \exp(iX\theta)$ composed with an amplitude damping channel $\mathcal{E}_p$ with noise parameter $p$, defined by Kraus operators $K_1 = \mathrm{diag}(1, \sqrt{1-p})$, $K_2 = \sqrt{p}\,|0\rangle\langle 1|$. We computed the respective cost functions for a range of values of $p$ and $\theta$. For noise $p = 0.1$ we see that when the noise channel follows the gate, there is no difference between the three quantities (Figure 6.2). However, if the noise channel acts before the unitary, both our monotones show a reduced value, whereas $\mathcal{R}_{\mathrm{CPR}}$ increases. Recall that the reduction in sample complexity for $t$ uses

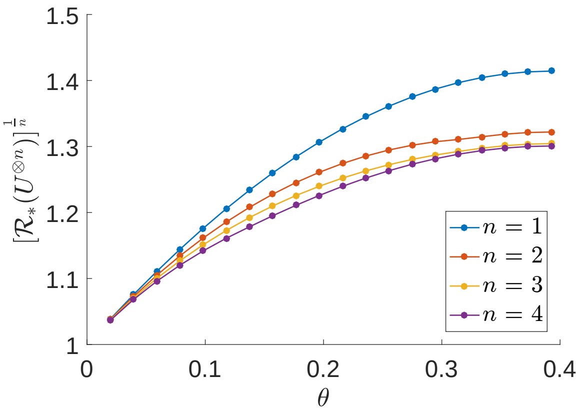


**Figure 6.3:** Normalised channel robustness $[\mathcal{R}_*(U^{\otimes n})]^{1/n}$ plotted for $Z$-rotations $U(\theta) = \exp[iZ\theta]$ and for $n$ qubits, up to $n = 4$. Figure reproduced from our article Ref. [1].

of the channel is given by $(\mathcal{R}_*/\mathcal{R}_{\text{CPR}})^{2t}$.This suggests that for intermediate-sized circuits using this noise model, the better strategy with respect to sample complexity would be to choose the ordering $\mathcal{U}(\theta) \circ \mathcal{E}_p$, and either our static or dynamic channel simulator in preference to the Oak Ridge simulator (with the caveat that the dynamic simulator requires additional per-sample runtime to compute intermediate decompositions).

### 6.1.2 Reduced overhead from submultiplicativity

We define *normalised* channel robustness for single-qubit gates $U$, defined as $[\mathcal{R}_*(U^{\otimes n})]^{1/n}$. This allows us to quantify the per-gate savings in sample complexity that can be achieved by grouping single-qubit rotations in $n$-qubit blocks. In Figure 6.3 we present results for qubit $Z$-rotations $U = \exp[iZ\theta]$, up to four qubits. We find that strict submultiplicativity is observed for all values of $\theta$, with significant reductions between the $n = 2$ and $n = 4$ cases for a wide range of angles. For the static channel simulator there are therefore large performance improvements to be gained by blocking together gates acting on different qubits in order to obtain an optimal decomposition for the composite channel.

### 6.1.3 Multi-qubit controlled-phase gates

Recall from the sandwich theorem for RoM-based monotones (Theorem 4.30) that $\mathcal{R}(\Phi_{\mathcal{E}}) \leq \mathcal{C}_{\mathcal{R}}(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$ for any channel $\mathcal{E}$. We found numerically that for all amplitude-damped Pauli rotations studied, $\mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ up to solver precision, and in the absence of noise we had full equality $\mathcal{R}(\Phi_{\mathcal{E}}) = \mathcal{C}_{\mathcal{R}}(\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$. Theorem 4.30 also states that all measures are equal for gates from the third level of the Clifford hi-

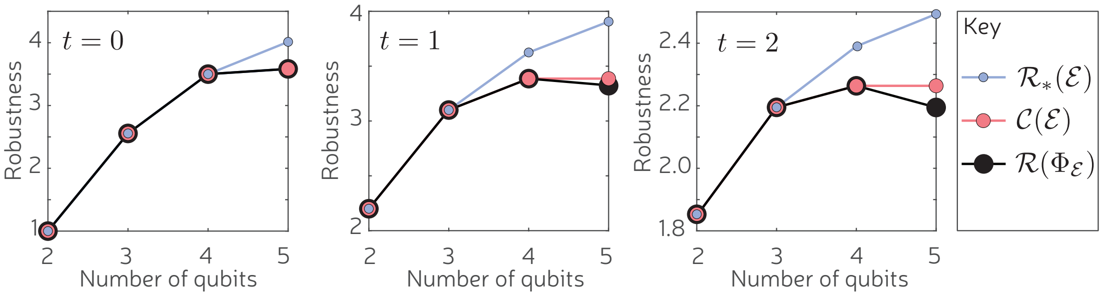


**Figure 6.4:** Comparison of quantities for multicontrol phase gates (see equation 6.5). Left: Multicontrol-$Z$ gates (t=0). Middle: Multicontrol-$S$ gates (t=1). Right: Multicontrol-$T$ gates (t=2). Figure reproduced from our article Ref. [1].

erarchy. Under what conditions does this equality persist for more general multi-qubit operations? In the most general case, our monotones can only be computed for two-qubit channels, since optimising over all six-qubit stabiliser Choi states is intractable. However, we showed in Section 4.4 that $\mathcal{R}_*$ and $\mathcal{C}$ can be evaluated for diagonal channels on up to five qubits. As a special case we consider multicontrol phase gates of the form,

$$M_{t,n} = \mathrm{diag}\big(\exp\big(i\pi/2^t\big), 1, \ldots, 1\big), \quad t \in \mathbb{Z} \tag{6.5}$$

where $n$ denotes the number of qubits. The family includes familiar gates such as $CZ$ ($t = 0$, $n = 2$), $CCZ$ ($t = 0$, $n = 3$), multicontrol-$S$ ($t = 1$) and -$T$ ($t = 2$).

We find that the inequalities are tight for the $n = 2$ and $n = 3$ cases, but that this does not persist for larger system sizes (Figure 6.4). The $t = 0$ case (the family of multicontrol-$Z$ gates) turns out to be a special case (Figure 6.4, left panel). Here we find equality for all three quantities up to $n = 4$. For the $t = 0$, $n = 5$ case, $\mathcal{R}\big(\Phi_{M_{0,5}}\big) = \mathcal{C}_{\mathcal{R}}\big(M_{0,5}\big)$ holds, but $\mathcal{R}_*\big(M_{0,5}\big)$ is strictly greater than both. Note also that for $t = 0$, all three quantities increase with each increment in $n$.

The families of gates with $t > 0$ follow a pattern qualitatively similar to each other. The results for the $t = 1$ (multicontrol-$S$) and $t = 2$ (multicontrol-$T$) cases are shown in the middle and right panels of Figure 6.4. For $n = 4$, $t > 0$, the same situation holds as for $n = 5$, $t = 0$, as we find $\mathcal{R}\big(\Phi_{M_{t,4}}\big) = \mathcal{C}_{\mathcal{R}}(M_{t,4}) < \mathcal{R}_*(M_{t,4})$. At $n = 5$, all three quantities separate. In contrast with the multicontrol-$Z$, we see that $\mathcal{R}\big(\Phi_{M_{t,n}}\big)$ decreases as we go from four to five qubits, while $\mathcal{C}_{\mathcal{R}}(M_{t,n})$ levels off. We see similar behaviour for all non-zero values of $t$ investigated numerically. Our current techniques limit us to

| Linear subspace dimension, $k$ | Number of qubits, $n$ | | | |
|---|---|---|---|---|
| | 2 | 3 | 4 | 5 |
| 1 | 1.414 | 1.414 | 1.414 | 1.414 |
| 2 | **1.849** | 1.849 | 1.849 | 1.849 |
| 3 | - | **2.195** | 2.195 | 2.195 |
| 4 | - | - | **2.264** | **2.264** |
| 5 | - | - | - | 2.195 |

**Table 6.1:** Final robustness after multicontrol-$T$ gate applied to input stabiliser states $|\mathcal{L}\rangle$ with $k = \dim(\mathcal{L})$. In each column, the maximum robustness (i.e. the capacity) is highlighted red.

five-qubit gates, but we conjecture that the capacity will remain level for $n > 5$. This is a special case of the following conjecture.

**Conjecture 6.1.** *For any fixed $t$, the maximum increase in robustness of magic for $M_{t,n}$ is achieved at some finite number of qubits $n = K$ by acting on the state $\left|+^K\right\rangle$. Therefore $\mathcal{C}_\mathcal{R}(M_{t,n}) = \mathcal{R}\left(M_{t,K}\left|+^K\right\rangle\right)$ for all $n \geq K$.*

This conjecture is partially justified by Observation 4.37 from Chapter 4, which stated that for stabiliser states defined by the linear subspace $\mathcal{L}$ with dimension $k < n$, we have $\mathcal{R}(M_{t,n}|\mathcal{L}\rangle) = \mathcal{R}(M_{t,k}\left|+^k\right\rangle)$. Consider the maximisation over initial stabiliser states performed to calculate the capacity $\mathcal{C}$. In Section 4.4, we showed that for the family of gates $M_{t,n}$, we only need to calculate robustness for one representative initial stabiliser state for each possible *dimension* of linear subspace; that is, for $M_{t,n}$ we only need to compute the final robustness for $n$ initial states. In Table 6.1 we present the relevant values for the family of multicontrol-$T$ gates ($t = 2$) and make two observations. First, notice that in each row of Table 6.1 (i.e. for fixed $k = \dim(\mathcal{L})$), the RoM is constant with $n$. Indeed, this is a generic feature of the $M_{t,n}$ gates as formalised by Observation 4.37. Second, in the last column of Table 6.1, we see that $\mathcal{R}(M_{t,n}|\mathcal{L}\rangle)$ increases with $k$ for $k \leq 4$, but at $k = 5$ the value drops. With a little thought we can see that this is necessarily the case if $\mathcal{R}\left(\Phi_{M_{t,5}}\right) < \mathcal{C}_\mathcal{R}\left(M_{t,5}\right)$; we saw earlier that for diagonal gates $U$ the Choi state robustness is equal to $\mathcal{R}(U|+^n\rangle)$, and $|+^n\rangle$ is a representative state for the $k = n$ case.

Our current techniques limit us to five-qubit operations, so we are unable to confirm numerically whether $\mathcal{R}(M_{t,n}|\mathcal{L}\rangle)$ continues to decrease with increasing $\dim(\mathcal{L})$. An intuition for why a decrease is plausible goes as follows. A stabiliser state $|\mathcal{L}\rangle$ with $\dim(\mathcal{L}) = k$ will have $2^k$ equally weighted terms when written in the computational basis, so will have a normalisation factor of $2^{-k/2}$. The non-stabiliser state $M_{t,n}|\mathcal{L}\rangle$ is identical

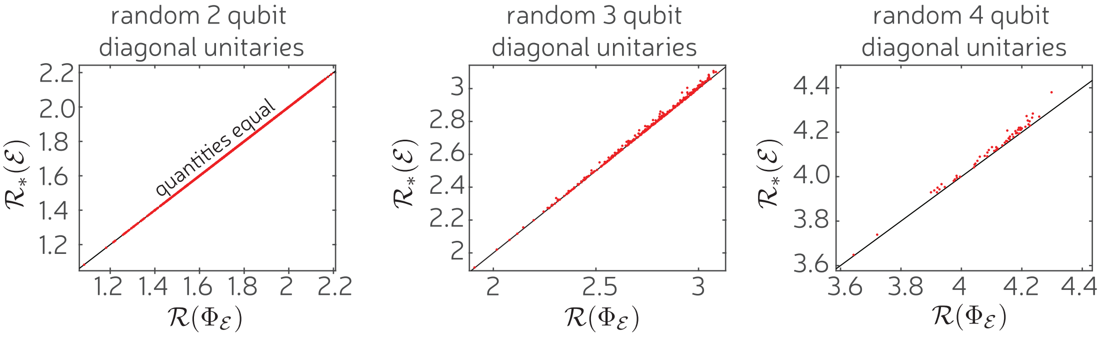


**Figure 6.5:** Channel robustness against robustness of Choi state for random $n$-qubit diagonal gates, up to $n = 4$. Black line indicates $\mathcal{R}_* = \mathcal{R}(\Phi)$. Each red dot represents the data point for an individual gate. Fewer points were calculated for larger $n$ due to the increased time to calculate each value. 1000 data points were calculated for $n = 2$, 300 for $n = 3$, and 60 for $n = 4$. Reproduced from our article Ref. [1].

to $|\mathcal{L}\rangle$ apart from the phase on the all-zero term $|0\ldots0\rangle$. As $k$ becomes large, the amplitude of the term $\frac{e^{i\pi/2^t}}{2^{k/2}}|0\ldots0\rangle$ becomes very small, so that $M_{t,n}|\mathcal{L}\rangle$ has increasingly large overlap with the stabiliser state $|\mathcal{L}\rangle$. We would therefore expect $M_{t,n}|\mathcal{L}\rangle$ to have a small robustness if $k$ is large. In summary, for few-qubit gates, control by additional qubits tends to increase the magic capacity. For example, $CZ$ is a Clifford gate but $CCZ$ is not. In contrast, our conjecture suggests that for gates controlled on many qubits, this effect vanishes.

### 6.1.4 Random diagonal unitary operations

Next we compute channel monotone values for diagonal gates $U = \sum_x e^{i\theta_x}|x\rangle\langle x|$, with $\theta_x$ chosen uniformly at random. We are particularly interested in understanding when $\mathcal{R}(\Phi_\mathcal{E}) \leq C(\mathcal{E}) \leq \mathcal{R}_*(\mathcal{E})$ is tight or loose. In Figure 6.5 we compare the Choi state robustness with the channel robustness. For every 2-qubit gate tested we observed that $\mathcal{R}(\Phi_\mathcal{E}) = \mathcal{R}_*(\mathcal{E})$ up to numerical precision. For 3- and 4-qubit gates we typically saw that $\mathcal{R}(\Phi_\mathcal{E}) < \mathcal{R}_*(\mathcal{E})$, though the gap is not often large. While the difference is slight for a single gate, these quantities influence the rate of exponential scaling when considering $N$ uses of the gate, and lead to a large gap for modest $N$.

We compared the RoM of the Choi state with the magic capacity but do not plot this data as it was equal up to solver precision for every random instance we observed. This is curious since there exist diagonal gates $U$ for which $\mathcal{R}(\Phi_U) < \mathcal{C}_\mathcal{R}(U)$. While such gates exist, our random sampling does not tend to provide examples. We can explain this by a concentration effect. By Observation 4.35, for an $n$-qubit stabiliser state $|\mathcal{K}\rangle$ specified by affine space $\mathcal{K}$ of dimension $k$, and for any diagonal $U$, there exists a $k$-qubit state $|\psi\rangle$ such that $\mathcal{R}(U|\mathcal{K}\rangle) = \mathcal{R}(\psi)$. The $n$-qubit random diagonal gates concentrate (with high probability) within a narrow range of values for the magic capacity, close to the maximum possible magic capacity for an $n$-qubit diagonal gate. Recall that for diagonal gates, $\mathcal{R}(\Phi_U) = \mathcal{R}(U|+^n\rangle)$. If $\mathcal{R}(\Phi_U) < \mathcal{C}_{\mathcal{R}}(U)$ then by Theorem 4.33 we must have that $C(U) = \mathcal{R}(U|\mathcal{K}\rangle\langle\mathcal{K}|U^\dagger)$ for some affine space $\mathcal{K}$ of non-maximal dimension $k < n$, since up to phases, $|+^n\rangle$ is the unique stabiliser state with affine dimension $n$. However, $U|\mathcal{K}\rangle\langle\mathcal{K}|U^\dagger$ is Clifford equivalent to some $k$-qubit state $U'|\phi'\rangle$, where $U'$ is diagonal, $|\phi'\rangle$ is a stabiliser state, and $k < n$. Then $\mathcal{R}(U|\mathcal{K}\rangle\langle\mathcal{K}|U^\dagger)$ would be upper bounded by the maximum $\mathcal{C}_{\mathcal{R}}(\mathcal{E})$ for k-qubit diagonal unitary operations. But if $\mathcal{C}_{\mathcal{R}}(\mathcal{E})$ is close to the maximum possible for $n$-qubit diagonal unitary operations, then it is strictly larger than the maximum magic capacity for $k$-qubit diagonal gates, so it is impossible for $U|\mathcal{K}\rangle\langle\mathcal{K}|U^\dagger$ to achieve the magic capacity.

## 6.2 Channel decompositions for bit-string sampling

The stabiliser channel extent $\Xi_*$ for unital channels, defined in Section 4.2.3 as

$$\Xi_*(\mathcal{E}) = \min\{\sum_j p_j \xi(U_j) : \mathcal{E} = \sum_j p_j \mathcal{U}_j\}, \quad \text{where} \quad \mathcal{U}_j(\cdot) = U_j(\cdot)U_j^\dagger. \tag{6.6}$$

is not a true channel monotone, as it is not defined for non-unital channels. Nevertheless in Section 5.3.4 we showed that when such decompositions exist, they can be employed in a classical algorithm which simulates sampling from the output distribution of a quantum circuit. For a sequence of channels $\mathcal{E} = \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(1)}$, the average runtime of this simulator scales with $\prod_{t=1}^T \Xi_*(\mathcal{E}^{(t)})$ in the optimal case. Since the optimisation is over all possible decompositions, and there are an infinite number of unitary operations, computing the channel extent exactly is not tractable in the general case. However, we can upper bound the quantity by optimising over a finite set of unitary operations. The following observation will also be useful.

**Lemma 6.2** (Lower bound for channel extent)**.** *Suppose $\mathcal{E}$ can be written as a convex mixture of unitary channels. Then the channel extent is lower bounded by the dyadic channel negativity, $\Xi_*(\mathcal{E}) \geq \Lambda_*(\mathcal{E})$.*

*Proof.* Suppose an optimal decomposition of $\mathcal{E}$ is written $\mathcal{E} = \sum_j p_j \mathcal{U}_j$, so that it has channel extent $\Xi_*(\mathcal{E}) = \sum_j p_j \xi(U_j)$. For each $\mathcal{U}_j$ there exists an optimal (with respect to unitary extent) decomposition $U_j = \sum_k c_{j,k} V_{j,k}$ where $V_{j,k}$ are Clifford gates, so that $\sum_k |c_{j,k}| = \sqrt{\xi(U_j)}$. Then we can rewrite the decomposition of the channel as:

$$\mathcal{E} = \sum_j p_j U_j(\cdot) U_j^\dagger = \sum_{j,k,m} p_j c_{j,k} c_{j,m}^* V_{j,k}(\cdot) V_{j,m}^\dagger. \tag{6.7}$$

This now has the form of a complex linear combination of dyadic Clifford maps $V_{j,k}(\cdot)V_{j,m}^\dagger$, with $\ell_1$-norm equal to

$$\sum_{j,k,m} |p_j c_{j,k} c_{j,m}^*| = \sum_j p_j \sum_k |c_{j,k}| \sum_m |c_{j,m}| = \sum_j p_j \xi(U_j) = \Xi_*(\mathcal{E}). \tag{6.8}$$

But recall that the dyadic Clifford maps are a subset of the stabiliser dyadic channels $\{\mathcal{T}_j\}$, which form the vertices for the dyadic channel negativity problem, $\Lambda_*(\mathcal{E}) = \min\left\{ \|\alpha\|_1 : \sum_j \alpha_j \mathcal{T}_j = \mathcal{E} \right\}$ (see Definitions 4.21, 4.23 and 4.24 in Section 4.2.5). Therefore $\Xi_*(\mathcal{E}) \geq \Lambda_*(\mathcal{E})$. □

Since the dyadic channel negativity is lower bounded by the dyadic negativity of the Choi state, $\Lambda(\Phi_{\mathcal{E}})$ (Section 4.2.5), this immediately leads to the following.

**Theorem 6.3** (Sandwich theorem for channel extent and dyadic channel negativity)**.** *Suppose $\mathcal{E}$ is a channel where a decomposition $\mathcal{E} = \sum_j p_j \mathcal{U}_j$ exists for unitary operations $\mathcal{U}_j$. Then,*

$$\Lambda(\Phi_{\mathcal{E}}) \leq \Lambda_*(\mathcal{E}) \leq \Xi_*(\mathcal{E}) \leq \sum_j p_j \xi(U_j), \tag{6.9}$$

*where each unitary operator has optimal decomposition $U_j = \sum_k c_{j,k} V_{j,k}$. If $\sum_j p_j \xi(U_j) = \Lambda(\Phi_{\mathcal{E}})$ then the decomposition is optimal, and*

$$\Lambda(\Phi_{\mathcal{E}}) = \Lambda_*(\mathcal{E}) = \Xi_*(\mathcal{E}) = \sum_j p_j \xi(U_j), \tag{6.10}$$

*Moreover, when this holds, the dyadic Clifford decomposition*

$$\mathcal{E} = \sum_{j,k,m} p_j c_{j,k} c_{j,m}^* V_{j,k}(\cdot) V_{j,m}^\dagger \tag{6.11}$$

*is optimal with respect to dyadic channel negativity.*

It is also always the case that $\Xi_*(\mathcal{E}) \geq \Xi(\Phi_{\mathcal{E}})$, where $\Xi$ is the density-operator extent, since the optimisation on the RHS is over a strictly larger set (all pure states) than the LHS (restricted to unitary operations). It is not clear whether these quantities are always equal. Nevertheless if we have some subset of channels $\mathcal{S}$ where, for each $\mathcal{E} \in \mathcal{S}$, we can find an optimal (w.r.t. density-operator extent) decomposition $\Phi_{\mathcal{E}} = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ such that each $|\psi_j\rangle$ satisfies two conditions: (i) $|\psi_j\rangle$ corresponds to a unitary $U_j$ and (ii) unitary extent is equal to the pure-state extent of the Choi state, $\xi(U_j) = \xi(|U_j\rangle)$; then we can deploy arguments simular to those used for magic states [2] in Section 4.1 to assist with the channel problem. We will show that this strategy is successful for single-qubit Pauli rotations subject to dephasing noise. Subsequently we use heuristic techniques to obtain decompositions for rotations subject to depolarising noise.

### 6.2.1 Decompositions for dephased single-qubit Z-rotations

Consider the channel $\mathcal{E} = \mathcal{E}_p \circ \mathcal{U}_\phi$, where $\mathcal{U}_\phi(\cdot) = U_\phi(\cdot)U_\phi^\dagger$, for $U_\phi = \exp[-i\phi Z]$ and $\mathcal{E}_p$ is the dephasing channel, $\mathcal{E}_p = (1-p)\mathbb{1} + p\mathcal{Z}$. Here $\mathbb{1}$ is the identity map, and $\mathcal{Z}(\cdot) = Z(\cdot)Z$. Then the composed channel is written:

$$\mathcal{E} = (1-p)\mathcal{U}_\phi + p\mathcal{Z} \circ \mathcal{U}_\phi = (1-p)\mathcal{U}_\phi + p\mathcal{U}_{\phi+\pi/2} \tag{6.12}$$

where we used the fact that $Z = iU_{\pi/2}$. We could simulate the dephased gate operation using the algorithm of Section 5.3.4 with the same overhead as for the ideal gate, simply by sampling either $\mathcal{U}_\phi$ or $\mathcal{U}_{\phi+\pi/2}$ with probabilities $(1-p)$ and $p$ respectively. However, we would like to reduce the overhead by finding an optimal decomposition for the noisy channel. Single-qubit diagonal CPTP maps have a one-to-one mapping with single-qubit states in the equatorial plane of the Bloch sphere, via the relation $\Phi_{\mathcal{T}} = \mathcal{T}(|+\rangle\langle+|)$. For our channel we have

$$\begin{aligned}\Phi_{\mathcal{E}} &= (1-p)\mathcal{U}_\phi(|+\rangle\langle+|) + p\mathcal{U}_{\phi+\pi/2}(|+\rangle\langle+|) && (6.13)\\ &= (1-p)\,|\psi_\phi\rangle\langle\psi_\phi| + p\,|\psi_{\phi+\pi/2}\rangle\langle\psi_{\phi+\pi/2}| && (6.14)\end{aligned}$$

where

$$|\psi_\alpha\rangle = U_\alpha|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} e^{-i\alpha} \\ e^{i\alpha}\end{pmatrix}, \quad |\psi_\alpha\rangle\langle\psi_\alpha| = \frac{1}{2}\begin{pmatrix} 1 & e^{-2i\alpha} \\ e^{2i\alpha} & 1\end{pmatrix}. \tag{6.15}$$

So

$$\Phi_{\mathcal{E}} = \frac{1}{2}\begin{pmatrix} 1 & (1-p)e^{-2i\phi} + pe^{-2i(\phi+\pi/2)} \\ (1-p)e^{2i\phi} + pe^{2i(\phi+\pi/2)} & 1 \end{pmatrix}. \tag{6.16}$$

Consider the following ansatz for a decomposition of the Choi state,

$$\Phi_{\mathcal{E}} = q\,|\psi_\theta\rangle\langle\psi_\theta| + (1-q)\,|\psi_{\pi/4-\theta}\rangle\langle\psi_{\pi/4-\theta}|\,. \tag{6.17}$$

where $0 < q < 1$. We can assume without loss of generality that when $\Phi_{\mathcal{E}}$ is a magic state and $p > 0$, both $\phi$ and $\theta$ are in the range $(0, \pi/8)$, as all other cases are equivalent up to Clifford rotations. This decomposition is equimagical since $|\psi_\theta\rangle$ and $|\psi_{\pi/4-\theta}\rangle$ are Clifford-equivalent. We will show it is optimal with respect to channel extent. The argument will be similar to that used for density-operator extent [2] (Section 4.1.4), but the geometry here is somewhat simpler since we only need consider the equatorial plane. Any pure state $|\psi_\theta\rangle$ in the equatorial plane is Clifford-equivalent to one in the $x$-$z$ plane, via a rotation by $\pi/2$ about the $X$-axis,

$$|\psi'_\theta\rangle = e^{i\pi/4}\exp\left[\frac{-iX\pi}{4}\right]|\psi_\theta\rangle = \cos\left(\frac{\pi}{4}-\theta\right)|0\rangle + \sin\left(\frac{\pi}{4}-\theta\right)|1\rangle\,. \tag{6.18}$$

In BBCCGH [3], it was shown that states of this form have optimal decomposition

$$|\psi'_\theta\rangle = \left(\cos\left(\frac{\pi}{4}-\theta\right) - \sin\left(\frac{\pi}{4}-\theta\right)\right)|0\rangle + \sqrt{2}\sin\left(\frac{\pi}{4}-\theta\right)|+\rangle\,, \tag{6.19}$$

One can check that $H\,|\psi'_\alpha\rangle = |\psi'_{\pi/4-\alpha}\rangle$, so they have the same extent, and the optimal decomposition for $|\psi'_{\pi/4-\theta}\rangle$ is simply the expression (6.19) with $|+\rangle$ and $|0\rangle$ interchanged. So for both states the extent is

$$\xi(\psi_\theta) = \xi(\psi_{\pi/4-\theta}) = \left(\cos\left(\frac{\beta}{2}\right) + (\sqrt{2}-1)\sin\left(\frac{\beta}{2}\right)\right)^2. \tag{6.20}$$

Recall from Section 4.1.4 that an optimal witness $|\omega'\rangle$ must be tight against each term in the optimal decomposition, so if $|\langle\omega'|\psi_\theta\rangle|^2 = \xi(\psi_\theta)$, then $|\langle\omega'|0\rangle| = |\langle\omega'|+\rangle| = 1$. We have assumed that $0 < \theta \leq \pi/8$, so both coefficients in equation (6.19) are positive. It follows that $\langle\omega'|0\rangle = \langle\omega'|+\rangle$. Then since the optimal decomposition for $|\psi'_{\pi/4-\theta}\rangle$ is the same but with $|0\rangle$ and $|+\rangle$ permuted, $|\omega'\rangle$ must also be an optimal decomposition for that state. Rotating back to the equatorial plane, $|\psi_\theta\rangle$ and $|\psi_{\pi/4-\theta}\rangle$ must share an optimal witness $|\omega\rangle$. Recall from Lemma 4.9 that since $\Phi_{\mathcal{E}}$ is in the convex hull of two states that share the same optimal $\omega$-witness, it follows that $\Xi(\Phi_{\mathcal{E}}) = \langle\omega|\Phi_{\mathcal{E}}|\omega\rangle = \xi(\psi_\theta)$. Therefore the ansatz is optimal with respect to density-operator extent. But since $\Xi(\Phi_{\mathcal{E}}) \leq \Xi_*(\mathcal{E})$, it follows that the corresponding decomposition of the map $\mathcal{E}$ is optimal for channel extent.

It remains to determine the values of $q$ and $\theta$ that equations (6.17) and (6.16) are consistent. In the matrix representation, our ansatz can be written

$$\Phi_{\mathcal{E}} = \frac{1}{2}\begin{pmatrix} 1 & qe^{-2i\theta} + (1-q)e^{-2i(\pi/4-\theta)} \\ qe^{2i\theta} + (1-q)e^{2i(\pi/4-\theta)} & 1 \end{pmatrix}. \tag{6.21}$$

By equating the expression (6.16) with (6.21), we find

$$(1-p)e^{2i\phi} + pe^{2i(\phi+\pi/2)} = qe^{2i\theta} + (1-q)e^{2i(\pi/4-\theta)} \tag{6.22}$$

$$\implies \quad (1-2p)e^{2i\phi} = qe^{2i\theta} + i(1-q)e^{-2i\theta}. \tag{6.23}$$

We then solve for $q$ and $\theta$. Then, using straightforward algebra and trigonometric identities, one finds that the equation has a unique solution in the range $0 < \theta < \pi/8$,

$$\theta(\phi, p) = \frac{1}{2}\arcsin[(1-2p)\sin(2\phi+\pi/4)] - \frac{\pi}{8}. \tag{6.24}$$

We can use the solution above to check the noise level $p$ that negates all magic for the rotation $U_\phi$. The extent goes to 1 when $\theta(\phi, p) = 0$. So from equation (6.24),

$$p = \frac{1}{2} - \frac{1}{2[\sin(2\phi)+\cos(2\phi)]}. \tag{6.25}$$

In Figures 6.6b and 6.6a we plot channel extent for a range of values of $\phi$ and error rate $p$. In Figure 6.7 we compute the ratio $\Xi(\mathcal{E}_p \circ \mathcal{U}_\phi)^N/\Xi(\mathcal{U}_\phi)^N$. This quantifies the factor reduction in runtime for the bit-string simulator that we obtain by using an optimal decomposition of the dephased rotation, as compared to the naive decomposition given in equation (6.12).

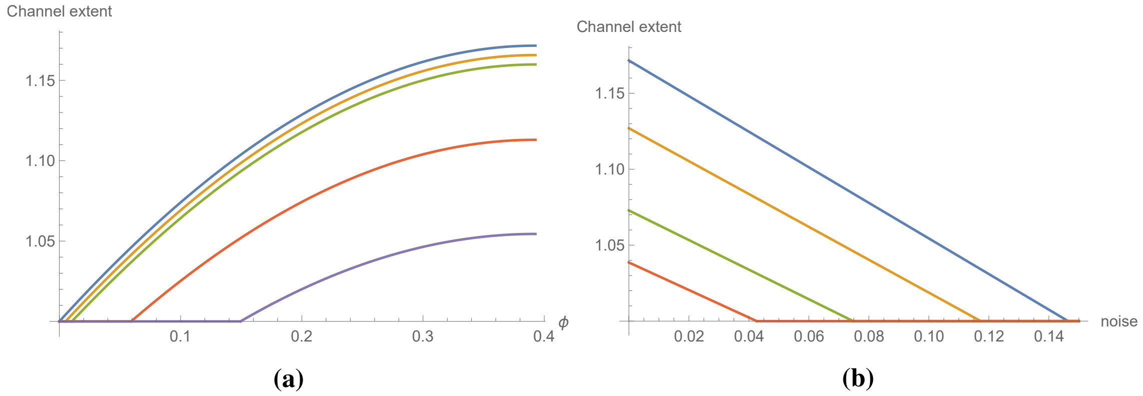


**Figure 6.6:** (a) Channel extent of dephased $Z$-rotation $\mathcal{E}_p \circ \mathcal{U}_\phi$, as a function of $\phi$, plotted for different values of the noise parameter $p$: $p = 0$ (blue), $p = 0.005$ (brown), $p = 0.01$ (green), $p = 0.05$ (red), $p = 0.1$ (purple). (b) Extent for the same channel as a function of error rate $p$, plotted for different values of $\phi$: $\phi = \pi/8$ (blue), $\phi = \pi/16$ (brown), $\phi = \pi/32$ (green), $\phi = \pi/64$ (red).

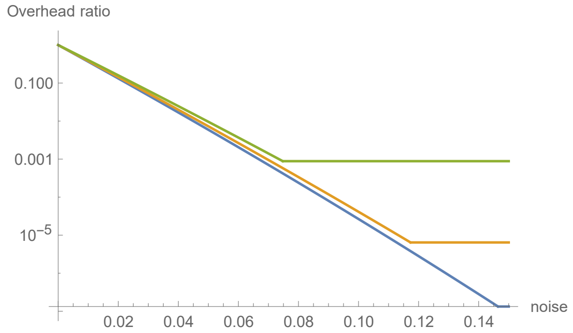


**Figure 6.7:** Here we plot the ratio $\Xi(\mathcal{E}_p \circ \mathcal{U}_\phi)^N / \Xi(\mathcal{U}_\phi)^N$ against noise parameter $p$, for $N = 100$. That is, we show the ratio of the simulation overhead for $N$ uses of the noisy channel, to that for N uses of the ideal rotation $\mathcal{U}_\phi$. The ratio is plotted against $p$ for $\phi = \pi/32$ (green), $\phi = \pi/16$ (brown) and $\phi = \pi/8$ (blue).

### 6.2.2 Decompositions for depolarised single-qubit rotations

We now consider $Z$-rotations subject to the depolarising channel,

$$\mathcal{E}_p = (1-p)\mathbb{1} + \frac{p}{3}(\mathcal{X} + \mathcal{Y} + \mathcal{Z}), \tag{6.26}$$

where $\mathcal{X}(\cdot) = X(\cdot)X$ for Pauli $X$, and so on. In this case, the channel $\mathcal{E}_p \circ U_\phi$ is not diagonal, so cannot be modelled using single-qubit states as in the previous case. However, we find that provided $U_\phi$ is a Pauli rotation, optimal decompositions for the noisy channel can be obtained numerically using a relatively modest library of unitary operations. Given any finite set of unitary operations $\mathbb{S}$, the following gives an upper bound for the

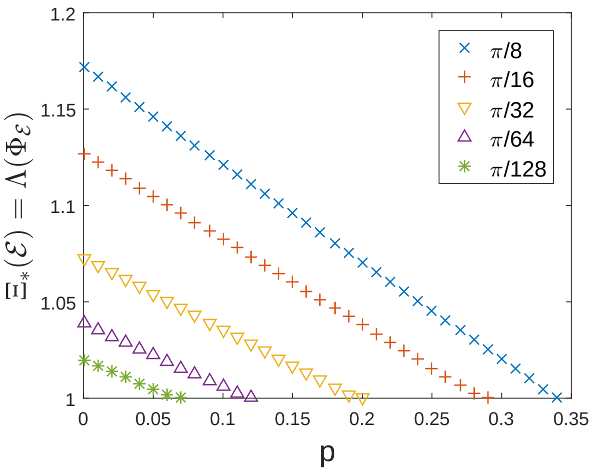


**Figure 6.8:** Channel extent for depolarised $Z$-rotations $\mathcal{E} = \mathcal{E}_p \circ \mathcal{U}_\alpha$, where $U = \exp(-iZ\alpha)$ for $\alpha \in \{\pi/8, \pi/16, \pi/32, \pi/64, \pi/128\}$. The channel extent $\Xi_*(\mathcal{E})$ is plotted against error rate $p$, and is found numerically to be equal to the dyadic negativity of the Choi state, $\Lambda(\Phi_\mathcal{E})$, up to solver precision, for all data points shown.

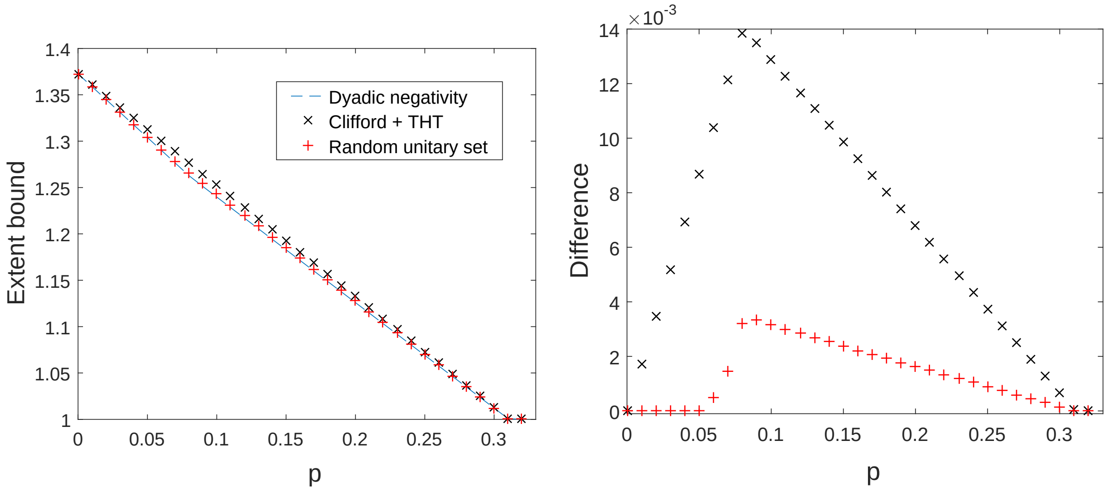
channel extent.

$$\Xi_\mathbb{S}(\mathcal{E}) = \min\{\sum_j p_j \xi(U_j) : \mathcal{E} = \sum_j p_j \mathcal{U}_j, \mathcal{U}_j \in \mathbb{S}\} \geq \Xi_*(\mathcal{E}) \tag{6.27}$$

We call this quantity the $\mathbb{S}$-extent. By Lemma 6.2, we then have that for any channel in the convex hull of $\mathbb{S}$, the stabiliser channel extent $\Xi_*$ is bounded by

$$\Lambda(\Phi_\mathcal{E}) \leq \Xi_*(\mathcal{E}) \leq \Xi_\mathbb{S}(\mathcal{E}). \tag{6.28}$$

Let $\mathbb{S}_\phi = \mathrm{Cl}_n \cup \{\mathcal{U}_\phi\}$, the set of Clifford operations supplemented by a single noiseless non-Clifford gate $U_\phi$. We compute the quantity $\Xi_{\mathbb{S}_\phi}(\mathcal{E}_p \circ \mathcal{U}_\phi)$ for $Z$-rotation $\mathcal{U}_\phi$ and depolarising noise $\mathcal{E}_p$ for a range of values of the parameters $\phi$ and $p$. We find that in all instances evaluated, we have $\Xi_{\mathbb{S}_\phi}(\mathcal{E}_p \circ U_\phi) = \Lambda(\Phi_{\mathcal{E}_p \circ U_\phi})$, up to the precision of the solver (error of less than $10^{-9}$ in the computed quantities). We plot this for a range of parameters in 6.8. From equation (6.28) this implies that the computed decompositions are optimal up to solver precision, and provides evidence that $\Xi_*(\mathcal{E}) = \Lambda_*(\mathcal{E}) = \Lambda(\Phi_\mathcal{E})$ when $\mathcal{E}$ is a depolarised rotation about a Pauli axis. We note this alignment between $\Xi_*(\mathcal{E})$ and the upper bound computed from the set $\Xi_{\mathbb{S}_\phi}$ is not obtained for rotations about non-Pauli axes. For example, for the depolarised $THT$-gate we find a gap between $\Xi$ and the upper bound based on the set $\mathrm{Cl}_1 \cup \{THT\}$. One can obtain tighter bounds by



**(a)** Bounds on channel extent.

**(b)** Gap between upper and lower bounds.

**Figure 6.9:** (a) Bounds on channel extent for depolarised $THT$-gate, $\mathcal{E} = \mathcal{E}_p \circ THT$, where $p$ denotes the error rate. The channel extent is lower bounded by dyadic negativity of the Choi state $\Lambda(\Phi_{\mathcal{E}})$ (Dashed blue line). Upper bounds are computed by searching for optimal convex mixtures derived from finite sets of unitaries. These are (i) the Clifford gates supplemented by the noiseless gate, $\mathbb{S}_{THT} = \mathrm{Cl}_1 \cup \{THT\}$ (black $\times$) and (ii) the union of $\mathbb{S}_{THT}$, Clifford orbit, and 10000 Haar-random single-qubit unitary operations (red $+$). (b) Difference between each upper bound and $\Lambda(\Phi_{\mathcal{E}})$.

optimising over a set of unitary gates generated randomly according to the Haar measure. In Figure 6.9 we show bounds on channel extent for the depolarised THT state, using (i) $\mathbb{S}_{THT} = \mathrm{Cl}_1 \cup \{THT\}$, and (ii) a set of 10000 Haar-random unitaries supplemented by $\mathbb{S}_{THT}$ and the Clifford orbit of THT. We also computed the bound using the union of $\mathbb{S}_{THT}$ and the Clifford orbit, but found this does no better than $\mathbb{S}_{THT}$ alone. In contrast adding Haar-random unitary operations to the set provides a tighter bound. The final set has 10168 elements, and the linear program remains easily tractable.

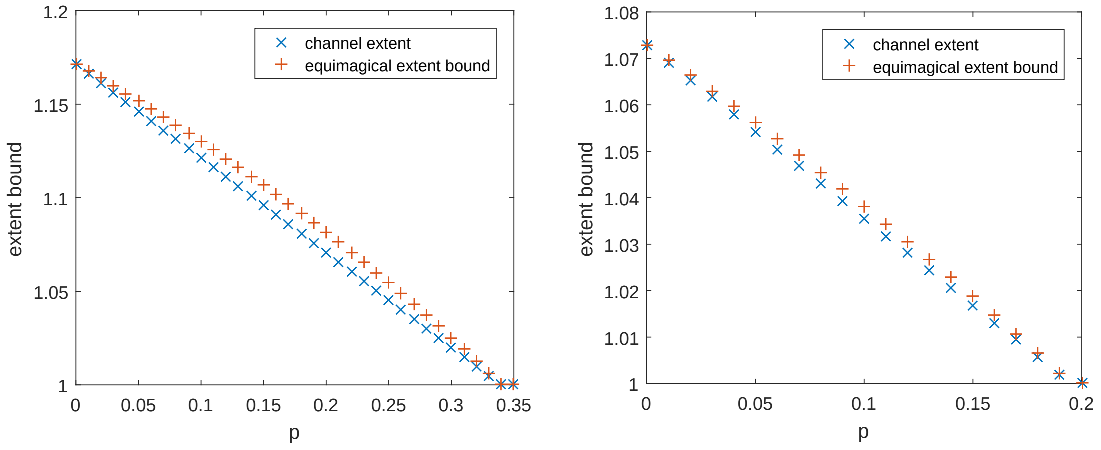
Since the decompositions obtained numerically using $\mathbb{S}_\phi$ are not equimagical, employing them in the sampling simulator (Section 5.3.4) leads to a random runtime $\tau$. For a sequence of $L$ depolarised $Z$-rotations $\mathcal{E}^{(t)}$, the mean runtime is proportional to $\prod_{t=1}^{L} \Xi_*(\mathcal{E}^{(t)})$. When drawing a single bit-string, each noisy gate contributes a factor $\mathcal{O}(1)$ to the runtime if a Clifford gate is sampled, or $\mathcal{O}\big(\xi(U_\phi)\big)$ otherwise. One may prefer to obtain a guarantee of worst-case runtime smaller than $\mathcal{O}\big(\xi(U_\phi)\big)$. To this end, we propose a heuristic algorithm (Algorithm 18) that searches for (possibly suboptimal) equimagical decompositions. The key idea is to construct the Clifford orbit of $Z$-rotation $U_\alpha$ with progressively larger extent $\xi(U_\alpha)$, and use standard convex optimisation to check whether the target channel $\mathcal{E}$ lies within the convex hull of this orbit.

The variable $E_t$ tracks a candidate value for the extent, starting with the dyadic negativity of the Choi state, and for each value of $E_t$, step 5 computes the parameter $\alpha$ that specifies a $Z$-rotation $\mathcal{U}_\alpha$ such that $\xi(\mathcal{U}_\alpha) = E_t$. The formula used in this step is straightforwardly derived from the analysis of extent given by Bravyi et al. [3]. For completeness we give the derivation in Appendix E.4. The following step computes the Clifford orbit $\mathbb{O}_t$ of this gate. Since the unitary extent is Clifford-invariant, every element of $\mathbb{O}_t$ has the same extent. In step 7, the quantity $Q$ is in effect the robustness with respect to $\mathbb{O}_t$. When this evaluates to 1, it means that the target channel is within the convex hull of the elements of $\mathbb{O}_t$. This step can be formulated as a linear program where the polytope $\mathrm{conv}(\mathbb{O}_t)$ has at most $24^2 = 576$ vertices, so can be computed relatively quickly, and it is practical to choose $\Delta$ to be fairly small. Clearly one could optimise the algorithm by adaptively changing the interval size close to $Q = 1$. If in the final iteration one has $Q = 1$ and $\alpha_t < \phi$, then the algorithm will have found a possibly suboptimal equimagical decomposition with $\sum_j p_{j,k}\xi(V_{j,k}) = \xi(|U_{\alpha_T}\rangle) = E_T < \xi(U_\phi)$.

We call the quantity $\Xi_{\mathbb{O}}(\mathcal{E})$ output from Algorithm 18 the equimagical extent bound. In practice, we find that the technique is successful in finding equimagical decompositions that are close to optimal. In Figure 6.10a we plot channel extent and the equimagical bound against the error rate $p$ for noisy $T$-gates. We find that the discrepancy between the heuristic bound and the channel extent is small ($\sim 10^{-3}$) when the error rate is very low

**Algorithm 18** Equimagical decomposition search

```
Input: Target depolarised $Z$-rotation $\mathcal{E}_p \circ \mathcal{U}_\phi$, interval $\Delta > 0$.
Output: Equimagical decomposition of the channel $\mathcal{E}$, comprising elements of some set
    $\mathbb{O}$. Computed value of the $\mathbb{O}$-extent, $\Xi_{\mathbb{O}}(\mathcal{E})$.
 1: Compute dyadic negativity of the Choi state $E_0 = \Lambda(\Phi_{\mathcal{E}})$.
 2: $t \leftarrow 0$, $Q \leftarrow 2$, $\alpha \leftarrow 0$.        ▷ We need to initialise $Q > 1$, the precise value is
    unimportant.
 3: while $Q > 1$ AND $\alpha \leq \phi$ do
 4:     $E_t \leftarrow E_0 + \Delta t$
 5:     $\alpha \leftarrow \arcsin\left(\sqrt{E_t}\sin\left(\frac{3\pi}{8}\right)\right)$ .
 6:     Compute Clifford orbit $\mathbb{O}_t = \left\{\mathcal{V}_{j,k} : V_{j,k} = C_j U_\alpha C_k, \quad C_j, C_k \in \mathrm{Cl}_1\right\}$
 7:     Compute $Q = \min\left\{\|\mathbf{q}\|_1 | \sum_{j,k} q_{j,k}\mathcal{V}_{j,k} = \mathcal{E}, \mathcal{V}_{j,k} \in \mathbb{O}_t\right\}$, storing optimal distribu-
    tion $\mathbf{q}$.
 8:     $t \leftarrow t + 1$
 9: end while
10: return $\mathbb{O}_{t-1}$, $E_{t-1}$, $\mathbf{q}$.
```



**(a)** Depolarised $T$-gate.

**(b)** Depolarised $\pi/32$-gate.

**Figure 6.10:** Stabiliser channel extent and equimagical upper bound for depolarised single-qubit gates. Magic measures plotted against error rate $p$ for (a) $T$-gate subject to depolarising noise $\mathcal{E}_p \circ T$, and (b) noisy $\pi/32$-gate, $\mathcal{E}_p \circ \exp(-i(\pi/32)Z)$.

($p \leq 1\%$) or very high ($p \geq 33\%$), just before the channel becomes stabiliser-preserving). For intermediate values of $p$, the gap between $\Xi_*$ and $\Xi_{\mathbb{O}}$ was largest ($\sim 0.0112$) at $p \approx 20\%$. For small-angle rotations the gap is narrower. Results for depolarised rotations $\exp(-i(\pi/32)Z)$ are shown in Figure 6.10b. Here the gap is maximised in the range $9\% \leq p \leq 11\%$, but only reaches $\sim 0.0026$. Since the gap is so narrow, the suboptimal equimagical decompositions produced by Algorithm 18 may be used in many practical scenarios with only minor runtime penalties, provided the amount of magic gates is not very large (for example more than $\sim 200$ $T$-gates or $\sim 700$ $\pi/32$-gates).

## 6.3 Runtime comparison between dyadic frame and stabiliser rank

In Section 5.2.3.2 we briefly discussed the fact that in addition to bit-string sampling, sparsification of density-operator extent decompositions can be used to estimate stabiliser observables (see Appendix E.3 for technical details). Given magic state $\rho$, completely stabiliser-preserving channel $\mathcal{E}$ and stabiliser projector $\Pi$, then with probability at least $(1-p_{\text{fail}})$, given optimal decomposition of the state $\rho$ we can compute an estimate $\eta$ for Born rule probability $\text{Tr}[\Pi\mathcal{E}(\rho)]$ satisfying,

$$|\eta - \text{Tr}[\Pi\mathcal{E}(\rho)]| \leq 3\varepsilon_{\text{FN}} \text{Tr}[\Pi\mathcal{E}(\rho)] + \delta' + \mathcal{O}(\delta'^2). \tag{6.29}$$

This estimate is produced in average runtime

$$\mathbb{E}(\tau_{\mathrm{SR}}) = \mathcal{O}\left(\Xi(\rho) n^3 \varepsilon_{\mathrm{FN}}^{-2} \delta'^{-3} \log\left(\frac{2}{p_{\mathrm{fail}}}\right)\right). \tag{6.30}$$

For the dyadic frame simulator with additive error $\delta$, the worst-case runtime is

$$\tau_{\mathrm{DF}} = \mathcal{O}\left(\frac{\Lambda(\rho)^2}{\delta^2} n^3 \log\left(\frac{2}{p_{\mathrm{fail}}}\right)\right). \tag{6.31}$$

Now assume that $\rho$ is a tensor product of single-qubit magic states, so that $\Lambda(\rho) = \Xi(\rho)$, and $\rho$ has an equimagical decomposition, so that the expression (6.30) gives the worst-case runtime (Theorems 4.8 and 4.15). At first sight, it appears that the stabiliser rank simulator would perform better, since it depends only linearly on the monotone value rather than quadratically. However, the stabiliser rank runtime $\tau_{\mathrm{SR}}$ has worse scaling with target precision in the estimate than the dyadic frame runtime $\tau_{DF}$. Despite the exponential scaling with monotone value, there may be parameter regimes where the dyadic frame simulator has the advantage.

The optimal choice of algorithm will depend on the fine details of the circuit to be simulated, and when the runtimes are of the same order of magnitude, constant factors and lower order terms will play a role. Here we make some simplifying assumptions to approximately characterise the regimes where one or the other simulator is favoured. We assume that we have no knowledge of the true value of $\mathrm{Tr}[\Pi\mathcal{E}(\rho)]$, so that it could take any value between 0 and 1. Then the bound on the estimate for the stabiliser rank simulator becomes

$$|\eta - \mathrm{Tr}[\Pi\mathcal{E}(\rho)]| \leq 3\varepsilon_{\mathrm{FN}} + \delta' + \mathcal{O}\left(\delta'^2\right). \tag{6.32}$$

Splitting the error budget between the two first order contributions, $\delta = 3\varepsilon_{\mathrm{FN}} + \delta'$, and optimising the runtime, we find that we should set $\delta' = (3/5)\delta$ and $\varepsilon_{\mathrm{FN}} = (2/15)\delta$. We can neglect the $n^3$ and $\log\left(2p_{\mathrm{fail}}^{-1}\right)$ factors, since these are the same for both algorithms. Absorbing all constant factors into $K_{\mathrm{SR}}$, to first order we can express the runtime for the stabiliser rank algorithm as $\tau_{\mathrm{SR}} = K_{\mathrm{SR}}\Xi(\rho)\delta^{-5}$. Meanwhile making the same simplifications for the dyadic frame simulator, we obtain runtime $\tau_{\mathrm{DF}} = K_{\mathrm{DF}}\Lambda(\rho)^2\delta^{-2}$. We have assumed that $\rho$ is a tensor product state, $\rho = \otimes_{j=1}^{n}\rho_j$, where each $\rho_j$ is a single qubit state. Then $\Lambda(\rho) = \Xi(\rho)$, and we can directly compare the runtimes, up to constant

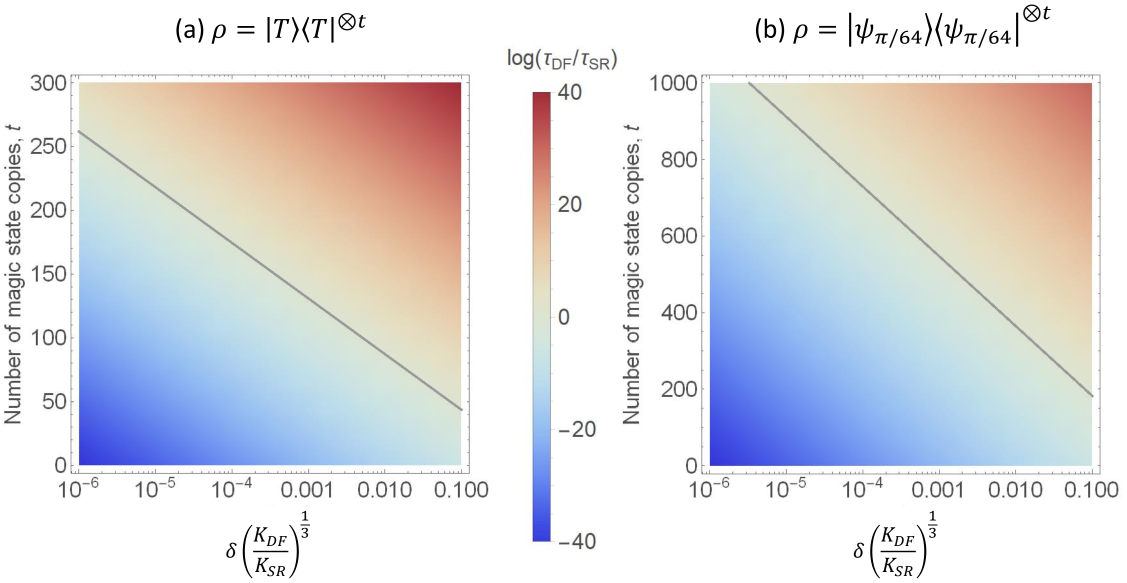


**Figure 6.11:** Runtime comparison for dyadic frame and stabiliser rank methods for initial magic state $\rho = \rho_M^{\otimes t}$, plotted against number of state copies $t$ and scaled target precision $\delta/(K_{SR}/K_{DF})^{1/3}$. Runtime ratio $\tau_{DF}/\tau_{SR}$ is plotted with logarithmic scale. (a) Comparison when each state copy is the $T$-gate resource state $\rho_M = |T\rangle\langle T|$. (b) Comparison for $\rho_M = |\psi_{\pi/64}\rangle\langle\psi_{\pi/64}|$, where $|\psi_{\pi/64}\rangle = \cos(\pi/64)|0\rangle + \sin(\pi/64)|1\rangle$.

factors. The ratio of runtimes is

$$\frac{\tau_{DF}}{\tau_{SR}} = \frac{K_{DF}\Lambda(\rho)^2\delta^{-2}}{K_{SR}\Lambda(\rho)\delta^{-5}} = \frac{K_{DF}}{K_{SR}}\Lambda(\rho)\delta^3. \tag{6.33}$$

To make the discussion more concrete we study some specific cases. Assume that the system is initialised in the state $\rho = \sigma \otimes \rho_M^{\otimes t}$, where $\sigma$ is an $(n-t)$-qubit stabiliser state $\sigma$, and $\rho_M$ are single-qubit magic states. By faithfulness and multiplicativity of dyadic negativity, we have $\Lambda(\rho) = \Lambda(\rho_M^{\otimes t}) = \Lambda(\rho_M)^t$. Consider the case where $\rho_M = |T\rangle\langle T|$, the resource state for the $T$-gate. Here we have $\Lambda(\rho) \approx (1.1716)^t$. In Figure 6.11(a) we plot the runtime ratio for up to up to 300 magic states. The horizontal axis shows target precision $\delta$, scaled by the ratio of constant factors to the power $\frac{1}{3}$. We note that the constant factors for each simulator are of similar size, so we can conservatively estimate that $0.1 \leq (K_{\mathrm{SR}}/K_{\mathrm{DF}})^{(1/3)} \leq 10$. Consequently we can expect that if we were to compute the runtime exactly, and plot the unscaled error $\delta$, the values on the horizontal axis would be shifted by no more than one order of magnitude in either direction. Despite the quadratic scaling of the dyadic frame runtime with $\Lambda$, there is a surprisingly large region where it retains the advantage over the mixed-state stabiliser rank simulator. Even for low precision ($\delta/(K_{\mathrm{SR}}/K_{\mathrm{DF}})^{1/3} \approx 0.1$), the dyadic frame simulator is favoured for circuits including up to approximately 44 copies of the $T$ state. For sharper precision, $\delta/(K_{\mathrm{SR}}/K_{\mathrm{DF}})^{1/3} \leq 0.005$, this regime extends to over a hundred copies. Compare this to the family of circuits where the single-qubit magic states are $\rho_M = |\psi_{\pi/64}\rangle\langle\psi_{\pi/64}|$, where $|\psi_{\pi/64}\rangle = \cos(\pi/64)|0\rangle + \sin(\pi/64)|1\rangle$ (Figure 6.11(b)). Then the initial state has dyadic negativity $\Lambda(\rho) \approx (1.0386)^t$. As we would expect, the region where the dyadic frame is preferred extends to a significantly greater number of magic states when the per-state negativity is small; the region reaches hundreds of state copies for any precision $\delta/(K_{\mathrm{SR}}/K_{\mathrm{DF}})^{1/3} \leq 0.1$.

Since the channel extent simulator (Section 5.3.4) works on the same principles as the mixed-state stabiliser rank simulator, it can be adapted in exactly the same way to estimate Born rule estimation probabilities, with the same runtime scaling as equation (6.29), except that the density-operator extent $\Xi(\rho)$ is replaced by the product of the channel extent for each circuit element, $\prod_{t=1}^{L} \Xi_*(\mathcal{E}^{(t)})$. One can then carry out a similar comparison with the dyadic channel simulator. We omit a full analysis here, as a sequence of $t$ $Z$-rotations $\exp(-iZ\alpha)$ decomposed individually results in the same numerics as $t$ copies of $|\psi_\alpha\rangle = \cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle$. We leave more fine-grained analyses of specific circuit instances for future work.

## 6.4 Summary and conclusions

We began this chapter by highlighting Pauli axis rotations as an important primitive for quantum simulation. While long sequences of fault-tolerant logical gates are needed for conventional quantum simulation methods such as estimation of energy spectra via phase estimation, recent variational algorithms designed for NISQ-era devices only require low depth circuits, and errors are mitigated against rather than corrected. The analysis of noisy rotation gates using our framework is a first step towards classical simulation of such NISQ-tailored algorithms.

We first evaluated our RoM-based channel monotones for amplitude-damped gates, showing that our simulators achieve significant speedup over the Oak Ridge simulator [45]. We computed the same monotones for multi-qubit controlled-phase gates and random diagonal gates to investigate the bounds derived analytically in Chapter 4, finding them to be tight in a number of scenarios. We conjectured that for controlled-phase gates

$M_{t,n}$ from the family defined in Section 6.1.3, the capacity to generate magic does not increase indefinitely as the number of qubits $n$ increases. Instead we expect that the capacity is capped at some finite $n$, and we provided arguments and numerical evidence for this claim.

In Section 6.2 we showed how to obtain optimal decompositions analytically for dephased Pauli rotations, and using heuristic search methods for depolarised gates. We showed that both cases also yield Clifford dyadic decompositions that are optimal (up to solver precision) with respect to the dyadic channel negativity. These channel decompositions can be used directly in classical simulations. We then compared the runtime $\tau_{\mathrm{SR}}$ of the density-operator extent simulator with that of the dyadic frame simulator, $\tau_{\mathrm{DF}}$, for estimation of Born rule probabilities. Despite $\tau_{\mathrm{DF}}$ having a poorer scaling with the amount of magic, the dyadic frame is faster at producing high precision estimates if the number of qubits is not too high. For large numbers of highly magic states, stabiliser rank simulators retain the advantage. These results can be straightforwardly generalised to channel simulators.

There are of course myriad reasons for selecting one simulator over another. The dyadic frame algorithm is limited to additive error estimates, and as such is unsuitable for bit-string sampling; stabiliser rank methods are better suited to that task. However, the stabiliser rank simulator for channels (Section 5.3.4) can be used only when each magic channel is expressible as a convex combination of unitary operations. So if the task is to estimate an observable for a circuit involving a non-unital noise model, for example, the dyadic channel or channel robustness simulator must be used. Since computing optimal dyadic channel decompositions is difficult in the general case, it may be beneficial to hybridise the method with easier to compute channel robustness decompositions. The two types of decomposition are fully compatible, since the completely stabiliser-preserving channels are a strict subset of the dyadic stabiliser channels. We anticipate that the tools and results presented in this chapter for computing the resource value of elementary operations will be useful for developing highly accurate estimates of classical simulation overhead for specific circuit instances in future work.

# Chapter 7

# Conclusions

In the first part of this thesis, we set out the prior state-of-the-art in the resource theory of magic and stabiliser-based classical simulation techniques for $n$-qubit circuits, and the connections between the two. Previous classical simulators were largely divided into those based on sampling from quasiprobability distributions [45, 103, 124], and those based on stabiliser rank decompositions of pure states or unitary operations [3, 59, 60]. In the past, some similarities had been noted between the two strands; for example, $\ell_1$-minimisation is used to compute resource monotones associated with both quasiprobability and stabiliser rank methods [3, 103]. However, beyond these isolated observations, the two types of simulator were previously treated as unrelated in the literature, and the connection between them was not understood. The previous literature also lacked comprehensive study of the classical simulation of general quantum channels using stabiliser techniques. Amongst authors on quasiprobability methods, Bennink et al. [45] and Rall et al. [142] had designed algorithms that would admit general CPTP maps. A natural expectation for a stabiliser-based simulator of general $n$-qubit circuits is that it should be able to efficiently simulate any sequence of few-qubit stabiliser channels, whereas for each of these algorithms there exist relatively simple stabiliser operations that can only be simulated inefficiently. Meanwhile, Bravyi et al. [3] introduced powerful sparsification techniques for simulating pure magic states and unitary non-Clifford operations with reduced runtime, but did not show how the simulator could be extended to noisy circuits.

In the main part of the thesis, we presented our novel techniques and results, addressing the issues described above. In Chapter 4 we made several contributions to the resource theory of magic, focusing on resource monotones that quantify the performance of classical algorithms we introduce later in the thesis. We first discussed a new family of magic state monotones, elucidating the link between quasiprobability decompositions of density operators and stabiliser rank decompositions of pure state vectors. In particular we saw that for pure states, the generalised robustness and dyadic negativity, which are closely related to quasiprobability simulators, are equal to the stabiliser extent, which is directly connected to the sparsification technique introduced by Bravyi et al. [3] for stabiliser rank simulation. The new state monotones are multiplicative for single-qubit mixed states, and pure states on up to three qubits, making them efficiently computable for many important cases. In the rest of the chapter we extended the framework to the channel picture, defining magic monotones for general CPTP multi-qubit maps. These can broadly be divided into monotones directly based on channel decompositions of various kinds, and those that measure the maximum increase in magic over all input states. We studied the relationships between these monotones and presented techniques for easing their computation in the case of diagonal channels.

We began Chapter 5 by introducing a subroutine for simulation of completely stabiliser-preserving channels, showing that they are efficiently simulable subject to reasonable constraints on the number of Kraus operators. This had not previously been proven in the literature, and recent work due to Heimendahl, Heinrich and Gross has shown that our result is strictly stronger than the Gottesman-Knill theorem [63]. We then studied the connection between magic resource theory and classical simulation by developing a number of algorithms whose performance was measured by the quantifiers of magic introduced in the previous chapter. We first presented a trio of simulators associated with the new magic state monotones described in Section 4.1. These lend practical significance to the fact that the new monotones are significantly tighter than robustness of magic; for example our dyadic frame simulator is provably faster than previously known qubit quasiprobability simulators for product states, by a factor exponential in the number of magic states. In addition, we modified the sparsification technique introduced by Bravyi et al. [3], yielding a signficant reduction in runtime for bit-string simulation, and extending the method to mixed states.

We then introduced a suite of simulation methods for CPTP maps on qubits. We divided these into static simulators, based on stabiliser decompositions of channels that can be precomputed prior to running the simulation proper, and dynamic simulators, associated with our capacity monotones, where few-qubit magic state decompositions are computed on the fly as the algorithm steps through the circuit. The latter can achieve reduced sample complexity for circuits with many highly non-stabiliser channels, at the cost of increased runtime per sample. Our channel simulators open the way for classical simulation of NISQ-era quantum circuits, including noise models, using stabiliser-based methods, as well as reducing overhead for non-Clifford operations by avoiding gadgetisation.

A key advantage of classical algorithms associated with computable magic monotones is that one can meaningfully estimate classical simulation overhead for a given circuit, or family of circuits, without having to run the simulation. This has clear practical benefits since we can separate circuits into those which are tractable or intractable given particular classical computing resources, and can provide insight into which types of circuit are suitable candidates for demonstrating quantum advantage. In Chapter 6 we discussed such estimates for a number of practically important classes of quantum operation. We showed how to obtain optimal decompositions for Z-rotations subject to dephasing or depolarising noise, which can be used in simulated sampling scenarios.

We ended the chapter by showing that while performance scaling with monotone value plays a central role in determining which simulator is the best choice for a given circuit, other factors such as target precision also come into play. In particular we compared dyadic frame and stabiliser rank techniques for estimating observables for circuits using many copies of single-qubit magic states. In this setting, the results of Section 4.1 imply that the respective monotones quantifying runtime for each simulator take the same value. The stabiliser rank simulator is preferred for circuits consuming very large amounts of magic, since it is quadratically faster in the monotone value. However, we showed that when high-precision estimates are required, there is nevertheless a significant parameter regime comprising intermediate-sized circuits where the dyadic frame simulator will have the advantage, due to its more modest error scaling.

## 7.1 Outlook and future work

The work presented in Chapters 4 and 5 leaves open a number of directions for future investigation. In Section 4.2.3 we introduced the stabiliser channel extent, and explained that while it is defined only for CPTP maps that can be expressed as a convex mixture of unitary operations, it nevertheless has practical application, for example, in the simulation of unitary gates subject to depolarising or dephasing noise, as studied in Chapter 6. This type of decomposition is of particular interest, as it can be used to simulate sampling from the output distribution of a quantum circuit (Section 5.3.4). This is in contrast to the dyadic channel simulator (Section 5.3.3), which is only able to estimate Born rule probabilities up to additive error. It would be useful to extend the definition of the channel extent to admit decomposition of general CPTP maps. This would allow simulation of bit-string sampling for circuits involving non-Clifford gates subject to non-unital noise models such as amplitude-damping, or those involving non-stabiliser measurements with adaptivity.

Most of the channel monotones we have discussed can be computed exactly for general few-qubit CPTP maps, along with the associated channel decompositions. The channel extent is one exception, and the dyadic channel negativity may be tractable only for single-qubit maps. We showed in Sections 6.2.1 and 6.2.2 that we can compute optimal channel extent solutions for depolarised or dephased phase gates, which also yield optimal dyadic channel decompositions. However, it is not known how to efficiently compute optimal decompositions for more general channels. In Section 4.2.5 we proposed a method to upper bound the dyadic channel negativity using dyadic projective channels. It is an open question whether bounds computed in this way are tight, and evaluating even this bound becomes intractable for two-qubit channels. Methods to compute these quantities more efficiently would allow classical simulation of a larger class of quantum circuit. One route to achieving this could be to use symmetry reduction techniques similar to those Heinrich and Gross [133] for the case of RoM; such techniques can likely also be applied for other magic monotones. For example, dyadic negativity for states can currently only be computed on a standard desktop or laptop for up to three qubits, and the channel robustness can only be evaluated for two-qubit channels, unless the channel can be diagonalised by Clifford gates. Due to the submultiplicativity of the magic monotones considered, the ability to find optimal decompositions for larger numbers of qubits could reduce simulation overhead, as it would allow gates acting in parallel or in sequence to be decomposed in a single block.

Ultimately, stabiliser-based classical simulators will find their main application in the simulation of realistic intermediate-sized circuits. In this thesis, we have focused on developing the resource-theoretic foundations upon which simulators can be built, leaving implementation for future work. During the final stages of the preparation of this thesis, several papers have emerged demonstrating how stabiliser-based simulation methods can be employed in practical settings [187, 188], and pointing the way to new directions of enquiry for improving performance. This research direction opens up the possibility of simulating error-correction protocols implemented on real-world quantum computing hardware where noise may look very different to the depolarising or dephasing noise model often assumed in earlier work. Other sub-circuits important for fault-tolerance include magic state distillation and injection protocols based on realistic chip layouts [189] and these are also amenable to simulation using stabiliser methods. For these reasons, we believe that stabiliser-based simulation techniques will remain relevant beyond the NISQ era, and further refinement and optimisation of these techniques will be highly beneficial in extending their reach. We expect that accurate simulations using realistic noise models will be invaluable in helping to inform design decisions for the fault-tolerant quantum computing architectures of the future.

# Appendix A

# Clifford circuits

## A.1 Generating the Clifford group

Let $\langle \mathcal{G} \rangle$ be the group generated by the gate set, $\mathcal{G} = \{H, S, CNOT\}$ and let $\langle \mathcal{G}_n \rangle$ be the restriction to $n$-qubit circuits. We want to show that $\langle \mathcal{G}_n \rangle \equiv \mathrm{Cl}_n$, where $\mathrm{Cl}_n$ is the $n$-qubit Clifford group. It is manifestly clear that $\mathcal{G}_n \subseteq \mathrm{Cl}_n$. We now argue that the sets are equivalent by showing that any $U \in \mathrm{Cl}_n$ can be decomposed as a sequence of gates from $\mathcal{G}$. We first prove a simple lemma for the two-qubit case.

**Lemma A.1.** *For any two-qubit Pauli operator with real sign, $P \otimes Q \in \mathcal{P}_{2,\pm}$, (except $P \otimes Q = \pm \mathbb{1}_2$) there exists $V \in \langle \mathcal{G}_2 \rangle$ such that $V(P \otimes Q)V^\dagger = Z \otimes \mathbb{1}$. Similarly there always exists $V' \in \langle \mathcal{G}_2 \rangle$ such that $V'(P \otimes Q)V'^\dagger = X \otimes \mathbb{1}$.*

*Proof.* First we consider the case that $P \neq \mathbb{1}$ and $Q \neq \mathbb{1}$. From equations (1.37) and (1.38) we can map $HXH = Z$ and $SYS^\dagger = X$, so we can always map $P \otimes Q$ to $\pm Z \otimes Z$. Now we can apply $CNOT$ from 2 to 1, $C_{2\to1}$, to obtain $\pm Z \otimes \mathbb{1}$.

Next consider the case when $Q = \mathbb{1}$. In this case we can simply apply $H$ or $S$ to obtain $\pm Z \otimes \mathbb{1}$. Finally, if $P = \mathbb{1}$, we first map $\mathbb{1} \otimes Q \to \pm \mathbb{1} \otimes Z$, then apply $C_{1\to2}$ followed by $C_{2\to1}$ to map $\pm \mathbb{1} \otimes Z \to \pm Z \otimes Z \to \pm Z \otimes \mathbb{1}$. Now note that $S^2 = Z$, and $HS^2H = X$, so that $X$ and $Z$ are themselves in $\langle \mathcal{G}_1 \rangle$. Then if we have obtained a negative sign on the $Z$ operator we append $X$ to our sequence of gates to flip the sign,

$$(X \otimes \mathbb{1})(-Z \otimes \mathbb{1})(X^\dagger \otimes \mathbb{1}) = (Z \otimes \mathbb{1})(X^2 \otimes \mathbb{1}) = Z \otimes \mathbb{1}. \tag{A.1}$$

This proves the result for $Z \otimes \mathbb{1}$. Then we can simply apply the Hadamard to the first qubit to obtain $X \otimes \mathbb{1}$. □

Now we prove the general result.

**Theorem A.2** (Generating the Clifford group)**.** *The Clifford group is generated by* $\mathcal{G} = \{H, S, CNOT\}$.

*Proof.* Since the Pauli operators form a complete operator basis, any unitary transformation is fully specified (up to a phase) by its action on this basis. But from the earlier discussion, we know that a Clifford update to any Pauli operator is fixed by how the unitary operation maps each $Z_j$ and $X_j$ [30]. We can therefore uniquely (up to a phase) describe any Clifford unitary $U \in \mathrm{Cl}_n$ by a list of Pauli operators,

$$\mathbb{L}_U = \{P_1, P_2, \ldots, P_n; Q_1, Q_2, \ldots, Q_n\}, \tag{A.2}$$

$$\text{where} \quad P_j = UZ_jU^\dagger, \quad Q_j = UX_jU^\dagger \tag{A.3}$$

Note that since unitary operations preserve hermiticity, $P_j$, $Q_j$ cannot have imaginary phase. We will show that for any $\mathbb{L}_U$ for $U \in \mathrm{Cl}_n$ there exists a circuit $V \in \langle \mathcal{G}_n \rangle$ that maps each $P_j$ to $Z_j$ and each $Q_j$ to $X_j$. It then follows that $\mathbb{L}_{V^\dagger} = \mathbb{L}_U$, and hence $U = V^\dagger$, up to a phase. But $V^\dagger \in \langle \mathcal{G}_n \rangle$, which will prove that any $U \in \mathrm{Cl}_n$ is in the group generated from the gate set $\mathcal{G}$.

First, consider $P_1 = P_1^{(1)} \otimes P_1^{(2)} \otimes \ldots \otimes P_1^{(n)}$, where $P_1^{(j)}$ is some single-qubit Pauli operator acting on the $j$-th qubit. We will consider in turn each pair $P_1^{(j-1)} \otimes P_1^{(j)}$, beginning with $j = n$. By Lemma A.1, we can use gates from $\langle \mathcal{G}_n \rangle$ to map $P_1^{(j-1)} \otimes P_1^{(j)} \to Z \otimes \mathbb{1}$. Repeating this process for each pair of qubits, we are able to find a sequence of gates $V_1$ such that $P_1' = V_1 P_1 V_1^\dagger = Z_1$. Now this circuit will have mapped $Q_1$ to some other Pauli operator $Q_1' = V_1 Q_1 V_1^\dagger$, but is constrained by commutation relations. Recall the commutation relations for $Z_j$ and $X_k$,

$$[Z_j, Z_k] = 0, \quad [X_j, X_k] = 0, \quad \{Z_j, X_j\} = 0 \quad \forall j,k \tag{A.4}$$

$$[Z_j, X_k] = 0 \quad j \neq k. \tag{A.5}$$

That is, $Z_j$ and $X_k$ anticommute if and only if $j = k$. Since unitary operations preserve commutation relations, analogous relations must hold for the $P_j$ and $Q_j$, and any subsequent unitary transformations. Therefore $P_1' = Z_1$ and $Q_1'$ must anticommute. Since $Z_1$ has the identity on every qubit except the first, $Q_1'$ must have either $\pm X$ or $\pm Y$ in position

1. If $\pm Y$, we can simply apply the $S$ gate, which maps $Y$ to $X$ but leaves $Z$ invariant, so without loss of generality, we can assume that $Q'_1 = \pm X \otimes Q'^{(2)}_1 \otimes \ldots Q'^{(2)}_n$. Then applying Lemma 1 repeatedly as we did for $P_1$, we obtain, for some $V'_1 \in \langle \mathcal{G}_n \rangle$,

$$V'_1 Q'_1 V_1 = X \otimes X \otimes \mathbb{1}_{n-2}. \tag{A.6}$$

Note that since $V'_1$ acts non-trivially only on the last $n-1$ qubits, $Z_1$ is left unchanged. Then we apply the $CNOT$ from 1 to 2, so that we have,

$$C_{1\to 2} Z_1 C^{\dagger}_{1\to 2} = Z_1, \quad C_{1\to 2} X_1 X_2 C^{\dagger}_{1\to 2} = X_1. \tag{A.7}$$

Thus we have found a sequence of gates $V''_1 \in \langle \mathcal{G}_n \rangle$ that maps $P_1$ to $Z_1$ and $Q_1$ to $X_1$. Now, all the other Pauli operators in the list $\mathbb{L}_U$ will have been transformed as well, $P_j \to P''_j$, $Q_j \to Q''_j$. But the commutation relations must still hold, so that each $P''_j$ must commute with both $Z_1$ and $X_1$. This is only possible if they all have the identity in the first position, $P''_j = \mathbb{1} \otimes \ldots$, and similarly we must have $Q''_j = \mathbb{1} \otimes \ldots$. Therefore any future operations need only be applied to the last $n-1$ qubits, leaving $Z_1$ and $X_1$ invariant. We then follow the same procedure to map $P_2 \to Z_2$ and $Q_2 \to X_2$ using only gates from $\langle \mathcal{G}_n \rangle$. We repeat this iteratively for all qubits until we have found a sequence of gates $V = V''_n \ldots V''_2 V''_1 \in \langle \mathcal{G}_n \rangle$ that maps all $P_j \to Z_j$ and $Q_j \to X_j$. But we earlier argued that finding such a sequence implies that $V^{\dagger} = U$ up to a phase, so this completes the proof. □

## A.2 Canonical forms

Since any stabiliser group can be prepared using Clifford operations, any stabiliser state can be specified by giving the (non-unique) Clifford circuit $U$ that prepares it from a standard initial state,

$$|\phi\rangle = U |0^n\rangle. \tag{A.8}$$

It is useful to decompose Clifford circuits [31, 61] in a canonical form. This is clearly relevant to the synthesis of more general circuits, but also has application in classical simulation. In Ref. [31], Aaronson and Gottesman showed that any unitary Clifford operation can be decomposed as an 11-layer circuit of the form H-C-P-C-P-C-H-P-C-P-

C, where H means a layer of Hadamards, P is a layer of single-qubit phase gates, and C is a layer of CNOT gates. Hadamard and phase layers contain at most $\mathcal{O}(n)$ gates, since $S^4 = H^2 = \mathbb{1}$. It was previously known that for any CNOT circuit on $n$ qubits, there is an equivalent CNOT circuit consisting of at most $\mathcal{O}\big(n^2/\log(n)\big)$ CNOT gates [190]. Applying this to Aaronson and Gottesman's construction, it follows that Clifford circuits can be put in canonical form comprising $\mathcal{O}\big(n^2/\log(n)\big)$ gates from $\{H, S, CNOT\}$. While this asymptotic scaling can be shown to be optimal with respect to gate count, other authors have sought to improve the construction in various ways [58, 62]. For example, Maslov et al. [62] reduced the number of layers to 7 by allowing layers of $CZ$ gates, so that the canonical form is C-CZ-P-H-P-CZ-C, leading to circuits with two-qubit gate depth of $14n-4$. Van den Nest [61] showed that any stabiliser state $|\phi\rangle$ can be prepared with a circuit of the form H-C-P-CZ, but did not give an algorithm for computing this decomposition for an arbitrary $|\phi\rangle$. Garcia, Markov and Cross [58] gave an explicit algorithm for computing a 5-layer preparation circuit of form H-C-CZ-P-H, given the stabiliser generators for $|\phi\rangle$, leading to an efficient algorithm for computing stabiliser inner products.

## A.3 CH-form and Clifford simulation

Here we review a standard form for representing stabiliser states introduced by Bravyi et al. in Ref. [3], and sketch how it can be used for classical simulation tasks. The CH-form builds on the work on canonical Clifford circuits described above, but is geared toward a compact data format for classical simulation of stabiliser circuits, rather than Clifford circuit synthesis. Using results from the sequence of papers described in the previous paragraph, Bravyi et al. note that any $n$-qubit stabiliser state preparation can be decomposed as a computational basis state preparation $|\mathbf{s}\rangle$, followed by a layer of Hadamard-type ("H-type") gates, of the form $U_H = \bigotimes_{j=1}^{N} U_j$, where $U_j$ is either the identity or the Hadamard, and finally a layer of control-type ("C-type") gates, which are those Clifford operations satisfying $U_C\,|0^n\rangle = |0^n\rangle$. For full generality, Bravyi et al. include a complex number $\omega$, so that any vector proportional to a stabiliser state can be expressed $|\phi\rangle = \omega U_C U_H\,|s\rangle$. The Hadamard-type gates can be described by a binary vector $\mathbf{v} \in \{0,1\}^n$ specifying on which qubits to apply a Hadamard, so that,

$$U_H = H^{v_1} \otimes H^{v_2} \otimes \ldots \otimes H^{v^n}. \tag{A.9}$$

The C-type operators are specified by stabiliser tableaux specifying how $U_C$ maps each single-qubit $X$ and $Z$ operator. Note that since any C-type unitary $U_C$ leaves invariant the state $|0^n\rangle$, which has stabiliser group $\mathcal{S} = \langle Z_1, Z_2, \ldots, Z_n\rangle$, $Z$ operators are never mapped to $X$ or $Y$ operators. If we express the transformations as

$$U_C^\dagger Z_p U_C = \prod_{j=1}^{n} Z_j^{G_{p,j}}, \quad U_C^\dagger X_p U_C = i^{\gamma_p} \prod_{j=1}^{n} X_j^{F_{p,j}} Z_n^{M_{p,j}}, \tag{A.10}$$

then $U_C$ can be encoded in a tableau comprising three $n \times n$ binary matrices $G$, $F$ and $M$, and an $n$-element $\mathbb{Z}_4$-valued vector $\boldsymbol{\gamma}$. Therefore any stabiliser state can be specified by the data format $(F, G, M, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{s}, \omega)$. The space requirement is then $3n^2 + 4n + \mathcal{O}(1)$ classical bits, where the constant term comes from the space needed to store one complex number to some fixed precision.

An advantage of the CH-form is that the updates for the $S$, $CZ$ and $CNOT$ gates are particularly straightforward to compute. If $\Gamma$ is one of these three gates, then we can simply write $\Gamma|\phi\rangle = \omega(\Gamma U_C)U_H|\phi\rangle$, so that $U_C' = \Gamma U_C$, and $F$, $G$, $M$, and $\boldsymbol{\gamma}$ are updated using tableau operations similar to those already described, in time $\mathcal{O}(n)$.

Hadamard gates require a few more steps. First, a Hadamard $H_p$ applied to qubit $p$ is written in the Pauli basis as $H_p = (X_p + Z_p)/\sqrt{2}$. This allows it to be commuted through the $U_C$ layer using the known relations (A.10), eg. $Z_p U_C = U_C \prod_{j=1}^{n} Z_j^{G_{p,j}}$. Thus $H_p$ is transformed to a linear combination of two Pauli operators, determined by $\gamma_p$ and the $p$-th row of $F$, $G$ and $M$. Then this operator is commuted through the H-layer using the somewhat simpler relations $Z_p H_p = H_p X_p$. This first step leaves $\omega$, $U_C$ and $U_H$ unchanged, but has the effect of splitting $|\mathbf{s}\rangle$ into a superposition of two terms,

$$H_p(\omega U_C U_H)|\mathbf{s}\rangle = \omega U_C U_H \left( \frac{(-1)^\alpha |\mathbf{t}\rangle + i^{\gamma_p}(-1)^\beta |\mathbf{u}\rangle}{\sqrt{2}} \right), \tag{A.11}$$

where $\alpha$, $\beta$, $\mathbf{t}$ and $\mathbf{u}$ depend on $p$ and the initial data $(F, G, M, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{s})$. For the case where $\mathbf{t} = \mathbf{u}$, any phase is simply absorbed into $\omega$. Bravyi et al. [3] show that when $\mathbf{t} \neq \mathbf{u}$, the superposition has a CH-form,

$$U_H(|\mathbf{t}\rangle + i^{\delta}|\mathbf{u}\rangle) = \omega' W_C W_H |\mathbf{s}'\rangle \tag{A.12}$$

where $W_C$ is composed only of gates from $\{S, CZ, CNOT\}$, and which can be computed in time $\mathcal{O}(n)$. Then $W_H$ becomes the new H-layer, and $W_C$ is combined into the C-layer using standard tableau rules. There are at most $\mathcal{O}(n)$ elementary gates in $W_C$, and each tableau update takes $\mathcal{O}(n)$ time, so the total time to compute the updated data for a Hadamard gate is $\mathcal{O}(n^2)$.

To compute the inner product $\langle \mathbf{x}|\phi\rangle$, first note that $|\mathbf{x}\rangle = \prod_{j=1}^{n} X_j^{x_j}|0^n\rangle$. Then recall that C-type operators are defined by the relation $U_C|0^n\rangle = |0^n\rangle$, so we can write

$$\langle \mathbf{x}|\phi\rangle = \langle 0^n|\,(\prod_{j=1}^{n} U_C^{\dagger} X_j U_C) U_H |\mathbf{s}\rangle\,, \tag{A.13}$$

where each $U_C^{\dagger} X_j U_C$ is known from the initial data $(F, G, M, \boldsymbol{\gamma})$ (see equation (A.10)). The problem then reduces to a product of $n$ single-qubit inner products of the form either $i^{\mu_j}\langle u_j|s_j\rangle$, where $u_j, s_j \in \{0,1\}$, or $i^{\mu_j}\langle u_j|\pm\rangle$. This can be computed in time $\mathcal{O}(n^2)$. The expression (A.13) can also be used to infer the necessary probabilities to draw a string $\mathbf{x}$ with probability $|\langle \mathbf{x}|\phi\rangle|^2$, but we omit the details here [3].

Finally, a Pauli projection $\Pi = (\mathbb{1}_n + P)/2$ can be commuted through the C- and H-layers in a similar way to the Pauli decomposition of the Hadamard, again in time $\mathcal{O}(n^2)$. This completes the list of stabiliser operations.

## Appendix B

# Further results on stabiliser-preserving maps

In Chapter 1 we defined the set of completely stabiliser-preserving CPTP channels. These were shown in [1] to be exactly those maps $\mathcal{E}$ whose Choi state $\Phi_{\mathcal{E}}$ is a stabiliser state that satisfies $\mathrm{Tr}_A(\Phi_{\mathcal{E}}) = \mathbb{1}/d$, where $d$ is the dimension of the system. The second condition is just the generic requirement for a map to be trace-preserving, and by removing this condition, one is left with the maps that are not necessarily trace-preserving, but completely stabiliser-preserving in the sense that for all $n+k$ qubit stabiliser states $\rho$, and for any $k$, $\mathcal{E} \otimes \mathbb{1}_k(\rho)$ is proportional to a stabiliser state. In Chapter 4 we defined channel monotones with respect to the set of completely stabiliser-preserving maps $\mathrm{SP}_{n,n}$, and the efficient simulability of maps in $\mathrm{SP}_{n,n}$ is key to the algorithms presented in Chapter 5. However, alternative perspectives on what constitutes a stabiliser operation have been proposed, and in this appendix we give some examples and results related to these approaches.

## B.1 SP but not completely SP maps

Beyond the well-known set of standard stabiliser operations (Clifford gates, Pauli measurements and classical co-processing), the simplest notion of a stabiliser map is perhaps the set $\mathrm{SP}_{n,0}$ first mentioned in Chapter 1. This is the set of maps $\mathcal{E}$ that are stabiliser-preserving on an $n$-qubit system in the sense that $\mathcal{E}(\rho) \in \mathrm{STAB}_n$ for all $\rho \in \mathrm{STAB}_n$. It is well-known that there exist maps $\mathcal{T}$ that are positive but not completely positive [4], so that $\mathcal{T}(A)$ remains positive for any positive operator $A$, while the same does not hold for $\mathcal{T} \otimes \mathbb{1}(A)$. Famously, the transpose is one such positive but not completely map, and the negative partial transpose can be used as a test for witnessing entanglement [191, 192]. Here we show that there exist analogous stabiliser-preserving but not completely-stabiliser-preserving maps.

Consider the single-qubit channel $\mathcal{E}_T$ defined by the Kraus operators $\{|0\rangle\langle T|, |1\rangle\langle T_\perp|\}$, where $|T\rangle = T|+\rangle$ and $|T_\perp\rangle = T|-\rangle$. Operationally, this channel could be implemented by measuring in the basis $\{|T\rangle, |T_\perp\rangle\}$, then resetting the state to $|0\rangle$ or $|1\rangle$ conditioned on the outcome. Clearly, applied to any single-qubit state, the output will be some probabilistic mixture of $|0\rangle$ and $|1\rangle$, and so must have $\mathcal{R} = 1$, so $\mathcal{E}_T \in \mathrm{SP}_{1,0}$. But if $\mathcal{E}_T$ is applied to one qubit in a Bell pair, we obtain:

$$(\mathcal{E}_T \otimes \mathbb{1})(|\Phi_+\rangle\langle\Phi_+|) = \frac{1}{2}(|0T^*\rangle\langle 0T^*| + |1T_\perp^*\rangle\langle 1T_\perp^*|), \tag{B.1}$$

where $|T^*\rangle = T^\dagger|+\rangle$, $|T_\perp^*\rangle = T^\dagger|-\rangle$. From this output state, we can deterministically recover a pure magic state on qubit 2 using only stabiliser operations, by making a $Z$-measurement on qubit 1 and then performing a rotation on qubit 2 conditioned on the outcome. The output state has robustness $\mathcal{R}((\mathcal{E}_T \otimes \mathbb{1})(|\Phi_+\rangle\langle\Phi_+|)) = \mathcal{R}(|T\rangle) = \sqrt{2}$. This example easily extends to show that for any $n$, there exist maps that are in $\mathrm{SP}_{n,0}$ but not $\mathrm{SP}_{n,1}$ (and hence not in $\mathrm{SP}_{n,n}$). Strikingly, this example shows that given access to $\mathrm{SP}_{1,0}$ as well as the standard stabiliser operations, we can deterministically prepare a perfect magic state, and then perform state injection to implement the $T$ gate. This would allow universal quantum computation. We can summarise the result in the following theorem.

**Theorem B.1.** *The set* $\mathrm{SP}_{n,0}$ *is strictly larger than* $\mathrm{SP}_{n,n}$. *Moreover provided we have access to a system with at least* $n+1$ *qubits, the standard stabiliser operations are promoted to universality by* $\mathrm{SP}_{n,0}$.

Of course, the standard stabiliser operations are themselves contained in $\mathrm{SP}_{n,0}$. For state injection, we need at least one entangling Clifford gate, which would be contained in $\mathrm{SP}_{n,0}$ for $n \geq 2$. Therefore we have the following corollary.

**Corollary B.2.** *For* $n \geq 2$, *access to* $\mathrm{SP}_{n,0}$ *allows universal quantum computation for systems with at least* $n+1$ *qubits.*

It is therefore clear that we should not consider channels in $\mathrm{SP}_{n,0}$ to be free operations from a computational point of view, since they allow stabiliser circuits to be promoted to universality.

## B.2 Stinespring dilation of stabiliser channels

Some authors have argued that the study of the resourcefulness of unitary operations is useful as a proxy for more general channels [115], as any CPTP map can be implemented by performing some unitary operation on a larger system via Stinespring dilation [50]. That is, we can express any channel as $\mathcal{E}(\rho) = \mathrm{Tr}_R[U_{AR}(\rho \otimes |\phi\rangle\langle\phi|)U_{AR}^\dagger]$ where $A$ is the system of interest, $R$ is an $m$-qubit reference system, $U_{AR}$ is some unitary operation on the joint system $AR$, and $|\phi\rangle$ is some pure state. Note that without loss of generality, we can assume that $|\phi\rangle$ is a stabiliser state, since any $\mathbb{1}_A \otimes V_R$ such that $V^\dagger |\psi\rangle \in \mathrm{STAB}_m$ can be absorbed into $U_{AR}$.

As an example of how this might be useful, one could seek to quantify the magic of a quantum channel by employing some well-behaved measure of magic $\mathcal{M}_\mathcal{U}$ defined for unitaries and minimising this quantity over all minimal dilations,

$$\mathcal{M}_\mathrm{S}(\mathcal{E}) = \inf_{U,|\phi\rangle} \{\mathcal{M}(U) : \mathcal{E}(\cdot) = \mathrm{Tr}_R[U_{AR}(\cdot \otimes |\phi\rangle\langle\phi|)U_{AR}^\dagger]\}. \tag{B.2}$$

The operational intuition is that if one had access to such a resourceful $U$ and a large enough system initialised in a stabiliser state, one could implement $\mathcal{E}$ without any further access to magic resource.

However, we argue that this is problematic as a means to measure magic, as it fails the test of faithfulness. In this case faithfulness would mean that all completely stabiliser-preserving channels have a Clifford Stinespring dilation (i.e. the unitary operation $U_{AR}$ would be a Clifford gate), while all resourceful channels do not. Suppose that the monotone $\mathcal{M}_\mathcal{U}$ *is* faithful, such that $\mathcal{M}_\mathcal{U}(C) = 1$ for any Clifford gate $C$ (for example, this is true of the unitary extent [3]). Certainly there do exist channels for which $\mathcal{M}_\mathrm{S} = 1$, since trivially this is true for any Clifford gate. Moreover, for any channel $\mathcal{E}$ with the ability to generate magic, it must be the case that $\mathcal{M}_\mathrm{S} > 1$, since we assume that the reference system is initialised as a stabiliser state, and the partial trace preserves stabiliser states, so the optimal $U$ must be non-Clifford. However, it is not the case that $\mathcal{M}_\mathrm{S} = 1$ for arbitrary completely stabiliser-preserving channels. This is proved by the following theorem.

**Theorem B.3.** *There exist completely stabiliser-preserving channels $\mathcal{E} \in \mathrm{SP}_{n,n}$ for which* *it is impossible to construct a Clifford Stinespring dilation of the form:*

$$\mathcal{E}_S(\rho) = \mathrm{Tr}_R[U_{AR}(\rho \otimes |\phi\rangle\langle\phi|)U_{AR}^{\dagger}], \tag{B.3}$$

*such that the reference system $|\phi\rangle$ is a stabiliser state and $U_{AR}$ is a Clifford gate and $\mathcal{E}_S = \mathcal{E}$. Moreover there are stabiliser channels for which it is not possible to do this even approximately, in the sense that $\exists \mathcal{E} \in \mathrm{SP}_{n,n}$ such that:*

$$\|\mathcal{E} - \mathcal{E}_S\|_{\diamond} \geq \frac{1}{2}, \tag{B.4}$$

*for all Clifford Stinespring dilations $\mathcal{E}_S$.*

*Proof.* Recall that every completely stabiliser-preserving $n$-qubit channel $\mathcal{E} \in \mathrm{SP}_{n,n}$ has a stabiliser Choi state:

$$\Phi_{\mathcal{E}} = (\mathcal{E}_A \otimes \mathbb{1}_B)(|\Phi\rangle\langle\Phi|_{AB}) \in \overline{\mathrm{STAB}}_{2n}, \quad \text{where} \quad |\Phi\rangle = \frac{1}{2^{\frac{n}{2}}} \sum_{j=0^n}^{2^n-1} |j\rangle_A |j\rangle_B \,. \tag{B.5}$$

and for trace-preserving maps, $\mathrm{Tr}_A(\Phi_{\mathcal{E}}) = \mathbb{1}/2^n$. Conversely, every normalised state satisfying the above conditions is the Choi state for some completely stabiliser-preserving CPTP map. Now, note that $|\Phi\rangle$ is itself a stabiliser state. Let $\mathcal{E}_S$ be a generic Clifford Stinespring implementation of some channel, where we assume that the reference state $|\phi\rangle$ is a stabiliser state and the global unitary operation $U$ is a Clifford gate. Then from equation (B.3), we get

$$\Phi_{\mathcal{E}_S} = \mathrm{Tr}_R[U_{AR} \otimes \mathbb{1}_B(|\Phi\rangle\langle\Phi|_{AB} \otimes |\phi\rangle\langle\phi|_R)U_{AR}^{\dagger} \otimes \mathbb{1}_B] = \mathrm{Tr}_R[|\phi'\rangle\langle\phi'|_{ABR}] \tag{B.6}$$

where $|\phi'\rangle_{ABR} = (U_{AR} \otimes \mathbb{1}_B)\,|\Phi\rangle_{AB} \otimes |\phi\rangle_R$ is a stabiliser state.

Consider the channel $\mathcal{E}$ with Choi state:

$$\Phi_{\mathcal{E}} = p\,|0\rangle\langle 0|_A \otimes \frac{\mathbb{1}_B}{2} + (1-p)\,|1\rangle\langle 1|_A \otimes \frac{\mathbb{1}_B}{2}. \tag{B.7}$$

This is a completely stabiliser-preserving CPTP map for any $p$ between 0 and 1. If it is possible to implement this channel as some Clifford Stinespring dilation $\mathcal{E}_S$, so that $\Phi_{\mathcal{E}} = \Phi_{\mathcal{E}_S}$, then there must exist a pure stabiliser state on some larger system $ABR$ whose partial trace over $R$ gives us the RHS of equation (B.7). Comparing measurement statistics, we see that this cannot be true for arbitrary $p$. The expectation value for $Z_1 = Z_A \otimes \mathbb{1}_B$ on the state in (B.7) is:

$$\mathrm{Tr}[Z_1 \Phi_{\mathcal{E}}] = p - (1-p) = 2p - 1. \tag{B.8}$$

Whereas from equation (B.6), we obtain $\mathrm{Tr}\left[Z_1 \Phi_{\mathcal{E}_S}\right] = \mathrm{Tr}[(Z_A \otimes \mathbb{1}_{BR})(|\phi'\rangle\langle\phi'|_{ABR})]$. But $\mathrm{Tr}\left[Z_1 \Phi_{\mathcal{E}_S}\right] \in \{0, \pm 1\}$ since $Z \otimes \mathbb{1}$ is a Pauli operator and $|\phi'\rangle$ is assumed to be a stabiliser state. Conversely, from equation (B.8) $2p-1$ can take any real value in the range $[-1,1]$. Therefore if $(2p-1) \in (0,1)$ or $(2p-1) \in (-1,0)$, then $\mathcal{E} \neq \mathcal{E}_S$ unless $U$ is non-Clifford or $|\phi\rangle$ is a magic state.

Now let us fix the channel $\mathcal{E}$ to be the instance with $p = 1/4$, and consider how well $\mathcal{E}_S$ can approximate $\mathcal{E}$. Considering the diamond norm distance, we have:

$$\begin{aligned} \|\mathcal{E} - \mathcal{E}_S\|_\diamond &= \sup \|\mathcal{E} \otimes \mathbb{1}(\rho) - \mathcal{E}_S \otimes \mathbb{1}(\rho)\|_1 & \text{(B.9)} \\ &\geq \left\|\Phi_{\mathcal{E}} - \Phi_{\mathcal{E}_S}\right\|_1 & \text{(B.10)} \\ &= 2D_{\mathrm{Tr}}(\Phi_{\mathcal{E}}, \Phi_{\mathcal{E}_S}) & \text{(B.11)} \end{aligned}$$

where $D_{\mathrm{Tr}}(\cdot,\cdot)$ is the trace-norm distance. But the trace-norm distance upper bounds differences in measurement statistics, $D_{\mathrm{Tr}}(\rho,\sigma) = \max_\Pi |\mathrm{Tr}[\Pi(\rho-\sigma)]_1|$ where $\Pi$ are projectors. Let $\Pi_Z = (\mathbb{1}_2 + Z \otimes \mathbb{1})/2$ be the projector onto the positive outcome of the $Z_1$ measurements. Above we saw that for $\mathcal{E}$ we have

$$\mathrm{Tr}[\Pi_Z \Phi_{\mathcal{E}}] = \frac{1 + (2p-1)}{2} = p, \tag{B.12}$$

whereas for a Clifford Stinespring dilation we must have $\mathrm{Tr}\left[\Pi_Z \Phi_{\mathcal{E}_S}\right] = 0,\ 1/2$ or $1$. Therefore for any $\mathcal{E}_S$, with $p = 1/4$,

$$|\mathrm{Tr}\left[\Pi_Z(\Phi_{\mathcal{E}} - \Phi_{\mathcal{E}_S})\right]| \geq \left|\frac{1}{4} - 0\right| = \frac{1}{4}. \tag{B.13}$$

It follows that for $\mathcal{E}$ with $p = 1/4$, we have $\|\mathcal{E} - \mathcal{E}_S\|_\diamond \geq 2D_{\mathrm{Tr}}(\Phi_{\mathcal{E}}, \Phi_{\mathcal{E}_S}) \geq 1/2$ for any Clifford Stinespring dilation $\mathcal{E}_S$. □

Note that the argument does not rely on the non-unitality of the channel described in the proof. For example, take the following bit-flip error channel $\mathcal{E}_X$: with probability

$(1-p)$ do nothing, with probability $p$ apply an $X$ gate. This has the Choi state:

$$\Phi_{\mathcal{E}_X} = (1-p)\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + p\left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| \tag{B.14}$$

where $|\Phi^+\rangle$ and $|\Psi^+\rangle$ are Bell states with even and odd parity, respectively. Then if we replace $Z_1$ with $-Z_1Z_2 = Z_A \otimes Z_B$ in equation (B.8) and what follows, the argument still goes through.

We have shown that there exist completely stabiliser-preserving channels $\mathcal{E}$ whose Stinespring dilations cannot be even approximately Clifford. It follows that for these channels $\mathcal{M}_{\mathrm{S}}(\mathcal{E})$ must be strictly larger than 1, so this cannot be considered a faithful measure of magic. This does, however, give us an interesting perspective on the status of classical randomness and processing as stabiliser operations, compared to the "fully quantum" stabiliser operations such as Clifford gates and Pauli measurements. The proof rested on the fact that in setting up the channel $\mathcal{E}$ we can set $p$ to be any real number, as classical randomness with conditioning preserves the stabiliser polytope. On the other hand, a global Clifford unitary followed by a Pauli measurement cannot reproduce the same statistics. Therefore classical processing can be seen as a qualitatively different kind of computational resource that cannot be recovered from Stinespring dilation of Clifford operations. Nevertheless it is appropriate to include classical operations as part of the set of free operations, precisely because their effect on stabiliser states can be efficiently computed.

## B.3 Incomplete Kraus representations

Our main focus has been on completely stabiliser-preserving CPTP maps, but one can similarly define completely stabiliser-preserving maps that are not trace-preserving. In particular we can define completely positive trace-contractive (CPTC) maps that are also CSP, i.e maps such that $\mathrm{Tr}[\mathcal{E}(\rho)] \leq \mathrm{Tr}[\rho]$. Just as all $n$-qubit CPTP maps can be expressed by some complete set of Kraus operators $\{K_j\}$, such that $\sum_j K_j^\dagger K_j = \mathbb{1}_n$, any CPTC map $\mathcal{T}$ can be represented by an *incomplete* set such that $\sum_j K_j^\dagger K_j \leq 1$. For any CPTC $\mathcal{T}$ we can always find another map $\mathcal{T}_\perp$ such that $\mathcal{E} = \mathcal{T} + \mathcal{T}_\perp$ is a CPTP map. In general these completions to a CPTP map are non-unique, and we say that $\mathcal{T}_\perp$ is the complement of $\mathcal{T}$ with respect to $\mathcal{E}$.

Since finding a complementary $\mathcal{T}_\perp$ for a trace-contractive CP map $\mathcal{T}$ is always possible, and there is often considerable freedom in this choice, one might expect that if $\mathcal{T}$ is also completely stabiliser-preserving, we could also find a complementary $\mathcal{T}_\perp$ that is also completely stabiliser-preserving. Here we show that in fact this is not always possible.

**Theorem B.4.** *Let TC be the set of all trace-contractive, completely stabiliser-preserving and completely positive maps. There exists a map $\mathcal{T} \in TC$ such that, for any completion to a CPTP map, $\mathcal{E} = \mathcal{T} + \mathcal{T}_\perp$, the complement $\mathcal{T}_\perp$ can never be completely stabiliser-preserving.*

*Proof.* From the Choi-Jamiolkowski isomorphism, we have a condition on the reduced state on subsystem B. A map $\mathcal{E}$ is trace-preserving if and only if $\rho_B = \mathrm{Tr}_A[\Phi_{\mathcal{E}}] = \frac{\mathbb{1}}{d}$. Now let $\rho_{B,0} = \mathrm{Tr}_A[\Phi_{\mathcal{T}}]$ and $\rho_{B,1} = \mathrm{Tr}_A[\Phi_{\mathcal{T}_\perp}]$. The operators $\rho_{B,0}$ and $\rho_{B,1}$ are both positive semidefinite because $\mathcal{T}$ and $\mathcal{T}_\perp$ each have a Kraus representation. Using the Choi-Jamiolkowski isomorphism we have $\Phi_{\mathcal{E}} = \Phi_{\mathcal{T}} + \Phi_{\mathcal{T}_\perp}$. Taking the partial trace over subsystem $A$,

$$\frac{\mathbb{1}}{d} = \rho_{B,0} + \rho_{B,1}, \tag{B.15}$$

where $\rho_{B,j} = \mathrm{Tr}_A[\Phi_{\mathcal{T}_j}]$ Although there could be many possible maps $\mathcal{T}_\perp$ that complete $\mathcal{T}$ to a CPTP channel, the reduced state $\rho_{B,1}$ on subsystem B is unique and fully determined by $\rho_{B,0}$,

$$\rho_{B,1} = \mathrm{Tr}_A[\Phi_{\mathcal{T}_1}] = \frac{\mathbb{1}}{d} - \rho_{B,0}. \tag{B.16}$$

If $\rho_{B,1}$ is proportional to a stabiliser state, then we can always find some unnormalised stabiliser state $\Phi_{\mathcal{T}_1}$ whose partial trace is $\rho_{B,1}$, which means the CPTP channel $\Phi_{\mathcal{E}}$ can be stabiliser-preserving. But since partial trace is stabiliser-preserving, if $\rho_{B,1}$ is not a stabiliser state, then $\Phi_{\mathcal{T}_\perp}$ must also be non-stabiliser, meaning $\mathcal{T}_\perp$ can generate magic.

By assumption $\mathcal{T}$ is a completely-stabiliser-preserving map. $\Phi_{\mathcal{T}}$ must be proportional to a stabiliser-preserving map, and since it does not represent a complete CPTP channel it can be unnormalised, but for it to be trace-contractive, it must satisfy certain constraints. First, we clearly must have $\mathrm{Tr}[\Phi_{\mathcal{T}}] \leq 1$, but the eigenvalues of its reduced state $\rho_{B,0}$ are also constrained. The map $\mathcal{T}$ is trace-contractive if and only if its Choi state satisfies $\frac{\mathbb{1}}{2} - \mathrm{Tr}_A[\Phi_{\mathcal{T}}]^T \geq 0$. For a positive matrix, the eigenvalues are invariant under transpose, so the condition is equivalent to $\frac{\mathbb{1}}{d} - \rho_{B,0} \geq 0$. The unnormalised state has a spectral decomposition over some basis $\{|j\rangle\}$, so using this basis to write $\mathbb{1} = \sum_j |j\rangle\langle j|$, we have:

$$\frac{1}{2}\sum_j |j\rangle\langle j| - \sum_j p_j\, |j\rangle\langle j| \geq 0 \tag{B.17}$$

where $p_j \geq 0$ since $\rho_{B,0} \geq 0$. It then follows that $(1/d) - p_j \geq 0$ for all $j$, so $\mathcal{T}$ is trace-contractive provided no eigenvalue of $\rho_{B,0}$ is larger than $1/2$.

Let $\Phi_{\mathcal{T}}$ have the form $\Phi_{\mathcal{T}} = c(|00\rangle\langle 00| + |0+\rangle\langle 0+|)$. This is proportional to a mixed stabiliser state, so corresponds to a completely stabiliser-preserving map. Taking partial trace:

$$\rho_{B,0} = c(|0\rangle\langle 0| + |+\rangle\langle +|) = \frac{c}{2}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}. \tag{B.18}$$

Solving this, we find $\rho_{B,0}$ has eigenvalues $\lambda_\pm = c(1 \pm \frac{1}{\sqrt{2}})$. We can ensure that the constraints on trace-contractive maps are satisfied by choosing $c$ so that $\lambda_+ \leq \frac{1}{2}$, which holds provided $c \leq 1/(2+\sqrt{2})$. We choose the maximal allowed value, $c = 1/(2+\sqrt{2})$. Then we have that

$$\mathrm{Tr}_A[\Phi_{\mathcal{T}_\perp}] = \rho_{B,1} = \frac{\mathbb{1}}{2} - \frac{|0\rangle\langle 0| + |+\rangle\langle +|}{2+\sqrt{2}}. \tag{B.19}$$

We can check that this is proportional to a non-stabiliser state. In fact, it turns out to be a pure magic state, with robustness of magic $\sqrt{2}$. Therefore $\mathcal{T}_\perp$ is not completely stabiliser-preserving. To summarise, the channel $\mathcal{T}$ with Choi state $\Phi_{\mathcal{T}}$ as defined in equation (B.18) and $c = 1/(2+\sqrt{2})$, is a completely positive, completely stabiliser-preserving trace contractive map. We have shown that for any CPTP map $\mathcal{E}$ of which $\mathcal{T}$ is a sub-channel, the complement of $\mathcal{T}$ in $\mathcal{E}$ cannot be completely stabiliser-preserving. This completes the proof. □

# Appendix C

# Robustness of magic technical details

## C.1 Robustness of the Choi state

Here we confirm that the robustness of the Choi state, $\mathcal{R}(\Phi)$ has the properties convexity and submultiplicativity under tensor product. We then give an example to show that it is not submultiplicative under composition.

**Convexity:** This follows immediately from convexity of robustness of magic. Consider a real linear combination of $n$-qubit channels: $\mathcal{E} = \sum_k q_k \mathcal{E}_k$. The Choi state for $\mathcal{E}$ is

$$\Phi_{\mathcal{E}} = \sum_k q_k (\mathcal{E}_k \otimes \mathbb{1}_n)(|\Omega_n\rangle\langle\Omega_n|) = \sum_k q_k \Phi_{\mathcal{E}_k}, \tag{C.1}$$

where $\Phi_{\mathcal{E}_k}$ is the Choi state for $\mathcal{E}_k$. Then by convexity of robustness of magic:

$$\mathcal{R}(\Phi_{\mathcal{E}}) \leq \sum_k |q_k| \mathcal{R}\left(\Phi_{\mathcal{E}_k}\right), \tag{C.2}$$

which shows $\mathcal{R}(\Phi_{\mathcal{E}})$ is convex in $\mathcal{E}$.

**Submultiplicativity under tensor product:** The maximally entangled state $|\Omega_{n+m}\rangle^{AA'|BB'}$ as defined by equation (1.28) in the main text can be factored as $|\Omega_{n+m}\rangle^{AA'BB'} = |\Omega_n\rangle^{A|B} |\Omega_m\rangle^{A'|B'}$. So the Choi state for a channel $\mathcal{E}^{AA'} = \mathcal{E}^A \otimes \mathcal{E}'^{A'}$, where $\mathcal{E}^A$ and $\mathcal{E}'^{A'}$ are respectively $n$-qubit and $m$-qubit channels, can be written:

$$\Phi_{\mathcal{E}} = \left(\mathcal{E}^A \otimes \mathcal{E}'^{A'} \otimes \mathbb{1}_{n+m}\right)(|\Omega_{n+m}\rangle\langle\Omega_{n+m}|^{AA'|BB'}) \tag{C.3}$$

$$= \left(\mathcal{E}^A \otimes \mathbb{1}_n\right)(|\Omega_n\rangle\langle\Omega_n|^{A|B}) \otimes \left(\mathcal{E}'^{A'} \otimes \mathbb{1}_m\right)(|\Omega_m\rangle\langle\Omega_m|^{A'|B'}) = \Phi_{\mathcal{E}^A} \otimes \Phi_{\mathcal{E}'^{A'}}. \tag{C.4}$$

Then by submultiplicativity of robustness of magic for states, we have:

$$\mathcal{R}\left(\Phi_{\mathcal{E}^A \otimes \mathcal{E}'^{A'}}\right) \leq \mathcal{R}(\Phi_{\mathcal{E}^A})\mathcal{R}(\Phi_{\mathcal{E}'^{A'}}),$$

which is the desired property.

**Failure of submultiplicativity under composition:** Let $\mathcal{E}_1$ be the single-qubit $Z$-reset channel defined by Kraus operators $\{|0\rangle\langle 0|, |0\rangle\langle 1|\}$, and let $\mathcal{E}_2$ be the conditional channel defined by $\{|T\rangle\langle 0|, |1\rangle\langle 1|\}$, where $|T\rangle = T|+\rangle$. These channels respectively have Choi states $\Phi_{\mathcal{E}_1} = |0\rangle\langle 0| \otimes \frac{\mathbb{1}}{2}$, and $\Phi_{\mathcal{E}_2} = \frac{1}{2}(|T0\rangle\langle T0| + |11\rangle\langle 11|)$, with robustness of magic $\mathcal{R}(\Phi_{\mathcal{E}_1}) = 1$ and $\mathcal{R}(\Phi_{\mathcal{E}_2}) \approx 1.207$.

The composed channel $\mathcal{E}_2 \circ \mathcal{E}_1$ has a Kraus representation $\{|T\rangle\langle 0|, |T\rangle\langle 1|\}$, and so has a Choi state $\Phi_{\mathcal{E}_2 \circ \mathcal{E}_1} = |T\rangle\langle T| \otimes \frac{\mathbb{1}}{2}$, with $\mathcal{R}(\Phi_{\mathcal{E}_2 \circ \mathcal{E}_1}) \approx 1.414 > \mathcal{R}(\Phi_{\mathcal{E}_2})\mathcal{R}(\Phi_{\mathcal{E}_1})$. So it is not the case that the robustness of the Choi state is submultiplicative under composition.

More intuitively, such counterexamples arise for channels $\mathcal{E}$ where the stabiliser state $|\phi_*\rangle$ that results in maximal final robustness $\mathcal{R}[(\mathcal{E} \otimes \mathbb{1}_n)(|\phi_*\rangle\langle\phi_*|)]$ is not the maximally entangled state $|\Omega_n\rangle$, as then we can always boost the output robustness by using a stabiliser-preserving operation to prepare $|\phi_*\rangle$ before applying $\mathcal{E}$.

## C.2 Robustness of magic and stabiliser frames

In this appendix we show how the robustness of magic [103] can be directly related to a Wigner-function-like quasiprobability decomposition over some frame [124]. We say that a frame $\mathcal{F} = \{F(\lambda)\}$ is a stabiliser projector frame if it is paired with a dual frame $\mathcal{G} = \{G(\lambda) = |\phi_\lambda\rangle\langle\phi_\lambda|\}_{\lambda\in\Lambda}$ where each $\lambda$ labels a unique stabiliser state, and $|\Lambda| = |\mathrm{STAB}_n|$. Note that the terminology we use here should not be confused with that introduced by Garcia, Markov and Cross [58, 193], who use the term stabiliser frame to refer to an orthornormal basis for a stabiliser subspace. It can be shown that a valid choice of stabiliser projector frame can be constructed by setting $F(\lambda) = \alpha|\phi_\lambda\rangle\langle\phi_\lambda| - \beta\mathbb{1}_n$ for a suitable choice of $\alpha$ and $\beta$ [194]. Starting from any such valid pairing of frame and dual frame, then we must have, for any density operator $\rho$,

$$\rho = \sum_\lambda G(\lambda)\,\mathrm{Tr}[F(\lambda)\rho]. \tag{C.5}$$

On the other hand, there exists an optimal decomposition with respect to the robustness of magic such that [135]

$$\rho = \sum_\lambda q_\lambda G(\lambda) \tag{C.6}$$

where in general $q_\lambda$ can be negative. Then from equations (C.5) and (C.6), we have

$$0 = \sum_\lambda G(\lambda)(q_\lambda - \mathrm{Tr}[F(\lambda)\rho]). \tag{C.7}$$

Now let $\Delta_\lambda = q_\lambda - \mathrm{Tr}[F(\lambda)\rho]$, and let $\mathcal{F}' = \{F'(\lambda)\}$, where $F'(\lambda) = F(\lambda) + \Delta_\lambda \mathbb{1}$. We can check that $\mathcal{F}'$ is still a valid choice of frame for dual frame $\mathcal{G}$. For any Hermitian operator $A$,

$$\begin{aligned}\sum_\lambda G(\lambda)\,\mathrm{Tr}\big[F'(\lambda)A\big] &= \sum_\lambda G(\lambda)\,\mathrm{Tr}[F(\lambda)A] + \sum_\lambda G(\lambda)\Delta_\lambda\,\mathrm{Tr}[\mathbb{1}A] && \text{(C.8)}\\ &= A + \mathrm{Tr}[A]\sum_\lambda G(\lambda)(q_\lambda - \mathrm{Tr}[F(\lambda)\rho]). && \text{(C.9)}\end{aligned}$$

But from equation (C.7), the summation in the second term is equal to zero, so $A = \sum_\lambda G(\lambda)\,\mathrm{Tr}[F'(\lambda)A]$ for any $A$. Now if we consider the state of interest $\rho$, we have

$$W_\rho(\lambda) = \mathrm{Tr}\big[F'(\lambda)\rho\big] = \mathrm{Tr}[F(\lambda)\rho + \Delta_\lambda\rho] = \mathrm{Tr}[F(\lambda)\rho] + \Delta_\lambda = q_\lambda. \tag{C.10}$$

Therefore $\mathcal{F}'$ is a valid frame with dual frame $\mathcal{G}$, such that the quasiprobability function $W'_\rho$ yields the same weight as the decomposition that is optimal with respect to robustness of magic.

# Appendix D

# Additional channel monotone technical details

## D.1 Linear program for channel robustness

In this section we show explicitly how computing the channel robustness may be formulated as a linear optimisation problem, without explicitly enumerating the extreme points of $\mathrm{SP}_{n,n}$. In Howard and Campbell [103], the optimisation problem for calculating robustness of magic for states was specified as follows.

$$\begin{aligned} \textbf{minimise} \quad & \|\mathbf{q}\|_1 \\ \textbf{subject to} \quad & A\mathbf{q} = \mathbf{b}, \end{aligned}$$

where $\mathbf{q}$ is a vector of coefficients, $\mathbf{b}$ is the vector of Pauli expectation values for the target state $\Phi_{\mathcal{E}}$, and $A$ is a matrix whose columns are the Pauli vectors for the stabiliser states. For $n$-qubit channels, we have $2n$-qubit Choi states, so the number of generalised Paulis is $N_P = 4^{2n}$, and the number of stabiliser states is $N_S = 2^{2n}\prod_{j=1}^{2n}(2^j+1)$ [103]. Then $\mathbf{b}$ has $N_P$ entries, $\mathbf{q}$ has $N_S$ entries, and the dimension of $A$ is $(N_P \times N_S)$. From this construction we can recover optimal decompositions of the form $\Phi_{\mathcal{E}} = \sum_j q_j \left|\phi_j\right\rangle\!\left\langle\phi_j\right|$, where $\sum_j q_j = 1$ and $\left|\phi_j\right\rangle$ are the pure stabiliser states.

We want to restrict the problem to decompositions of the form $\Phi_{\mathcal{E}} = (1+p)\rho_+ - p\rho_-$, where $p \geq 0$ and $\rho_\pm$ correspond to trace-preserving channels, and can in general be mixed. Rather than enumerating all the extreme points of the set of stabiliser states corresponding to maps in $\mathrm{SP}_{n,n}$, it is more convenient to retain the same $A$ matrix and modify the constraints. We still need to start from a finite set of extreme points, i.e. pure stabiliser states, so first rewrite as:

$$\Phi_{\mathcal{E}} = \sum_j q_{j_+}\rho_j + \sum_j q_{j_-}\rho_j = \sum_j p_{j_+}\rho_j - \sum_j p_{j_-}\rho_j, \tag{D.1}$$

where $q_{j_+}$ are the positive quasiprobabilities, $q_{j_-}$ are the negative quasiprobabilities, and $p_{j_\pm} = |q_{j_\pm}|$. In the Pauli vector picture we can write this as $\mathbf{b} = A\mathbf{p}_+ - A\mathbf{p}_-$, where all the entries of $\mathbf{p}_\pm$ are non-negative. We define a new variable vector $\mathbf{p}$ which will have twice the length of the previous $\mathbf{q}$, i.e. $2N_S$ entries:

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_+ \\ \mathbf{p}_-, \end{pmatrix} \tag{D.2}$$

and define a new $(N_P \times 2N_S)$ matrix $A'$ in block form, $A' = \begin{pmatrix} A & -A \end{pmatrix}$. Then we have:

$$A'\mathbf{p} = \begin{pmatrix} A & -A \end{pmatrix} \begin{pmatrix} \mathbf{p}_+ \\ \mathbf{p}_- \end{pmatrix} = A\mathbf{p}_+ - A\mathbf{p}_- = \mathbf{b}. \tag{D.3}$$

So now we need to minimise $\|\mathbf{p}\|_1 = \sum_j p_j$ subject to $A'\mathbf{p} = \mathbf{b}$ and $\mathbf{p} \geq 0$.

Next, we need the trace-preserving condition. Provided $\mathcal{E}$ is CPTP, if one part of the decomposition is trace-preserving, then the other will be as well, so we only need enforce the constraint on one of $\rho_+$ or $\rho_-$. Assume that we check $\rho_+$. The condition for a Choi state $\Phi^{AB} = \mathcal{E}^A \otimes \mathbb{1}^B(|\Phi\rangle\langle\Phi|^{AB})$ to be trace-preserving is:

$$\mathrm{Tr}_A(\Phi^{AB}) = \frac{\mathbb{1}}{d}, \tag{D.4}$$

where $d$ is the dimension of the subsystem. We need to convert this to a constraint on the vector $\mathbf{b}_+$ corresponding to $\phi_+$, which is given by $\mathbf{b}_+ = A\mathbf{p}_+$. First, note that all Paulis are traceless except for the identity $P_0 = \mathbb{1}$, so for the maximally mixed state:

$$\langle P_j \rangle = \mathrm{Tr}\left(P_j \frac{\mathbb{1}}{d}\right) = \frac{\mathrm{Tr}(P_j)}{d} = \delta_{j,0}, \tag{D.5}$$

so if the first entry in a Pauli vector is always $\langle \mathbb{1} \rangle$, the maximally mixed state has Pauli vector:

$$\mathbf{b}_B = \begin{pmatrix} 1 \\ \vec{0} \end{pmatrix}. \tag{D.6}$$

where $\vec{0}$ is the zero vector. However, we need this to hold just for the reduced state on $B$ rather than the full Pauli vector. Consider that if the whole state is written $\Phi^{AB} = \sum_{j,k} r_{j,k} P_j \otimes P_k$ for some set of coefficients $r_{j,k}$, then the expectation values are given by:

$$\langle P_l \otimes P_m \rangle = \sum_{j,k} r_{j,k} \operatorname{Tr}\big(P_l P_j \otimes P_m P_k\big) = \sum_{j,k} r_{j,k} d^2 \delta_{j,l} \delta_{m,k} = d^2 r_{l,m}. \tag{D.7}$$

The reduced state is:

$$\operatorname{Tr}_A(\Phi^{AB}) = \sum_{j,k} r_{j,k} d \delta_{j,0} P_k = d \sum_k r_{0,k} P_k. \tag{D.8}$$

and the entries of the reduced Pauli vector will be:

$$\langle P_m \rangle = d \sum_k r_{0,k} \operatorname{Tr}\{P_m P_k\} = d^2 r_{0,m} = \langle P_0 \otimes P_m \rangle. \tag{D.9}$$

So for condition (D.4) to hold for the reduced state on $B$, we combine equations (D.5) and (D.9) to get $\langle P_m \rangle = \langle P_0 \otimes P_m \rangle = \delta_{m,0}$. That is, we just need to look at the entries of $\mathbf{b}_+$ corresponding to Paulis of the form $\mathbb{1} \otimes P_j$. These should all be zero except the first entry, which corresponds to $\langle \mathbb{1} \otimes \mathbb{1} \rangle$. Note that $\mathbf{b}_+ = A\mathbf{p}_+$ will in general not be normalised, but this does not matter, since we are only interested in whether or not entries are zero. We can use a binary matrix $M$ to pick out the values of interest. As an example we consider the two-qubit case, and assume that the entries are ordered as:

$$\mathbf{b}_+ = \begin{pmatrix} \langle \mathbb{1} \otimes \mathbb{1} \rangle \\ \langle \mathbb{1} \otimes X \rangle \\ \langle \mathbb{1} \otimes Y \rangle \\ \langle \mathbb{1} \otimes Z \rangle \\ \langle X \otimes \mathbb{1} \rangle \\ \vdots \\ \langle Z \otimes Z \rangle \end{pmatrix}. \tag{D.10}$$

Here, we are only interested in the 2nd, 3rd and 4th entries. We form a new vector $\mathbf{c}$ by left multiplying with $M$:

$$\mathbf{c} = M\mathbf{b}_+ = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \mathbf{b}_+ = \begin{pmatrix} \langle \mathbb{1} \otimes X \rangle \\ \langle \mathbb{1} \otimes Y \rangle \\ \langle \mathbb{1} \otimes Z \rangle \end{pmatrix}. \tag{D.11}$$

Then the condition we need is just $\mathbf{c} = 0$. To convert this to a condition on the $2N_S$-entry variable $\mathbf{p} = \begin{pmatrix} \mathbf{p}_+ \\ \mathbf{p}_- \end{pmatrix}$, we first pad $A$ with zeroes: $A_+ = \begin{pmatrix} A & \bar{0} \end{pmatrix}$, where $\bar{0}$ is the $(N_P \times N_S)$ zero matrix. We then have

$$\mathbf{b}_+ = A\mathbf{p}_+ = A\mathbf{p}_+ + \bar{0}\mathbf{p}_- = \begin{pmatrix} A & \bar{0} \end{pmatrix} \begin{pmatrix} \mathbf{p}_+ \\ \mathbf{p}_- \end{pmatrix} = A_+\mathbf{p}, \tag{D.12}$$

so that $\mathbf{c} = M\mathbf{b}_+ = MA_+\mathbf{p}$. Therefore, $\rho_+$ is trace-preserving when $MA_+\mathbf{p} = 0$. We can therefore specify the new optimisation problem as:

$$\begin{aligned} \textbf{minimise} \quad & \|\mathbf{p}\|_1 = \sum_j p_j \\ \textbf{subject to} \quad & A'\mathbf{p} = \mathbf{b}, \\ & \mathbf{p} \geq 0, \\ & MA_+\mathbf{p} = 0 \end{aligned}$$

where $A' = \begin{pmatrix} A & -A \end{pmatrix}$, and $A_+ = \begin{pmatrix} A & \bar{0} \end{pmatrix}$, with $A$ and $\mathbf{b}$ having the same definitions as previously, $\bar{0}$ is the zero matrix with dimension the same as $A$, and with $M$ being the binary matrix that picks out the $\langle \mathbb{1} \otimes P_j \rangle$ entries from the vector $A_+\mathbf{p}$. Most of this is straightforward to implement. The step that requires some care is in correctly constructing the matrix $M$, as it will depend on the choice of ordering of Pauli operators in the construction of $A$ and $\mathbf{b}$. If the $B$ subsystem has $n$ qubits, then we will need to constrain $4^n - 1$ non-trivial $\langle \mathbb{1} \otimes P_j \rangle$ expectation values to zero, so $M$ should have dimension $((4^n - 1) \times N_P)$. If the Paulis are ordered as in the example given above for 2-qubit Choi states, then the construction is just $M = \begin{pmatrix} \vec{0} & \mathbb{1}' & \vec{0} & \cdots & \vec{0} \end{pmatrix}$, where $\mathbb{1}'$ is the $((4^n - 1) \times (4^n - 1))$ identity, and $\vec{0}$ denotes a column of zeroes. We have implemented this linear program in MATLAB, using the convex optimisation package CVX

[132], and have made the code available from the repository Ref. [185].

## D.2 Dyadic Clifford negativity properties

In Section 4.2.5 we defined the dyadic Clifford negativity, a quantity that can be used to upper bound the dyadic channel negativity,

$$\Lambda_{\mathrm{Cl}}(\mathcal{E}) = \min\left\{ \|\alpha\|_1 : \sum_{jk} \alpha_{jk}\mathcal{C}_{jk} = \mathcal{E} \right\} \tag{D.13}$$

where we have use the shorthand $\mathcal{C}_{jk}$ to mean $\mathcal{C}_{jk}(\ldots) = C_j(\ldots)C_k^\dagger$, where $C_j$ and $C_k$ are Clifford gates. Here we show that this quantity cannot be a channel monotone; we show that, while it is convex and submultiplicative under composition on its domain, it cannot be a faithful monotone for completely stabiliser-preserving CPTP maps.

**Convexity:** Suppose we have a set of channels with optimal decompositions:

$$\mathcal{E}_s = \sum_{jk} \alpha_{jk}^{(s)} \mathcal{C}_{jk}^{(s)}, \tag{D.14}$$

such that $\Lambda_{\mathrm{Cl}}(\mathcal{E}_s) = \|\alpha^{(s)}\|$. Then if we form a linear combination of these,

$$\mathcal{E} = \sum_s q_s \mathcal{E}_s = \sum_s q_s \sum_{j,k} \alpha_{j,k}^{(s)} \mathcal{C}_{jk}^{(s)}. \tag{D.15}$$

This is a valid decomposition of $\mathcal{E}$ over dyadic Clifford elements $\mathcal{C}_{jk}^{(s)}$, with $\ell_1$-norm:

$$\begin{aligned} \sum_{s,j,k} |q_s \alpha_{j,k}^{(s)}| &= \sum_{s,j,k} |q_s|\,|\alpha_{j,k}^{(s)}| && \text{(D.16)} \\ &= \sum_s |q_s| \sum_{j,k} |\alpha_{j,k}^{(s)}| && \text{(D.17)} \\ &= \sum_s |q_s| \Lambda_{\mathrm{Cl}}(\mathcal{E}_s). && \text{(D.18)} \end{aligned}$$

However, this may not be an optimal decomposition for $\mathcal{E}$, so $\Lambda_{\mathrm{Cl}}(\mathcal{E}) \leq \sum_s |q_s| \Lambda_{\mathrm{Cl}}(\mathcal{E}_s)$.

**Submultiplicativity under composition:** For any composition of a pair of dyadic Clifford elements, we have:

$$\mathcal{C}_{lm} \circ \mathcal{C}_{jk}(\ldots) = C_l C_j(\ldots) C_k^\dagger C_m^\dagger = C_p(\ldots)C_q^\dagger = \mathcal{C}_{p,q} \tag{D.19}$$

for some new pair of Clifford gates $C_p$ and $C_q$. So when composing two optimal dyadic Clifford decompositions:

$$\mathcal{E}_\alpha = \sum_{j,k} \alpha_{j,k} \mathcal{C}_{j,k}, \quad \Lambda_{\mathrm{Cl}}(\mathcal{E}_\alpha) = \|\alpha\|_1, \tag{D.20}$$

$$\mathcal{E}_\beta = \sum_{l,m} \beta_{l,m} \mathcal{C}_{l,m}, \quad \Lambda_{\mathrm{Cl}}(\mathcal{E}_\beta) = \|\beta\|_1, \tag{D.21}$$

we obtain:

$$\mathcal{E}_\beta \circ \mathcal{E}_\alpha = \sum_{j,k} \alpha_{j,k} \mathcal{E}_\beta \circ \mathcal{C}_{j,k} = \sum_{j,k,l,m} \alpha_{j,k} \beta_{l,m} \mathcal{C}_{(j,l),(k,m)}. \tag{D.22}$$

This dyadic Clifford decomposition has $\ell_1$-norm given by:

$$\begin{aligned} \sum_{j,k,l,m} |\alpha_{j,k}\beta_{l,m}| &= \sum_{j,k,l,m} |\alpha_{j,k}||\beta_{l,m}| & \text{(D.23)} \\ &= \sum_{j,k} |\alpha_{j,k}| \sum_{l,m} |\beta_{l,m}| & \text{(D.24)} \\ &= \Lambda_{\mathrm{Cl}}(\mathcal{E}_\alpha)\Lambda_{\mathrm{Cl}}(\mathcal{E}_\beta). & \text{(D.25)} \end{aligned}$$

Again, this decomposition may not be optimal, so:

$$\Lambda_{\mathrm{Cl}}(\mathcal{E}_\beta \circ \mathcal{E}_\alpha) \leq \Lambda_{\mathrm{Cl}}(\mathcal{E}_\alpha)\Lambda_{\mathrm{Cl}}(\mathcal{E}_\beta). \tag{D.26}$$

**Measure is not faithful for stabiliser-preserving CPTP maps:** Clearly, we have that for any Clifford operation $\mathcal{U}(\ldots) = U(\ldots)U^\dagger$, we have $\Lambda_{\mathrm{Cl}}(\mathcal{U}) = 1$, since the channel can be trivially represented by a single dyadic element $\mathcal{C}_{j,j}$, for some $j$. It immediately follows that any CPTP map $\mathcal{E}$ that can be represented as a probabilistic mixture of Clifford operations (for example, the depolarising channel) also has $\Lambda_{\mathrm{Cl}}(\mathcal{E}) = 1$. However, numerically we find examples of completely stabiliser-preserving maps that have $\Lambda_{\mathrm{Cl}}(\mathcal{E}) > 1$. In particular we find that the Pauli reset channel $\mathcal{E}_Z$ defined by Kraus operators $\{|0\rangle\langle 0|, |0\rangle\langle 1|\}$ has $\Lambda_{\mathrm{Cl}}(\mathcal{E}_Z) = 2$. In contrast, this channel is a free operation for our static simulator, and even for the more limited Bennink et al. CPR simulator. However we have also seen numerically that there exist non-stabiliser operations for which the Clifford dyadic simulator does better than the static simulator.

## D.3 Dyadic stabiliser channels are norm-contractive

Suppose we have a normalised initial dyad $|\phi_L\rangle\langle\phi_R|$, and a map $\mathcal{T} \in \mathrm{DSP}_{n,n}$ (see Definition 4.23). Define real-valued vectors $\mathbf{v}$ and $\mathbf{w}$ such that $v_j = \left\|L_j|\phi_L\rangle\right\|$, and $w_j = \left\|R_j|\phi_R\rangle\right\|$. Since $\{L_j\}_j$ is a complete set of Kraus operators, we have $\mathbf{v}\cdot\mathbf{v} = \sum_j \left\|L_j|\phi_L\rangle\right\|^2 = 1$, and similarly for $\mathbf{w}$. Now consider the norm of the dyad after $\mathcal{T}$ has been applied:

$$\|\mathcal{T}(|\phi_L\rangle\langle\phi_R|)\|_1 = \|\sum_j L_j|\phi_L\rangle\langle\phi_R|R_j^\dagger\|_1 \tag{D.27}$$

$$\leq \sum_j \left\|L_j|\phi_L\rangle\right\|\cdot\left\|R_j|\phi_R\rangle\right\| \tag{D.28}$$

$$= \mathbf{v}\cdot\mathbf{w}, \tag{D.29}$$

where we used the triangle inequality in the second line. Then by the Cauchy-Schwarz inequality

$$\|\mathcal{T}(|\phi_L\rangle\langle\phi_R|)\|_1 \leq \mathbf{v}\cdot\mathbf{w} \leq \sqrt{(\mathbf{v}\cdot\mathbf{v})(\mathbf{w}\cdot\mathbf{w})} = 1. \tag{D.30}$$

We can then see that $\mathcal{T} \in \mathrm{DSP}_{n,n}$ are norm-contractive in the more general case. We can express any $n$-qubit operator $A$ as the singular value decomposition $A = UDV^\dagger$, where $U$ and $V$ are unitary, and $D = \sum_j \sigma_j|j\rangle\langle j|$, with $\sigma_j$ being the singular values, for some basis $\{|j\rangle\}$. But then we can express it as $A = \sum_j \sigma_j U|j\rangle\langle j|V^\dagger = \sum_j \sigma_j \left|\psi_j\right\rangle\left\langle\phi_j\right|$ where $\left|\psi_j\right\rangle\left\langle\phi_j\right|$ are normalised dyads, since $U$ and $V$ are unitary. Then $\mathcal{T}(A) = \sum_j \sigma_j \mathcal{T}(\left|\psi_j\right\rangle\left\langle\phi_j\right|)$. So we have:

$$\|\mathcal{T}(A)\|_1 \leq \sum_j \sigma_j \|\mathcal{T}(\left|\psi_j\right\rangle\left\langle\phi_j\right|)\|_1 \leq \sum_j \sigma_j = \|A\|_1, \tag{D.31}$$

where the last step follows since the Schatten 1-norm is the sum of the singular values.

## D.4 Towards Choi matrix criteria for dyadic channels

In this appendix, we give a partial characterisation of Choi matrices for dyadic channels as defined by (4.75) in the main text, restricted to a subset defined as follows.

**Definition D.1** (Bifurcated channels)**.** *We say that a map $\mathcal{T}$ is a* ***bifurcated channel*** *if it can be written in the form:*

$$\mathcal{T} = \mathcal{U}\circ\mathcal{E}\circ\mathcal{V} \tag{D.32}$$

*where $\mathcal{E}$ is a CPTP map, and $\mathcal{U}$ and $\mathcal{V}$ are dyadic unitary maps, i.e. $\mathcal{U}(\cdot) = U_L(\cdot)U_R^\dagger$, and where $U_L$ and $U_R$ are unitary operators, and similarly for $\mathcal{V}$. We use* $\mathrm{BC}_n$ *to denote the set of n-qubit instances of such channels.*

The term bifurcated refers to the fact that a standard CPTP channel has been split into two asymmetric 'branches' by composition with dyadic unitary maps. This class of maps is a subset of the complete dyadic maps $\mathrm{BC}_n \subseteq \mathrm{CD}_n$, and a strict superset of the following stabiliser-preserving subclass.

**Definition D.2** (Bifurcated stabiliser channels)**.** *We say that a map $\mathcal{T}$ is a **bifurcated stabiliser channel** if it has the form $\mathcal{T} = \mathcal{U} \circ \mathcal{E} \circ \mathcal{V}$, where $\mathcal{E}$ is a CPTP map in $\mathrm{SP}_{n,n}$, and $\mathcal{U}$ and $\mathcal{V}$ are dyadic Clifford maps. We denote this class of channels $\mathrm{BSP}_{n,n}$.*

Notice that $\mathrm{BSP}_{n,n}$ is a superset of $\mathrm{PDSP}_{n,n}$ defined in Section 4.2.5.2, as we simply require that $\mathcal{E}$ is a CPTP map, making no assumptions about the weight, orthogonality, or number of Kraus operators. We also have $\mathrm{SP}_{n,n} \subseteq \mathrm{BSP}_{n,n}$ but it is not clear whether all dyadic stabiliser channels can be expressed as a convex mixture of maps in BSP. We can make the following statement about the Choi matrices corresponding to bifurcated channels $\mathcal{T} \in \mathrm{BC}_n$. This applies to non-stabiliser maps as well as maps in $\mathrm{BSP}_{n,n}$.

**Theorem D.3** (Choi matrix criteria for bifurcated channels)**.** *Let $\mathcal{T}$ be a linear n-qubit map. Then $\mathcal{T}$ is a bifurcated channel, $\mathcal{T} \in \mathrm{BC}_n$, if and only if its Choi matrix $\Phi_{\mathcal{T}}^{AB}$ has a polar decomposition $\Phi_{\mathcal{T}}^{AB} = WP$ satisfying:*

1. $\mathrm{Tr}_A[P] = \mathbb{1}_n/2^n$;
2. *$W$ is a tensor product of a pair of n-qubit unitary operators $W^{AB} = U^A \otimes V^B$.*

Note that the first condition is straightforward to check if we know the matrix $\Phi_{\mathcal{T}}$, since $P = \sqrt{\Phi_{\mathcal{T}}^\dagger \Phi_{\mathcal{T}}}$. In other words it depends only on the unique positive operator $P$, and not on the unitary operator $W$. The second condition may be less easy to check, since if $\Phi_{\mathcal{T}}$ is not invertible, $W$ is not unique, and we would need to check whether or not there exists a suitable $W$ that is non-entangling across the partition $A|B$. We will prove Theorem D.3 shortly, but first remark that we can immediately specialise this to criteria for bifurcated stabiliser channels.

**Theorem D.4** (Choi matrix criteria for bifurcated stabiliser channels)**.** *Let $\mathcal{T}$ be a linear n-qubit map. Then $\mathcal{T}$ is a bifurcated* **stabiliser** *channel, $\mathcal{T} \in \mathrm{BSP}_{n,n}$, if and only if its Choi matrix $\Phi_{\mathcal{T}}^{AB}$ has a polar decomposition $\Phi_{\mathcal{T}}^{AB} = WP$ satisfying:*

1. $\mathrm{Tr}_A[P] = \mathbb{1}_n/2^n$*;*

2. *$W$ is a tensor product of a pair of n-qubit Clifford operators $W^{AB} = U^A \otimes V^B$;*

3. *$P$ is a stabiliser state, $P \in \overline{\mathrm{STAB}}_{2n}$.*

*Moreover it is also the case that $\Phi_{\mathcal{T}}$ must be a convex mixture of stabiliser dyads, $\Phi_{\mathcal{T}} = \sum_j p_j e^{i\phi} |L_j\rangle\langle R_j|$.*

We now prove Theorem D.3.

*Proof.* By the arguments of Section 4.2.5.2, we can rewrite any map of the form $\mathcal{T} = \mathcal{U} \circ \mathcal{E} \circ \mathcal{V}$ as follows:

$$\mathcal{T}(\cdot) = \sum_j U_L K_j V_L (\cdot) V_R^\dagger K_j^\dagger U_R^\dagger \tag{D.33}$$

$$= \sum_j U' K_j' V' (\cdot) K_j'^\dagger, \tag{D.34}$$

where $\{K_j'\}$ is a complete set of Kraus operators for some CPTP map $\mathcal{E}$, and $U'$ and $V'$ are unitary operators. Therefore without loss of generality, we can assume that maps in $\mathrm{BC}_n$ take the form in line (D.34). Henceforth we drop the prime to simplify the notation. Consider the Choi matrix for a map of this form,

$$\Phi_{\mathcal{T}} = \sum_j (U K_j V \otimes \mathbb{1}_n) |\Phi_n\rangle\langle\Phi_n| (K_j^\dagger \otimes \mathbb{1}_n). \tag{D.35}$$

Using the fact that $(M \otimes \mathbb{1}_n)|\Phi_n\rangle = (\mathbb{1}_n \otimes M^T)|\Phi_n\rangle$ for any matrix $M$ [46], we have:

$$\Phi_{\mathcal{T}} = (U \otimes V^T) \sum_j (K_j \otimes \mathbb{1}_n) |\Phi_n\rangle\langle\Phi_n| (K_j^\dagger \otimes \mathbb{1}_n) = (U \otimes V^T)\Phi_{\mathcal{E}}, \tag{D.36}$$

where it is important to be clear that $(U \otimes V^T)$ is not a map acting on $\Phi_{\mathcal{E}}$, but a matrix multiplying it on the left. But since $\Phi_{\mathcal{E}}$ is positive semidefinite, and $(U \otimes V^T)$ is a unitary operator, equation (D.36) is the polar decomposition of $\Phi_{\mathcal{T}}$, so:

$$\Phi_{\mathcal{E}} = \sqrt{\Phi_{\mathcal{T}}^{\dagger}\Phi_{\mathcal{T}}}. \tag{D.37}$$

Therefore if $\mathcal{E}$ is a CPTP map, it follows that:

$$\mathrm{Tr}_A\left[\sqrt{\Phi_{\mathcal{T}}^{\dagger}\Phi_{\mathcal{T}}}\right] = \frac{\mathbb{1}_n}{2^n}, \quad \text{and} \quad \mathrm{Tr}\left[\sqrt{\Phi_{\mathcal{T}}^{\dagger}\Phi_{\mathcal{T}}}\right] = 1. \tag{D.38}$$

Conversely, suppose $\mathcal{T}$ is an $n$-qubit linear map with Choi matrix $\Phi_{\mathcal{T}}$, and suppose it has a polar decomposition $\Phi_{\mathcal{E}} = WP$, where $P = \sqrt{\Phi_{\mathcal{T}}^{\dagger}\Phi_{\mathcal{T}}}$ satisfies $\mathrm{Tr}_A P = \mathbb{1}_n/2^n$, and the unitary $W$ can be written as a tensor product of product of two $n$-qubit unitary operators $W = U \otimes V$. We immediately have that $\mathrm{Tr}[P] = \mathrm{Tr}_B[\mathbb{1}_n/2^n] = 1$, and $P$ is always Hermitian and positive semidefinite by construction, so it is a valid, normalised density operator. Since it satisfies the condition on the partial trace, it must be the Choi state for some CPTP map, $\mathcal{E}$, so that $P = \Phi_{\mathcal{E}}$. The channel must have some Kraus representation so:

$$\Phi_{\mathcal{E}} = \sum_j (K_j \otimes \mathbb{1}_n) |\Phi_n\rangle\langle\Phi_n| (K_j^{\dagger} \otimes \mathbb{1}_n) \tag{D.39}$$

for some complete set of Kraus operators $\{K_j\}_j$. Then if the unitary $W$ is a tensor product, the Choi matrix for $\Phi_{\mathcal{T}}$ has the form:

$$\begin{aligned}
\Phi_{\mathcal{T}} &= (U \otimes V) \sum_j (K_j \otimes \mathbb{1}_n) |\Phi_n\rangle\langle\Phi_n| (K_j^{\dagger} \otimes \mathbb{1}_n) && \text{(D.40)}\\
&= \sum_j (UK_jV^T \otimes \mathbb{1}_n) |\Phi_n\rangle\langle\Phi_n| (K_j^{\dagger} \otimes \mathbb{1}_n). && \text{(D.41)}
\end{aligned}$$

Then since the Choi matrix is unique, the map $\mathcal{T}$ can be written:

$$\mathcal{T}(\cdot) = \sum_j UK_jV^T(\cdot)K_j^{\dagger}. \tag{D.42}$$

Therefore $\mathcal{T}$ is a bifurcated channel, $\mathcal{T} \in \mathrm{BC}_n$. □

## D.5 Counting projective dyadic stabiliser channels

For a single qubit, counting the number of maps in PDSP, the set of projective dyadic stabiliser channels (recall Section 4.2.5.2) is relatively straightforward. First we include the dyadic Clifford elements as a special case. As discussed above, we require that Kraus operators have the form $U_j\Pi_j$, where $U_j$ are Cliffords and $\left\{\Pi_j\right\}_j$ is a set of orthogonal projectors. If we take this to be the trivial set containing a single trivial projector $\mathbb{1}$, then we have the Clifford operations. By equation (4.100) the operators acting on the right are also Clifford gates, so we have the dyadic Clifford elements. For a single qubit, there are 576 of these.

Now we count the maps in PDSP corresponding to channels with Kraus rank 2. For the single-qubit case, there are only three possible sets of orthonormal stabiliser-preserving projectors (before we include adaptivity),

$$\left\{\Pi_j^{(X)}\right\} = \left\{\frac{\mathbb{1}\pm X}{2}\right\} = \{|+\rangle\langle+|\,,|-\rangle\langle-|\}, \tag{D.43}$$

$$\left\{\Pi_j^{(Y)}\right\} = \left\{\frac{\mathbb{1}\pm Y}{2}\right\} = \{|+i\rangle\langle+i|\,,|-i\rangle\langle-i|\}, \tag{D.44}$$

$$\left\{\Pi_j^{(Z)}\right\} = \left\{\frac{\mathbb{1}\pm Z}{2}\right\}, = \{|0\rangle\langle0|\,,|1\rangle\langle1|\}. \tag{D.45}$$

Let us consider the set $\left\{\Pi_j^{(Z)}\right\}$. The left operators $\left\{K'_j\right\} = \left\{C_j\Pi_j^{(Z)}\right\}$ take the form:

$$K'_0 = C_0\,|0\rangle\langle0| = |\phi_0\rangle\langle0|\,, K'_1 = C_1\,|1\rangle\langle1| = |\phi_1\rangle\langle1| \tag{D.46}$$

where $|\phi_0\rangle$ and $|\phi_1\rangle$ are some pure single-qubit stabiliser states. Clearly there is redundancy among the Clifford gates in terms of how they map $|0\rangle$ and $|1\rangle$, so we only need to count all combinations of $|\phi_0\rangle$ and $|\phi_1\rangle$: there are 36 in total (neglecting phase, which we will consider shortly), but we need to divide these into three cases. For each $|\phi_0\rangle$, we can have either $\langle\phi_0|\phi_1\rangle = 1$, $\langle\phi_0|\phi_1\rangle = 0$ or $|\langle\phi_0|\phi_1\rangle| = 1/\sqrt{2}$. For each $|\phi_0\rangle$ there is 1 of the first case, 1 of the second, and 4 of the third.

Now for each pair of left operators $K'_0, K'_1$ we need to count the distinct pairs of right operators taking the form $UK'_jV^\dagger = \left|\phi'_j\right\rangle\left\langle\phi''_j\right|$, recalling that for each map $\mathcal{T}$, we use the same Clifford gates $U$ and $V$ for both values of $j$, so that orthogonality must be preserved. For the case where $|\phi_0\rangle = |\phi_1\rangle$, we must have $U\,|\phi_0\rangle = U\,|\phi_1\rangle$, so once $\left|\phi'_0\right\rangle$ is chosen from the 6 possible choices, $|\phi'_1\rangle$ is already fixed. Similarly, when $\langle\phi_0|\phi_1\rangle = 0$, once we fix $\left|\phi'_0\right\rangle$, there is only one stabiliser state $|\phi'_1\rangle$ orthogonal to it, so there are 6 choices here. In the case where $|\langle\phi_0|\phi_1\rangle| = 1/\sqrt{2}$, for each choice of $\left|\phi'_0\right\rangle$, we can map $|\phi'_1\rangle$ to any of the four states such that $|\langle\phi_0'|\phi_1'\rangle| = 1/\sqrt{2}$. Then for each of the possibilities just mentioned, we have 6 possible choices for $\langle 0| V^\dagger = \langle\phi_0''|$. Since we always have $\langle\phi_0''|\phi_1''\rangle = 0$, there is then no further freedom in the choice of $|\phi_1''\rangle$.

We summarise this counting in Table D.1. Following the same counting argument for the other stabiliser projector sets given in (D.43) and (D.44), we obtain another 3888 each, so neglecting phase, in total we have 11664 maps $\mathcal{T}$ of Kraus rank 2.

| Case | Choices for: | | | | | Total choices |
|---|---|---|---|---|---|---|
| | $\|\phi_0\rangle$ | $\|\phi_1\rangle$ | $\|\phi_0'\rangle$ | $\|\phi_1'\rangle$ | $\langle\phi_0''\|$ | |
| $\langle\phi_0\|\phi_1\rangle = 1$ | 6 | 1 | 6 | 1 | 6 | 216 |
| $\langle\phi_0\|\phi_1\rangle = 0$ | 6 | 1 | 6 | 1 | 6 | 216 |
| $\|\langle\phi_0'\|\phi_1'\rangle\| = 1/\sqrt{2}$ | 6 | 4 | 6 | 4 | 6 | 3456 |
| | | | | | | 3888 |

**Table D.1:** Counting the number of elements $\mathcal{T}$ corresponding to the set of orthogonal projectors $\left\{\Pi_j^{(Z)}\right\}$, neglecting phase.

What happens if take phase into account? We want to use these maps in the context of a decomposition $\mathcal{E} = \sum_j c_j \mathcal{T}_j$. Clearly for the dyadic Clifford elements, phase is unimportant as it can be absorbed into the coefficient $c_j$. Similarly for maps with Kraus rank 2, a 'global' phase (i.e. the same phase on both terms $K_j'(\ldots)K_j''$ ) can be absorbed. However, a 'relative' phase between terms $K_0'(\ldots)K_0''$ and $K_1'(\ldots)K_1''$ is non-trivial. For example, the CPTP map $\mathcal{E}_Z$ defined by Kraus operators $|0\rangle\langle 0|, |1\rangle\langle 1|$ is clearly not the same as the dyadic map:

$$\mathcal{T}(\ldots) = |0\rangle\langle 0|(\ldots)|0\rangle\langle 0| - |1\rangle\langle 1|(\ldots)|1\rangle\langle 1| \tag{D.47}$$

which is obtained from $\mathcal{E}_Z$ by setting, for example, $U = Z$ and $V = \mathbb{1}$ in equation (4.100). Let's assume we fix the phase convention for our matrix representation of the 24 Clifford gates (so that for example the Hadamard gives $H|0\rangle = |0\rangle$ etc) . Then the global phases arising from applying a given Clifford to each pure stabiliser state will be of the form $e^{i\pi r/4}$, where $r$ is an integer from 0 to 7. For example, in the standard matrix representation, $H|0\rangle = |+\rangle$, but $H|+i\rangle = e^{i\pi/4}|-i\rangle$. Without loss of generality, we can then assume that for the Kraus rank 2 maps $\mathcal{T}$ the first term $K_0'(\ldots)K_0''^\dagger$ in the decomposition $\mathcal{T}(\ldots) = \sum_s K_s'(\ldots)K_s''^\dagger$ has phase $+1$ (since any phase can be absorbed by the coefficient), while the second term $K_1'(\ldots)K_1''^\dagger$ has phase $e^{i\pi r/4}$, $r \in \mathbb{Z}_8$. Including this factor of 8, this brings the total number of elements $\mathcal{T}$ with Kraus rank 2 to $8 \cdot 11664 = 93312$. Including the Clifford dyadic elements, the total number of distinct maps for the single-qubit case is 93888. Clearly this represents a much larger optimisation problem compared to considering the 576 dyadic Clifford elements, but it would still be feasible to enumerate these maps and solve the problem on a standard laptop or desktop PC. For two qubit channels, the number of elements we need to consider would be astronomical, so the problem is probably not tractable without significant simplification.

## Appendix E

# Sparsification and bit-string sampling technical details

## E.1 Trace norm error for BBCCGH sparsification

As discussed in Section 3.4, the BBCCGH sparsification lemma [3, Lem. 6] entails that, given a pure state with exact stabiliser decomposition $|\psi\rangle = \sum_j c_j \left|\phi_j\right\rangle$, one can randomly generate a $k$-term sparsification $|\Omega\rangle$, such that

$$\mathbb{E}(\| \, |\psi\rangle - |\Omega\rangle \, \|^2) \leq \frac{\|c\|_1^2}{k}, \tag{E.1}$$

where $\|\cdot\|$ is the standard vector norm. In order to compare with our new sparsification result, which deals with density operators, we need to translate this in terms of the trace norm. In this appendix we prove the following simple corollary to the BBCCGH sparsification lemma.

**Corollary 3.8** (BBCCGH sparsification trace-norm error)**.** *Given a normalised state $|\psi\rangle = \sum_j c_j \left|\phi_j\right\rangle$, for any $k > 0$, one can sample from a distribution of sparsified vectors $|\Omega\rangle = (\|c\|_1/k)\sum_{\alpha=1}^{k} |\omega_\alpha\rangle$, where $|\omega_\alpha\rangle$ are stabiliser states, such that:*

$$\mathbb{E}(\| \, |\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega| \, \|_1) \leq 2\frac{\|c\|_1}{\sqrt{k}} + \frac{\|c\|_1^2}{k}. \tag{E.2}$$

*Proof.* Let $|\Delta\rangle = |\psi\rangle - |\Omega\rangle$. Then for any particular $|\Omega\rangle$ we have,

$$
\begin{aligned}
|\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega| = |\psi\rangle\langle\psi| - (|\psi\rangle\langle\psi| + |\Delta\rangle\langle\Delta| \\
- |\Delta\rangle\langle\psi| - |\psi\rangle\langle\Delta|) \qquad \text{(E.3)} \\
= |\Delta\rangle\langle\psi| + |\psi\rangle\langle\Delta| - |\Delta\rangle\langle\Delta|\,. \qquad \text{(E.4)}
\end{aligned}
$$

Using the triangle inequality,

$$
\begin{aligned}
\|\,|\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega|\,\|_1 &\leq 2\|\,|\Delta\rangle\langle\psi|\,\|_1 + \|\,|\Delta\rangle\langle\Delta|\,\|_1 && \text{(E.5)} \\
&= 2\||\Delta\rangle\| \cdot \||\psi\rangle\| + \||\Delta\rangle\|^2 && \text{(E.6)} \\
&= 2\||\Delta\rangle\| + \||\Delta\rangle\|^2, && \text{(E.7)}
\end{aligned}
$$

where the last line follows because $|\psi\rangle$ is normalised. Since the above is true for any $|\Omega\rangle$ taken from the distribution, it follows that,

$$
\mathbb{E}(\|\,|\psi\rangle\langle\psi| - |\Omega\rangle\langle\Omega|\,\|_1) \leq 2\mathbb{E}(\||\Delta\rangle\|) + \mathbb{E}(\||\Delta\rangle\|^2). \qquad \text{(E.8)}
$$

For the second term, the BBCCGH sparsification lemma [3, Lem. 6] tells us that we have $\mathbb{E}(\||\Delta\rangle\|^2) \leq \|c\|_1^2/k$.

This leaves the first term. From Jensen's inequality, for any random variable $X$, we have that $\mathbb{E}(X) \leq \sqrt{\mathbb{E}(X^2)}$. So,

$$
\begin{aligned}
\mathbb{E}(\|\Delta\|) &\leq \sqrt{\mathbb{E}(\|\Delta\|^2)} && \text{(E.9)} \\
&\leq \frac{\|c\|_1}{\sqrt{k}}, && \text{(E.10)}
\end{aligned}
$$

where the second line again follows from Ref.[3, Lem. 6]. Substituting into the inequality (E.8), we obtain the result. □

## E.2 Bit-string sampling simulator technical details

In this appendix, we give full pseudocode for the bit-string sampling simulator presented in Section 5.2.3 (Algorithm 19), prove its validity as a method to classically emulate sampling from the quantum distribution $P(\mathbf{x}) = \mathrm{Tr}[\Pi_{\mathbf{x}}\rho]$, and analyze its runtime. This constitutes a proof of Theorem 5.9. The material in this appendix first appeared in our article

**Algorithm 19** Bit-string sampling for mixed states

```
Input: Decomposition $\rho = \sum_j p_j \left|\psi_j\right\rangle\!\left\langle\psi_j\right|$, where for each $\psi_j$ we have a known stabiliser state decomposition $\left|\psi_j\right\rangle = \sum_r c_r^{(j)} \left|\phi_r^{(j)}\right\rangle$. Real numbers $\delta_S$, $p_{\mathrm{FN}}$ and $\varepsilon_{\mathrm{FN}}$. Number of bits $w$.
Output: String $\mathbf{x}$ of length $w$, sampled from a distribution $P''(\mathbf{x}) = \sum_j p_j P_j(\mathbf{x})$, which approximates $P(\mathbf{x}) = \mathrm{Tr}(\Pi_{\mathbf{x}}\rho)$, where $\Pi_{\mathbf{x}} = |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \mathbb{1}_{n-w}$.
Select index $j$ with probability $p_j$.
$k \leftarrow \lceil 4\|c^{(j)}\|_1^2 \delta_S^{-1} \rceil$
$|\Omega'\rangle \leftarrow$ SPARSIFY($\left|\psi_j\right\rangle$,$k$)
$W \leftarrow$ FASTNORM($|\Omega'\rangle$,$p_{\mathrm{FN}}$,$\varepsilon_{\mathrm{FN}}$)
$|\Omega\rangle \leftarrow |\Omega'\rangle/\sqrt{W}$
$\mathbf{x} \leftarrow ()$ (initialise empty string)
$P_{\mathbf{x}} \leftarrow 1$
for $b \leftarrow 1$ to $w$ do
    $P_{(\mathbf{x},0)} \leftarrow$    ▷ $(\mathbf{x},0)$ denoted the $b$-bit string formed by appending 0 to the $(b-1)$-bit string $\mathbf{x}$. FASTNORM($\Pi_{(\mathbf{x},0)}\,|\Omega\rangle$,$p_{\mathrm{FN}}$,$\varepsilon_{\mathrm{FN}}$)
    $P(x_b = 0|\mathbf{x}) \leftarrow P_{(\mathbf{x},0)}/P_x$
    if $P_{x_b=0} < 1/2$ then
        $P(x_b = 1|\mathbf{x}) \leftarrow 1 - P_{x_b=0}$
    else
        $P_{(\mathbf{x},1)} \leftarrow$ FASTNORM($\Pi_{(\mathbf{x},1)}\,|\Omega\rangle$,$p_{\mathrm{FN}}$,$\varepsilon_{\mathrm{FN}}$)
        $P(x_b = 1|\mathbf{x}) \leftarrow P_{(\mathbf{x},1)}/P_{\mathbf{x}}$
        $P(x_b = 0|\mathbf{x}) \leftarrow 1 - P_{x_b=1}$
    end if
    Select $y \in \{0,1\}$ with probability $P(x_b = y|\mathbf{x})$, then $x_b \leftarrow y$.
    $P_{(\mathbf{x},x_b)} \leftarrow P_{\mathbf{x}} \times P(x_b|\mathbf{x})$
    $\mathbf{x} \leftarrow (\mathbf{x}, x_b)$
end for
return $\mathbf{x}$
```

Ref. [2]. As described in Section 5.2.3, Algorithm 19 draws bit strings $\mathbf{x}$ from a classical distribution $P_{\mathrm{sim}}(\mathbf{x})$, using two subroutines from Ref. [3], SPARSIFY and FASTNORM. As sketched in the main text, our strategy is to define an idealised algorithm EXACT where calls to FASTNORM are replaced by an oracle which can compute $\left\|\Pi_{\vec{y}}\,|\Omega\rangle\right\|$ exactly for any un-normalised $|\Omega\rangle$ and bit string $\vec{y}$. The algorithm EXACT draws from a distribution $P_{\mathrm{ex}}(\mathbf{x})$. We first show that $P_{\mathrm{ex}}$ is $\delta_S$-close to the quantum distribution $P$. We then argue that the distribution $P_{\mathrm{sim}}$ that Algorithm 19 draws from is $\varepsilon$-close to $P_{\mathrm{ex}}$. Finally we optimise the choice of $\delta_S$ and $\varepsilon$ and analyze the runtime.

EXACT is identical to our Algorithm 19, except where our algorithm estimates probabilities $\left\|\Pi_{\vec{y}}\,|\Omega\rangle\right\|^2$ using FASTNORM, EXACT computes them exactly. Therefore EX-

ACT first samples a state $\left|\psi_j\right\rangle$ from the ensemble with probability $p_j$, and chooses a sparsification $|\Omega\rangle = \left|\Omega_{j,l}\right\rangle$ with probability $q_{j,l} = \Pr\left(\Omega_{j,l}|\psi_j\right)$. Given the selected $|\Omega\rangle$, a bit string is sampled by choosing each bit in turn via a series of conditional probabilities:

$$\begin{aligned}\Pr(\mathbf{x}|\Omega) &= \Pr(x_1)\Pr(x_2|\mathbf{x}_1)\ldots\Pr(x_w|\mathbf{x}_{w-1}) && \text{(E.11)}\\ &= \frac{\|\Pi_{\mathbf{x}}|\Omega\rangle\|^2}{\||\Omega\rangle\|^2} = \operatorname{Tr}\left[\Pi_{\mathbf{x}}\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}\right]. && \text{(E.12)}\end{aligned}$$

Here we use the notation $\mathbf{x}_m$ to denote the string comprised of the first $m$ bits of $\mathbf{x}$, so that $\Pi_{\mathbf{x}_m} = \bigotimes_{j=1}^{m}\left|x_j\right\rangle\left\langle x_j\right| \otimes \mathbb{1}_{n-m}$. We take $\mathbf{x}_0$ to be the empty string, so that $\Pi_{\mathbf{x}_0} = \mathbb{1}$. The probability of choosing $y \in 0,1$ for the $m$-th bit, given $m-1$ bits already sampled, is computed as:

$$\Pr(y|\mathbf{x}_{m-1}) = \left\|\Pi_{(x_1,\ldots,x_{m-1},y)}|\Omega\rangle\right\|^2 / \left\|\Pi_{\mathbf{x}_{m-1}}|\Omega\rangle\right\|. \tag{E.13}$$

Thus EXACT outputs bit strings X sampled from a distribution:

$$\begin{aligned}P_{\mathrm{ex}}(\mathbf{x}) &= \sum_j\sum_l p_j q_{j,l}\frac{\|\Pi_{\mathbf{x}}|\Omega_{j,l}\rangle\|^2}{\||\Omega_{j,l}\rangle\|^2} && \text{(E.14)}\\ &= \sum_j\sum_l p_j q_{j,l}\frac{\operatorname{Tr}\left[\Pi_{\mathbf{x}}|\Omega_{j,l}\rangle\langle\Omega_{j,l}|\right]}{\langle\Omega_{j,l}|\Omega_{j,l}\rangle}\\ &= \operatorname{Tr}\left[\Pi_{\mathbf{x}}\sum_j p_j\sum_l q_{j,l}\frac{|\Omega_{j,l}\rangle\langle\Omega_{j,l}|}{\langle\Omega_{j,l}|\Omega_{j,l}\rangle}\right]\\ &= \operatorname{Tr}\left[\Pi_{\mathbf{x}}\sum_j p_j\mathbb{E}\left(\frac{|\Omega_j\rangle\langle\Omega_j|}{\langle\Omega_j|\Omega_j\rangle}\right)\right] = \operatorname{Tr}\left[\Pi_{\mathbf{x}}\rho'\right],\end{aligned}$$

where $\rho' = \sum_j p_j\rho_1^{(j)}$, and each $\rho_1^{(j)}$ given by

$$\rho_1^{(j)} := \sum_{\Omega}\Pr\left(\Omega|\psi_j\right)\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}. \tag{E.15}$$

In other words $\rho_1^{(j)}$ is the expected sparsification given target pure state $\left|\psi_j\right\rangle$, as defined in Eq. (5.48). In step 2, $k$ is chosen so that by our improved sparsification lemma (Theorem 5.6 in Section 5.2.3.1), we have $\left\|\rho_1^{(j)} - \left|\psi_j\right\rangle\left\langle\psi_j\right|\right\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2)$, provided $\delta_S \geq \delta_c$, where $\delta_c$ is the critical precision. We will return to the $\delta_S < \delta_c$ case at the end of this appendix. By the triangle inequality we have

$$\left\|\rho' - \rho\right\|_1 = \left\| \sum_j p_j \rho_1^{(j)} - \sum_j p_j \left|\psi_j\right\rangle\left\langle\psi_j\right| \right\|_1 \tag{E.16}$$

$$\leq \sum_j p_j \left\| \rho_1^{(j)} - \left|\psi_j\right\rangle\left\langle\psi_j\right| \right\|_1 \tag{E.17}$$

$$\leq \sum_j p_j [\delta_S + \mathcal{O}(\delta_S^2)] = \delta_S + \mathcal{O}(\delta_S^2). \tag{E.18}$$

Since $P_{\text{ex}}(\mathbf{x}) = \text{Tr}[\Pi_{\mathbf{x}}\rho']$ and for the quantum distribution we have $P(\mathbf{x}) = \text{Tr}[\Pi_{\mathbf{x}}\rho']$, It follows that $\|P_{\text{ex}} - P\|_1 \leq \delta_S + \mathcal{O}(\delta_S^2)$.

It remains to show that using a sequence of calls to FASTNORM, Algorithm 19 generates probability distributions $P_{\text{sim}}(x)$ that well approximate $P_{\text{ex}}(\mathbf{x})$, where

$$P_{\text{sim}}(\mathbf{x}) = \sum_j p_j q_{j,l} Q_{j,l}(\mathbf{x}). \tag{E.19}$$

Here each $Q_{j,l}(\mathbf{x})$ is the probability of Algorithm 19 returning $\mathbf{x}$ given the sparsification $\left|\Omega_{j,l}\right\rangle$. We now drop the subscript as we consider a single sparsification $|\Omega\rangle$. Recall that FASTNORM takes as input error parameters $p_{\text{FN}}$ and $\varepsilon_{\text{FN}}$, and un-normalised vectors $\Pi_{\vec{y}}|\Omega\rangle$ with known $k$-term stabiliser decomposition. Then with probability $(1 - p_{\text{FN}})$ it returns a random variable $\eta$ that approximates $\left\|\Pi_{\vec{y}}|\Omega\rangle\right\|^2$ to within a multiplicative error of $\varepsilon_{\text{FN}}$,

$$(1 - \varepsilon_{\text{FN}})\left\|\Pi_{\vec{y}}|\Omega\rangle\right\|^2 \leq \eta \leq (1 + \varepsilon_{\text{FN}})\left\|\Pi_{\vec{y}}|\Omega\rangle\right\|^2. \tag{E.20}$$

Algorithm 19 approximates the chain of conditional probabilities E.11 by calls to FASTNORM. The probability of choosing $y \in \{0, 1\}$ for the $m$-th bit of $\mathbf{x}$, conditioned on the first $m-1$ bits being $\mathbf{x}_{m-1}$ is therefore bounded as:

$$\varepsilon_- \frac{\left\|\Pi_{(\mathbf{x}_{m-1}, y)}|\Omega\rangle\right\|^2}{\left\|\Pi_{\mathbf{x}_{m-1}}|\Omega\rangle\right\|^2} \leq \Pr(y|\mathbf{x}_{m-1}) \leq \varepsilon_+ \frac{\left\|\Pi_{(\mathbf{x}_{m-1}, y)}|\Omega\rangle\right\|^2}{\left\|\Pi_{\mathbf{x}_{m-1}}|\Omega\rangle\right\|^2},$$

with probability $(1 - p_{\text{FN}})^2$, where

$$\varepsilon_\pm = \frac{1 \pm \varepsilon_{\text{FN}}}{1 \mp \varepsilon_{\text{FN}}}. \tag{E.21}$$

So, given a particular sparsification $|\Omega\rangle$, the $w$-bit string $\mathbf{x}$ is sampled from a distribution

$Q(\mathbf{x})$ which satisfies

$$\prod_{m=1}^{w} \varepsilon_{-} \frac{||\Pi_{\mathbf{x}_m}|\Omega\rangle||^2}{\left\|\Pi_{\mathbf{x}_{m-1}}|\Omega\rangle\right\|^2} \leq Q(\mathbf{x}) \leq \prod_{m=1}^{w} \varepsilon_{+} \frac{||\Pi_{\mathbf{x}_m}|\Omega\rangle||^2}{\left\|\Pi_{\mathbf{x}_{m-1}}|\Omega\rangle\right\|^2}$$

with probability at least $(1-p_{\mathrm{FN}})^{2w}$. This simplifies to

$$\frac{(1-\varepsilon_{\mathrm{FN}})^w||\Pi_{\mathbf{x}}|\Omega\rangle||^2}{(1+\varepsilon_{\mathrm{FN}})^w|||\Omega\rangle||^2} \leq Q(\mathbf{x}) \leq \frac{(1+\varepsilon_{\mathrm{FN}})^w||\Pi_{\mathbf{x}}|\Omega\rangle||^2}{(1-\varepsilon_{\mathrm{FN}})^w|||\Omega\rangle||^2}. \tag{E.22}$$

One can check that $(1+\varepsilon_{\mathrm{FN}})^w/(1-\varepsilon_{\mathrm{FN}})^w \leq 1+3w\varepsilon_{\mathrm{FN}}$, whenever $\varepsilon_{\mathrm{FN}} \leq 1/5$, and the analogous result holds for the lower bound. Therefore $Q_{j,l}(\mathbf{x})$ approximates $\left\|\Pi_{\mathbf{x}}\left|\Omega_{j,l}\right\rangle\right\|^2/\left\|\left|\Omega_{j,l}\right\rangle\right\|^2$ up to multiplicative error $3w\varepsilon_{\mathrm{FN}}$. Comparing (E.14) with (E.19), we therefore obtain:

$$(1-3w\varepsilon_{\mathrm{FN}})P_{\mathrm{ex}}(\mathbf{x}) \leq P_{\mathrm{sim}}(\mathbf{x}) \leq (1+3w\varepsilon_{\mathrm{FN}})P_{\mathrm{ex}}(\mathbf{x}) \tag{E.23}$$

If we want to bound the total multiplicative error due to the sequence of calls to FASTNORM to $\varepsilon$, then we must set $\varepsilon_{\mathrm{FN}} = \varepsilon/(3w)$. It then follows that

$$||P_{\mathrm{sim}} - P_{\mathrm{ex}}||_1 \leq \varepsilon. \tag{E.24}$$

In the first part of the proof we showed that $\|P_{\mathrm{ex}} - P\|_1 \leq \delta_S + O(\delta_S^2)$ (provided we are above the critical precision threshold $\delta_c$). Combined with Eq. (E.24), we obtain

$$||P_{\mathrm{sim}} - P||_1 \leq \varepsilon + \delta_S + \mathcal{O}(\delta_S^2), \tag{E.25}$$

where $P(\mathbf{x}) = \mathrm{Tr}[\Pi_{\mathbf{x}}\rho]$.

Similarly the error bound given above is only obtained with probability $(1 - p_{\mathrm{FN}})^{2w} \approx 1 - 2wp_{\mathrm{FN}}$, so to obtain the above closeness in $\ell_1$-norm, with failure probability at most $p_{\mathrm{fail}}$, we must set $p_{\mathrm{FN}} = p_{\mathrm{fail}}/(2w)$. If we select the state $\left|\psi_j\right\rangle$ in step 1, then $k \leq 4\left\|c^{(j)}\right\|_1^2 \delta_S^{-1} + 1$. To return a single bit-string $\mathbf{x}$ there are at most $2w$ calls to FASTNORM, so the runtime is $\mathcal{O}(wkn^3\varepsilon_{\mathrm{FN}}^{-2}\log p_{\mathrm{FN}}^{-1}) = \mathcal{O}(w^3n^3\|c^{(j)}\|_1^2\delta_S^{-1}\varepsilon^2\log(w/p_{\mathrm{fail}}))$. Recall that the statement of the theorem defined the quantity $\widetilde{\Xi} = \sum_j p_j\|c^{(j)}\|_1^2$, so that the time $T$ to obtain a single bit string is non-deterministic. The *expected* (average-case) runtime is $\mathcal{O}(w^3 n^3 \widetilde{\Xi} \delta_S^{-1} \varepsilon^2 \log(w/p_{\text{fail}}))$. If the decomposition is optimal with respect to the monotone $\Xi$, then we have $\widetilde{\Xi} = \Xi(\rho)$ and the average-case runtime is $\mathcal{O}(\Xi(\rho))$. For equimagical states, $\Xi(\rho) = \xi(\psi_j)$ for all $j$, and this expression becomes the worst-case runtime.

We now optimise the choice of $\delta_S$ and $\varepsilon$. Setting the total error budget $\delta = \delta_S + \varepsilon$, by inspecting the runtime we find that the best constant is obtained by setting $\delta_S = \delta/3$ and $\varepsilon = 2\delta/3$. Substituting the optimal choice of $\delta_S$ and $\varepsilon$ into the expected runtime, we obtain

$$\mathbb{E}(T) = \mathcal{O}(w^3 n^3 \widetilde{\Xi} \delta^{-3} \log(w/p_{\text{fail}})). \tag{E.26}$$

The above holds for the case where the sparsification error $\delta_S$ is no smaller than a critical value $\delta_c = 8(C_j - 1)/\|c^{(j)}\|_1^2$, where $C_j = \left\|c^{(j)}\right\|_1 \sum_r |c_r| |\langle \psi_j | \phi_r \rangle|^2$ is defined for the randomly chosen pure state $|\psi_j\rangle$. Therefore, to ensure we are above the critical error regime for any $|\psi_j\rangle$, we can require that $\delta_S \geq 8D$, where $D = \max\{(C_j - 1)/\|c^{(j)}\|_1^2\}$. This entails $\delta \geq 24D$ for the overall precision.

Now suppose that we want to achieve arbitrary precision, $\delta < 24D$. In this regime, one can amend the expression for $k$ in step 2 to achieve any desired precision, at the cost of slightly poorer scaling in the runtime. We first use Lemmata 5.7 and 5.8 to obtain a sharpened bound on the sparsification error:

$$\delta_S \leq 2\frac{\|c^{(j)}\|_1^2}{k} + \sqrt{\frac{\|c^{(j)}\|_1^2}{k}} \sqrt{4D + 2\frac{\|c^{(j)}\|_1^2}{k} + \mathcal{O}\left(\frac{1}{k^2}\right)}. \tag{E.27}$$

When $\delta_S \ll 8D$, we can achieve a precision of $\delta_S$ by choosing

$$k \approx 4\|c^{(j)}\|_1^2 \left(\frac{D}{\delta_S^2} + \frac{1}{\delta_S}\right) + \mathcal{O}(1). \tag{E.28}$$

Substituting the revised expression for $k$ into the expected runtime, with $\delta_S = \delta/3$ and $\varepsilon = 2\delta/3$, we obtain:

$$\mathbb{E}(T) = \mathcal{O}(w^3 n^3 \widetilde{\Xi} (\delta^{-3} + 3D\delta^{-4}) \log(w/p_{\text{fail}})). \tag{E.29}$$

Here we recover the same asymptotic $\delta^{-4}$ scaling as derived from the original BBC-CGH sparsification lemma [3]. However, the prefactor from this prior work was two, whereas our prefactor $D$ is typically exponentially small in the number of qubits (see Section 5.2.3.1). Therefore, at intermediate precision, the $\delta^{-3}$ term may still dominate. When the target precision $\delta$ is too small, our bound on the required $k$ exceeds the number of terms in the exact decomposition of $|\psi\rangle$ (i.e. the decomposition achieving the stabiliser rank $\chi(\psi)$). In this scenario, using a sparsified approximation in both our approach and in [3] has no benefit, and one should instead use an exact decomposition without any sparsification.

## E.3 Observable estimation for sparsified mixed states

**Algorithm 20** Observable estimator using sparsification and fast norm estimation

```
Input: Decomposition $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, where for each $\psi_j$ we have a known stabiliser state decomposition $|\psi_j\rangle = \sum_r c_r^{(j)} |\phi_r^{(j)}\rangle$; number of samples $M$; sparsification parameter $\delta_S$; fast norm error $\varepsilon_{\mathrm{FN}}$ and failure probability $p_{\mathrm{FN}}$.
Output: Additive error estimate for $\mathrm{Tr}[\Pi\rho]$.
$\eta' \leftarrow 0$
for $m \leftarrow 1$ to $M$ do
    Select index $j$ with probability $p_j$.
    $k \leftarrow \lceil 4\|c^{(j)}\|_1^2 \delta_S^{-1} \rceil$
    $|\Omega\rangle \leftarrow$ SPARSIFY($|\psi_j\rangle$,$k$)
    $W \leftarrow$ FASTNORM($|\Omega\rangle$,$p_{\mathrm{FN}}$,$\varepsilon_{\mathrm{FN}}$)
    $V \leftarrow$ FASTNORM($\Pi|\Omega\rangle$,$p_{\mathrm{FN}}$,$\varepsilon_{\mathrm{FN}}$)
    $\eta_m \leftarrow V/W$                ▷ Unless $W = 0$, but this is vanishingly unlikely.
    $\eta' \leftarrow \eta' + \eta_m$
end for
return $\eta \leftarrow \eta'/M$
```

In this appendix we explain how one can use sparsification and stabiliser rank techniques to estimate observables rather than sample bit strings from a distribution. Let

$$\rho' = \sum_j \sum_\Omega p_j \Pr(\Omega|j) \frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle} = \sum_\Omega \Pr(\Omega) \frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}. \tag{E.30}$$

where $\Pr(\Omega|j)$ is the probability of SPARSIFY outputting the unnormalised vector $|\Omega\rangle$, given sampled pure state $|\psi_j\rangle$, and $\Pr(\Omega) = \sum_j p_j \Pr(\Omega|j)$ is the total probability of obtaining $|\Omega\rangle$ given target state $\rho$. Therefore $\rho'$ is the ensemble from which our classical algorithm draws $|\Omega\rangle/\|\Omega\|$. By the same argument as for the bit-string simulator,

$$\|\rho - \rho'\|_1 \leq \delta_S + O(\delta_S^2) \quad \Longrightarrow \quad \left|\mathrm{Tr}[\Pi\rho] - \mathrm{Tr}[\Pi\rho']\right| \leq \frac{\delta_S}{2} + O(\delta_S^2) \tag{E.31}$$

We want to show the algorithm constitutes a (possibly biased) estimator for $\mathrm{Tr}[\Pi\rho']$, which is $\delta_S$-close to the quantum mean value.

There is a complication in that there is a small possibility that fast norm estimation fails badly. To deal with this, we add the following fix to step 8. We assume that $\varepsilon_{\mathrm{FN}} \leq 1/5$ so that if $\eta_m > 1+3\varepsilon_{\mathrm{FN}}$, we know for certain that FASTNORM has failed for at least one of $|||\Omega\rangle||^2$ or $||\Pi|\Omega\rangle||^2$. Therefore if this happens we cap the estimate, $\eta_m = 1+3\varepsilon_{\mathrm{FN}}$. This includes the case where $W = 0$. This fix may have the effect of biasing the procedure, but since failure probability is exponentially suppressed, the bias can be made arbitrarily small with negligible cost.

With this establised, let us check the mean value of the random variable $\eta$. Since all samples are produced in the same way $\mathbb{E}(\eta) = \mathbb{E}(\eta_m)$. In each iteration, we sample $|\Omega\rangle$ from the ensemble $\rho'$ then compute the fast norm estimate, which is accurate up to relative error with probability $(1-p') = (1-p_{\mathrm{FN}})^2$. Then with probability $p'$ the fast norm estimate could be faulty. Let $\eta_\Omega$ be the random variable $\eta_m$ conditioned on a particular sparsification $|\Omega\rangle$. Let $\eta_{\mathrm{FAIL}}$ be the random variable output when the fast norm estimation fails. We do not know the mean value of this variable but using our cap we do know that it satisfies $0 \leq \mathbb{E}(\eta_{\mathrm{FAIL}}) \leq 1+3\varepsilon_{\mathrm{FN}}$. If $\eta_{\mathrm{SUCCESS}}$ is the random variable conditioned on success and a particular $|\Omega\rangle$, then we know that:

$$\frac{1-\varepsilon_{\mathrm{FN}}}{1+\varepsilon_{\mathrm{FN}}}\frac{||\Pi|\Omega\rangle||^2}{|||\Omega\rangle||^2} \leq \eta_{\mathrm{SUCCESS}} \leq \frac{1+\varepsilon_{\mathrm{FN}}}{1-\varepsilon_{\mathrm{FN}}}\frac{||\Pi|\Omega\rangle||^2}{|||\Omega\rangle||^2} \tag{E.32}$$

The expected value of the random variable $\eta_\Omega$ is

$$\mathbb{E}(\eta_\Omega) = (1-p')\mathbb{E}(\eta_{\mathrm{SUCCESS}}) + p'\mathbb{E}(\eta_{\mathrm{FAIL}}). \tag{E.33}$$

So we can bound $\mathbb{E}(\eta_\Omega)$ as,

$$(1-p')\left(\frac{1-\varepsilon_{\mathrm{FN}}}{1+\varepsilon_{\mathrm{FN}}}\right)\frac{||\Pi|\Omega\rangle||^2}{|||\Omega\rangle||^2} \leq \mathbb{E}(\eta_\Omega) \leq (1-p')\left(\frac{1+\varepsilon_{\mathrm{FN}}}{1-\varepsilon_{\mathrm{FN}}}\right)\frac{||\Pi|\Omega\rangle||^2}{|||\Omega\rangle||^2} + p'(1+3\varepsilon_{\mathrm{FN}}). \tag{E.34}$$

To bound the expected value of the final output of the algorithm, we write

$\mathbb{E}(\eta) = \sum_\Omega \Pr(\Omega)\eta_\Omega$. So,

$$\varepsilon_- \sum_\Omega \Pr(\Omega) \frac{\|\Pi\,|\Omega\rangle\|^2}{\||\Omega\rangle\|^2} \le \mathbb{E}(\eta) \le \varepsilon_+ \left( \Pr(\Omega) \frac{\|\Pi\,|\Omega\rangle\|^2}{\||\Omega\rangle\|^2} \right) + p'(1+3\varepsilon_{\mathrm{FN}}) \tag{E.35}$$

where $\varepsilon_\pm = (1-p')(1 \pm \varepsilon_{\mathrm{FN}})/(1 \mp \varepsilon_{\mathrm{FN}}) \approx 1 \pm 3\varepsilon_{\mathrm{FN}}$.

But,

$$\sum_\Omega \Pr(\Omega) \frac{\|\Pi\,|\Omega\rangle\|^2}{\||\Omega\rangle\|^2} = \sum_\Omega \Pr(\Omega) \operatorname{Tr}\left[\Pi \frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle}\right] = \operatorname{Tr}\left[\Pi\rho'\right]. \tag{E.36}$$

So we have

$$\varepsilon_- \operatorname{Tr}\left[\Pi\rho'\right] \le \mathbb{E}(\eta) \le \varepsilon_+ \operatorname{Tr}\left[\Pi\rho'\right] + p'(1+3\varepsilon_{\mathrm{FN}}) \tag{E.37}$$

We can also place bounds on the absolute value of each sample, since if fast norm is successful, the estimate will always be between 0 and $1+3\varepsilon_{\mathrm{FN}}$, and due to our cap, the same is true for the case when FASTNORM fails. Using the Hoeffding bound,

$$\Pr\left[|\eta - \mathbb{E}(\eta)| \ge \varepsilon\right] \le 2\exp\left(-\frac{2M\varepsilon^2}{(1+3\varepsilon_{\mathrm{FN}})^2}\right) \tag{E.38}$$

Fix a required success probability $p_H$. Then $\eta$ is $\varepsilon$-close to $\mathbb{E}(\eta)$, provided we set the number of samples to

$$M = \left\lceil \frac{1}{2\varepsilon^2}(1+3\varepsilon_{\mathrm{FN}})^2 \log\left(\frac{2}{p_H}\right) \right\rceil \tag{E.39}$$

Since $(1+3\varepsilon_{\mathrm{FN}})^2 \le 4$, we can ignore this factor when considering the runtime scaling.

Let us assume that we have an optimal decomposition, so that the average runtime for each call to SPARSIFY is $\mathcal{O}\left(\Xi\delta_S^{-1}\right)$, and average runtime for each call to FASTNORM is $\mathcal{O}\left(\Xi\delta_S^{-1} n^3 \varepsilon_{\mathrm{FN}}^{-2} \log\left(p_{\mathrm{FN}}^{-1}\right)\right)$. There are $M$ calls in total, so on average the total runtime is

$$\tau = \mathcal{O}\left( \frac{\Xi n^3}{\varepsilon^2 \varepsilon_{\mathrm{FN}}^2 \delta_S} \log\left(\frac{2}{p_H}\right) \log\left(\frac{1}{p_{\mathrm{FN}}}\right) \right). \tag{E.40}$$

This is the runtime to produce an estimate for $\operatorname{Tr}[\Pi\rho']$. But from our sparsification lemma we know that this is $\delta_S$-close to the quantum mean value. So the output from the algorithm satisfies:

$$\varepsilon_- \operatorname{Tr}[\Pi\rho] - \mathcal{O}(\delta_S) - \varepsilon \le \eta \le \varepsilon_+ \operatorname{Tr}\left[\Pi\rho'\right] + p'(1+3\varepsilon_{\mathrm{FN}}) + \mathcal{O}(\delta_S) + \varepsilon \tag{E.41}$$

Since $p' \approx 2p_{\mathrm{FN}}$, the term $p'(1+3\varepsilon_{\mathrm{FN}})$ can be made arbitrarily small with logarithmic cost in the runtime. The final step is to balance the error budget between additive errors $\delta_S$ and $\varepsilon$ and the relative error $\varepsilon_{\pm}$, in order to optimise runtime. We can straightforwardly combine $\delta_S$ and $\varepsilon$ - using similar analysis as for the bit-string simulator, we should set $\delta_S \propto \varepsilon \propto \delta$, for target additive error $\delta$. How we balance with the relative error depends on the size of $\mathrm{Tr}[\Pi\rho]$. If close to 1, it would be optimal to set $\varepsilon_{\mathrm{FN}}$ to the same order as $\delta$. If $\mathrm{Tr}[\Pi\rho]$ is known to be close to zero, the optimal strategy would be to spend more of the runtime budget on reducing the additive error. Neglecting small error terms that can be exponentially suppressed, the upshot is that we can use this algorithm to get an estimate $\eta$ for $\mathrm{Tr}[\Pi\rho]$ satisfying

$$|\eta - \mathrm{Tr}[\Pi\rho]| \lessapprox 3\varepsilon_{\mathrm{FN}}\,\mathrm{Tr}[\Pi\rho] + \delta \tag{E.42}$$

with high probability in average runtime $\mathcal{O}\left(\Xi n^3 \varepsilon_{\mathrm{FN}}^{-2}\delta^{-3}\right)$. This should be compared with runtime $\mathcal{O}\left(n^3\Lambda^2\delta^{-2}\right)$ for the dyadic frame simulator, or $\mathcal{O}\left(n^2\mathcal{R}^2\delta^{-2}\right)$ for robustness of magic. The polynomial scaling in $n$ is based on the assumption that we simply estimate a single Pauli projector for input state $\rho$. For the fast norm and dyadic frame algorithms, we always need to compute stabiliser inner products, which take time $\mathcal{O}\left(n^3\right)$, whereas for quasiprobability sampling with stabiliser projectors, we only need to simulate Pauli measurements for each sampled pure stabiliser state $|\phi\rangle\langle\phi|$, which takes time $\mathcal{O}\left(n^2\right)$. For general stabiliser code projectors we would recover the $n^3$ scaling for the robustness of magic simulator as well.

## E.4 Computing Z-rotation angle for given unitary extent

Suppose we are given the value of $\xi(U_\alpha) = \xi(|\psi_\alpha\rangle)$, where $U_\alpha = \exp(-iZ\alpha)$ and $|\psi_\alpha\rangle = U_\alpha|+\rangle$. Assume we want to find unknown $\alpha$, satisfying $0 \leq \alpha \leq \pi/8$. This task is relevant for Algorithm 18 in Section 6.2.2, which searches for an equimagical decomposition for a depolarised qubit rotation. Recall that the pure state extent for the single-qubit state $|\psi\rangle$ has an analytic expression [3],

$$\xi(|\psi_\alpha\rangle) = \left(\cos\left(\frac{\pi}{4} - \alpha\right) + (\sqrt{2}-1)\sin\left(\frac{\pi}{4} - \alpha\right)\right)^2. \tag{E.43}$$

Using standard trigonometric identities,

$$\sqrt{\xi(|\psi_\alpha\rangle)} = \cos\left(\frac{\pi}{4}-\alpha\right) + (\sqrt{2}-1)\sin\left(\frac{\pi}{4}-\alpha\right) \tag{E.44}$$

$$= \frac{2}{\sqrt{2+\sqrt{2}}}\left(\frac{\sqrt{2+\sqrt{2}}}{2}\cos\left(\frac{\pi}{4}-\alpha\right) + \frac{\sqrt{2-\sqrt{2}}}{2}\sin\left(\frac{\pi}{4}-\alpha\right)\right) \tag{E.45}$$

$$= \left(\sin\left(\frac{3\pi}{8}\right)\right)^{-1}\left[\sin\left(\frac{3\pi}{8}\right)\cos\left(\frac{\pi}{4}-\alpha\right) + \sin\left(\frac{\pi}{4}-\alpha\right)\cos\left(\frac{3\pi}{8}\right)\right] \tag{E.46}$$

$$= \left(\sin\left(\frac{3\pi}{8}\right)\right)^{-1}\sin\left(\frac{5\pi}{8}-\alpha\right) \tag{E.47}$$

$$= \left(\sin\left(\frac{3\pi}{8}\right)\right)^{-1}\sin\left(\alpha+\frac{3\pi}{8}\right). \tag{E.48}$$

Then in the interval $0 \leq \alpha \leq \pi/8$, there is a unique solution for $\alpha$,

$$\alpha = \arcsin\left(\sqrt{\xi(|\psi_\alpha\rangle)}\sin\left(\frac{3\pi}{8}\right)\right). \tag{E.49}$$

# Appendix F

# Validity and runtime analysis for dyadic channel simulator

In this appendix we give a full proof of Theorem 5.12, which affirms the validity and runtime scaling of the static dyadic channel simulator. For convenience we restate the pseudocode in Algorithm 21 here. We have assumed in the description of

**Algorithm 21** Static dyadic channel simulator

**Input:** Circuit description $\left\{\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \ldots, \mathcal{E}^{(T)}\right\}$, where each $\mathcal{E}^{(t)}$ has a known decomposition with $\ell_1$-norm $B^{(t)}$ as per equation (5.125), and each dyadic stabiliser subchannel $\mathcal{T}_k^{(j)} \in DSP$ is specified by a simulable decomposition $(\mathbb{L}_k^{(t)}, \mathbb{R}_k^{(t)})$; stabiliser observable $E$; initial stabiliser state $\sigma^{(0)} = |\phi^{(0)}\rangle\langle\phi^{(0)}|$; number of samples $M$.

**Output:** Estimate $\widehat{E}$ for expectation value $\langle E\rangle = \mathrm{Tr}\left[E\mathcal{E}(|\phi^{(0)}\rangle\langle\phi^{(0)}|)\right]$.

```
Set $\widetilde{E} \leftarrow 0$, $\theta \leftarrow 0$, $B \leftarrow \prod_{t=1}^{T} B^{(t)}$.        ▷ Initialise phase angle $\theta$.
for $j = 1$ to $M$ do
    Prepare representation of initial state $\sigma^{(0)}$.
    for $t = 1$ to $T$ do
        Sample $k_t$ with probability $|\beta_{k_t}^{(t)}|/B^{(t)}$.
        $\theta \leftarrow \theta + \arg\beta_{k_t}^{(t)}$        ▷ Update phase angle.
        $\sigma^{(t)} \leftarrow \textsc{dyadicMapUpdate}(\sigma^{(t-1)}, (\mathbb{L}_{k_t}^{(t)}, \mathbb{R}_{k_t}^{(t)}))$
        if $\sigma^{(t)} = 0$ then        ▷ Terminate trajectory if "zero" selected.
            $\sigma^{(T)} \leftarrow 0$
            break
        end if
    end for
    $\widetilde{E}_j \leftarrow \mathrm{Re}\left\{Be^{i\theta}\,\mathrm{Tr}\left[E\sigma^{(T)}\right]\right\}$        ▷ Computed with CH-simulator.
    $\widetilde{E} \leftarrow \widetilde{E} + \widetilde{E}_j$
end for
return $\widehat{E} \leftarrow \widetilde{E}/M$
```

DYADICMAPUPDATE that the Kraus operators comprising the decomposition $(\mathbb{L}_k^{(t)}, \mathbb{R}_k^{(t)})$ are supplied in polar form. Here for clarity we simplify the notation and assume that each dyadic stabiliser channel is decomposed as

$$\mathcal{T}_k^{(t)} = \sum_r \mathcal{K}_{k,r}^{(t)}. \tag{F.1}$$

where each $\mathcal{K}_{k,r}^{(t)}(\cdot) = L_j(\cdot)R_j^\dagger$ is a dyadic map (absorbing any prefactors), which we will call a Kraus map, that takes stabiliser dyads either to (possibly sub-normalised) stabiliser dyads, or the zero operator. Then let $\mathbf{k} = (k_1, \ldots, k_T)$ denote a vector of indices labelling a sequence of dyadic maps $\mathcal{T}_{k_t}^{(t)}$ through the circuit, and let $\mathbf{r} = (r_1, \ldots, r_T)$ specify a choice of Kraus map $\mathcal{K}_{k,r}^{(t)}$ for each step $t$. Thus a particular trajectory through the circuit is specified by a pair of vectors $(\mathbf{k}, \mathbf{r})$. Note that as for the dyadic frame simulator, some trajectories would lead to the stabiliser dyad being mapped to zero, so these trajectories have zero probability of being sampled, and we again let $\mathbb{P}$ denote the set of trajectories $(\mathbf{k}, \mathbf{r})$ where $\Pr(\mathbf{k}, \mathbf{r}) \neq 0$. Let $\sigma_{(\mathbf{k},\mathbf{r})}^{(t)}$ denote the dyad output by DYADICMAPUPDATE after the $t$-th Kraus map along the trajectory $(\mathbf{k}, \mathbf{r})$. The probability of selecting a trajectory $(\mathbf{k}, \mathbf{r}) \in \mathbb{P}$ is the product of the probabilities, at each step $t$, of selecting the dyadic channel $\mathcal{T}_{k_t}^{(t)}$ and of DYADICMAPUPDATE then selecting the Kraus map $\mathcal{K}_{k_t,r_t}^{(t)}$ from that channel, that is:

$$\begin{aligned}
\Pr(\mathbf{k}, \mathbf{r}) &= \prod_{t=1}^{T} \Pr(k_t) \Pr\left(r_t | \sigma_{(\mathbf{k},\mathbf{r})}^{(t-1)}\right) && \text{(F.2)}\\
&= \prod_{t=1}^{T} \frac{|\beta_{k_t}^{(t)}|}{B^{(t)}} \|\mathcal{K}_{k_t,r_t}^{(t)}(\sigma_{(\mathbf{k},\mathbf{r})}^{(t-1)})\|_1 && \text{(F.3)}\\
&= \frac{1}{B} \prod_{t=1}^{T} |\beta_{k_t}^{(t)}| \, \|\mathcal{K}_{k_t,r_t}^{(t)}(\sigma_{(\mathbf{k},\mathbf{r})}^{(t-1)})\|_1 && \text{(F.4)}
\end{aligned}$$

The set of all outcomes where the algorithm terminating with $\sigma^{(t)} = 0$ for some $t$ is the complement of the set of all trajectories in $\mathbb{P}$. So the probability of the simulator terminating with a zero outcome is $\Pr\left(\sigma^{(T)} = 0\right) = 1 - \sum_{(\mathbf{k},\mathbf{r}) \in \mathbb{P}} \Pr(\mathbf{k}, \mathbf{r})$. Then the expected value of the random variable $\widetilde{E}_j$ is:

$$\mathbb{E}(\widetilde{E}_j) = \Pr\Big(\sigma^{(T)} = 0\Big)\cdot 0 + \sum_{(\mathbf{k},\mathbf{r})\in\mathbb{P}} \Pr(\mathbf{k},\mathbf{r})\cdot \mathrm{Re}\Big\{Be^{i\theta_{\mathbf{k}}}\,\mathrm{Tr}\Big[E\sigma^{(T)}_{(\mathbf{k},\mathbf{r})}\Big]\Big\} \tag{F.5}$$

$$= \sum_{(\mathbf{k},\mathbf{r})\in\mathbb{P}} \left(\frac{1}{B}\prod_{t=1}^{T}|\beta^{(t)}_{k_t}|\,\|\mathcal{K}^{(t)}_{k_t,r_t}(\sigma^{(t-1)}_{(\mathbf{k},\mathbf{r})})\|_1\right)\mathrm{Re}\Big\{Be^{i\theta_{\mathbf{k}}}\,\mathrm{Tr}\Big[E\sigma^{(T)}_{(\mathbf{k},\mathbf{r})}\Big]\Big\}, \tag{F.6}$$

where $e^{i\theta_{\mathbf{k}}}$ is the phase accumulated in step 6 along the trajectory $(\mathbf{k},\mathbf{r})$, which depends only on the indices $k_t$ sampled from the quasiprobability distribution in step 5, and factorises as $e^{i\theta_{\mathbf{k}}} = \prod_{t=1}^{T}\frac{\beta^{(t)}_{k_t}}{|\beta^{(t)}_{k_t}|}$, so that:

$$\left(\prod_{t=1}^{T}|\beta^{(t)}_{k_t}|\right)e^{i\theta_{\mathbf{k}}} = \prod_{t=1}^{T}\beta^{(t)}_{k_t} \tag{F.7}$$

For non-zero-probability trajectories:

$$\sigma_{(\mathbf{k},\mathbf{r})^{(T)}} = \frac{\mathcal{K}^{(T)}_{k_T,r_T}\circ\ldots\circ\mathcal{K}^{(1)}_{k_1,r_1}(\sigma^{(0)})}{\prod_{t'=1}^{T}\|\mathcal{K}^{(t)}_{k_t,r_t}(\sigma^{(t-1)}_{(\mathbf{k},\mathbf{r})})\|_1} \tag{F.8}$$

The product of 1-norms in the denominator cancels with the same product in equation (F.6). Then since for zero-probability trajectories $(\mathbf{k},\mathbf{r})\notin\mathbb{P}$, it is also the case that $\sigma^{(T)}_{(\mathbf{k},\mathbf{r})} = \mathcal{K}^{(T)}_{k_T,r_T}\circ\ldots\circ\mathcal{K}^{(1)}_{k_1,r_1}(\sigma^{(0)}) = 0$, we can freely add them to the sum. So,

$$\mathbb{E}(\widetilde{E}_j) = \sum_{(\mathbf{k},\mathbf{r})}\left(\prod_{t=1}^{T}|\beta^{(t)}_{k_t}|\right)\mathrm{Re}\Big\{e^{i\theta}\,\mathrm{Tr}\Big[E\mathcal{K}^{(T)}_{k_T,r_T}\circ\ldots\circ\mathcal{K}^{(1)}_{k_1,r_1}(\sigma^{(0)})\Big]\Big\} \tag{F.9}$$

$$= \sum_{\mathbf{k}}\mathrm{Re}\left\{\left(\prod_{t=1}^{T}\beta^{(t)}_{k_t}\right)\sum_{\mathbf{r}}\mathrm{Tr}\Big[E\mathcal{K}^{(T)}_{k_T,r_T}\circ\ldots\circ\mathcal{K}^{(1)}_{k_1,r_1}(\sigma^{(0)})\Big]\right\} \tag{F.10}$$

$$= \sum_{\mathbf{k}}\mathrm{Re}\left\{\left(\prod_{t=1}^{T}\beta^{(t)}_{k_t}\right)\mathrm{Tr}\left[E\sum_{r_T}\mathcal{K}^{(T)}_{k_T,r_T}\circ\ldots\circ\sum_{r_t}\mathcal{K}^{(t)}_{k_t,r_t}\circ\ldots\circ\sum_{r_1}\mathcal{K}^{(1)}_{k_1,r_1}(\sigma^{(0)})\right]\right\}$$

$$= \sum_{\mathbf{k}}\mathrm{Re}\left\{\left(\prod_{t=1}^{T}\beta^{(t)}_{k_t}\right)\mathrm{Tr}\Big[E\mathcal{T}^{(T)}_{k}\circ\ldots\circ\mathcal{T}^{(t)}_{k}\circ\ldots\circ\mathcal{T}^{(1)}_{k}(\sigma^{(0)})\Big]\right\} \tag{F.11}$$

where in the second line we used equation (F.7) to recover the complex quasiprobabilities $\beta_{k_t}^{(t)}$. Reuniting these coefficients with the dyadic stabiliser channels $\mathcal{T}_k^{(t)}$ we obtain:

$$\mathbb{E}(\widetilde{E}_j) = \mathrm{Re}\left\{ \mathrm{Tr}\left[ E \sum_{k_T} \beta_{k_T}^{(T)} \mathcal{T}_k^{(T)} \circ \ldots \circ \sum_{k_t} \beta_{k_t}^{(t)} \mathcal{T}_k^{(t)} \circ \ldots \circ \sum_{k_1} \beta_{k_1}^{(1)} \mathcal{T}_k^{(1)} (\sigma^{(0)}) \right] \right\} \tag{F.12}$$

$$= \mathrm{Tr}\left[ E \mathcal{E}^{(T)} \circ \ldots \circ \mathcal{E}^{(t)} \circ \ldots \circ \mathcal{E}^{(1)} (\sigma^{(0)}) \right] \tag{F.13}$$

$$= \mathrm{Tr}\left[ E \mathcal{E}(\sigma^{(0)}) \right] = \langle E \rangle \,. \tag{F.14}$$

This shows that the random variable $\widetilde{E}_j$ is an unbiased estimator for the target mean value $\langle E \rangle$, and the usual arguments for quasiprobability apply [128]. We have $|\widetilde{E}_j| \leq B$, so we obtain the usual $\ell_1$-norm-dependence for the number of samples required to achieve fixed precision and success probability. Theorem 5.12 follows.

# Bibliography

[1] James R. Seddon and Earl T. Campbell. “Quantifying magic for multi-qubit operations”. In: *Proc. Royal Soc. A* 475 (2019), p. 20190251. DOI: 10.1098/rspa.2019.0251. arXiv: 1901.03322.

[2] James R. Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T. Campbell. “Quantifying quantum speedups: improved classical simulation from tighter magic monotones”. In: *Phys. Rev. X Quantum* 2 (2021), p. 010345. DOI: 10.1103/PRXQuantum.2.010345. arXiv: 2002.06181.

[3] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. “Simulation of quantum circuits by low-rank stabilizer decompositions”. In: *Quantum* 3 (2019), p. 181. DOI: 10.22331/q-2019-09-02-181. arXiv: 1808.00128.

[4] M.A. Nielsen and I. L. Chuang. *Quantum Computation and Communication*. 10th ed. Cambridge University Press, 2010.

[5] Edward Fredkin and Tommaso Toffoli. “Conservative logic”. In: *Int. J. Theor. Phys.* 21 (1982), pp. 219–253. DOI: 10.1007/BF01857727.

[6] Alonzo Church. “An unsolvable problem of elementary number theory”. In: *Am. J. Math.* 58 (1936), pp. 345–363.

[7] A. M. Turing. “On computable numbers, with an application to the entscheidungsproblem”. In: *Proc. Lond. Math. Soc.* s2-42 (1937), pp. 230–265. DOI: 10.1112/plms/s2-42.1.230.

[8] Juris Hartmanis and R. Stearns. “On the computational complexity of algorithms”. In: *Trans. Am. Math. Soc.* 117 (1965), pp. 285–285. DOI: 10.2307/1994208.

[9] Stephen A. Cook. “The complexity of theorem-proving procedures”. In: *Proceedings of the Third Annual ACM Symposium on Theory of Computing*. STOC. Shaker Heights, Ohio, USA: Association for Computing Machinery, 1971, pp. 151–158. DOI: 10.1145/800157.805047.

[10] I. M. Georgescu, S. Ashhab, and Franco Nori. “Quantum simulation”. In: *Rev. Mod. Phys.* 86 (2014), pp. 153–185. DOI: 10.1103/RevModPhys.86.153.

[11] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. “Quantum computational chemistry”. In: *Rev. Mod. Phys.* 92 (2020), p. 015003. DOI: 10.1103/RevModPhys.92.015003.

[12] Richard P. Feynman. “Simulating physics with computers”. In: *Int. J. Theor. Phys.* 21 (1982), pp. 467–488.

[13] Yuri Manin. *Computable and Uncomputable (in Russian), Sovetskoye Radio.* (1980).

[14] Robert Raussendorf and Hans J. Briegel. “A one-way quantum computer”. In: *Phys. Rev. Lett.* 86 (2001), pp. 5188–5191. DOI: 10.1103/PhysRevLett.86.5188.

[15] Robert Raussendorf, Daniel E. Browne, and Hans J. Briegel. “Measurement-based quantum computation on cluster states”. In: *Phys. Rev. A* 68 (2003), p. 022312. DOI: 10.1103/PhysRevA.68.022312.

[16] Olivier Pfister. “Continuous-variable quantum computing in the quantum optical frequency comb”. In: *J. Phys. B: Atom. Mol. Opt. Phys.* 53 (2020), p. 012001. DOI: 10.1088/1361-6455/ab526f.

[17] John Preskill. “Quantum computing in the NISQ era and beyond”. In: *Quantum* 2 (2018), p. 79. DOI: 10.22331/q-2018-08-06-79. arXiv: 1801.00862.

[18] Aram W. Harrow and Ashley Montanaro. “Quantum computational supremacy”. In: *Nature* 549 (2017), pp. 203–209. DOI: 10.1038/nature23458.

[19] Andrew M. Childs, Dmitri Maslov, Yunseong Nam, Neil J. Ross, and Yuan Su. “Toward the first quantum simulation with quantum speedup”. In: *Proc. Natl. Acad. Sci. Unit. States. Am.* 115 (2018), pp. 9456–9461. DOI: 10.1073/pnas.1801723115. arXiv: 1711.10980.

[20] Cupjin Huang, Michael Newman, and Mario Szegedy. *Explicit lower bounds on strong simulation of quantum circuits in terms of T-gate count*. Pre-print. (2019). arXiv: 1902.04764.

[21] Tomoyuki Morimae and Suguru Tamaki. “Fine-grained quantum computational supremacy”. In: *Quantum Information and Computation* 19 (2019), pp. 1089–1115. arXiv: 1901.01637.

[22] Ethan Bernstein and Umesh Vazirani. “Quantum complexity theory”. In: *SIAM J. Comput.* 26 (1997), pp. 1411–1473. DOI: 10.1137/S0097539796300921.

[23] Ashley Montanaro. “Quantum algorithms: an overview”. In: *npj Quant. Inf.* 2 (2016), p. 15023. DOI: 10.1038/npjqi.2015.23.

[24] Sergey Bravyi, David Gosset, and Robert König. “Quantum advantage with shallow circuits”. In: *Science* 362 (2018), pp. 308–311. DOI: 10.1126/science.aar3106. arXiv: 1704.00690.

[25] Daniel Gottesman. “The Heisenberg representation of quantum computers”. In: *22nd International Colloquium on Group Theoretical Methods in Physics.* (1998), pp. 32–43. arXiv: quant-ph/9807006.

[26] Leslie G. Valiant. “Quantum circuits that can be simulated classically in polynomial time”. In: *SIAM J. Comput.* 31 (2002), pp. 1229–1254. DOI: 10.1137/S0097539700377025.

[27] Barbara M. Terhal and David P. DiVincenzo. “Classical simulation of noninteracting-fermion quantum circuits”. In: *Phys. Rev. A* 65 (2002), p. 032325. DOI: 10.1103/PhysRevA.65.032325. arXiv: quant-ph/0108010.

[28] Richard Jozsa and Akimasa Miyake. “Matchgates and classical simulation of quantum circuits”. In: *Proc. Royal Soc. A* 464 (2008), pp. 3089–3106. DOI: 10.1098/rspa.2008.0189. arXiv: 0804.4050.

[29] Y.-Y. Shi, L.-M. Duan, and G. Vidal. “Classical simulation of quantum many-body systems with a tree tensor network”. In: *Phys. Rev. A* 74 (2006), p. 022320. DOI: 10.1103/PhysRevA.74.022320. arXiv: quant-ph/0511070.

[30] Daniel Gottesman. “Stabilizer Codes and Quantum Error Correction”. PhD thesis. California Institute of Technology, (1997). arXiv: quant-ph/9705052.

[31] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". In: *Phys. Rev. A* 70 (2004), p. 052328. DOI: 10.1103/PhysRevA.70.052328. arXiv: quant-ph/0406196.

[32] Bryan Eastin and Emanuel Knill. "Restrictions on transversal encoded quantum gate sets". In: *Phys. Rev. Lett.* 102 (2009), p. 110502. DOI: 10.1103/PhysRevLett.102.110502. arXiv: 0811.4262.

[33] Earl T. Campbell, Barbara M. Terhal, and Christophe Vuillot. "Roads towards fault-tolerant universal quantum computation". In: *Nature* 549 (2017), pp. 172–179. DOI: 10.1038/nature23460. arXiv: 1612.07330.

[34] Emanuel Knill. "Quantum computing with realistically noisy devices". In: *Nature* 434 (2005), pp. 39–44. DOI: 10.1038/nature03350. arXiv: quant-ph/0410199.

[35] Sergey Bravyi and Alexei Kitaev. "Universal quantum computation with ideal clifford gates and noisy ancillas". In: *Phys. Rev. A* 71 (2005), p. 022316. DOI: 10.1103/PhysRevA.71.022316. arXiv: quant-ph/0403025.

[36] Ben W. Reichardt. "Quantum Universality from Magic States Distillation Applied to CSS Codes". In: *Quantum Inf. Process.* 4 (2005), pp. 251–264. DOI: 10.1007/s11128-005-7654-8. arXiv: quant-ph/0411036.

[37] Sergey Bravyi and Jeongwan Haah. "Magic-state distillation with low overhead". In: *Phys. Rev. A* 86 (2012), p. 052329. DOI: 10.1103/PhysRevA.86.052329. arXiv: 1209.2426.

[38] Cody Jones. "Multilevel distillation of magic states for quantum computing". In: *Phys. Rev. A* 87 (2013), p. 042305. DOI: 10.1103/PhysRevA.87.042305. arXiv: 1210.3388.

[39] Colin J. Trout and Kenneth R. Brown. "Magic state distillation and gate compilation in quantum algorithms for quantum chemistry". In: *Int. J. Quant. Chem.* 115 (2015), pp. 1296–1304. DOI: 10.1002/qua.24856. arXiv: 1501.01298.

[40] Jeongwan Haah, Matthew B. Hastings, D. Poulin, and D. Wecker. "Magic State Distillation with Low Space Overhead and Optimal Asymptotic Input Count".

In: *Quantum* 1 (2017), p. 31. DOI: 10.22331/q-2017-10-03-31. arXiv: 1703.07847.

[41] Matthew B. Hastings and Jeongwan Haah. "Distillation with Sublogarithmic Overhead". In: *Physical Review Letters* 120 (2018), p. 050504. DOI: 10.1103/PhysRevLett.120.050504. arXiv: 1709.03543.

[42] Anirudh Krishna and Jean-Pierre Tillich. "Towards low overhead magic state distillation". In: *Phys. Rev. Lett.* 123 (2019), p. 070507. DOI: 10.1103/PhysRevLett.123.070507. arXiv: 1811.08461.

[43] Earl T. Campbell and Mark Howard. "Magic state parity-checker with pre-distilled components". In: *Quantum* 2 (2018), p. 56. DOI: 10.22331/q-2018-03-14-56. arXiv: 1709.02214.

[44] Xin Wang, Mark M. Wilde, and Yuan Su. "Efficiently computable bounds for magic state distillation". In: *Phys. Rev. Lett.* 124 (2020), p. 090505. DOI: 10.1103/PhysRevLett.124.090505. arXiv: 1812.10145.

[45] 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". In: *Phys. Rev. A* 95 (2017), p. 062337. DOI: 10.1103/PhysRevA.95.062337. arXiv: 1703.00111.

[46] Michael Wolf. *Quantum Channels and Operations - Guided Tour (lecture notes).* (2012). URL: https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf (visited on 12/09/2021).

[47] J.R. Johansson, P.D. Nation, and Franco Nori. "Qutip 2: a python framework for the dynamics of open quantum systems". In: *Comp. Phys. Comm.* 184 (2013), pp. 1234–1240. DOI: 10.1016/j.cpc.2012.11.019. arXiv: 1211.6518.

[48] Joschka Roffe. "Quantum error correction: an introductory guide". In: *Contemporary Physics* 60 (2019), pp. 226–245. DOI: 10.1080/00107514.2019.1667078. arXiv: 1907.11157.

[49] Suguru Endo, Simon C. Benjamin, and Ying Li. "Practical quantum error mitigation for near-future applications". In: *Phys. Rev. X* 8 (2018), p. 031027. DOI: 10.1103/PhysRevX.8.031027. arXiv: 1712.09271.

[50] W. Forrest Stinespring. “Positive functions on C* -algebras”. In: *Proc. Am. Math. Soc.* 6 (1955), p. 211. DOI: 10.2307/2032342.

[51] Man-Duen Choi. “Completely positive linear maps on complex matrices”. In: *Lin. Algebra Appl.* 10 (1975), pp. 285–290. DOI: 10.1016/0024-3795(75)90075-0.

[52] Andrzej Jamiolkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. In: *Rep. Math. Phys.* 3 (1972), pp. 275–278. DOI: 10.1016/0034-4877(72)90011-0.

[53] Min Jiang, Shunlong Luo, and Shuangshuang Fu. “Channel-state duality”. In: *Phys. Rev. A* 87 (2013), p. 022310. DOI: 10.1103/PhysRevA.87.022310.

[54] Jeroen Dehaene and Bart De Moor. “Clifford group, stabilizer states, and linear and quadratic operations over GF(2)”. In: *Phys. Rev. A* 68 (2003), p. 042318. DOI: 10.1103/PhysRevA.68.042318. arXiv: quant-ph/0304125.

[55] David Gross and Maarten Van den Nest. “The LU-LC conjecture, diagonal local operations and quadratic forms over GF(2)”. In: *Quant. Inf. Comput.* 8 (2008), pp. 263–281. arXiv: 0707.4000.

[56] Hakop Pashayan, Stephen D. Bartlett, and David Gross. “From estimation of quantum probabilities to simulation of quantum circuits”. In: *Quantum* 4 (2020), p. 223. DOI: 10.22331/q-2020-01-13-223.

[57] David Fattal, Toby S. Cubitt, Yoshihisa Yamamoto, Sergey Bravyi, and Isaac L. Chuang. *Entanglement in the stabilizer formalism.* Pre-print. (2004). arXiv: quant-ph/0406168.

[58] Hector J. Garcia, Igor L. Markov, and Andrew W. Cross. *Efficient Inner-product Algorithm for Stabilizer States*. Pre-print. (2013). arXiv: 1210.6646.

[59] Sergey Bravyi, Graeme Smith, and John A. Smolin. “Trading classical and quantum computational resources”. In: *Phys. Rev. X* 6 (2016), p. 021043. DOI: 10.1103/PhysRevX.6.021043. arXiv: 1506.01396.

[60] Sergey Bravyi and David Gosset. “Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates”. In: *Phys. Rev. Lett.* 116 (2016), p. 250501. DOI: 10.1103/PhysRevLett.116.250501. arXiv: 1601.07601.

[61] Maarten van Den Nest. “Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond”. In: *Quant. Inf. Comput.* 10 (2010), pp. 0258–0271. arXiv: 0811.0898.

[62] D. Maslov and M. Roetteler. “Shorter stabilizer circuits via bruhat decomposition and quantum circuit transformations”. In: *IEEE Transactions on Information Theory* 64 (2018), pp. 4729–4738. DOI: 10.1109/TIT.2018.2825602. arXiv: 1705.09176.

[63] Arne Heimendahl, Markus Heinrich, and David Gross. *The axiomatic and the operational approaches to resource theories of magic do not coincide*. Pre-print. (2020). arXiv: 2011.11651.

[64] Carsten Damm. “Problems complete for $\oplus$L”. In: *Aspects and Prospects of Theoretical Computer Science*. Ed. by Jürgen Dassow and Jozef Kelemen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 130–137.

[65] David Deutsch. “Quantum computational networks”. In: *Proc. Royal Soc. A* 425 (1989), pp. 73–90. DOI: 10.1098/rspa.1989.0099.

[66] David P. DiVincenzo. “Two-bit gates are universal for quantum computation”. In: *Phys. Rev. A* 51 (1995), pp. 1015–1022. DOI: 10.1103/PhysRevA.51.1015. arXiv: cond-mat/9407022.

[67] Dorit Aharonov and Michael Ben-Or. “Fault-tolerant quantum computation with constant error rate”. In: *SIAM J. Comput.* 38 (2008), pp. 1207–1282. DOI: 10.1137/S0097539799359385.

[68] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. “Asymptotically optimal approximation of single qubit unitaries by Clifford and T circuits using a constant number of ancillary qubits”. In: *Phys. Rev. Lett.* 110, 190502 (2013), p. 190502. DOI: 10.1103/PhysRevLett.110.190502. arXiv: 1212.0822 [quant-ph].

[69] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. “Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and T gates”. In: *Quant. Inf. Comput.* 13 (2013), pp. 607–630. arXiv: 1206.5236.

[70] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. "Practical approximation of single-qubit unitaries by single-qubit quantum clifford and t circuits". In: *IEEE Trans. Comput.* 65 (2016), pp. 161–172. DOI: 10.1109/TC.2015.2409842. arXiv: 1212.6964.

[71] Neil J. Ross and Peter Selinger. "Optimal ancilla-free Clifford+T approximation of z-rotations". In: *Quant. Inf. Comput.* 16 (2016), pp. 901–953. arXiv: 1403.2975.

[72] P. Oscar Boykin, Tal Mor, Matthew Pulver, Vwani Roychowdhury, and Farrokh Vatan. *On Universal and Fault-Tolerant Quantum Computing*. Pre-print. (1999). arXiv: quant-ph/9906054.

[73] Emanuel Knill. *Approximation by quantum circuits*. Los Alamos National Lab report LAUR-95-2225. (1995). arXiv: quant-ph/9508006.

[74] A. Yu Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52 (1997), pp. 1191–1249. DOI: 10.1070/rm1997v052n06abeh002155.

[75] Christopher M. Dawson and Michael A. Nielsen. "The solovay-kitaev algorithm". In: *Quant. Inf. Comput.* 6 (2006), pp. 81–95. arXiv: quant-ph/0505030.

[76] Aram W. Harrow, Benjamin Recht, and Isaac L. Chuang. "Efficient discrete approximations of quantum gates". In: *J. Math. Phys.* 43 (2002), pp. 4445–4451. DOI: 10.1063/1.1495899. arXiv: quant-ph/0111031.

[77] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. "Hybrid quantum-classical algorithms and quantum error mitigation". In: *J. Phys. Soc. Jpn.* 90 (2021), p. 032001. DOI: 10.7566/JPSJ.90.032001. arXiv: 2011.01382.

[78] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'Brien. "A variational eigenvalue solver on a photonic quantum processor". In: *Nat. Comm.* 5 (2014), p. 4213. DOI: 10.1038/ncomms5213. arXiv: 1304.3061.

[79] Ying Li and Simon C. Benjamin. “Efficient variational quantum simulator incorporating active error minimization”. In: *Phys. Rev. X* 7 (2017), p. 021050. DOI: 10.1103/PhysRevX.7.021050. arXiv: 1611.09301.

[80] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error mitigation for short-depth quantum circuits”. In: *Phys. Rev. Lett.* 119 (2017), p. 180509. DOI: 10.1103/PhysRevLett.119.180509. arXiv: 1612.02058.

[81] Filip B. Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography”. en-GB. In: *Quantum* 4 (2020), p. 257. DOI: 10.22331/q-2020-04-24-257. arXiv: 1907.08518.

[82] Ryuji Takagi. “Optimal resource cost for error mitigation”. In: *Phys. Rev. Research* 3 (2021), p. 033178. DOI: 10.1103/PhysRevResearch.3.033178. arXiv: 2006.12509.

[83] Matthew Otten and Stephen K. Gray. “Accounting for errors in quantum algorithms via individual error reduction”. en. In: *npj Quant. Inf.* 5 (2019), pp. 1–6. DOI: 10.1038/s41534-019-0125-3. arXiv: 1804.06969.

[84] Peter W. Shor. “Scheme for reducing decoherence in quantum computer memory”. In: *Phys. Rev. A* 52 (1995), R2493–R2496. DOI: 10.1103/PhysRevA.52.R2493.

[85] Peter W. Shor. “Fault-tolerant quantum computation”. In: *Annual Symposium on Foundations of Computer Science - Proceedings*. IEEE, 1996, pp. 56–65. arXiv: quant-ph/9605011.

[86] Emanuel Knill. *Non-binary Unitary Error Bases and Quantum Codes*. Los Alamos National Lab report LAUR-96-2717. 1996. arXiv: quant-ph/9608048.

[87] Emanuel Knill, Raymond Laflamme, and Lorenza Viola. “Theory of quantum error correction for general noise”. In: *Phys. Rev. Lett.* 84 (2000), pp. 2525–2528. DOI: 10.1103/PhysRevLett.84.2525. arXiv: quant-ph/9908066.

[88] Daniel Gottesman. “Class of quantum error-correcting codes saturating the quantum hamming bound”. In: *Phys. Rev. A* 54 (1996), pp. 1862–1868. DOI: 10.1103/PhysRevA.54.1862. arXiv: quant-ph/9604038.

[89] Daniel Gottesman and Isaac L. Chuang. “Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations”. In: *Nature* 402 (1999), pp. 390–393. DOI: 10.1038/46503. arXiv: quant-ph/9908010.

[90] Xinlan Zhou, Debbie W. Leung, and Isaac L. Chuang. “Methodology for quantum logic gate construction”. In: *Phys. Rev. A* 62 (2000), p. 052316. DOI: 10.1103/PhysRevA.62.052316. arXiv: quant-ph/0002039.

[91] Joe O’Gorman and Earl T. Campbell. “Quantum computation with realistic magic-state factories”. In: *Phys. Rev. A* 95 (2017), p. 032338. DOI: 10.1103/PhysRevA.95.032338. arXiv: 1605.07197.

[92] Craig Gidney and Austin G. Fowler. “Efficient magic state factories with a catalyzed $|CCZ\rangle$ to $2|T\rangle$ transformation”. In: *Quantum* 3 (2019), p. 135. DOI: 10.22331/q-2019-04-30-135. arXiv: 1812.01238.

[93] Michal Horodecki and Jonathan Oppenheim. “(quantumness in the context of) resource theories”. In: *Int. J. Mod. Phys. B B* 27 (2013), p. 1345019. DOI: 10.1142/S0217979213450197. arXiv: 1209.2162.

[94] Fernando G. S. L. Brandao and Gilad Gour. “Reversible framework for quantum resource theories”. In: *Phys. Rev. Lett.* 115 (2015), p. 070503. DOI: 10.1103/PhysRevLett.115.070503. arXiv: 1502.03149.

[95] Bartosz Regula. “Convex geometry of quantum resource quantification”. In: *J. Phys. A: Math. Theor.* 51 (2017), p. 045303. DOI: 10.1088/1751-8121/aa9100. arXiv: 1707.06298.

[96] Martin B. Plenio and Shashank Virmani. “An introduction to entanglement measures”. In: *Quant. Inf. Comput.* 7 (2007), pp. 1–51. arXiv: quant-ph/0504163.

[97] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. “Quantum entanglement”. In: *Rev. Mod. Phys.* 81 (2009), pp. 865–942. DOI: 10.1103/RevModPhys.81.865. arXiv: quant-ph/0702225.

[98] Carmine Napoli, Thomas R. Bromley, Marco Cianciaruso, Marco Piani, Nathaniel Johnston, and Gerardo Adesso. “Robustness of coherence: an operational and observable measure of quantum coherence”. In: *Phys. Rev. Lett.*

116 (2016), p. 150502. DOI: 10.1103/PhysRevLett.116.150502. arXiv: 1601.03781.

[99] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik. “Quantifying contextuality”. In: *Phys. Rev. Lett.* 112 (2014), p. 120401. DOI: 10.1103/PhysRevLett.112.120401. arXiv: 1209.3745.

[100] Fernando G. S. L. Brandão, Michał Horodecki, Jonathan Oppenheim, Joseph M. Renes, and Robert W. Spekkens. “Resource theory of quantum states out of thermal equilibrium”. In: *Phys. Rev. Lett.* 111 (2013), p. 250404. DOI: 10.1103/PhysRevLett.111.250404. arXiv: 1111.3882.

[101] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson. “Negative quasi-probability as a resource for quantum computation”. In: *New J. Phys.* 14 (2012), p. 113011. DOI: 10.1088/1367-2630/14/11/113011. arXiv: 1201.1256.

[102] Victor Veitch, S. A. Hamed Mousavian, Daniel Gottesman, and Joseph Emerson. “The resource theory of stabilizer quantum computation”. In: *New J. Phys.* 16 (2014), p. 013009. DOI: 10.1088/1367-2630/16/1/013009. arXiv: 1307.7171.

[103] Mark Howard and Earl Campbell. “Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing”. In: *Phys. Rev. Lett.* 118 (2017), p. 090501. DOI: 10.1103/PhysRevLett.118.090501. arXiv: 1609.07488.

[104] Mehdi Ahmadi, Hoan Bui Dang, Gilad Gour, and Barry C. Sanders. “Quantification and manipulation of magic states”. In: *Phys. Rev. A* 97 (2018), p. 062332. DOI: 10.1103/PhysRevA.97.062332. arXiv: 1706.03828.

[105] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Kliuchnikov. “Lower bounds on the non-Clifford resources for quantum computations”. In: *Quantum Science and Technology* 5 (2020), p. 035009. DOI: 10.1088/2058-9565/ab8963. arXiv: 1904.01124.

[106] Bartosz Regula, Kaifeng Bu, Ryuji Takagi, and Zi-Wen Liu. “Benchmarking one-shot distillation in general quantum resource theories”. In: *Phys. Rev. A* 101 (2020), p. 062315. DOI: 10.1103/PhysRevA.101.062315. arXiv: 1909.11677.

[107] Guifré Vidal. “Entanglement monotones”. In: *J. Mod. Opt.* 47 (2000), pp. 355–376. DOI: 10.1080/09500340008244048. arXiv: quant-ph/9807077.

[108] V. Vedral. “The role of relative entropy in quantum information theory”. In: *Rev. Mod. Phys.* 74 (2002), pp. 197–234. DOI: 10.1103/RevModPhys.74.197. arXiv: quant-ph/0102094.

[109] Paweł Horodecki, Michał Horodecki, and Ryszard Horodecki. “Bound entanglement can be activated”. In: *Phys. Rev. Lett.* 82 (1999), pp. 1056–1059. DOI: 10.1103/PhysRevLett.82.1056. arXiv: quant-ph/9806058.

[110] Robert Raussendorf, Juani Bermejo-Vega, Emily Tyhurst, Cihan Okay, and Michael Zurel. “Phase-space-simulation method for quantum computation with magic states on qubits”. In: *Phys. Rev. A* 101 (2020), p. 012350. DOI: 10.1103/PhysRevA.101.012350. arXiv: 1905.05374.

[111] Martin B. Plenio. “Logarithmic negativity: a full entanglement monotone that is not convex”. In: *Phys. Rev. Lett.* 95 (2005), p. 090503. DOI: 10.1103/PhysRevLett.95.090503. arXiv: quant-ph/0505071.

[112] Xin Wang, Mark M Wilde, and Yuan Su. “Quantifying the magic of quantum channels”. In: *New J. Phys.* 21 (2019), p. 103002. DOI: 10.1088/1367-2630/ab451d. arXiv: 1903.04483.

[113] Eugene Wigner. “On the quantum correction for thermodynamic equilibrium”. In: *Phys. Rev.* 40 (1932), pp. 749–759. DOI: 10.1103/PhysRev.40.749.

[114] Anatole Kenfack and Karol Zyczkowski. “Negativity of the Wigner function as an indicator of non-classicality”. In: *Journal of Optics B: Quantum and Semiclassical Optics* 6 (2004), pp. 396–404. DOI: 10.1088/1464-4266/6/10/003. arXiv: quant-ph/0406015.

[115] Ryuji Takagi and Quntao Zhuang. “Convex resource theory of non-gaussianity”. In: *Phys. Rev. A* 97 (2018), p. 062337. DOI: 10.1103/PhysRevA.97.062337. arXiv: 1804.04669.

[116] R.L. Hudson. “When is the Wigner quasi-probability density non-negative?” In: *Rep. Math. Phys.* 6 (1974), pp. 249–252. DOI: https://doi.org/10.1016/0034-4877(74)90007-X.

[117] Francisco Soto and Pierre Claverie. "When is the Wigner function of multidimensional systems nonnegative?" In: *Journal of Mathematical Physics* 24 (1983), pp. 97–100. DOI: 10.1063/1.525607.

[118] D. Gross. "Hudson's theorem for finite-dimensional quantum systems". In: *Journal of Mathematical Physics* 47 (2006), p. 122107. DOI: 10.1063/1.2393152. arXiv: quant-ph/0602001.

[119] Kathleen S. Gibbons, Matthew J. Hoffman, and William K. Wootters. "Discrete phase space based on finite fields". In: *Phys. Rev. A* 70 (2004), p. 062101. DOI: 10.1103/PhysRevA.70.062101. arXiv: quant-ph/0401155.

[120] Ernesto F. Galvão. "Discrete Wigner functions and quantum computational speedup". In: *Phys. Rev. A* 71 (2005), p. 042302. DOI: 10.1103/PhysRevA.71.042302. arXiv: quant-ph/0405070.

[121] Cecilia Cormick, Ernesto F. Galvão, Daniel Gottesman, Juan Pablo Paz, and Arthur O. Pittenger. "Classicality in discrete Wigner functions". In: *Phys. Rev. A* 73 (2006), p. 012301. DOI: 10.1103/PhysRevA.73.012301. arXiv: quant-ph/0506222.

[122] Earl T. Campbell. "Catalysis and activation of magic states in fault-tolerant architectures". In: *Phys. Rev. A* 83 (2011), p. 032317. DOI: 10.1103/PhysRevA.83.032317. arXiv: 1010.0104.

[123] Andrea Mari and Jens Eisert. "Positive Wigner functions render classical simulation of quantum computation efficient". In: *Phys. Rev. Lett.* 109 (2012), p. 230503. DOI: 10.1103/PhysRevLett.109.230503. arXiv: 1208.3660.

[124] Hakop Pashayan, Joel J. Wallman, and Stephen D. Bartlett. "Estimating outcome probabilities of quantum circuits using quasiprobabilities". In: *Phys. Rev. Lett.* 115 (2015), p. 070501. DOI: 10.1103/PhysRevLett.115.070501. arXiv: 1503.07525.

[125] Christopher Ferrie and Joseph Emerson. "Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations". In: *Journal of Physics A: Mathematical and Theoretical* 41 (2008), p. 352001. DOI: 10.1088/1751-8113/41/35/352001. arXiv: 0711.2658.

[126] Christopher Ferrie and Joseph Emerson. “Framed Hilbert space: hanging the quasi-probability pictures of quantum theory”. In: *New J. Phys.* 11 (2009), p. 063040. DOI: 10.1088/1367-2630/11/6/063040. arXiv: 0903.4843.

[127] Christopher Ferrie. “Quasi-probability representations of quantum theory with applications to quantum information science”. In: *Rep. Progr. Phys.* 74 (2011), p. 116001. DOI: 10.1088/0034-4885/74/11/116001. arXiv: 1010.2701.

[128] Wassily Hoeffding. “Probability inequalities for sums of bounded random variables”. In: *J. Am. Stat. Assoc.* 58 (1963), pp. 13–30.

[129] Nicolas Delfosse, Philippe Allard Guerin, Jacob Bian, and Robert Raussendorf. “Wigner function negativity and contextuality in quantum computation on rebits”. In: *Phys. Rev. X* 5 (2015), p. 021003. DOI: 10.1103/PhysRevX.5.021003. arXiv: 1409.5170.

[130] Robert Raussendorf, Dan E. Browne, Nicolas Delfosse, Cihan Okay, and Juan Bermejo-Vega. “Contextuality and Wigner-function negativity in qubit quantum computation”. In: *Phys. Rev. A* 95 (2017), p. 052334. DOI: 10.1103/PhysRevA.95.052334. arXiv: 1511.08506.

[131] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. 1st ed. Cambridge University Press, (2004).

[132] Michael Grant and Stephen Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1.* (2014). URL: http://cvxr.com/cvx (visited on 12/09/2021).

[133] Markus Heinrich and David Gross. “Robustness of magic and symmetries of the stabiliser polytope”. In: *Quantum* 3 (2019), p. 132. DOI: 10.22331/q-2019-04-08-132. arXiv: 1807.10296.

[134] Shigeo Hakkaku and Keisuke Fujii. “Comparative study of sampling-based simulation costs of noisy quantum circuits”. In: *Phys. Rev. Applied* 15 (2021), p. 064027. DOI: 10.1103/PhysRevApplied.15.064027. arXiv: 2011.06233.

[135] Earl Campbell. private communication. 2018.

[136] Simon Kochen and E. P. Specker. “The problem of hidden variables in quantum mechanics”. In: *J. Math. Mech.* 17 (1967), pp. 59–87.

[137] Asher Peres. "Incompatible results of quantum measurements". In: *Phys. Lett. A* 151 (1990), pp. 107–108. DOI: 10.1016/0375-9601(90)90172-K.

[138] N. David Mermin. "Hidden variables and the two theorems of John Bell". In: *Rev. Mod. Phys.* 65 (1993), pp. 803–815. DOI: 10.1103/RevModPhys.65.803. arXiv: 1802.10119.

[139] Robert W. Spekkens. "Contextuality for preparations, transformations, and unsharp measurements". In: *Phys. Rev. A* 71 (2005), p. 052108. DOI: 10.1103/PhysRevA.71.052108. arXiv: quant-ph/0406166.

[140] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson. "Contextuality supplies the 'magic' for quantum computation". In: *Nature* 510 (2014), pp. 351–355. DOI: 10.1038/nature13460. arXiv: 1401.4174.

[141] Nicolas Delfosse, Cihan Okay, Juan Bermejo-Vega, Dan E. Browne, and Robert Raussendorf. "Equivalence between contextuality and negativity of the wigner function for qudits". In: *New J. Phys.* 19 (2017), p. 123024. DOI: 10.1088/1367-2630/aa8fe3. arXiv: 1610.07093.

[142] Patrick Rall, Daniel Liang, Jeremy Cook, and William Kretschmer. "Simulation of qubit quantum circuits via Pauli propagation". In: *Phys. Rev. A* 99 (2019), p. 062337. DOI: 10.1103/PhysRevA.99.062337. arXiv: 1901.09070.

[143] Hector J. Garcia, Igor L. Markov, and Andrew W. Cross. "On the geometry of stabilizer states". In: *Quant. Inf. Comput.* 14 (2014), pp. 683–720. arXiv: 1711.07848.

[144] Lucas Kocia. *Improved strong simulation of universal quantum circuits*. Preprint. (2021). arXiv: 2012.11739.

[145] Hammam Qassim, Hakop Pashayan, and David Gosset. *Improved upper bounds on the stabilizer rank of magic states*. Pre-print. (2021). arXiv: 2106.07740.

[146] Shir Peleg, Amir Shpilka, and Ben Lee Volk. *Lower Bounds on Stabilizer Rank*. Pre-print. (2021). arXiv: 2106.03214.

[147] Arne Heimendahl, Felipe Montealegre-Mora, Frank Vallentin, and David Gross. "Stabilizer extent is not multiplicative". In: *Quantum* 5 (2021), p. 400. DOI: 10.22331/q-2021-02-24-400. arXiv: 2007.04363.

[148] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". In: *Phys. Rev. A* 80 (2009), p. 012304. DOI: 10.1103/PhysRevA.80.012304. arXiv: quant-ph/0606161.

[149] Richard Kueng and David Gross. *Qubit stabilizer states are complex projective 3-designs*. Pre-print. (2015). arXiv: 1510.02767.

[150] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theor. Comput. Sci.* 43 (1986), pp. 169–188. DOI: 10.1016/0304-3975(86)90174-X.

[151] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. "Mixed-state entanglement and quantum error correction". In: *Phys. Rev. A* 54 (1996), pp. 3824–3851. DOI: 10.1103/PhysRevA.54.3824. arXiv: quant-ph/9604024.

[152] Jens Eisert, Fernando G. S. L. Brandão, and Koenraad M. R. Audenaert. "Quantitative entanglement witnesses". In: *New J. Phys.* 9 (2007), p. 46. DOI: 10.1088/1367-2630/9/3/046. arXiv: quant-ph/0607167.

[153] Oliver Rudolph. "A new class of entanglement measures". In: *J. Math. Phys.* 42 (2001), pp. 5306–5314. DOI: 10.1063/1.1398062. arXiv: math-ph/0005011.

[154] Oliver Rudolph. "Further results on the cross norm criterion for separability". In: *Quant. Inf. Proc.* 4 (2005), pp. 219–239. DOI: 10.1007/s11128-005-5664-1. arXiv: quant-ph/0202121.

[155] Tillmann Baumgratz, Marcus Cramer, and Martin B. Plenio. "Quantifying coherence". In: *Phys. Rev. Lett.* 113 (2014), p. 140401. DOI: 10.1103/PhysRevLett.113.140401. arXiv: 1311.0275.

[156] Guifré Vidal and Rolf Tarrach. "Robustness of entanglement". In: *Phys. Rev. A* 59 (1999), pp. 141–155. DOI: 10.1103/PhysRevA.59.141. arXiv: quant-ph/9806094.

[157] Michael Steiner. "Generalized robustness of entanglement". In: *Phys. Rev. A* 67 (2003), p. 054305. DOI: 10.1103/PhysRevA.67.054305. arXiv: quant-ph/0304009.

[158] Marc Hein, Jens Eisert, and Hans J. Briegel. “Multiparty entanglement in graph states”. In: *Phys. Rev. A* 69 (2004), p. 062311. DOI: 10.1103/PhysRevA.69.062311. arXiv: quant-ph/0307130.

[159] Gaurav Saxena and Gilad Gour. *Quantifying dynamical magic with completely stabilizer preserving operations as free*. Pre-print. (2022). arXiv: 2202.07867.

[160] Earl T. Campbell. “Optimal entangling capacity of dynamical processes”. In: *Phys. Rev. A* 82 (2010), p. 042314. DOI: 10.1103/PhysRevA.82.042314. arXiv: 1007.1445.

[161] Dan Stahlke. “Quantum interference as a resource for quantum speedup”. In: *Phys. Rev. A* 90 (2014), p. 022302. DOI: 10.1103/PhysRevA.90.022302. arXiv: 1305.2186.

[162] Joel J Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. In: *Phys. Rev. A* 94 (2016), p. 052325. DOI: 10.1103/PhysRevA.94.052325. arXiv: 1512.01098.

[163] Earl Campbell. “Shorter gate sequences for quantum computing by mixing unitaries”. In: *Phys. Rev. A* 95 (2017), p. 042306. DOI: 10.1103/PhysRevA.95.042306. arXiv: 1612.02689.

[164] Matthew B. Hastings. “Turning gate synthesis errors into incoherent errors”. In: *Quant. Inf. Comput.* 17 (2017), pp. 488–494. arXiv: 1612.01011.

[165] Earl Campbell. “Random compiler for fast Hamiltonian simulation”. In: *Phys. Rev. Lett.* 123 (2019), p. 070503. DOI: 10.1103/PhysRevLett.123.070503. arXiv: 1811.08017.

[166] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P.D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik. “Quantum chemistry in the age of quantum computing”. In: *Chem. Rev.* 119 (2019), pp. 10856–10915. DOI: 10.1021/acs.chemrev.8b00803. arXiv: 1812.09976.

[167] Pascual Jordan and Eugene Wigner. “Über das paulische äquivalenzverbot”. In: *Z. Phy* 47 (1928), pp. 631–651.

[168] Rolando Somma, Gerardo Ortiz, James E. Gubernatis, Emanuel Knill, and Raymond Laflamme. “Simulating physical phenomena by quantum networks”. In: *Phys. Rev. A* 65 (2002), p. 042323. DOI: 10.1103/PhysRevA.65.042323. arXiv: quant-ph/0108146.

[169] Sergey B. Bravyi and Alexei Yu. Kitaev. “Fermionic quantum computation”. In: *Ann. Phys.* 298 (2002), pp. 210–226. DOI: 10.1006/aphy.2002.6254. arXiv: quant-ph/0003137.

[170] F. Verstraete and J. I. Cirac. “Mapping local Hamiltonians of fermions to local Hamiltonians of spins”. In: *J. Stat. Mech. Theor. Exp.* 2005 (2005), P09012. DOI: 10.1088/1742-5468/2005/09/p09012. arXiv: cond-mat/0508353.

[171] Kanav Setia and James D. Whitfield. “Bravyi-Kitaev Superfast simulation of electronic structure on a quantum computer”. In: *J. Chem. Phys.* 148 (2018), p. 164104. DOI: 10.1063/1.5019371. arXiv: 1712.00446.

[172] Charles Derby, Joel Klassen, Johannes Bausch, and Toby Cubitt. “Compact fermion to qubit mappings”. In: *Phys. Rev. B* 104 (2021), p. 035118. DOI: 10.1103/PhysRevB.104.035118.

[173] Hale F Trotter. “On the product of semi-groups of operators”. In: *Proc. Am. Math. Soc.* 10 (1959), pp. 545–551. DOI: 10.2307/2033649.

[174] Masuo Suzuki. “General theory of fractal path integrals with applications to many body theories and statistical physics”. In: *J. Math. Phys.* 32 (1991), pp. 400–407. DOI: 10.1063/1.529425.

[175] Nathan Wiebe, Dominic Berry, Peter Høyer, and Barry C Sanders. “Higher order decompositions of ordered operator exponentials”. In: *J. Phys. Math. Theor.* 43 (2010), p. 065203. DOI: 10.1088/1751-8113/43/6/065203. arXiv: 0812.0562.

[176] Dominic W Berry, Graeme Ahokas, Richard Cleve, and Barry C Sanders. “Efficient quantum algorithms for simulating sparse Hamiltonians”. In: *Commun. Math. Phys.* 270 (2007), pp. 359–371. DOI: 10.1007/s00220-006-0150-x. arXiv: quant-ph/0508139.

[177] Nathan Wiebe, Dominic W Berry, Peter Høyer, and Barry C Sanders. "Simulating quantum dynamics on a quantum computer". In: *J. Phys. Math. Theor.* 44 (2011), p. 445308. DOI: 10.1088/1751-8113/44/44/445308. arXiv: 1011.3489.

[178] Andrew M. Childs, Aaron Ostrander, and Yuan Su. "Faster quantum simulation by randomization". In: *Quantum* 3 (2019), p. 182. DOI: 10.22331/q-2019-09-02-182. arXiv: 1805.08385.

[179] Jonathan Romero, Ryan Babbush, Jarrod R. McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik. "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz". In: *Quantum Science and Technology* 4 (2018), p. 014008. DOI: 10.1088/2058-9565/aad3e4. arXiv: 1701.02691.

[180] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 549 (2017), pp. 242–246. DOI: 10.1038/nature23879. arXiv: 1704.05018.

[181] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". In: *New J. Phys.* 18 (2016), p. 023023. DOI: 10.1088/1367-2630/18/2/023023. arXiv: 1509.04279.

[182] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. "Scalable quantum simulation of molecular energies". In: *Phys. Rev. X* 6 (2016), p. 031007. DOI: 10.1103/PhysRevX.6.031007. arXiv: 1512.06860.

[183] Earl T. Campbell and Mark Howard. "Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost". In: *Phys. Rev. A* 95 (2017), p. 022316. DOI: 10.1103/PhysRevA.95.022316. arXiv: 1606.01904.

[184] MATLAB. *version 9.0.0 (R2016a)*. Natick, Massachusetts, USA: The MathWorks Inc., 2016.

[185] James R. Seddon. *Channel magic - tools for calculating measures of magic for channels using MATLAB and CVX*. 2019-2022. URL: https://github.com/jamesrseddon/channel_magic (visited on 01/10/2021).

[186] Toby Cubitt. *Toby 'qubit' Cubitt - Matlab code*. 2004-2020. URL: http://www.dr-qubit.org/matlab.html (visited on 10/01/2021).

[187] Hakop Pashayan, Oliver Reardon-Smith, Kamil Korzekwa, and Stephen D. Bartlett. *Fast estimation of outcome probabilities for quantum circuits*. Pre-print. (2021). arXiv: 2101.12223.

[188] Shigeo Hakkaku, Kosuke Mitarai, and Keisuke Fujii. *A sampling-based quasi-probability simulation for fault-tolerant quantum error correction on the surface codes under coherent noise*. Pre-print. (2021). arXiv: 2105.04478.

[189] Daniel Litinski. "A game of surface codes: Large-scale quantum computing with lattice surgery". en-GB. In: *Quantum* 3 (2019), p. 128. DOI: 10.22331/q-2019-03-05-128. arXiv: 1808.02892.

[190] Ketan N. Patel, Igor L. Markov, and John P. Hayes. "Optimal synthesis of linear reversible circuits". In: *Quant. Inf. Comput.* 8 (2008), pp. 282–294. arXiv: quant-ph/0302002.

[191] Asher Peres. "Separability criterion for density matrices". In: *Phys. Rev. Lett.* 77 (1996), pp. 1413–1415. DOI: 10.1103/PhysRevLett.77.1413. arXiv: quant-ph/9604005.

[192] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Separability of mixed states: necessary and sufficient conditions". In: *Phys. Lett. A* 223 (1996), pp. 1–8. DOI: https://doi.org/10.1016/S0375-9601(96)00706-2. arXiv: quant-ph/9605038.

[193] Hector J. Garcia and Igor L. Markov. "Simulation of quantum circuits via stabilizer frames". In: *IEEE Trans. Computers* 64 (2015), pp. 2323–2336. DOI: 10.1109/TC.2014.2360532. arXiv: 1712.03554.

[194] Mark Howard. private communication. 2018.