IMPERIAL COLLEGE LONDON

DEPARTMENT OF PHYSICS

CENTRE FOR DOCTORAL TRAINING IN CONTROLLED QUANTUM DYNAMICS

Quasi-probability representations of quantum computing

Author: Nikolaos Koukoulekidis

November 2022

Supervised by: Dr. David Jennings & Prof. Myungshik Kim

Submitted in partial fulfillment of the requirements for
the PhD degree at Imperial College London

Abstract

If universal quantum computing is Tartarus, the mythical underworld where Titans are tormented, then magic is Charon, the ferryman tasked to get you there.

Shifting perspective often reveals simpler solutions to hard problems. In this thesis, we shift our view of quantum computing from the Hilbert space picture to a geometric picture of a discrete phase space on which computational elements can be represented through quasi-probability distributions. In this new picture, we recognize new ways to refine the theory of quantum computing.

Magic states play a crucial role in upgrading fault-tolerant computational frameworks beyond classically efficient capabilities and simulation techniques. Theories of magic have so far attempted to quantify this computational element via coarse-grained monotones and determine how these states may be efficiently transformed into useful forms. Using a quasi-probability representation of quantum states on a discrete phase space, it is known that we can identify useful magic states by the presence of negative probabilities. This thesis utilizes this representation to develop a novel statistical mechanical framework that provides a more fine-grained description of magic state transformations as well as to develop classical algorithms that simulate quantum circuits containing magic states more efficiently.

We show that majorization allows us to quantify disorder in the Wigner representation, leading to entropic bounds on magic distillation rates. The bounds are shown to be more restraining than previous bounds based on monotones and can be used to incorporate features of the distillation protocol, such as invariances of CSS protocols, as well as hardware physics, such as temperature dependence and system Hamiltonians. We also show that a subset of single-shot Rényi entropies remain well-defined on quasi-probability distributions, are fully meaningful in terms of data processing and can acquire negative values that signal magic.

Moreover, we propose classical sub-routines that reduce the sampling overhead for important classical samplers with run-times that depend on the negativity present in the Wigner representation. We show that the run-times of our sub-routines scale polynomially in circuit size and gate depth. We also demonstrate numerically that our methods provide improved scaling in the sampling overhead for random circuits with Clifford+T and Haar-random gates, while the performance of our methods compares favorably with prior simulators based on quasi-probability representations as the number of non-Clifford gates increases.

To my father, who is the better doctor.

Declaration of copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of originality

I hereby declare that this thesis is composed of my own original work and that of my acknowledged collaborators, developed during my PhD at Imperial College. All published and unpublished work of others has been appropriately cited.

List of publications

This thesis is based upon the following publications,

- Alexander R., Gvirtz-Chen S., **Koukoulekidis N.**, Jennings, D. Upper and lower bounds on CSS code dimensions for magic state distillation protocols. *PRX Quantum* **4**, 020359 (Jun. 2023) [1];
- **Koukoulekidis, N.**, Kwon, H., Jee, H., Jennings, D. & Kim M. Faster Born probability estimation via gate merging and frame optimisation. *Quantum* **6**, 838 (Oct. 2022) [2];
- **Koukoulekidis, N.** & Jennings, D. Constraints on magic state protocols from the statistical mechanics of Wigner negativity. *npj Quantum Information* **8**, 42 (Apr. 2022) [3].

During my PhD, I have also been involved in the following manuscripts which are not included in this thesis,

- **Koukoulekidis, N.**, Wang, S., O'Leary, T., Bultrini, D., Cincio, L., Czarnik, P. A framework of partial error correction for intermediate-scale quantum computers. *arXiv* (Jun. 2023) [4];
- **Koukoulekidis, N.**, Alexander, R., Hebdige, T. & Jennings, D. The geometry of passivity for quantum systems and a novel elementary derivation of the Gibbs state. *Quantum* **5**, 411 (Mar. 2021) [5];
- Dive, B., **Koukoulekidis, N.**, Mousafeiris, S. & Mintert, F. Characterization of multilevel quantum coherence without ideal measurements. *Physical Review Research* **2**, 013220 (Feb. 2020) [6].

Acknowledgments

First and foremost, I thank my primary PhD supervisor, David, for his patience with my notation, his responsiveness to my emails, and his general guidance that over the years transformed me from a caveman to a researcher. As evidence of his academic influence, I use the oxford comma, and American English throughout this thesis. I also thank my secondary PhD supervisor, Myungshik, with whom I was very fortunate to have a great deal of interactions throughout my PhD. Among the roles of collaborator, teaching assistant, conference funding collector, and attendant of his regular group meetings, I have enjoyed one slightly more than the rest. Finally, I thank Florian, who was the first to introduce me to quantum research prior to the start of my PhD.

After the start of the pandemic, I have enjoyed some time outside London, and I am indebted to the numerous friends who offered to host me during my various visits back. I would also like to thank Mara, Panaitsa and Yorgos who were determined to keep me well-nourished during my stays in Greece.

My time at Imperial as a PhD student has been brilliant mostly due to the people surrounding me: all the students at the CDT and the people with whom I shared the office (and the broken kettle which I barely managed to outlast). A special mention to Rhea, who has been an incredible collaborator, and to Ale, Dominika and Jacopo for not believing in me when I announced my plan to spend a month on a tropical island for the entire composition of this thesis.

I finally thank all the friends who put up with my rants on the dark sides of academia. In particular, Dimitris is the only witness of my thesis writing, so you have to contact him for behind-the-scenes footage and bloopers.

Contents

Chapter 1

Introduction

1.1 How does this thesis shake up the quantum world?

I applied for a PhD because I was promised that quantum computers will outperform their classical counterparts, resulting in my financial prosperity. Nonetheless, I have come to realize that the validity of this claim also constitutes a rich question of fundamental and practical significance within quantum theory. Specifically, the exact boundary between quantum and classical computational power is far from being fully characterized [7–11]. This boundary can be approached from both sides. On one hand, clever people have demonstrated the difficulty in simulating certain quantum processes classically [11–21]. Such results hint towards the ingredients that may be sufficient to achieve quantum advantage. On the other hand, it is possible to keep oneself busy with the search for efficient methods to classically simulate families of quantum circuits [22–28]. Such results provide insights on what ingredients are necessary for quantum advantage.

The question of efficient classical simulation of quantum circuits has recently received vast attention due to the ongoing rapid development of quantum devices aiming to supersede classical capabilities (e.g. [29–32]). Aided by powerful error mitigation techniques [33–37], noisy intermediate-scale quantum (NISQ) [38] devices aim to deliver computational advantages, therefore fast and accurate outcome probability estimation is a necessity for quantitative benchmarking of the devices [37,39,40]. For example, Google's recent experimental realization of a quantum speed-up [30] relies on classical estimation methods to predict statistical features of the outcome probabilities.

It is expected that exact classical simulation of arbitrary quantum systems is inefficient in general, as the resource overhead exponentially grows with the size of the system. Nevertheless, there are restricted classes of quantum circuits for which

exact classical simulation is possible [41]. The most notable example is given by circuits composed only with stabilizer states and gates in the Clifford group, which can be efficiently simulated classically via the Gottesman-Knill theorem [42]. In this thesis, we show progress in *Born probability estimation* of more general circuits with efficient methods.

On the other hand, there has been rapid progress towards the goal of experimentally achieving quantum computational fault-tolerance [43–57], promising to realize quantum advantages. Work towards fault-tolerance is increasingly bridging the gap between abstract theory and experiment. Extensive work is being done on error mitigation [36, 37, 58–60] and the incorporation of hardware physics into the theoretical models [34, 35, 61–63]. For example, the XZZX code [33] is a variant of the surface code that incorporates noise bias explicitly and has been shown to attain the hashing bound of random codes [64].

Nevertheless, many challenges remain and there is increasing need for theory to take into account physical limitations of the hardware involved. The surface code [65–68] is a leading framework for fault-tolerance with very high error thresholds. Within this scheme, Clifford unitaries can be implemented in a robust, fault-tolerant way. However, we also know that it is impossible to have a universal set of transversal gates [69] and although Clifford unitaries can be realized transversally [70, 71], one needs to find ways around this restriction for gates that do not admit efficient simulation. This can be achieved by injecting in quantum states, called magic states, which promote the Clifford group to universal quantum computing [72]. The obstacle to this is that these states are invariably noisy, so magic distillation protocols involving stabilizer states and Clifford operations must be employed to purify many copies of the magic states and improve the overall performance of the induced quantum gates [72–74]. A key question then arises about the overhead on purifying many copies of a magic state into less noisy forms.

To address this, concrete distillation protocols have been developed, such as the Bravyi-Haah qubit protocol that provides a quadratic reduction in noise per cycle [75] as well as protocols based on CSS codes [70, 71]. Such distillation rates have been improved in more recent works [73, 76–78] and there have been proposals of protocols for qudits of odd dimension [79–83] as well as of protocols within a full architectural analysis [84, 85].

There is also analysis of magic protocols from the perspective of magic theories. A number of magic theories exist [86–89], where magic states are viewed as computational resource states with respect to a class of quantum operations that are considered free [90]. One natural class of free operations are those obtained from

Clifford unitaries, Pauli measurements and the ability to discard quantum systems. However, there are also other candidates [91–93]. In any theory of magic, one route to bounding distillation rates is through a magic monotone. A magic monotone M is a real-valued function of any quantum state in a magic theory, so that M is monotonically non-increasing under the free operations of the theory, reflecting the property that free operations cannot generate magic. More precisely, given a magic theory with a set of free states $\mathcal{F}$ and free operations $\mathcal{O}$, we define a *magic monotone* as a real-valued function M such that

$$\begin{aligned} &1.\ M(\sigma) = 0, \text{ for all } \sigma \in \mathcal{F}, \text{ and} \\ &2.\ M(\rho') \leq M(\rho) \text{ whenever there exists } \mathcal{E} \in \mathcal{O} \text{ such that } \rho' = \mathcal{E}(\rho), \text{ for all } \rho, \rho'. \end{aligned} \tag{1.1}$$

In this thesis, we go beyond monotone descriptions of magic, by developing instead a thermodynamic approach to *magic state distillation* which allows for a more powerful description of magic state transformations.

1.2 Thesis outline

We begin by discussing a discrete phase space picture of quantum mechanics in chapter 2. We then establish a mathematical framework of statistical mechanics in the presence of negative probabilities in chapter 3. We use this framework to derive novel results that directly relate to quantum computing in the following three chapters (chapter 4, chapter 5, chapter 6). Finally, we propose ideas for future research in chapter 7. A slightly more detailed description of the thesis' main body and the original work within each chapter now ensues.

Chapter 2: We start by justifying the use of a phase space picture with Wigner representations for quantum objects rather than the Hilbert space picture. We proceed by identifying geometric properties of discrete phase spaces that describe quantum mechanical states and dynamics. We then construct the discrete phase space, discussing important distinctions between odd and even dimensions. We finally define Wigner representations for quantum states and operations on discrete phase spaces, firstly for odd dimensions and then for qubit systems. We identify families of operations that are important in magic distillation and classical simulation and discuss how they connect to the geometry of the phase space. All major results in this chapter are taken from the literature.

Chapter 3: We identify majorization theory, and its resulting entropic analysis, as the natural method of describing transformations of state representations within discrete phase spaces. This requires a novel extension of the scope of majorization

from classical to quantum phase spaces, where negative probabilities necessarily arise. Crucially, we develop a mathematical framework where majorization results can be extended from proper probability distributions to quasi-probability distributions (henceforth quasi-distributions for brevity), such as the Wigner representations introduced in the previous section. We first introduce non-monotonic Lorenz curves and establish relative majorization conditions for quasi-distributions. We then proceed by developing Rényi entropies and divergences on quasi-distributions and prove that they retain desired properties such as the data-processing inequality. All results that extend majorization to quasi-distributions constitute original work.

Chapter 4: We are now in a position to apply our majorization results on the Wigner representations of odd dimensional systems. We show how this gives rise to a generalized family of bounds on the rate of magic distillation for protocols of increasing complexity. We first consider protocols that possess a fixed point, such as the large family of unital protocols that leave the maximally mixed state invariant. We then discuss protocols with sub-linear correlations that admit nice descriptions in the asymptotic limit, before finally addressing general protocols. The bounds we derive take into account physical properties of the system such as Hamiltonian dependence, beyond just properties of the magic states involved. All majorization analysis, including the bounds derived using it, is novel in the context of magic.

Chapter 5: We extend the results to qubit systems. We spend the first part of this chapter motivating a family of channels that is important for qubit magic distillation, but also appropriate for our majorization framework. We then derive lower and upper bounds on qubit magic distillation rates that incorporate operational properties of the protocols considered and provide tighter constraints than magic monotones. The stochasticity results for magic distillation protocols, as well as the majorization analysis and bounds are novel.

Chapter 6: We discuss improvements in classical sampling of quantum circuits via well-known samplers that are based on the Wigner representation of the circuit components. We generalize the Wigner representation into a family of representations, controlled by parameters that adjust how large the negativities within the representation can be for any given circuit component. We develop classical algorithmic routines that minimize this negativity sampling overhead by locally varying the representations in polynomial run-time. Then, we use our sub-routines to numerically demonstrate a reduced exponential negativity scaling for important circuit architectures. The main sampling algorithm appears in the literature, while the two sub-routines improving the sampling overhead are original.

Chapter 2

Discrete Wigner representations

Central to the results of this thesis is the correspondence between the Hilbert space and phase space picture of quantum mechanics, a concept which is usually familiar from classical statistical mechanics. In this chapter, we discuss representations for quantum states and quantum operations of finite system dimension d in the form of quasi-probability distributions on a discrete phase space [1, 94–98]. These constructions are discrete versions of the real-valued Wigner function that appears in continuous quantum mechanics and quantum optics [94, 99–101].

This chapter provides results that are mostly well-known in the literature, with a few additions or restatements that assist our results in the following chapters. In particular, we broadly motivate the need for a phase space picture in section 2.1, before moving on to section 2.2 to describe the geometric elements comprising a discrete phase space as well as the transformation that these elements admit. We then describe how Pauli operators can be viewed as displacement operators on the phase space and also explain why even and odd dimensional systems should be treated separately in section 2.3. Finally, in section 2.4, we describe how to represent systems of odd and even dimension on the phase space, following the construction of [97] and of [1, 98, 102] respectively.

2.1 Why should we think in the phase space picture?

There are two equivalent representations of continuous quantum mechanics: the density operator, and the quasi-distribution representation, as established in 1932 by Wigner [99]. The continuous quasi-distribution representation is now known as the Wigner function and it encodes the same information as the density operator of a state. Laws of quantum mechanics are expressible in both representations and, con-

versely, either representation may be better-suited for describing a certain law. For example, evolving the Wigner function of a classical particle in time directly reduces to the Liouville equation, elegantly revealing how classical statistical dynamics form a subset of quantum theory.

Discrete phase spaces were considered in Wootters' seminal work [94], aiming to extend the Wigner function so it describes discrete degrees of freedom of a quantum system, such as its spin. As is the case with continuous quantum mechanics, the discrete Wigner representation is arguably a more natural framework in describing the behavior of discrete systems than the density matrix within certain contexts. Pertinently, negativity in the Wigner representation is an indicator of superpolynomial advantage in quantum computing [41, 103–105]. This is the reason why working in the phase space picture is natural for the subsequent chapters that relate to magic. Therefore, in this chapter, we study the geometric elements of a Wigner representation, first developed by Wootters [94] and later expressed in slightly altered forms [97, 98, 102], then we derive the various distinguishing properties that stabilizer and magic states as well as Clifford channels possess on a discrete phase space.

We should take a step back and ponder over why it is necessary to deal with odd and even dimensional systems separately. Even though the Wigner representation is a natural indicator of non-classical advantage in the computational paradigm of magic state injection, this connection is even more satisfying in odd dimensions. Remarkably, negativity of the Wigner representation, contextuality and magic are three indicators of non-classicality that align exactly for the case of odd dimensional systems [106]. We recall that contextuality refers to the inability to model the inherent randomness of a quantum measurement via a combination of predetermined measurement outcomes [106].

This connection breaks down in even dimensions. The negativity of the Wigner representation is no longer an indicator of non-classical states and operations. In fact, the correspondence between stabilizer states and non-negative Wigner representation breaks down for qubits. Furthermore, contextuality is now a necessary but not sufficient condition for quantum advantage in computations with qubit magic [106]. Work attempting to reconcile the satisfying bidirectional connection between these notions is ongoing [98, 102, 105, 107, 108]. In section 2.3, we explain that we are able to retrieve this connection by choosing a different, non-Hermitian basis for even dimensional Wigner representations, at the cost of restricting the analysis to a subset of quantum states and channels.

2.2 Geometry on the phase space

We start by considering a d–dimensional quantum system with Hilbert space $\mathcal{H}_d$ for any finite dimension d. Let $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$ denote the standard computational basis, defined over $\mathbb{Z}_d = \{0, 1, \ldots, d-1\}$.

The *phase space* corresponding to $\mathcal{H}_d$ is defined as the 2–dimensional vector space $\mathcal{P}_d \coloneqq \mathbb{Z}_d \times \mathbb{Z}_d$. Any vector $\boldsymbol{x} \coloneqq (x_0, x_1) \in \mathcal{P}_d$ can be viewed as a pair of position and momentum coordinates on a $d \times d$ square grid, with $(0, 0)$ as the origin. It forms the core link between the Hilbert space picture and the geometrical picture we will discuss in the rest of the chapter. In order to describe the geometry of the phase space we first require rigorous notions of its basic elements, such as points and lines, and then examine the transformations between these elements, namely translations and rotations.

The phase space is defined over the finite field $\mathbb{Z}_d$, so it is endowed with the usual operations:

1. *(scalar multiplication)*. For any point $\boldsymbol{x} \in \mathcal{P}_d$ and $k \in \mathbb{Z}_d$, $k\boldsymbol{x} \coloneqq (kx_0, kx_1)$;
2. *(addition)*. For any two points $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_d$, $\boldsymbol{x} + \boldsymbol{y} \coloneqq (x_0 + y_0, x_1 + y_1)$;
3. *(inner product)*. For any two points $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_d$, $\boldsymbol{x} \cdot \boldsymbol{y} \coloneqq x_0 y_0 + x_1 y_1$.

We further define

$$J_1 \coloneqq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.1}$$

and equip $\mathcal{P}_d$ with the symplectic inner product $\langle \cdot\ ,\ \cdot \rangle : \mathcal{P}_d \times \mathcal{P}_d \mapsto \mathbb{Z}_d$,

$$\langle \boldsymbol{x}\ ,\ \boldsymbol{y} \rangle \coloneqq \boldsymbol{y}^T J_1 \boldsymbol{x} = x_1 y_0 - x_0 y_1. \tag{2.2}$$

The symplectic product satisfies the following properties:

1. *(anti-symmetry)*. $\langle \boldsymbol{x}\ ,\ \boldsymbol{y} \rangle = -\langle \boldsymbol{y}\ ,\ \boldsymbol{x} \rangle$;
2. *(bi-linearity)*. $\langle \alpha\boldsymbol{x} + \beta\boldsymbol{x}'\ ,\ \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}\ ,\ \boldsymbol{y} \rangle + \beta \langle \boldsymbol{x}'\ ,\ \boldsymbol{y} \rangle$;
3. *(alternation)*. $\langle \boldsymbol{x}\ ,\ \boldsymbol{x} \rangle = 0$.

2.2.1 Phase space elements for prime dimensions

The basic elements of the phase space are points and lines, as well as the possible transformations one can perform on them. We develop these elements here.

The geometric notion of a line can be intuitively introduced on $\mathcal{P}_d$,

$$\boldsymbol{a} \cdot \boldsymbol{x} = k, \boldsymbol{a} \in \mathcal{P}_d, k \in \mathbb{Z}_d, \tag{2.3}$$

where $\boldsymbol{a}$ is the direction vector.

Note that for non-prime dimensions, this notion of a line is not well-defined. As an example, consider $\mathcal{P}_4$ and the line $(1,2) \cdot \boldsymbol{x} = 0$. The dilation $\boldsymbol{x} \to 2\boldsymbol{x}$, leads to a different line $(1,0) \cdot \boldsymbol{x} = 0$, with different direction vector $\boldsymbol{a}$. This is a consequence of the general property of a field $\mathbb{Z}_n$ with composite $n = ab$, where elements a and b have no multiplicative inverse and instead multiply together to the identity of addition 0. In the specific example of $\mathbb{Z}_4$, the property manifests as $2 \cdot 2 = 0$.

For prime dimensions d, the direction vector $\boldsymbol{a}$ labels an equivalence class of parallel lines, while the constant k distinguishes between the lines of the class. There are $(d+1)$ distinct choices for $\boldsymbol{a}$, therefore $d(d+1)$ distinct lines in total. Note that in prime dimensions d, dilations of the form $\boldsymbol{x} \to c\boldsymbol{x}$, for $c \in \mathbb{Z}_d$, simply amount to shifting the points along the same line, $\boldsymbol{a} \to c^{-1}\boldsymbol{a} \propto \boldsymbol{a}$, since any element c has a multiplicative inverse. Therefore, from now on, we focus on phase spaces of prime dimension d and discuss composite systems in section 2.2.2.

We then examine transformations between points on $\mathcal{P}_d$. There are d^2 translations on $\mathcal{P}_d$ generated by addition of vectors, hence $\mathcal{P}_d$ can be considered as the Abelian *group of translations* when equipped with vector addition. We will denote this group $\mathcal{T}(\mathbb{Z}_d)$, or in shorthand notation $\mathcal{T}$, to avoid ambiguity with the set $\mathcal{P}_d$ of points on the phase space. Any translation $\boldsymbol{x} \to \boldsymbol{x} + \xi$ results in a change of constant $k \to k - \boldsymbol{a} \cdot \xi$ in Eq. (2.3), so that the orbit of a line under the action of translations is its equivalence class.

Rotations on $\mathcal{P}_d$ are represented by (2×2) matrices $S : \mathcal{P}_d \mapsto \mathcal{P}_d$ with $\det[S] = 1$. The non-zero determinant also guarantees an inverse rotation. There are $d(d^2-1)$ rotations and they form a group under matrix multiplication, known as the *symplectic group* $\mathrm{SP}(2, \mathbb{Z}_d)$. The name is justified based on the property that a symplectic rotation S preserves the symplectic product, $S^T J_1 S = J_1$. Any rotation $\boldsymbol{x} \to S\boldsymbol{x}$ results in the change of direction vector $\boldsymbol{a} \to S^{-1}\boldsymbol{a}$, so that the orbit of a line under the action of rotations passes through all equivalence classes once via the line with the same constant k.

We now establish that the translations and rotations as defined above form the group of *symplectic affine transformations*, which encompasses all possible transformations between lines on the phase space.

Claim 2.1. *The group of line transformations on $\mathcal{P}_d$ is the semi-direct product of translations and rotations $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ defined as the Cartesian product $\mathrm{SP}(2, \mathbb{Z}_d) \times \mathcal{T}$*

endowed with the composition rule

$$(S_2, \xi_2) \circ (S_1, \xi_1) = (S_2 S_1, \xi_2 + S_2 \xi_1), \tag{2.4}$$

Proof. We first prove that $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ is a group. Closeness and associativity follow from the closeness and associativity of $\mathrm{SP}(2, \mathbb{Z}_d)$ and $\mathcal{T}$. The identity is $(\mathbb{1}, 0)$ and every element $(S, \boldsymbol{\xi})$ has inverse $(S^{-1}, -S^{-1}\boldsymbol{\xi})$.

The group of translations $\mathcal{T}$ is a normal subgroup, since for any $(\mathbb{1}, \xi) \in \mathcal{T}$ and $(S, \boldsymbol{z}) \in \mathrm{SP}(2, \mathbb{Z}_d) \times \mathcal{T}$,

$$(S, \boldsymbol{z}) \circ (\mathbb{1}, \xi) \circ (S, \boldsymbol{z})^{-1} = (S, \boldsymbol{z}) \circ (\mathbb{1}, \xi) \circ (S^{-1}, -S^{-1}\boldsymbol{z}) = (\mathbb{1}, S\xi) \in \mathcal{T}. \tag{2.5}$$

The group of symplectic rotations $\mathrm{SP}(2, \mathbb{Z}_d)$ is thus a subgroup of $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ and its intersection with $\mathcal{T}$ is simply the identity, $\mathrm{SP}(2, \mathbb{Z}_d) \cap \mathcal{T} = (\mathbb{1}, \mathbf{0})$. Therefore, $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ splits over $\mathcal{T}$ and is the semi-direct product of $\mathrm{SP}(2, \mathbb{Z}_d)$ and $\mathcal{T}$. □

This crucial result will ensure a correspondence between the action of the Clifford group on stabilizer states and transformations of lines on a discrete phase space.

2.2.2 Phase space elements for composite dimensions

The formalism readily generalizes to a system of n particles. For such a system, the d–dimensional composite Hilbert space is

$$\mathcal{H}_d = \mathcal{H}_{d_1} \otimes \mathcal{H}_{d_2} \otimes \cdots \otimes \mathcal{H}_{d_n}, \tag{2.6}$$

where $d = d_1 d_2 \ldots d_n$ and d_i is the dimension of the i–th particle Hilbert space. The Hilbert space is defined over the Cartesian product of the finite fields $\mathcal{Z}_d \coloneqq \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_n}$. Therefore, the corresponding phase space is

$$\mathcal{P}_d \coloneqq \mathcal{Z}_d \times \mathcal{Z}_d \cong \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_n} \times \mathbb{Z}_{d_n}. \tag{2.7}$$

Any vector in $\mathcal{P}_d$ is of the form

$$\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \oplus \cdots \oplus \boldsymbol{x}_n \coloneqq (x_{1,0}, x_{2,0}, \ldots, x_{n,0}, x_{1,1}, x_{2,1}, , \ldots, x_{n,1}), \tag{2.8}$$

where $\boldsymbol{x}_i \coloneqq (x_{i,0}, x_{i,1})$ is the component vector in the *factor phase space* $\mathcal{P}_{d_i}$. The symplectic inner product $\langle \boldsymbol{x} \ , \ \boldsymbol{y} \rangle \coloneqq \boldsymbol{y}^T J_n \boldsymbol{x}$ is now defined by the generalized matrix

$$J_n \coloneqq \begin{pmatrix} \mathbb{O}_n & \mathbb{1}_n \\ -\mathbb{1}_n & \mathbb{O}_n \end{pmatrix}, \tag{2.9}$$

where $\mathbb{1}_n$ and $\mathbb{O}_n$ are the $n \times n$ identity and zero matrices respectively. Claim 2.2 reduces calculations on phase spaces of product dimensions to calculations on the factor phase spaces. For clarity it is stated in the context of two particles, but generalization to the multi-particle case is obvious.

Claim 2.2. *For any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{P}_d$ with product dimension $d = d_A d_B$, $\boldsymbol{x} = \boldsymbol{x}_A \oplus \boldsymbol{x}_B$ and $\boldsymbol{y} = \boldsymbol{y}_A \oplus \boldsymbol{y}_B$, the following hold:*

1. *(distribution over addition).* $\boldsymbol{x} + \boldsymbol{y} = (\boldsymbol{x}_A + \boldsymbol{y}_A) \oplus (\boldsymbol{x}_B + \boldsymbol{y}_B)$;
2. *(distribution over inner product).* $\boldsymbol{x} \cdot \boldsymbol{y} = (\boldsymbol{x}_A \cdot \boldsymbol{y}_A) + (\boldsymbol{x}_B \cdot \boldsymbol{y}_B)$;
3. *(distribution over symplectic product).* $\langle \boldsymbol{x} \ , \ \boldsymbol{y} \rangle = \langle \boldsymbol{x}_A \ , \ \boldsymbol{y}_A \rangle + \langle \boldsymbol{x}_B \ , \ \boldsymbol{y}_B \rangle$.

Proof. These geometric properties were first discussed in the context of discrete phase spaces in [94]. The first two statements are a direct consequence of Eq. (2.8), while the third statement is a direct consequence of Eq. (2.8) and Eq. (2.9). □

Translations and symplectic rotations can simply be written as $\boldsymbol{\xi} = \boldsymbol{\xi}_A \oplus \boldsymbol{\xi}_B$ and $S = S_A \oplus S_B$, while they satisfy $S\boldsymbol{x} = (S_A \oplus S_B)(\boldsymbol{x}_A \oplus \boldsymbol{x}_B) = (S_A \boldsymbol{x}_A) \oplus (S_B \boldsymbol{x}_B)$. The group $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ retains its composition rule, Eq. (2.4), since

$$\begin{aligned}
&(S_{A2} \oplus S_{B2}, \boldsymbol{\xi}_{A2} \oplus \boldsymbol{\xi}_{B2}) \circ (S_{A1} \oplus S_{B1}, \boldsymbol{\xi}_{A1} \oplus \boldsymbol{\xi}_{B1}) \\
=&\Big((S_{A2} S_{A1}) \oplus (S_{B2} S_{B2}), (\boldsymbol{\xi}_{A2} + S_{A2} \boldsymbol{\xi}_{A1}) \oplus (\boldsymbol{\xi}_{B2} + S_{B2} \boldsymbol{\xi}_{B1}) \Big) \\
=&(S_{A2} S_{A1}, \boldsymbol{\xi}_{A2} + S_{A2} \boldsymbol{\xi}_{A1}) \oplus (S_{B2} S_{B2}, \boldsymbol{\xi}_{B2} + S_{B2} \boldsymbol{\xi}_{B1}).
\end{aligned} \tag{2.10}$$

2.3 Displacement operators

Generalized Pauli matrices $X, Z : \mathcal{H}_d \mapsto \mathcal{H}_d$ can be defined by their respective roles as position and phase shift operators, acting on the basis states as follows,

$$X \left| k \right\rangle = \left| k + 1 \right\rangle \tag{2.11}$$

$$Z \left| k \right\rangle = \omega^k \left| k \right\rangle . \tag{2.12}$$

Here $\omega \coloneqq e^{2\pi i / d}$ is the d–th root of unity and addition is taken modulo d.

Claim 2.3. *For any dimension d, the Pauli operators X, Z satisfy:*

1. *(decomposition).* $X = \sum_{k=0}^{d-1} \left| k + 1 \right\rangle \left\langle k \right|$ *and* $Z = \sum_{k=0}^{d-1} \omega^k \left| k \right\rangle \left\langle k \right|$;
2. *(unitarity).* $X^\dagger = X^{-1}$ *and* $Z^\dagger = Z^{-1}$;
3. *(power rule).* $X^n \left| k \right\rangle = \left| k + n \right\rangle$ *and* $Z^n \left| k \right\rangle = \omega^{nk} \left| k \right\rangle$ *for all* $n \in \mathbb{Z}_d$;
4. *(commutation).* $ZX = \omega XZ$.

Proof. These well-known properties can be found in e.g. [96, 109].

1. Both decompositions follow from the orthonormality of the computational basis.
2. Using property 1, it is readily seen that $X^\dagger X = XX^\dagger = \mathbb{1}$ and $Z^\dagger Z = ZZ^\dagger = \mathbb{1}$.
3. It follows from induction on Eq. (2.11) and Eq. (2.12).
4. It follows from the decompositions in property 1.

□

The X, Z operators act as orthogonal unit displacement operators corresponding to the two orthogonal unit vectors on the phase space $\mathcal{P}_d$. We generalize this correspondence by assigning a displacement operator to every point of the phase space,

$$D_{\boldsymbol{x}} \coloneqq \chi(x_0, x_1) Z^{x_1} X^{x_0}, \tag{2.13}$$

where in general there is a phase factor $\chi(x_0, x_1)$ expressed as a function of the phase space point components (x_0, x_1).

The role of this phase factor is to ensure that the displacement operators satisfy certain desired properties. These properties manifest differently in odd and even dimensional systems, so we will treat these two cases separately. In particularly, for odd prime dimensional systems we choose $\chi(x_0, x_1) = \tau^{-x_0 x_1}$, where the phase $\tau \coloneqq -e^{i\pi/d}$ satisfies $\omega = \tau^2$. This choice of phase ensures Hermiticity of the displacement operators leading to a real-valued Wigner representation, where stabilizer states and Clifford operations are non-negatively represented. For qubit systems, we choose $\chi(x_0, x_1) = 1$, i.e. the phase factor is independent of the phase space coordinates. This choice of phase does not guarantee Hermiticity of the displacement operators leading to a complex-valued Wigner representation. However, it allows us to restrict the computational framework to a carefully selected subset of states and Clifford operations which admit real-valued Wigner representations, and where the stabilizer states and operations within this subset are represented non-negatively, a crucial property which fails in the qubit case if $\chi(x_0, x_1) = \tau^{-x_0 x_1}$.

We first describe the straightforward case of qudits for odd prime d, following the construction of [97], and then we summarize the case of qubits, following our own construction of [1] based on [98, 102].

2.4 Phase space in odd dimensions

We focus statements on odd prime dimensions, and when relevant we comment on their simple extensions to general odd dimensions via the tensor product construc-

tion.

Recall that the displacement operators are

$$D_{\boldsymbol{x}} := \tau^{x_0 x_1} X^{x_0} Z^{x_1}, \tag{2.14}$$

They satisfy the following properties:

Claim 2.4. *For any dimension d, the displacement operators satisfy:*

1. *(unitarity).* $D_{\boldsymbol{x}}^\dagger = D_{-\boldsymbol{x}} = D_{\boldsymbol{x}}^{-1}$;
2. *(closure under transposition).* $D_{(x_0,x_1)}^T = D_{(-x_0,x_1)}$;
3. *(vanishing trace).* $\mathrm{tr}[D_{\boldsymbol{x}}] = d\delta_{\boldsymbol{x},\boldsymbol{0}}$;
4. *(composition rule).* $D_{\boldsymbol{x}} D_{\boldsymbol{x}'} = \tau^{\langle \boldsymbol{x}\,,\,\boldsymbol{x}'\rangle} D_{\boldsymbol{x}+\boldsymbol{x}'}$;
5. *(orthogonality).* $\mathrm{tr}[D_{\boldsymbol{x}}^\dagger D_{\boldsymbol{x}'}] = d\delta_{\boldsymbol{x},\boldsymbol{x}'}$.

Proof. These properties of displacement operators also appear in e.g. [41, 96]. They are derived by use of the Pauli operator properties given in Claim 2.3 along with the definition of τ.

1. The unitarity of the Pauli operators leads to

$$D_{\boldsymbol{x}}^\dagger = (\tau^*)^{x_0 x_1} (Z^\dagger)^{x_1} (X^\dagger)^{x_0} = \tau^{-x_0 x_1} Z^{-x_1} X^{-x_0} = D_{\boldsymbol{x}}^{-1} \tag{2.15}$$

   and

$$D_{\boldsymbol{x}}^{-1} = \tau^{-x_0 x_1} \omega^{(-x_0)(-x_1)} X^{-x_0} Z^{-x_1} = \tau^{x_0 x_1} X^{-x_0} Z^{-x_1} = D_{-\boldsymbol{x}} \tag{2.16}$$

   because of the commutation property 4 of the Pauli operators and $\omega = \tau^2$.

2. Applying X repeatedly performs multiple unit shifts, $X^{x_0} = \sum_{k=0}^{d-1} |k + x_0\rangle \langle k|$ and, similarly, applying Z repeatedly adds multiple phases, $Z^{x_1} = \sum_{k=0}^{d-1} \omega^{x_1 k} |k\rangle \langle k|$. Then, it is readily seen that

$$D_{(x_0,x_1)}^T = D_{-(x_0,x_1)}^* = \tau^{-x_0 x_1} X^{-x_0} Z^{x_1} = D_{(-x_0,x_1)}. \tag{2.17}$$

3. Using the repeated calculation from above,

$$\mathrm{tr}[D_{\boldsymbol{x}}] = \tau^{x_0 x_1} \sum_{k=0}^{d-1} \omega^{x_1 k} \langle k | k + x_0 \rangle = \tau^{x_0 x_1} \sum_{k=0}^{d-1} \omega^{x_1 k} \delta_{x_0 0} = \tau^{x_0 x_1}\ d\delta_{x_1 0}\ \delta_{x_0 0} = d\delta_{\boldsymbol{x},\boldsymbol{0}}. \tag{2.18}$$

4. By repeated use of the commutation property 4 and explicit calculation of the phase factor in the last step,

$$\begin{aligned} D_{\boldsymbol{x}} D_{\boldsymbol{x}'} =& \tau^{x_0 x_1} X^{x_0} Z^{x_1} \tau^{x_0' x_1'} X^{x_0'} Z^{x_1'} \\ =& \tau^{x_0 x_1 + x_0' x_1'} \omega^{x_1 x_0'} X^{x_0 + x_0'} Z^{x_1 + x_1'} = \tau^{x_1 x_0' - x_0 x_1'} D_{\boldsymbol{x}+\boldsymbol{x}'}. \end{aligned} \tag{2.19}$$

5. Using the previous properties, $\mathrm{tr}[D_{\boldsymbol{x}}^\dagger D_{\boldsymbol{x}'}] = \tau^{(-x_1)x_0' - (-x_0)x_1'} \mathrm{tr}[D_{-\boldsymbol{x}+\boldsymbol{x}'}] = d\delta_{\boldsymbol{x},\boldsymbol{x}'}$.

□

The displacement operators thus form *the generalized Pauli group*,

$$\mathrm{GP} \coloneqq \{\tau^k D_{\boldsymbol{z}} : k \in \mathbb{Z}_d, \boldsymbol{z} \in \mathcal{P}_d\}. \tag{2.20}$$

Due to the composition property 4, GP is a projective representation of the group of translations $\mathcal{T}$. By construction, the center of the group is the set of possible phases $\{\tau^k \mathbb{1} : k \in \mathbb{Z}_d\}$ for odd dimensions (in the case of even d, the condition becomes $k \in \mathbb{Z}_{2d}$). There is an isomorphism between the group of translations and the group of displacement operators modulo phases, $\mathcal{T} \sim \mathrm{GP}/\{\tau^k \mathbb{1} : k \in \mathbb{Z}_d\}$.

For a system with composite Hilbert space, $\mathcal{H}_d = \mathcal{H}_{d_A} \otimes \mathcal{H}_{d_B}$, with $d = d_A d_B$, the displacement operators are defined as

$$D_{\boldsymbol{x}_A \oplus \boldsymbol{x}_B} \coloneqq D_{\boldsymbol{x}_A} \otimes D_{\boldsymbol{x}_B}, \tag{2.21}$$

where $\boldsymbol{x}_A \oplus \boldsymbol{x}_B \in \mathcal{P}_d$. Therefore, $D_{\boldsymbol{x}_A}, D_{\boldsymbol{x}_B}$ are completely uncorrelated and Claim 2.4 readily holds due to the composition properties laid out in Claim 2.2. We rederive the composition property 4 here as a demonstration of algebraic manipulations in composite spaces,

$$\begin{aligned} D_{\boldsymbol{x}_A \oplus \boldsymbol{x}_B} D_{\boldsymbol{x}'_A \oplus \boldsymbol{x}'_B} &= D_{\boldsymbol{x}_A} D_{\boldsymbol{x}'_A} \otimes D_{\boldsymbol{x}_B} D_{\boldsymbol{x}'_B} = \tau^{\langle \boldsymbol{x}_A \, , \, \boldsymbol{x}'_A \rangle} D_{\boldsymbol{x}_A + \boldsymbol{x}'_A} \otimes \tau^{\langle \boldsymbol{x}_B \, , \, \boldsymbol{x}'_B \rangle} D_{\boldsymbol{x}_B + \boldsymbol{x}'_B} \\ &= \tau^{\langle \boldsymbol{x}_A \, , \, \boldsymbol{x}'_A \rangle + \langle \boldsymbol{x}_B \, , \, \boldsymbol{x}'_B \rangle} D_{(\boldsymbol{x}_A + \boldsymbol{x}'_A) \oplus (\boldsymbol{x}_B + \boldsymbol{x}'_B)} \\ &= \tau^{\langle \boldsymbol{x}_A \oplus \boldsymbol{x}_B \, , \, \boldsymbol{x}'_A \oplus \boldsymbol{x}'_B \rangle} D_{(\boldsymbol{x}_A \oplus \boldsymbol{x}_B) + (\boldsymbol{x}'_A \oplus \boldsymbol{x}'_B)}. \end{aligned} \tag{2.22}$$

Eq. (2.21) determines the way an odd dimensional phase space is built on its prime dimensional factor phase spaces which act as building blocks. For n d–dimensional particles, we would write $\mathrm{GP}_n \coloneqq \{\tau^k D_{\boldsymbol{z}} : k \in \mathbb{Z}_d, \boldsymbol{z} \in \mathcal{P}_{d^n}\}$ with $\mathrm{GP}_1 = \mathrm{GP}$.

2.4.1 Phase-point operators

We seek a basis for Hermitian matrices, so that we can represent any density operator as a combination of this basis with real coefficients. Therefore, we define the phase-

point operators via the displacement operators,

$$A_{\mathbf{0}} \coloneqq \frac{1}{d} \sum_{\boldsymbol{z} \in \mathcal{P}_d} D_{\boldsymbol{z}}; \tag{2.23}$$

$$A_{\boldsymbol{x}} \coloneqq D_{\boldsymbol{x}} A_{\mathbf{0}} D_{\boldsymbol{x}}^{\dagger} = \frac{1}{d} \sum_{\boldsymbol{z} \in \mathcal{P}_d} \omega^{\langle \boldsymbol{x} \, , \, \boldsymbol{z} \rangle} D_{\boldsymbol{z}} \text{ for any } \boldsymbol{x} \in \mathcal{P}_d. \tag{2.24}$$

In the computational basis, the matrix elements of $A_{\boldsymbol{x}}$ are given by

$$\langle i | \, A_{\boldsymbol{x}} \, | j \rangle = \omega^{(i-j)x_1} \delta_{i+j, 2x_0}. \tag{2.25}$$

The phase-point operator at the origin also acts as the parity operator,

$$A_{\mathbf{0}} = \sum_{k \in \mathbb{Z}_d} |k\rangle\langle -k| \, . \tag{2.26}$$

The following properties ensure that the phase-point operators comprise a Hermitian orthonormal basis in the space of d–dimensional complex matrices and hence quantum states in particular.

Claim 2.5. *For odd prime dimension d, the phase-point operators satisfy:*

1. *(Hermiticity and unitarity).* $A_{\boldsymbol{x}}^{\dagger} = A_{\boldsymbol{x}} = A_{\boldsymbol{x}}^{-1}$;
2. *(closure under transposition).* $A_{(x_0, x_1)}^{T} = A_{(-x_0, x_1)}$;
3. *(unit trace).* $\mathrm{tr}[A_{\boldsymbol{x}}] = 1$;
4. *(completeness relation).* $\sum_{\boldsymbol{z} \in \mathcal{P}_d} A_{\boldsymbol{z}} = d\mathbb{1}$;
5. *(orthogonality).* $\mathrm{tr}[A_{\boldsymbol{x}}^{\dagger} A_{\boldsymbol{x}'}] = d\delta_{\boldsymbol{x}, \boldsymbol{x}'}$.

Proof. This Claim is a collection of results that also appear in e.g. [41, 96, 100]. Note that Eq. (2.26) can be reached with direct calculation from the definition Eq. (2.23) and readily implies that $A_{\mathbf{0}}$ is unitary and Hermitian.

1. The statement follows by Eq. (2.24) combined with the Hermiticity and unitarity of $A_{\mathbf{0}}$.
2. $A_{\boldsymbol{x}}^{T} = \frac{1}{d} \sum_{\boldsymbol{z} \in \mathcal{P}_d} \omega^{\langle \boldsymbol{x} \, , \, \boldsymbol{z} \rangle} D_{(z_0, -z_1)} = \sum_{\boldsymbol{z} \in \mathcal{P}_d} \omega^{\langle (-x_0, x_1) \, , \, \boldsymbol{z} \rangle} D_{(z_0, z_1)}$ due to the closure under transposition property 2 of displacement operators.
3. We observe that $\mathrm{tr}[A_{\mathbf{0}}] = \delta_{k, -k} = 1$, since for odd dimensions $k = -k$ in $\mathbb{Z}_d$ only when $k = 0$. Unitarity of the displacement operators (property 1) ensures that $\mathrm{tr}[A_{\boldsymbol{x}}] = 1$ for any $\boldsymbol{x}$.

4. We can prove this statement by direct calculation, using the definition of phase-point operators Eq. (2.24) and substituting Eq. (2.26).

5. Dropping the dagger due to Hermiticity and using the orthogonality of the displacement operators according to property 5, the trace becomes

$$\mathrm{tr}[A_{\boldsymbol{x}}^{\dagger}A_{\boldsymbol{x}'}] = \frac{1}{d^2}\mathrm{tr}\left[\sum_{\boldsymbol{z}\in\mathcal{P}_d}\omega^{\langle \boldsymbol{x}\,,\,\boldsymbol{z}\rangle}D_{\boldsymbol{z}}\sum_{\boldsymbol{z}'\in\mathcal{P}_d}\omega^{\langle \boldsymbol{x}'\,,\,\boldsymbol{z}'\rangle}D_{\boldsymbol{z}'}\right] = d\delta_{\boldsymbol{x},\boldsymbol{x}'}. \tag{2.27}$$

□

The phase-point operators extend easily to composite Hilbert spaces, once we identify that for $\mathcal{H}_d = \mathcal{H}_{d_A}\otimes\mathcal{H}_{d_B}$,

$$A_{\boldsymbol{0}\oplus\boldsymbol{0}} = \frac{1}{d}\sum_{\boldsymbol{x}\oplus\boldsymbol{y}\in\mathcal{P}_d^2}D_{\boldsymbol{x}\oplus\boldsymbol{y}} = \frac{1}{d_A d_B}\sum_{\boldsymbol{x}\in\mathcal{P}_d}\sum_{\boldsymbol{y}\in\mathcal{P}_d}D_{\boldsymbol{x}}\otimes D_{\boldsymbol{y}} = A_{\boldsymbol{0}}\otimes A_{\boldsymbol{0}}. \tag{2.28}$$

Then, $A_{\boldsymbol{x}\oplus\boldsymbol{y}} = A_{\boldsymbol{x}}\otimes A_{\boldsymbol{y}}$ and all properties in Claim 2.5 hold.

2.4.2 Wigner representation for states

We have established that any quantum state $\rho\in\mathcal{B}(\mathcal{H}_d)$ can be decomposed to phase-point operators,

$$\rho = \sum_{\boldsymbol{z}\in\mathcal{P}_d}W_\rho(\boldsymbol{z})A_{\boldsymbol{z}}. \tag{2.29}$$

The coefficient distribution defines the Wigner representation of state ρ,

$$W_\rho(\boldsymbol{x}) \coloneqq \frac{1}{d}\mathrm{tr}[A_{\boldsymbol{x}}\rho]. \tag{2.30}$$

Claim 2.6. *The Wigner representation of a state ρ is*

1. *(real-valued).* $W_\rho(\boldsymbol{x})\in\mathbb{R}$;
2. *(normalized).* $\sum_{\boldsymbol{z}\in\mathcal{P}_d}W_\rho(\boldsymbol{z}) = 1$;
3. *(bounded).* $|W_\rho(\boldsymbol{x})|\leq\frac{1}{d}$.

Proof. These properties can also be found in e.g. [41].

1. The trace is invariant under transposition, $W_\rho(\boldsymbol{x})^* = \frac{1}{d}\mathrm{tr}[(A_{\boldsymbol{x}}\rho)^\dagger]$, while the cyclic property of the trace and Hermiticity of the phase-point operators ensure that $\mathrm{tr}[(A_{\boldsymbol{x}}\rho)^\dagger] = \mathrm{tr}[A_{\boldsymbol{x}}\rho]$, concluding the proof.

2. The linearity of the trace and the completeness relation of the phase-point operators in Claim 4 immediately give the result.

3. Let $\{\lambda_i\}_{i\in\mathbb{Z}_d}$ be the non-negative eigenvalues of ρ, summing to 1. Let $\{\alpha_{\boldsymbol{x},i}\}_{i\in\mathbb{Z}_d}$ be the eigenvalues of $A_{\boldsymbol{x}}$. For any $\boldsymbol{x}, \alpha_{\boldsymbol{x},i} \in \{-1, 1\}$, due to the Hermiticity and unitarity of the phase-point operators. Then,

$$|W_\rho(\boldsymbol{x})| = \frac{1}{d}|\mathrm{tr}[A_{\boldsymbol{x}}\rho]| \leq \frac{1}{d}\left|\sum_i \alpha_{\boldsymbol{x},i}\lambda_i\right| \leq \frac{1}{d}\sum_i \lambda_i = \frac{1}{d}. \tag{2.31}$$

The first inequality follows from Theorem 1 of [110] for complex matrices, while the second is the triangle inequality. □

The Wigner representation is thus a bounded d^2–dimensional quasi-distribution over $\mathcal{P}_d$. The support of a state is the set of points on the phase space over which the Wigner representation is non-zero. Due to the the normalization property 2 and boundedness property 3 of Claim 2.6, no state can have support over fewer than d points on the phase space. Additionally, if a state has support over d points, its Wigner representation must be uniform and equal to $1/d$ over the support.

We also list additional intuitive properties of the Wigner representation that assist calculations.

Claim 2.7. *The Wigner representation satisfies:*

1. *(additive over mixing).* $W_{\sum_i p_i\rho_i}(\boldsymbol{x}) = \sum_i p_i W_{\rho_i}(\boldsymbol{x})$*;*

2. *(sharp for pure computational state).* $W_{|n\rangle\langle n|}(\boldsymbol{x}) = \frac{1}{d}\delta_{x_0,n}$ *for any* $n \in \mathbb{Z}_d$ *;*

3. *(multiplicative over tensoring).* $W_{\rho_A\otimes\rho_B}(\boldsymbol{x}_A \oplus \boldsymbol{x}_B) = W_{\rho_A}(\boldsymbol{x}_A)W_{\rho_B}(\boldsymbol{x}_B)$.

Proof. These statements simply follow from Claim 2.6.

1. This statement follows from the linearity of the trace.

2. This statement follows from direct calculation by substituting Eq. (2.30) into the definition of the Wigner representation Eq. (2.30).

3. For a state $\rho = \rho_A \otimes \rho_B \in \mathcal{H}_{d_A} \otimes \mathcal{H}_{d_B}$,

$$\begin{aligned} W_\rho(\boldsymbol{x}_A \oplus \boldsymbol{x}_B) &= \frac{1}{d}\mathrm{tr}[A_{\boldsymbol{x}_A\oplus\boldsymbol{x}_B}\rho] = \frac{1}{d_A}\mathrm{tr}_A[A_{\boldsymbol{x}_A}\rho_A]\frac{1}{d_B}\mathrm{tr}_B[A_{\boldsymbol{x}_B}\rho_B] \\ &= W_{\rho_A}(\boldsymbol{x}_A)W_{\rho_B}(\boldsymbol{x}_B). \end{aligned} \tag{2.32}$$

□

The Wigner representation thus acts as a joint quasi-distribution in a composite Hilbert space. For computational states, it is a uniform distribution over a vertical line for pure computational states. Expressing a state as a convex sum of computational states results in expressing the state representation as a convex sum of computational state representations.

One prominent magic monotone, as defined in Eq. (1.1), is the mana of a state [104], defined as

$$\mathcal{M}(\rho) \coloneqq \log(2\mathrm{sn}(\rho) + 1), \tag{2.33}$$

where the sum-negativity [104] is the sum of the negative components in the Wigner representation,

$$\mathrm{sn}(\rho) \coloneqq \sum_{\boldsymbol{z}:W_\rho(\boldsymbol{z})<0} |W_\rho(\boldsymbol{z})|. \tag{2.34}$$

Using the fact that $W_\rho(\boldsymbol{x})$ is a normalized quasi-distribution, we can also write $\mathcal{M}(\rho) = \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|$. Mana is an additive magic monotone, meaning that the contributions of two subsystems can be simply added together,

$$\mathcal{M}(\rho_1 \otimes \rho_2) = \mathcal{M}(\rho_1) + \mathcal{M}(\rho_2). \tag{2.35}$$

Because of this property, it is a useful monotone to characterize distillation of many-copy states and it will appear repeatedly in comparisons with our results in chapter 4 and chapter 5.

As an example of computing a Wigner representation, the qutrit zero computational state $|0\rangle$ has a Wigner representation consisting of three non-zero values at phase space points $(0,0)$, $(1,0)$ and $(2,0)$. For qutrit states, there exists a canonical magic state with maximal negativity in its Wigner representation, called the Strange state [104],

$$|S\rangle \coloneqq \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle). \tag{2.36}$$

Its representation $W_{|S\rangle\langle S|}(\boldsymbol{x})$ has a single negative value of $-1/3$ at $\boldsymbol{x} = (0,0)$ and the positive value $1/6$ at all other points. We depict the representation of some quantum states in Fig. (2.1).

2.4.3 Wigner representation for channels

A quantum channel $\mathcal{E}$ naturally admits a Wigner representation via its corresponding Choi-Jamiołkowski state $\mathcal{J}_\mathcal{E} \coloneqq (\mathbb{1} \otimes \mathcal{E}) \sum_{i,j} |ii\rangle\langle jj|$. Therefore, the Wigner representation of a CPTP operation $\mathcal{E} : \mathcal{B}(\mathcal{H}_{d_A}) \mapsto \mathcal{B}(\mathcal{H}_{d_B})$ is defined as

$$W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{x}) \coloneqq \frac{1}{d_B}\mathrm{tr}_B[A_{\boldsymbol{y}}\mathcal{E}(A_{\boldsymbol{x}})] = \frac{1}{d_B}\mathrm{tr}\left[\left(A_{\boldsymbol{x}}^T \otimes A_{\boldsymbol{y}}\right)\mathcal{J}_\mathcal{E}\right] = d_A^2 W_{\frac{1}{d_A}\mathcal{J}_\mathcal{E}}(\bar{\boldsymbol{x}} \oplus \boldsymbol{y}), \tag{2.37}$$

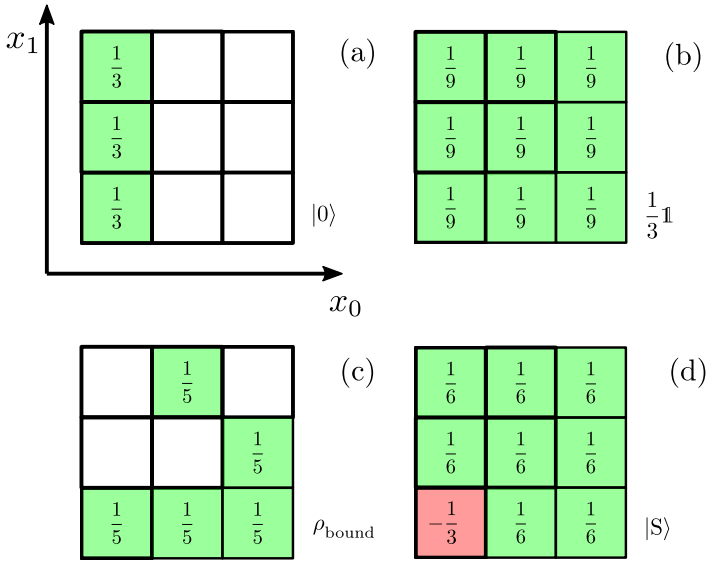

Figure 2.1: **State Wigner representations for qutrits.** (a) The computational zero state $|0\rangle$ corresponds to the vertical line $x_0 = 0$. (b) The maximally mixed state is an equal mixture of the computational states $\{|0\rangle, |1\rangle, |2\rangle\}$, so its representation is an equal mixture of all three vertical lines. (c) A Wigner positive, non-stabilizer state give in Eq. (2.59). (d) The canonical qutrit magic state, known as the Strange state.

where $\bar{\boldsymbol{x}} = (x_0, -x_1)$. Note that $\mathcal{J}_E \in \mathcal{B}(\mathcal{H}_{d_A} \otimes \mathcal{H}_{d_B})$ is a $(d_A d_B)$–dimensional state and has trace d_A, and so that $1/d_A \mathcal{J}_\mathcal{E}$ is a valid bipartite quantum state for any CPTP operation. The definition can be inverted to give $\mathcal{J}_E$ in terms of the operation Wigner representation,

$$\mathcal{J}_\mathcal{E} = \frac{1}{d_A} \sum_{\substack{\boldsymbol{x}\in\mathcal{P}_{d_A} \\ \boldsymbol{y}\in\mathcal{P}_{d_B}}} W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{x}) \left(A_{\boldsymbol{x}}^T \otimes A_{\boldsymbol{y}}\right). \tag{2.38}$$

For example, the two-qudit controlled-sum gate $C{-}SUM \coloneqq \sum_{i,j=0}^{d-1} |i\rangle\langle i| \otimes |i+j\rangle\langle j| \in \mathrm{CL}_2$ has distribution

$$W_{SUM}(\boldsymbol{y}_A \oplus \boldsymbol{y}_B | \boldsymbol{x}_A \oplus \boldsymbol{x}_B) = \delta_{y_{A0}, x_{A0}} \delta_{y_{A1}, x_{A1} - x_{B1}} \delta_{y_{B0}, x_{A0} + x_{B0}} \delta_{y_{B1}, x_{B1}}. \tag{2.39}$$

The definition in Eq. (2.37) allows for all the desired properties of the state Wigner representation to transfer to the channel Wigner representation.

Claim 2.8. *The Wigner representation of a* CPTP *operation* $\mathcal{E} : \mathcal{B}(\mathcal{H}_{d_A}) \mapsto \mathcal{B}(\mathcal{H}_{d_B})$ *is:*

1. *(real-valued).* $W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{x}) \in \mathbb{R}$*;*
2. *(normalized).* $\sum_{\boldsymbol{z}\in\mathcal{P}_{d_B}} W_\mathcal{E}(\boldsymbol{z}|\boldsymbol{x}) = 1$ *for any* $\boldsymbol{x} \in \mathcal{P}_{d_A}$*;*
3. *(bounded).* $|W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{x})| \leq \frac{d_A}{d_B}$*;*

4. *(transitive).* $W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_{d_A}} W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{z})W_{\rho}(\boldsymbol{z})$ *for any* $\boldsymbol{y} \in \mathcal{P}_{d_B}$*;*

5. *(distributive over tensoring).* $W_{\mathcal{E}\otimes\mathcal{F}}(\boldsymbol{x}\otimes\boldsymbol{y}) = W_{\mathcal{E}}(\boldsymbol{x})W_{\mathcal{F}}(\boldsymbol{y})$.

Proof. This Claim is reproduced from [93].

1. This statement is clear by the last expression of the definition in Eq. (2.37) and the real-valued property 1 of the state Wigner representation.

2. Linearity of the trace sets the left hand side equal to $\mathrm{tr}_B\left[\frac{1}{d_B}\sum_{\boldsymbol{z}\in\mathcal{P}_{d_B}} A_{\boldsymbol{z}}\mathcal{E}(A_{\boldsymbol{x}})\right]$. The completeness relation property 5 of Claim 2.5 simplifies the expression to $\mathrm{tr}_B\left[\mathcal{E}(A_{\boldsymbol{x}})\right] = \mathrm{tr}_A[A_{\boldsymbol{x}}] = 1$, because quantum operations leave the trace invariant.

3. This property is clear by the last expression in Eq. (2.37) and the boundedness property 3 of the state Wigner representation.

4. We can decompose any state into phase-point operators according to Eq. (2.29), so linearity of quantum operations and the trace imply that

$$W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_{d_A}} \frac{1}{d}\mathrm{tr}[A_{\boldsymbol{y}}\mathcal{E}(A_{\boldsymbol{z}})]W_{\boldsymbol{z}}(\rho).$$

   Identifying the coefficient of $W_{\boldsymbol{z}}(\rho)$ in the sum as the Wigner representation of operation $\mathcal{E}$ gives the result.

5. This property is clear by the first expression in Eq. (2.37) and the fact that phase-point operators are distributive over tensor products.

□

Having developed representations for quantum states and channels, we can now turn our attention to characterizing stabilizer states and Clifford channels which are computationally important. We will see that stabilizer states correspond to uniform lines on the phase space and Clifford channels correspond to symplectic affine transformations, shown in Claim 2.1 to be the group of permissible line transformations.

2.4.4 The stabilizer formalism on the phase space

The Clifford unitaries for odd dimension d are defined as the normalizer of the generalized Pauli group,

$$\mathrm{CL} \coloneqq \{U \in SU(d) : \forall \boldsymbol{x} \in \mathcal{P}_d, \exists (m, \boldsymbol{y}) \in \mathbb{Z}\times\mathcal{P}_d : U^\dagger D_{\boldsymbol{x}} U = \tau^m D_{\boldsymbol{y}}\}. \tag{2.40}$$

For n d–dimensional particles, we would write $\mathrm{CL}_n := \{U \in SU(d^n) : \forall \boldsymbol{x} \in \mathcal{P}_d, \exists (m, \boldsymbol{y}) \in \mathbb{Z} \times \mathcal{P}_{d^n} : U^\dagger D_{\boldsymbol{x}} U = \tau^m D_{\boldsymbol{y}}\}$ with $\mathrm{CL}_1 = \mathrm{CL}$. As already mentioned in section 2.2, the group of translations on the phase space and the group of displacement operators modulo phases are isomorphic. In other words, for a 1-qudit system of dimension d, the group of translations $\mathcal{T}$ on $\mathcal{P}_d$ admits a projective representation given by the displacement operators GP. Similarly, the Clifford group CL admits a projective representation given by the symplectic affine transformations $\mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{T}$ on $\mathcal{P}_d$ [111]. In particular, for an odd prime dimension d, there exists an isomorphism between the group of Clifford unitary transformations modulo phases on the Hilbert space $\mathcal{H}_d$ and the group of symplectic affine transformations on the corresponding phase space $\mathcal{P}_d$,

$$\mathrm{CL}/\{\tau^k \mathbb{1} : k \in \mathbb{Z}_d\} \sim \mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathcal{P}_d. \tag{2.41}$$

In order to explicitly express a Clifford unitary in terms of its isomorphic representation, we first define a projective representation of $\mathrm{SP}(2, \mathbb{Z}_d)$ [88],

$$F = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto V_F = \begin{cases} \frac{1}{\sqrt{d}} \sum_{j,k} \tau^{\beta^{-1}(\delta j^2 - 2jk + \alpha k^2)} |j\rangle\langle k| , & \beta \neq 0, \\ \sum_k \tau^{\alpha\gamma k^2} |\alpha k\rangle\langle k| , & \beta = 0. \end{cases} \tag{2.42}$$

Recall that $\det[F] = 1$, so $V_F^\dagger = V_F^{-1} = V_{F^{-1}}$ for all F. The unitaries V_F rotate displacement operators such that

$$V_F D_{\boldsymbol{x}} V_F^\dagger = D_{F\boldsymbol{x}}. \tag{2.43}$$

Therefore, any Clifford unitary U can be represented as

$$U_{F,\boldsymbol{z}} = D_{\boldsymbol{z}} V_F = V_F D_{F^{-1}\boldsymbol{z}}, \tag{2.44}$$

where the inverse is given by $U_{F,\boldsymbol{z}}^\dagger = U_{F,\boldsymbol{z}}^{-1} = V_{F^{-1}} D_{-\boldsymbol{z}} = D_{-F^{-1}\boldsymbol{z}} V_{F^{-1}} = U_{F^{-1},-F^{-1}\boldsymbol{z}}$.

Two important examples of Clifford unitaries are the single-qudit Hadamard gate H and phase gate S,

$$H := \frac{1}{\sqrt{d}} \sum_{j,k=1}^{d} \omega^{jk} |j\rangle\langle k| = V_F, \ F = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \tag{2.45}$$

$$S := \sum_{k=1}^{d} \tau^{k(k+1)} |k\rangle\langle k| = D_{\boldsymbol{z}'} V_{F'}, \ \boldsymbol{z}' = \left(0, 2^{-1}\right), \ F_S = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right). \tag{2.46}$$

They are Clifford because for any $\boldsymbol{x} \in \mathcal{P}_d$,

$$H D_{(x_0,x_1)} H^\dagger = D_{(-x_1,x_0)} \text{ and} \tag{2.47}$$

$$S D_{(x_0,x_1)} S^\dagger = \tau^{x_0} D_{(x_0,x_0+x_1)}, \tag{2.48}$$

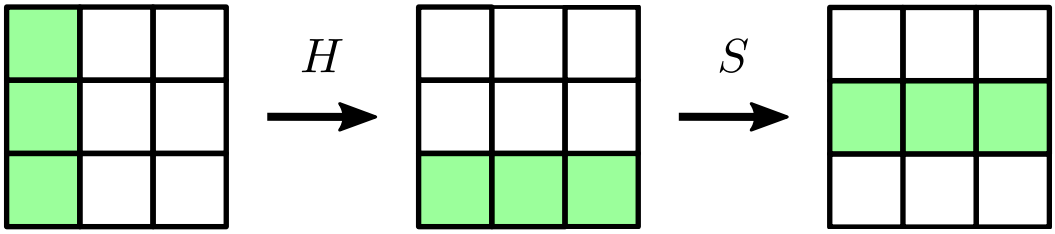

Figure 2.2: **Channel Wigner representations for qutrits.** The Hadamard gate H rotates state representations on the phase space by an angle of $\pi/2$ according to Eq. (2.49), while the phase gate S translates them vertically according to Eq. (2.50).

while they also correspond to symplectic affine transformations as shown in Eq. (2.45) and Eq. (2.46). These two matrices generate the d–dimensional group CL, for example $X = H^{-1}ZH$ and $Z = H^2S^{-1}H^2S$ [88].

Their Wigner representations can be evaluated as

$$W_H(\boldsymbol{y}|\boldsymbol{x}) = \delta_{y_0,-x_1}\delta_{y_1,x_0} \text{ and} \tag{2.49}$$

$$W_S(\boldsymbol{y}|\boldsymbol{x}) = \delta_{y_0,x_0}\delta_{y_1,x_0+x_1+2^{-1}} \tag{2.50}$$

which agree with the projective representations in Eq. (2.45) and Eq. (2.46). The action of these two gates on the phase space is depicted in Fig. (2.2).

The importance of the phase-point operators as the basis for quantum states can now become apparent due to the following result.

Claim 2.9 (Phase-point covariance [97])**.** *The phase-point operators are covariant under Clifford transformations. In particular,*

$$U_{F,\boldsymbol{z}}A_{\boldsymbol{x}}U_{F,\boldsymbol{z}}^{\dagger} = A_{F\boldsymbol{x}+\boldsymbol{z}}. \tag{2.51}$$

Proof. We use the unitary decomposotion in Eq. (2.44) for Clifford $U_{F,\boldsymbol{z}}$. The rotational part of the Clifford unitary on the operator gives

$$V_F A_{\boldsymbol{x}} V_F^{\dagger} = A_{F\boldsymbol{x}}. \tag{2.52}$$

Then, the translational part gives the final result,

$$D_{\boldsymbol{z}}A_{F\boldsymbol{x}}D_{-\boldsymbol{z}} = A_{F\boldsymbol{x}+\boldsymbol{z}}. \tag{2.53}$$

□

Claim 2.10 (Wigner distribution covariance [97])**.** *The Wigner representation of state ρ is covariant under a Clifford transformation $\rho \mapsto \rho' = U_{F,\boldsymbol{z}}\rho U_{F,\boldsymbol{z}}^{\dagger}$. In particular,*

$$W_\rho(\boldsymbol{x}) = W_{\rho'}(F\boldsymbol{x}+\boldsymbol{z}). \tag{2.54}$$

Proof. Using the cyclic property of the trace and Claim 2.9,

$$W_\rho(\boldsymbol{x}) = \frac{1}{d}\mathrm{tr}\left[A_{\boldsymbol{x}}\rho\right] = \frac{1}{d}\mathrm{tr}\left[U_{F,\boldsymbol{z}}A_{\boldsymbol{x}}U_{F,\boldsymbol{z}}^\dagger\rho'\right] = W_{\rho'}(F\boldsymbol{x}+\boldsymbol{z}). \tag{2.55}$$

□

The computational basis is taken as a subset of the stabilizer states STAB. We obtain the full set of stabilizer states by the action of the Clifford group CL on one computational basis state,

$$\mathrm{STAB} \coloneqq \mathrm{conv}\{U\left|0\right\rangle\!\left\langle 0\right|U^\dagger : U \in \mathrm{CL}\}. \tag{2.56}$$

We have already established in Claim 2.7 that the Wigner representation of any pure computational basis state is a uniform distribution over a vertical line ($x_0 = k \in \mathbb{Z}_d$) on the phase space $\mathcal{P}_d$. The definition in Eq. (2.56) makes it immediately apparent that the Wigner representation of any stabilizer state $\sigma = U\left|0\right\rangle\!\left\langle 0\right|U^\dagger$ is in fact a uniform distribution over a line on the phase space, obtained by applying the symplectic affine transformation corresponding to U on the line $x_0 = 0$.

Claim 2.11 (Stabilizer representation [97])**.** *Let $\sigma = \sum_i p_i U_{F_i,\boldsymbol{z}_i}\left|0\right\rangle\!\left\langle 0\right|U_{F_i,\boldsymbol{z}_i}^\dagger$ be any mixed stabilizer state in $\mathcal{H}_d$ with odd d. Then, its Wigner representation is*

$$W_\sigma(\boldsymbol{x}) = \frac{1}{d}\sum_i p_i\delta_{v_{i0},0} \geq 0, \tag{2.57}$$

with $\boldsymbol{v}_i \coloneqq F_i^{-1}(\boldsymbol{x}-\boldsymbol{z}_i)$ for any $\boldsymbol{x} \in \mathcal{P}_d$.

Proof. Here, we make use of Claim 2.9 and Claim 2.10.

$$\begin{aligned} W_\sigma(\boldsymbol{x}) &= \frac{1}{d}\mathrm{tr}\left[A_{\boldsymbol{x}}\sum_i p_i\, U_{F_i,\boldsymbol{z}_i}\left|0\right\rangle\!\left\langle 0\right|U_{F_i,\boldsymbol{z}_i}^\dagger\right] = \frac{1}{d}\sum_i p_i\,\mathrm{tr}\left[U_{F_i,\boldsymbol{z}_i}^\dagger A_{\boldsymbol{x}}U_{F_i,\boldsymbol{z}_i}\left|0\right\rangle\!\left\langle 0\right|\right] \\ &= \frac{1}{d}\sum_i p_i\,\mathrm{tr}\left[A_{F_i^{-1}(\boldsymbol{x}-\boldsymbol{z})}\left|0\right\rangle\!\left\langle 0\right|\right] = \sum_i p_i\, W_{\left|0\right\rangle\!\left\langle 0\right|}\left(F_i^{-1}(\boldsymbol{x}-\boldsymbol{z})\right) \\ &= \frac{1}{d}\sum_i p_i\delta_{v_{i0},0}. \end{aligned} \tag{2.58}$$

□

Therefore, all stabilizer states $\sigma \in \mathcal{F}$ are represented non-negatively by the Wigner representation. However, not all states with non-negative Wigner representations are stabilizer states [112]. For example, the qutrit state

$$\rho = \frac{1}{5}(A_{(0,0)} + A_{(0,1)} + A_{(0,2)} + A_{(1,2)} + A_{(2,1)}) \tag{2.59}$$

has a positive Wigner representation which is not a mixture of uniform distributions over lines.

Wigner representations of quantum channels act as transition matrices for Wigner representation of quantum states. In particular, channels that map a Hilbert space to itself and have non-negative Wigner representation, like Clifford channels, correspond to stochastic matrices that preserve the non-negativity of state Wigner representations.

A circuit consisting of stabilizer operations, including preparation of input stabilizer states and Clifford channels, can therefore be viewed as a process on a discrete phase space where probability distributions are updated stochastically. Magic states result in the stochastic processing of negativity.

2.5 Phase space in even dimensions

Having described the relatively straightforward case of odd dimensional systems in detail, we now introduce a new construction for even dimensions, explicitly remarking the differences.

2.5.1 Phase-point operators

Consider a n–qubit system with Hilbert space $\mathcal{H}_2^{\otimes n}$. The dimension of this system is $d \coloneqq 2^n$ and we define the corresponding phase space $\mathcal{P}_d \coloneqq \mathbb{Z}_d \times \mathbb{Z}_d$, where all vectors $\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \oplus \cdots \oplus \boldsymbol{x}_n \in \mathcal{P}_d$ satisfy the usual vector space operations as well as the symplectic inner product $\langle \boldsymbol{x} \ , \ \boldsymbol{y} \rangle$, with arithmetic carried out modulo 2.

Recall that the n–qubit displacement operators $D_{\boldsymbol{x}}$ are defined as

$$D_{\boldsymbol{x}} \coloneqq \bigotimes_{k=1}^{n} Z^{x_{k,1}} X^{x_{k,0}}, \tag{2.60}$$

where the phase factor simply equals 1. These displacement operators satisfy the same properties as in the odd dimensional case of Claim 2.4 with minor modifications in global phases, except that they are not closed under transposition as $D_{(1,1)}^T = -D_{(1,1)}$.

Using these displacement operators, we can now construct new phase-point oper-

ators defined as

$$A_{\mathbf{0}} \coloneqq \frac{1}{2^n} \sum_{\boldsymbol{z} \in \mathcal{P}_d} D_{\boldsymbol{z}}; \tag{2.61}$$

$$A_{\boldsymbol{x}} \coloneqq \frac{1}{2^n} \sum_{\boldsymbol{z} \in \mathcal{P}_d} (-1)^{\langle \boldsymbol{x} \, , \, \boldsymbol{z} \rangle} D_{\boldsymbol{z}} \text{ for any } \boldsymbol{x} \in \mathcal{P}_d. \tag{2.62}$$

For single qubit systems, these operators look like

$$\begin{aligned} A_{(0,0)} &= \frac{1}{2}(\mathbb{1} + X + Z + iY), & A_{(0,1)} &= \frac{1}{2}(\mathbb{1} - X + Z - iY), \\ A_{(1,0)} &= \frac{1}{2}(\mathbb{1} + X - Z - iY), & A_{(1,1)} &= \frac{1}{2}(\mathbb{1} - X - Z + iY). \end{aligned} \tag{2.63}$$

Note that these operators are not unitary or Hermitian. Note also that $A_{\mathbf{0}}$ is not equal to the parity operator $\sum_{k=0} \sum_{k \in \mathbb{Z}_2} |k\rangle\langle -k| = \mathbb{1}$ on qubits. We now prove the properties of phase-point operators that survive the transition from odd dimensions to n–qubit systems and are relevant in our construction of a qubit Wigner representation. Notably, the first property, that $A_{\boldsymbol{x}}$ is a real positive operator for qubit systems, replaces the property of Hermiticity and unitarity for odd dimensional systems.

Claim 2.12. *For systems of n qubits, the phase-point operators satisfy*

1. *(positivity).* $A_{\boldsymbol{x}} \geq 0$*;*
2. *(factorization).* $A_{\boldsymbol{x}} = \bigotimes_{i=1}^{n} A_{\boldsymbol{x}_j}$*, for* $\boldsymbol{x} = \bigoplus_{i=1}^{n} \boldsymbol{x}_j$*;*
3. *(unit trace).* $\mathrm{tr}[A_{\boldsymbol{x}}] = 1$*;*
4. *(completeness relation).* $\sum_{\boldsymbol{x} \in \mathcal{P}_d} A_{\boldsymbol{x}} = 2^n \mathbb{1}_n$*;*
5. *(orthogonality).* $\mathrm{tr}[A_{\boldsymbol{x}}^{\dagger} A_{\boldsymbol{y}}] = 2^n \delta_{\boldsymbol{x},\boldsymbol{y}}$*.*

Proof. The factorization property follows because $A_{\mathbf{0} \oplus \mathbf{0}} = A_{\mathbf{0}} \otimes A_{\mathbf{0}}$ according to the definition in Eq. (2.61), so it then follows that $A_{\boldsymbol{x} \oplus \boldsymbol{y}} = A_{\boldsymbol{x}} \otimes A_{\boldsymbol{y}}$ from Eq. (2.62).

$A_{\boldsymbol{x}}$ is a real operator as a consequence of the fact that $D_{\boldsymbol{x}} \in \mathbb{R}$ for all $\boldsymbol{x}$. Moreover, let $\{\alpha_{\boldsymbol{x},i}\}_{i \in \mathbb{Z}_d}$ be the eigenvalues of $A_{\boldsymbol{x}}$. For any $\boldsymbol{x}$, we see that $\alpha_{\boldsymbol{x},i} \in \{0, 1\}$, by direct calculation on the single qubit operators in Eq. (2.63) and the factorization property 2 of $A_{\boldsymbol{x}}$. Therefore, $A_{\boldsymbol{x}}$ is a positive operator for all $\boldsymbol{x}$.

The proofs of the rest of the properties follow the same steps as the proof of Claim 2.5, but we know make use of the definition of displacement operators in Eq. (2.60). □

2.5.2 Wigner representation of qubit states

We have now established that every n–qubit state $\rho \in \mathcal{H}_2^{\otimes n}$ can be decomposed to the new phase-point operators,

$$\rho = \sum_{\boldsymbol{z} \in \mathcal{P}_d} W_\rho(\boldsymbol{z}) A_{\boldsymbol{z}}, \tag{2.64}$$

as the distribution given by the coefficients in the decomposition,

$$W_\rho(\boldsymbol{x}) \coloneqq \frac{1}{2^n} \mathrm{tr}[A^\dagger(\boldsymbol{x})\rho], \tag{2.65}$$

is an informationally complete representation of n–qubit states. This follows from the properties outlined in Claim 2.12. Notice that the definition is in terms of the Hermitian conjugate of the phase-point operators.

The non-Hermiticity of the phase-point operators implies that in general the Wigner representation is a complex function, $W_\rho(\boldsymbol{z}) \in \mathbb{C}$. However, it turns out that the real and imaginary parts of $W_\rho(\boldsymbol{x})$ are related to the quantum state in a meaningful way. To show this, we first prove that useful algebraic properties from Claim 2.6 and Claim 2.7 transfer from odd to even dimensional systems.

Claim 2.13. *The Wigner representation of n–qubit systems satisfies:*

1. *(normalized).* $\sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z}) = 1$;
2. *(bounded).* $W_\rho(\boldsymbol{x}) \leq \frac{1}{2^n}$;
3. *(additive over mixing).* $W_{\sum_i p_i \rho_i}(\boldsymbol{x}) = \sum_i p_i W_{\rho_i}(\boldsymbol{x})$;
4. *(sharp for pure computational states).* $W_{|n\rangle\langle n|}(\boldsymbol{x}) = \frac{1}{2}\delta_{x_0,n}$ *for any* $n \in \mathbb{Z}_2$;
5. *(multiplicative over tensoring).* $W_{\rho_A \otimes \rho_B}(\boldsymbol{x}_A \oplus \boldsymbol{x}_B) = W_{\rho_A}(\boldsymbol{x}_A) W_{\rho_B}(\boldsymbol{x}_B)$.

Proof. The proofs of these statements are a reformulation of results in [102] and follow exactly as in Claim 2.6 and Claim 2.7, with the exception of the boundedness property, which we rederive here to highlight the differences:

Let $\{\lambda_i\}_{i \in \mathbb{Z}_d}$ be the non-negative eigenvalues of ρ, summing to 1. Let $\{\alpha_{\boldsymbol{x},i}\}_{i \in \mathbb{Z}_d}$ be the eigenvalues of $A_{\boldsymbol{x}}$. We have seen in the proof of Claim 2.12 that $\alpha_{\boldsymbol{x},i} \in \{0, 1\}$ for any $\boldsymbol{x} \in \mathcal{P}_d$. Then,

$$|W_\rho(\boldsymbol{x})| = \frac{1}{2^n} |\mathrm{tr}[A_{\boldsymbol{x}}\rho]| \leq \frac{1}{2^n} \left| \sum_i \alpha_{\boldsymbol{x},i} \lambda_i \right| \leq \frac{1}{2^n} \sum_i \lambda_i = \frac{1}{2^n}. \tag{2.66}$$

The first inequality follows because both ρ and $A_{\boldsymbol{x}}$ are positive semi-definite matrices, while the second is the triangle inequality. □

We can now characterize the significance of the real and imaginary parts of $W_\rho(\boldsymbol{x})$. Any n–qubit state ρ can be decomposed as

$$\rho = \left[\frac{1}{2}\left(\rho + \rho^T\right)\right] + i\left[\frac{-i}{2}\left(\rho - \rho^T\right)\right], \tag{2.67}$$

where the transposition is taken with respect to the computational basis. Because $\rho^* = \rho^T$ in any basis, we can identify

$$\rho^{(0)} \coloneqq \frac{1}{2}(\rho + \rho^T) = \mathrm{Re}[\rho] \text{ and } \rho^{(1)} \coloneqq i\left[\frac{-i}{2}(\rho - \rho^T)\right] = \mathrm{Im}[\rho], \tag{2.68}$$

i.e. $\rho^{(0)}$ and $\rho^{(1)}$ are respectively the real and imaginary components of the density matrix of ρ in the computational basis.

We will first prove Claim 2.14, which establishes a direct correspondence between the real and imaginary components of a state's Wigner representation to the real and imaginary components of its density matrix in the computational basis.

Claim 2.14. *Given $\rho \in \mathcal{H}_2^{\otimes n}$, we have that*

$$\mathrm{Re}[W_\rho(\boldsymbol{x})] = W_{\mathrm{Re}(\rho)}(\boldsymbol{x}) \tag{2.69}$$
$$\mathrm{Im}[W_\rho(\boldsymbol{x})] = W_{\mathrm{Im}(\rho)}(\boldsymbol{x}) \tag{2.70}$$

for all $\boldsymbol{x} \in \mathcal{P}_d$ with $d = 2^n$ and where the functions $\mathrm{Re}[\rho]$ and $\mathrm{Im}[\rho]$ are the real and imaginary parts of the density matrix ρ respectively when expressed in the computational basis.

Proof. Using the identification $\rho^{(0)} = \mathrm{Re}[\rho]$ and $\rho^{(1)} = \mathrm{Im}[\rho]$, we can then decompose $W_\rho(\boldsymbol{x})$ as

$$W_\rho(\boldsymbol{x}) = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^\dagger \rho^{(0)}\right] + i\frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^\dagger \rho^{(1)}\right], \tag{2.71}$$

which follows from the fact that $A_{\boldsymbol{x}}$ is always real. Since $\rho^{(0)}$ and $\rho^{(1)}$ are themselves real by their definition in Eq. (2.68), we conclude that $\mathrm{tr}[A_{\boldsymbol{x}}^\dagger \rho^{(0)}]$ and $\mathrm{tr}[A_{\boldsymbol{x}}^\dagger \rho^{(1)}]$ are both real for all $\boldsymbol{x} \in \mathcal{P}_d$. Therefore,

$$\mathrm{Re}[W_\rho(\boldsymbol{x})] = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^\dagger \rho^{(0)}\right] = W_{\rho^{(0)}}(\boldsymbol{x}) \tag{2.72}$$
$$\mathrm{Im}[W_\rho(\boldsymbol{x})] = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^\dagger \rho^{(1)}\right] = W_{\rho^{(1)}}(\boldsymbol{x}) \tag{2.73}$$

□

An n–rebit Wigner representation $W_\rho^{(0)}$ was introduced in [98], and is defined as

$$W_\rho^{(0)}(\boldsymbol{x}) \coloneqq \frac{1}{2^n}\mathrm{tr}[A_{\boldsymbol{x}}^{(0)}\rho], \tag{2.74}$$

for all $\boldsymbol{x} \in \mathcal{P}_d$, where

$$A_{\boldsymbol{x}}^{(0)} \coloneqq \frac{1}{2^n}\sum_{\boldsymbol{z}\in\mathcal{P}_d^{(0)}} (-1)^{\langle \boldsymbol{x}\,,\,\boldsymbol{z}\rangle} D_{\boldsymbol{z}} \tag{2.75}$$

is defined in terms of the subspace $\mathcal{P}_d^{(0)} \coloneqq \{\boldsymbol{x} \,:\, \sum_{k=1}^n x_{k,0}\cdot x_{k,1} = 0\}$.

By comparison between Eq. (2.65) and Eq. (2.74), we see that the difference between $W_\rho(\boldsymbol{x})$ and $W_\rho^{(0)}(\boldsymbol{x})$ comes down to the fact that $A_{\boldsymbol{x}}$ sums over all displacement operators defined over the phase space $\mathcal{P}_d$, whereas $A_{\boldsymbol{x}}^{(0)}$ only sums over displacement operators defined over the subspace $\mathcal{P}_d^{(0)}$.

By Eq. (2.60), we see that the condition $\boldsymbol{a}_x \cdot \boldsymbol{a}_z = 0$ implies $D_{\boldsymbol{x}}^\dagger = D_{\boldsymbol{x}}^T = D_{\boldsymbol{x}}$, i.e. that $D_{\boldsymbol{x}}$ is closed under transposition and, in particular, real symmetric, from which $A_{\boldsymbol{x}}^{(0)\dagger} = A_{\boldsymbol{x}}^{(0)T} = A_{\boldsymbol{x}}^{(0)}$, i.e. that $A_{\boldsymbol{x}}^{(0)}$ is also real symmetric, immediately follows. Each $D_{\boldsymbol{x}} \in \mathcal{P}_d^{(0)}$ lies inside the span of n–rebit states, since that is the vector space of all real $2^n \times 2^n$ symmetric matrices.

It will be helpful to introduce the complement of $\mathcal{P}_d^{(0)}$ in $\mathcal{P}_d$,

$$\mathcal{P}_d^{(1)} \coloneqq \{\boldsymbol{x} \,:\, \sum_{k=1}^n x_{k,0}\cdot x_{k,1} = 1\}, \tag{2.76}$$

and the set of real anti-symmetric phase-point operators

$$A_{\boldsymbol{x}}^{(1)} \coloneqq \frac{1}{2^n}\sum_{\boldsymbol{z}\in\mathcal{P}_d^{(1)}} (-1)^{\langle \boldsymbol{x}\,,\,\boldsymbol{z}\rangle} D_{\boldsymbol{z}} \text{ for any } \boldsymbol{x} \in \mathcal{P}_d. \tag{2.77}$$

The anti-symmetry, $A_{\boldsymbol{x}}^{(1)\dagger} = A_{\boldsymbol{x}}^{(1)T} = -A_{\boldsymbol{x}}$, follows from the anti-symmetry of the displacement operators $D_{\boldsymbol{x}}^\dagger = D_{\boldsymbol{x}}^T = -D_{\boldsymbol{x}}$ for $\boldsymbol{x} \in \mathcal{P}_d^{(1)}$.

We have seen that every displacement operator is either symmetric and anti-symmetric. The fact that $A_{\boldsymbol{x}}^{(0)}$ only sums over symmetric displacement operators defined on $\mathcal{P}_d$ is the reason why, unlike $A_{\boldsymbol{x}}$, it fails to be locally tomographic. For example, while $A_{(0,0)} = \frac{1}{2}(\mathbb{1} + X + Z + iY)$, $A_{(0,0)}^{(0)} = \frac{1}{2}(\mathbb{1} + X + Z)$. However, some global symmetric displacement operators are formed as a tensor product of anti-symmetric local displacement operators. For instance, $A_{(0,0)\oplus(0,0)} = A_{(0,0)} \otimes A_{(0,0)}$ sums over the global two-qubit symmetric displacement operator $(iY)\otimes(iY)$, which is formed from the anti-symmetric local displacement operator (iY) in each $A_{(0,0)}$. Because $A_{(0,0)\oplus(0,0)}^{(0)}$ also sums over $(iY)\otimes(iY)$, yet $A_{(0,0)}^{(0)}$ does not contain an (iY) component, it is not locally tomographic.

By Eq. (2.75) and Eq. (2.77), we see that each $A_{\boldsymbol{x}}$ splits up as

$$A_{\boldsymbol{x}} = A_{\boldsymbol{x}}^{(0)} + A_{\boldsymbol{x}}^{(1)} \tag{2.78}$$

We can correspondingly split up the Wigner representation of ρ as

$$W_\rho(\boldsymbol{x}) = \frac{1}{2^n}\mathrm{tr}\left[\left(A_{\boldsymbol{x}}^{(0)} + A_{\boldsymbol{x}}^{(1)}\right)^\dagger \rho\right] = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(0)\dagger}\rho\right] + \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(1)\dagger}\rho\right], \tag{2.79}$$

where

$$W_\rho^{(0)}(\boldsymbol{x}) \coloneqq \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(0)\dagger}\rho\right] = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(0)}\rho\right] \text{ and} \tag{2.80}$$

$$W_\rho^{(1)}(\boldsymbol{x}) \coloneqq \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(1)\dagger}\rho\right] = -\frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(1)}\rho\right]. \tag{2.81}$$

We can now prove the following result.

Claim 2.15. *Given any n–qubit state ρ, we have that*

$$W_\rho^{(0)}(\boldsymbol{x}) = \mathrm{Re}[W_\rho(\boldsymbol{x})] \text{ and} \tag{2.82}$$

$$W_\rho^{(1)}(\boldsymbol{x}) = i\mathrm{Im}[W_\rho(\boldsymbol{x})]. \tag{2.83}$$

Proof. For $k = 0, 1$,

$$\left[W_\rho^{(k)}(\boldsymbol{x})\right]^* = \frac{1}{2^n}\mathrm{tr}\left[A_{\boldsymbol{x}}^{(k)\dagger}\rho\right]^* \tag{2.84}$$

$$= \frac{1}{2^n}\mathrm{tr}\left[(A_{\boldsymbol{x}}^{(k)\dagger}\rho)^\dagger\right] \tag{2.85}$$

$$= \frac{1}{2^n}\mathrm{tr}\left[(-1)^k(A_{\boldsymbol{x}}^{(k)}\rho)\right] = (-1)^k W_\rho^{(k)}(\boldsymbol{x}), \tag{2.86}$$

which implies that $W_\rho^{(0)}(\boldsymbol{x}) = \mathrm{Re}[W_\rho(\boldsymbol{x})]$ is the real component, while $W_\rho^{(1)}(\boldsymbol{x}) = i\mathrm{Im}[W_\rho(\boldsymbol{x})]$. □

By combining Claim 2.14 and Claim 2.15, we arrive at

$$W_{\boldsymbol{x}}^{(0)}(\rho) = W_{\mathrm{Re}[\rho]}(\boldsymbol{x}). \tag{2.87}$$

When ρ is an n–rebit state, it holds that $\mathrm{Re}[\rho] = \rho$, so $W_\rho(\boldsymbol{x}) = W_{\boldsymbol{x}}^{(0)}(\rho)$.

In summary, the representation $W_\rho(\boldsymbol{x})$ of an n–qubit state ρ is a real quasi-distribution if and only if ρ is a *rebit state*, namely its density matrix contains only real elements when expressed in the computational basis. Examples of such states include the eigenstates of Pauli operators X and Z as well as the magic state $|H\rangle$, which is the $(+1)$–eigenstate of the qubit Hadamard operator,

$$|H\rangle \coloneqq \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle. \tag{2.88}$$

This state is equivalent to the canonical qubit magic state $|A\rangle \coloneqq \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{4}}|1\rangle$ up to a Clifford unitary [113], and thus can be used in a gadgetization circuit [72] to implement the T–gate $T \coloneqq \mathrm{diag}(1, e^{i\pi/4})$.

2.5.3 Wigner representation of qubit channels

As is the case in odd dimensions, the Wigner representation of qubit states induces a corresponding Wigner representation of qubit channels. Let $\mathcal{E}$ be an arbitrary channel from n to m qubits, and $\mathcal{J}(\mathcal{E}) = (\mathbb{1} \otimes \mathcal{E})(|\phi_n^+\rangle\langle\phi_n^+|)$ be its associated Choi-Jamiołkowski state [114], where $|\phi_n^+\rangle$ is the canonical maximally entangled state on two copies of the input system,

$$\left|\phi_n^+\right\rangle := \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}\in\{0,1\}^n} |\boldsymbol{k}\rangle \otimes |\boldsymbol{k}\rangle \right). \tag{2.89}$$

We can now define a Wigner representation [93] for a quantum channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_2^{\otimes n}) \mapsto \mathcal{B}(\mathcal{H}_2^{\otimes m})$ as

$$W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{x}) := 2^{2n} W_{\mathcal{J}(\mathcal{E})}(\boldsymbol{x} \oplus \boldsymbol{y}). \tag{2.90}$$

We now outline important properties of this representation.

Claim 2.16. *The Wigner representation of a* CPTP *operation* $\mathcal{E} : \mathcal{B}(\mathcal{H}_2^{\otimes n}) \mapsto \mathcal{B}(\mathcal{H}_2^{\otimes m})$ *is*

1. *(normalized).* $\sum_{\boldsymbol{z}\in\mathcal{P}_{2^m}} W_{\mathcal{E}}(\boldsymbol{z}|\boldsymbol{x}) = 1$ *for any* $\boldsymbol{x} \in \mathcal{P}_{2^n}$*;*
2. *(bounded).* $|W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{x})| \leq 2^{n-m}$*;*
3. *(transitive).* $W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_{2^n}} W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{z}) W_{\rho}(\boldsymbol{z})$ *for any* $\boldsymbol{y} \in \mathcal{P}_{2^m}$*;*
4. *(distributive over tensoring).* $W_{\mathcal{E}\otimes\mathcal{F}}(\boldsymbol{x} \otimes \boldsymbol{y}) = W_{\mathcal{E}}(\boldsymbol{x}) W_{\mathcal{F}}(\boldsymbol{y})$.

Proof. The properties of this Wigner representation were also considered in [102].

The proof makes use of the definition in Eq. (2.90) and the properties of the qubit state representation listed in Claim 2.13 and proceeds analogously to the odd dimensional case of Claim 2.8. □

The transitive property 3 and the tensoring property 4 ensure that the chosen representation respects sequential and parallel composition of processes, i.e.,

$$W_{\mathcal{E}\circ\mathcal{F}} = W_{\mathcal{E}} W_{\mathcal{F}}, \tag{2.91}$$

$$W_{\mathcal{E}\otimes\mathcal{F}} = W_{\mathcal{E}} \otimes W_{\mathcal{F}}. \tag{2.92}$$

It therefore follows that a quantum channel $\mathcal{E}$ from n qubits to m qubits is represented as a stochastic map if and only if $W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}, \boldsymbol{y}$. By inspection of

Eq. (2.90), we equivalently have that the quantum channel $\mathcal{E}$ is stochastically represented if and only if the Choi state $\mathcal{J}(\mathcal{E})$ on $n+m$ qubits is represented by a genuine probability distribution on the phase space $\mathcal{P}_{2^{n+m}}$.

We now proceed to characterize a subset of stabilizer states and Clifford operations that can be represented non-negatively by the qubit representation we have developed.

2.5.4 The CSS formalism on the phase space

Unlike the odd-dimensional case, not all stabilizer states and Clifford operations are represented non-negatively. For instance, the $(+1)$–eigenstate $|+i\rangle\langle+i|$ of Pauli operator Y is a stabilizer state, but has a complex representation as it is not a rebit. However, in this section, we describe an important subset of stabilizer states, called CSS states [70, 71], which are also a subset of rebits, therefore they admit a real, non-negative Wigner representation. Moreover, the class of quantum channel $\mathcal{E}$ with CSS $\mathcal{J}(\mathcal{E})$ is stochastically represented.

In this section, we use the notation $\langle \dots \rangle$ to denote the set of stabilizer generators for some state $|\psi\rangle$. The stabilizer generators generate the group of Clifford gates for which state $|\psi\rangle$ is a $(+1)$–eigenvalue. In Hilbert space $\mathcal{H}_2^{\otimes n}$, a set of n independent stabilizer generators define precisely one state [42].

A pure CSS state on n qubits is any stabilizer state whose stabilizer group can be generated by n Pauli observables that are individually of X–type or Z–type only. For instance, $|\phi^+\rangle \coloneqq \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ has the stabilizer group

$$\mathcal{S}(|\phi^+\rangle) = \langle X_1 X_2, Z_1 Z_2 \rangle, \tag{2.93}$$

and is therefore CSS, while $|\psi\rangle \coloneqq \mathbb{1} \otimes H\,|\phi^+\rangle$ is stabilized by

$$\mathcal{S}(|\psi\rangle) = \langle X_1 Z_2, Z_1 X_2 \rangle, \tag{2.94}$$

and therefore is *not* CSS, due to the stabilizer generators mixing X and Z operators.

It is shown in [98] that any pure n–rebit state is non-negatively represented if and only if it is CSS. Therefore, W_ρ is a valid probability distribution for all $\rho \in \mathrm{CSS}$, and we can extend this idea to quantum channels. A quantum channel $\mathcal{E}$ is CSS–preserving if $\mathcal{E}(\rho)$ is a CSS state for all $\rho \in \mathrm{CSS}$. It is completely CSS–preserving if $(\mathbb{1} \otimes \mathcal{E})(\rho_{AB})$ is CSS for any bipartite CSS state ρ_{AB}, regardless of the dimension of system A. We therefore reach the following result.

Claim 2.17. *A quantum channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_2^{\otimes n}) \mapsto \mathcal{B}(\mathcal{H}_2^{\otimes n})$ admits a stochastic representation $W_\mathcal{E}$ if $\mathcal{J}(\mathcal{E}) \in \mathrm{CSS}_{2n}$.*

Proof. $|\phi_n^+\rangle$ is a CSS state. For $i = 1, 2, \ldots, n$, we have that

$$Z_i Z_{n+i} \left|\phi_n^+\right\rangle = \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}\in\{0,1\}^n} Z_i \left|\boldsymbol{k}\right\rangle \otimes Z_i \left|\boldsymbol{k}\right\rangle \right)$$
$$= \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}\in\{0,1\}^n} (-1)^{k_i} \left|\boldsymbol{k}\right\rangle \otimes (-1)^{k_i} \left|\boldsymbol{k}\right\rangle \right) = \left|\phi_n^+\right\rangle . \tag{2.95}$$

Therefore, $|\phi_n^+\rangle$ is stabilized by $Z_i Z_{n+i}$ for $i = 1, 2, \ldots, n$. Furthermore, we have that $X_i \left|\boldsymbol{k}\right\rangle = \left|\boldsymbol{k}'\right\rangle$, where $\boldsymbol{k}'$ is identical to $\boldsymbol{k}$ except that its i–th bit has been flipped. Since the set of n–bit binary strings and the set of n–bit binary strings whose i–th bit has been flipped are identical, we conclude that

$$X_i X_{n+i} \left|\phi_n^+\right\rangle = \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}\in\{0,1\}^n} X_i \left|\boldsymbol{k}\right\rangle \otimes X_i \left|\boldsymbol{k}\right\rangle \right) = \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}\in\{0,1\}^n} \left|\boldsymbol{k}'\right\rangle \otimes \left|\boldsymbol{k}'\right\rangle \right)$$
$$= \frac{1}{\sqrt{2n}} \left(\sum_{\boldsymbol{k}'\in\{0,1\}^n} \left|\boldsymbol{k}'\right\rangle \otimes \left|\boldsymbol{k}'\right\rangle \right) = \left|\phi_n^+\right\rangle . \tag{2.96}$$

Therefore, $|\phi_n^+\rangle$ is stabilized by $X_i X_{n+i}$ for $i = 1, 2, \ldots, n$. As $Z_i Z_{n+i}$ and $X_j X_{n+j}$ commute for all $i, j = 1, 2, \ldots, n$, we have now found $2n$ commuting and independent stabilizers for $|\phi_n^+\rangle$. We therefore conclude that the stabilizer group of $|\phi_n^+\rangle$ is

$$\langle Z_i Z_{n+i}, X_i X_{n+i} \rangle_{i=1,\ldots,n}. \tag{2.97}$$

The proof is concluded by noticing that $W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{x})$ is non-negative due to its definition Eq. (2.90) and normalized, therefore stochastic. □

Claim 2.17 can be leveraged to identify stochastically-represented qubit Clifford operations in a systematic way. We recall that the maximally entangled state $|\phi_n^+\rangle$ over two sets of n qubits is CSS for all n. Therefore, if $\mathcal{E}$ is completely CSS–preserving, $\mathcal{J}(\mathcal{E})$ must also be a CSS state. By Claim 2.17, it follows that every completely CSS-preserving operation is stochastically represented.

To motivate the class of completely CSS–preserving operations as operationally significant, we highlight that they cover at least the following important stabilizer operations:

1. Preparation of a CSS state;

2. Any gate from the group of n–qubit completely CSS–preserving gates,

$$\mathcal{G}(n) \coloneqq \langle \mathrm{CNOT}_{i,j}, Z_i, X_i \rangle_{i,j=1,\ldots,n, i\neq j}; \tag{2.98}$$

3. Projective measurement of any X– or Z–type Pauli observable, followed by a completely CSS–preserving operation $\mathcal{E}_{\pm}$ conditioned on the outcome ± 1;

4. Tracing out;

5. Statistical mixtures of the above.

Therefore, all these operations are stochastically represented and we will prove this after making these operations precise in chapter 5.

We conclude this section by emphasizing the power of the set of CSS operations. Firstly, they can be promoted to universal quantum computing when supplemented with rebit magic states [98]. Moreover, the gate set $\mathcal{G}(n)$ constitute those that can be implemented fault tolerantly using defect braiding in surface codes constructions [68]. Finally, we will see in chapter 5 that they form the basis of many important existing protocols for magic state distillation based on CSS codes before proceeding to construct bounds on magic distillation via completely CSS–preserving operations, namely *CSS protocols*.

In this chapter, we have identified that stabilizer states and CSS states are positively represented on discrete phase spaces of odd and even dimensions respectively. Crucially, Claim 2.10 and Claim 2.17 respectively prove that Clifford operations in odd dimensions and CSS-preserving operations in even dimensions admit representations that act stochastically on Wigner state distributions. In particular, the transitivity property

$$W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_d} W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{z}) W_{\rho}(\boldsymbol{z}), \tag{2.99}$$

derived for odd dimensions in Claim 2.8 and for even dimensions in Claim 2.16 sets the scene for the development of a statistical mechanical framework that describes quasi-distributions on the phase space, as we imminently discuss in chapter 3.

Chapter 3

Majorization for quasi-distributions

Majorization [115, 116] is a collection of powerful tools that has found many applications in quantum information theory [117–124]. It describes the disorder of distributions that undergo stochastic transformations, and in its simplest form defines a pre-order on probability distributions. Therefore, it forms the basis for defining entropies and relative entropies between states of a system [125].

The development of majorization theory and all known applications thus far concern ordering of proper probability distributions. The aim of this chapter is to rigorously extend majorization to quasi-distributions by identifying that stochastic transformations are not restricted to proper probability distributions. This extension constitutes a technical novelty in the majorization literature. We are motivated by the phase space representation of quantum mechanics and more specifically by the Wigner representation of magic theories, although our majorization extension can find applications in other areas of physics, where negativity indicates non-classicality.

We begin this chapter in section 3.1 by providing a series of known definitions of majorization on probability distributions before motivating the extension of these definitions to the context of magic. We have seen that states are generally represented by quasi-distributions and free operations are represented stochastically, thus we naturally obtain the stochastic update of distributions that contain negativity,

$$W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_d} W_{\mathcal{E}}(\boldsymbol{y}|\boldsymbol{z}) W_{\rho}(\boldsymbol{z}). \tag{3.1}$$

This was proven as a transitivity property of quantum channel representations on the phase space in Claim 2.8 for odd dimensions and in Claim 2.16 for even dimensions. In section 3.1, we argue that Eq. (3.1) allows us to define majorization on Wigner distributions as long as $W_{\mathcal{E}}$ is stochastic.

The rest of the chapter concerns itself with rigorously extending important majorization properties from the subspace of proper probability distributions to the

quasi-distribution setting. We begin in section 3.2 by defining and extending the idea of a Lorenz curve, a visual method of representing the majorization pre-order. We then deal in section 3.3 with the most general form of majorization that we will later encounter in our discussion of magic in chapter 4 and chapter 5. Finally, in section 3.4, we show that relative majorization leads to a subset of well-behaved entropies and relative entropies on quasi-distributions. In the following chapters, we derive bounds on the resource cost of magic in terms of these entropic quantities.

3.1 Motivation in the context of magic

3.1.1 Definition of majorization

We start by presenting relevant definitions of majorization in order of increasing complexity. These definitions have been motivated and developed in various fields of quantum information, where describing order between non-negative probability distributions has been of interest.

The simplest definition of majorization is critical in describing state transformations in the context of bipartite entanglement [117]. Given two distributions $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$ over n outcomes, we say that $\boldsymbol{p}$ majorizes $\boldsymbol{q}$, denoted $\boldsymbol{p} \succ \boldsymbol{q}$, if there exists a bistochastic map $A = (A_{ij})$ such that $A\boldsymbol{p} = \boldsymbol{q}$, where bistochastic means that $A_{ij} \geq 0$ and $\sum_i A_{ij} = \sum_j A_{ij} = 1$. Later on, we show the known result [115] that the condition $\boldsymbol{p} \succ \boldsymbol{q}$ over probability distributions is equivalent to at most $n - 1$ inequalities (Claim 3.10), which can be checked efficiently.

There is a natural generalization, which is called $\boldsymbol{d}$–majorization [126], or in the context of thermodynamics, thermo-majorization [127]. For a fixed probability distribution $\boldsymbol{r} = (r_1, \ldots, r_n)$ with positive components, we define majorization relative to $\boldsymbol{r}$ as $\boldsymbol{p} \succ_{\boldsymbol{r}} \boldsymbol{q}$, if and only if there exists a stochastic map A such that $A\boldsymbol{r} = \boldsymbol{r}$ and $A\boldsymbol{p} = \boldsymbol{q}$. The original majorization condition between probability distributions corresponds to the case $\boldsymbol{r} = (1/n, \ldots, 1/n)$.

In fact, we can further generalize to relative majorization [116, 128–132], defined as an ordering between pairs of vectors and write

$$(\boldsymbol{p}, \boldsymbol{r}) \succ (\boldsymbol{q}, \boldsymbol{r}') \tag{3.2}$$

if and only if there is a stochastic map A such that $A\boldsymbol{r} = \boldsymbol{r}'$ and $A\boldsymbol{p} = \boldsymbol{q}$. We retrieve $\boldsymbol{d}$–majorization when $\boldsymbol{r} = \boldsymbol{r}'$.

We now proceed to motivate the use of majorization within magic theories, where proper probability distributions need to be replaced with quasi-distributions. How-

ever, as we see, the core aspect of majorization, i.e. the stochastic update of the distributions, remains intact.

3.1.2 Majorization in resource theories of magic

Within the Wigner representation for the rebit model of qubit computation as well as for computation with odd dimensional qudits, it is well-known that all positively represented states used in Clifford circuits admit an efficient classical simulation [98, 103], so negativity is a necessary resource for universal fault-tolerant quantum computing [41]. Hence, the free states in any magic theory are required to be a subset of

$$\mathcal{F} \coloneqq \{\rho : W_\rho(\boldsymbol{z}) \geq 0 \text{ for all } \boldsymbol{z} \in \mathcal{P}_d\}. \tag{3.3}$$

Our focus is on states with negativity, so the particular choice of free states is not critical for our analysis. The remaining component that defines any magic theory is the set of free quantum operations. The most basic assumption we require on free operations is that they send any free state to another free state.

Any magic state protocol will correspond to a quantum channel $\mathcal{E}$, so from Eq. (2.37) and Eq. (2.90) it admits a Wigner representation $W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{z})$ that acts as a transition matrix mapping phase space points $\boldsymbol{z} \to \boldsymbol{y}$. We have shown that the representation obeys the update property

$$W_{\mathcal{E}(\rho)}(\boldsymbol{y}) = \sum_{\boldsymbol{z}\in\mathcal{P}_d} W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{z}) W_\rho(\boldsymbol{z}), \tag{3.4}$$

for any ρ. Since the magic protocol sends free states to free states, $\mathcal{E}$ sends all positively represented quantum state to other positively represented quantum states. Therefore, if $\mathcal{E}$ is a free operation, its associated Wigner representation $W_\mathcal{E}(\boldsymbol{y}|\boldsymbol{z})$ must be a stochastic matrix. For example, we saw in section 2.4.4 all odd dimensional Clifford operations correspond to the symplectic transformations on the phase space, hence stochastic matrices in the Wigner representation.

We note however, that not all stochastic maps on the phase space correspond to valid quantum operations. The reason is that the maps must also respect the symplectic affine structure of the phase space, which is an additional non-trivial constraint.

In the context of magic distillation, we shall often assume that we have a magic theory $\mathcal{R} = (\mathcal{F}, \mathcal{O})$ in which the free states $\mathcal{F}$ admit non-negative Wigner representations, while the free operations $\mathcal{O}$ are stochastic maps in the Wigner representation. Our analysis applies majorization to magic states at the level of the associated

Wigner representations. The connection is clear when one rewrites Eq. (3.1) as

$$W_{\rho'} = W_{\mathcal{E}} W_{\rho}, \tag{3.5}$$

where $\rho' \coloneqq \mathcal{E}(\rho)$, and adds the constraint $\tau' = \mathcal{E}(\tau)$, for some chosen reference state $\tau \in \mathcal{F}$, implying that additionally

$$W_{\tau'} = W_{\mathcal{E}} W_{\tau}. \tag{3.6}$$

These two conditions in tandem define d–majorization $(W_\rho, W_\tau) \succ (W_{\rho'}, W_{\tau'})$. Their physical meaning in the context of magic is discussed at length in chapter 4 and chapter 5.

Since Wigner representations are in general quasi-distributions, it is important to first check how majorization can be extended to these cases and what differences quasi-distributions bring over genuine probability distributions. This is the topic of this chapter.

3.1.3 Technical aspects of Wigner quasi-distributions

Having identified that our quasi-distribution of interest is the Wigner state distribution, we derive some novel technical properties of these distributions that will later assist us in the definition of the Lorenz curve and explicit calculation of inequalities arising from majorization. The core idea of this section's contributions is to provide easier handling of the Wigner representation of a many-copy state. This is important later on, as magic state distillation is usually concerned with sending many copies of a noisy magic state to fewer copies of a less noisy magic state

We first define the *rescaled Wigner representation*

$$W_{\rho|\tau}(\boldsymbol{z}) \coloneqq \frac{W_\rho(\boldsymbol{z})}{W_\tau(\boldsymbol{z})}. \tag{3.7}$$

This quantity is required for the definition of a Lorenz curve in section 3.2.1 within the context of the Wigner representation. This distribution is well-defined if τ is a full-rank stabilizer state, or, equivalently, W_τ being a proper probability distribution with positive components. We can show that W_τ is multiplicative on tensor products of states.

Claim 3.1. *Let τ_A, τ_B be full rank stabilizer states on systems A and B, and let ρ_A, ρ_B be arbitrary states on A, B. Then, the rescaled quasi-distribution obeys*

$$W_{\rho_A \otimes \rho_B | \tau_A \otimes \tau_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B) = W_{\rho_A|\tau_A}(\boldsymbol{z}_A) W_{\rho_B|\tau_B}(\boldsymbol{z}_B). \tag{3.8}$$

Proof. This follows from the multiplicativity of the Wigner representation,

$$W_{\rho_A\otimes\rho_B|\tau_A\otimes\tau_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B) = \frac{W_{\rho_A\otimes\rho_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B)}{W_{\tau_A\otimes\tau_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B)}$$
$$= \frac{W_{\rho_A}(\boldsymbol{z}_A)W_{\rho_B}(\boldsymbol{z}_B)}{W_{\tau_A}(\boldsymbol{z}_A)W_{\tau_B}(\boldsymbol{z}_B)} = W_{\rho_A|\tau_A}(\boldsymbol{z}_A)W_{\rho_B|\tau_B}(\boldsymbol{z}_B). \tag{3.9}$$

□

We now proceed to discuss component-multiplicity pairs, which constitute a useful description of Wigner representations of many-copy states.

In general, a 1–copy d–dimensional state ρ is described by its d^2–dimensional Wigner representation W_ρ. The distribution W_ρ is defined on the phase space, but it can be convenient to re-express this using vector notation. We discuss this in terms of Wigner representations, but there is nothing to prevent the discussion from applying to rescaled Wigner representations as well.

To each Wigner representation $W_\rho(\boldsymbol{z})$ we can associate component-multiplicity pairs $\{(w_i, m_i)\}$ where the value w_i occurs in the distribution $W_\rho(\boldsymbol{z})$ with multiplicity m_i.

As an example, for the Strange state $|S\rangle$ with

$$W_{|S\rangle} = (-1/3, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6) \tag{3.10}$$

we have the component-multiplicity pairs: $\{(-1/3, 1), (1/6, 8)\}$. However, we might also wish more freedom and not require that the different w_i values are all distinct. For example, the component-multiplity pairs $\{(-1/3, 1), (1/6, 2), (1/6, 3), (1/6, 3)\}$ also describe $W_{|S\rangle}$.

This representation is more compact when a Wigner representation has a lot of multiplicities, and allows for simple handling of multiple copies via the following fact.

Claim 3.2. *Consider two Wigner representations $W_{\rho_A}(\boldsymbol{z}_A), W_{\rho_B}(\boldsymbol{z}_B)$ with component-multiplicity pairs*

$$\{(w_i, m_i)\} \text{ and } \{(w'_j, m'_j)\}, \tag{3.11}$$

respectively. Then, $\{(w_i w'_j, m_i m'_j)\}$ gives component-multiplicity pairs for $W_{\rho_A\otimes\rho_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B)$.

Proof. This result is true because all components of $W_{\rho_A\otimes\rho_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B)$ are of the form $w_i w'_j$ and

$$\sum_i \sum_j m_i m'_j = \sum_i m_i \sum_j m'_j = d_A^2 d_B^2, \tag{3.12}$$

where d_A, d_B are the dimensions of ρ_A, ρ_B respectively, and so the set $\{(w_i w'_j, m_i m'_j)\}$ contains exactly the Wigner components of $W_{\rho_A\otimes\rho_B}(\boldsymbol{z}_A \oplus \boldsymbol{z}_B)$. □

The above result also applies when we consider rescaled Wigner representations, crucially due to the multiplicativity result shown in Claim 3.1.

In the case of distillation protocols, we are interested in copies of distributions. For this, we have the following result, following from combinatorics, which allows us to simply characterize the Wigner representation of a multi-copy state in terms of the representations of the individual states.

Claim 3.3. *Suppose W_ρ has a set of D component-multiplicity pairs $\{(w_i, m_i)\}$. Then, $W_{\rho^{\otimes n}}$ has component-multiplicity pairs $\{(W_{\boldsymbol{q}}, M_{\boldsymbol{q}})\}$, with index $\boldsymbol{q}$ running through all vectors $(q_1, \ldots, q_D)$, where $q_1, \ldots, q_D$ are non-negative integers that sum to n, and*

$$W_{\boldsymbol{q}} = \prod_{i=1}^{D} w_i^{q_i}, \tag{3.13}$$

$$M_{\boldsymbol{q}} = \binom{n}{q_1, q_2, \ldots, q_d} \prod_{i=1}^{D} m_i^{q_i}. \tag{3.14}$$

The term outside the product in the expression for $M_{\boldsymbol{q}}$ is the generalized binomial coefficient,

$$\binom{n}{q_1, q_2, \ldots, q_d} \coloneqq \frac{n!}{q_1! \ldots q_D!}. \tag{3.15}$$

Proof. Denote by $C_D^n \coloneqq \{\boldsymbol{k}\}$ the set of all vectors $\boldsymbol{k} \coloneqq (k_1, \ldots, k_D)$ with non-negative integer components that sum to n, i.e.

$$0 \leq k_1, \ldots, k_D \leq n \text{ and } k_1 + \cdots + k_D = n.$$

We proceed by induction. Assume $n = 1$ and let $\boldsymbol{k}_i$ be the vector with its i-th component equal to 1 and 0's elsewhere. The set C_D^1 consists of all vectors of this form, i.e.

$$C_D^1 = \{\boldsymbol{k}_i\}_{i=1,\ldots,D} \tag{3.16}$$

It is also true by direct calculation that

$$(W_{\boldsymbol{k}_i}, M_{\boldsymbol{k}_i}) = (w_i, m_i). \tag{3.17}$$

Therefore, $\{(W_{\boldsymbol{k}}, M_{\boldsymbol{k}})\}_{\boldsymbol{k} \in C_D^1}$ is a complete set of component-multiplicity pairs for W_ρ.

Assume that $\{(W_{\boldsymbol{k}}, M_{\boldsymbol{k}})\}_{\boldsymbol{k} \in C_D^n}$ as given in Eq. (3.13) and Eq. (3.14) is a complete set of component-multiplicity pairs for the n–copy distribution $W_{\rho^{\otimes n}} = W_\rho^{\otimes n}$. By construction, $W_\rho^{\otimes (n+1)} = W_\rho^{\otimes n} \otimes W_\rho$, so it admits the complete set of component multiplicity pairs

$$\{(W_{\boldsymbol{k}} w_i, M_{\boldsymbol{k}} m_i)\},\ \boldsymbol{k} \in C_D^n \text{ and } i = 1, \ldots, D. \tag{3.18}$$

Consider the component sum of distribution $W_\rho^{\otimes(n+1)}$,

$$\sum_{\boldsymbol{k}\in C_D^n}\sum_{i=1}^{D} M_{\boldsymbol{k}} m_i W_{\boldsymbol{k}} w_i = \sum_{\boldsymbol{k}\in C_D^n} M_{\boldsymbol{k}} W_{\boldsymbol{k}} \sum_{i=1}^{D} m_i w_i =$$
$$\sum_{\boldsymbol{k}\in C_D^n} \frac{n!}{k_1!\dots k_D!} \prod_{i=1}^{D} {m_i}^{k_i} {w_i}^{k_i} \sum_{i=1}^{D} m_i w_i =$$
$$\left(\sum_{i=1}^{D} m_i w_i\right)^n \left(\sum_{i=1}^{D} m_i w_i\right) = \left(\sum_{i=1}^{D} m_i w_i\right)^{n+1} =$$
$$\sum_{\boldsymbol{q}\in C_D^{n+1}} M_{\boldsymbol{q}} W_{\boldsymbol{q}}, \tag{3.19}$$

where in the last expression, vectors $\boldsymbol{q} = (q_1, \dots, q_D)$ have non-negative integer components that sum to $(n+1)$ and

$$M_{\boldsymbol{q}} = \frac{(n+1)!}{q_1!\dots q_D!} \prod_{i=1}^{D} {m_i}^{q_i}, \tag{3.20}$$

$$W_{\boldsymbol{q}} = \prod_{i=1}^{D} {w_i}^{q_i}. \tag{3.21}$$

We have used the multinomial expansion to proceed with the second and third equalities in the derivation of Eq. (3.19).

We have achieved a regrouping of the distribution components. Every component $W_{\boldsymbol{q}}$ is of the form $W_{\boldsymbol{k}} w_i$ with $q_i = k_i + 1$ and $q_j = k_j$ for $j \neq i$ and

$$\sum_{\boldsymbol{q}\in C_D^{n+1}} M_{\boldsymbol{q}} = \sum_{\boldsymbol{q}\in C_D^{n+1}} \frac{(n+1)!}{q_1!\dots q_D!} \prod_{i=1}^{D} {m_i}^{q_i} = \left(\sum_{i=1}^{D} m_i\right)^{n+1} = d^{2(n+1)}, \tag{3.22}$$

which is the dimension of $W_\rho^{\otimes(n+1)}$.

Therefore, $\{(W_{\boldsymbol{q}}, M_{\boldsymbol{q}})\}_{\boldsymbol{q}\in C_D^{n+1}}$ contains exactly the components of $W_\rho^{\otimes n}$, completing the proof. □

The above result applies on rescaled Wigner representations as well and it will be of assistance in deriving explicit bounds for important examples of magic distillation processes in chapter 4.

3.2 Non-monotonic Lorenz curves

A Lorenz curve is a powerful tool that provides a visual and compact description of majorization. We begin our analysis by providing the well-known definition of the Lorenz curve in the general context of real vectors. The definition allows us to demonstrate in Claim 3.5 that comparison between Lorenz curves is equivalent to the full set of majorization conditions for quasi-distributions, much like it is already known to be the case for probability distributions [133]. Unlike a monotone function, Lorenz curves therefore contain the entire information that majorization can provide us with.

3.2.1 Definition of a Lorenz curve

We make use of the notion of a Lorenz curve of a vector $\boldsymbol{w} \in \mathbb{R}^n$ relative to some other vector $\boldsymbol{r} \in \mathbb{R}^n$. Given a vector $\boldsymbol{w}$ we define $\boldsymbol{w}^{\downarrow}$ to be the re-arrangement of the components of $\boldsymbol{w}$ into decreasing order. Given two n–component vectors $\boldsymbol{w}$ and $\boldsymbol{r}$, we first define $\widetilde{\boldsymbol{w}} = (\widetilde{w}_i)$, where $\widetilde{w}_i \coloneqq w_i/r_i$, as the vector of component-wise ratios between $\boldsymbol{w}$ and $\boldsymbol{r}$. We can now define the Lorenz curve of $\boldsymbol{w}$ relative to $\boldsymbol{r}$, denoted $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$, as the piece-wise linear function that passes through $(0,0)$ and the n points

$$(x_k, L_{\boldsymbol{w}|\boldsymbol{r}}(x_k)) = \left(\frac{1}{R}\sum_{i=1}^{k} r_{\pi(i)}, \sum_{i=1}^{k} w_{\pi(i)}\right), \tag{3.23}$$

where $R \coloneqq \sum_{i=1}^{n} r_i$ and π is the permutation on n objects mapping $\widetilde{\boldsymbol{w}}$ to $\widetilde{\boldsymbol{w}}^{\downarrow}$. The form of this requires that $\boldsymbol{r}$ has no zero components, which we shall assume without loss of generality as the rank of a quantum state is not operationally meaningful. We further define an *elbow point* of the Lorenz curve as any non-differentiable point of the curve. They can only be located at $(0,0)$ and some $x_k = \frac{1}{R}\sum_{i=1}^{k} r_{\pi(i)}$, as these points are connected between them via line segments. Conversely, it is possible that the slope at some x_k does not change due to degeneracy in the values of the vector components, hence the number of elbows is upper bounded by $n+1$. For example, elbow points are demonstrated in Fig. (3.1) as the points where the slopes of the curve change.

In the usual case where $\boldsymbol{w}$ and $\boldsymbol{r}$ are both probability distributions the Lorenz curve is defined on the interval $[0,1]$, and rises monotonically until it reaches the value 1 at $x = 1$. The value $L_{\boldsymbol{w}|\boldsymbol{r}}(x) = 1$ is a global maximum. Moreover, if $\boldsymbol{w}, \boldsymbol{w}', \boldsymbol{r}, \boldsymbol{r}'$ are all valid probability distributions with $\boldsymbol{r}, \boldsymbol{r}'$ having positive components, then

$$(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}') \text{ if and only if } L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x),$$

for all $x \in [0,1]$. This result is well-known; it is shown in [129] and we reproduce

it in Claim 3.7 for completeness. It provides a simple way of computing whether relative majorization holds between pairs of probability distributions.

However, if $\boldsymbol{w}$ is a quasi-distribution with negative values, and $\boldsymbol{r}$ a regular probability distribution things are different. Now the Lorenz curve is no longer monotonically increasing, but is a concave function that breaks through the $L_{\boldsymbol{w}|\boldsymbol{r}}(x) = 1$ barrier at an interior point and attains some non-trivial maximum $L_\star$ above the value 1, before decreasing monotonically to $L_{\boldsymbol{w}|\boldsymbol{r}}(x) = 1$ at the end-point $x = 1$.

3.2.2 Extending Lorenz curves to quasi-distributions

We first verify that the Lorenz curve condition applies to quasi-distributions. This can be done from first principles, but a simpler way is to use the equivalent result from probability distributions and work on the differences. We first need the following auxiliary result.

Claim 3.4. *Let $\boldsymbol{w}$ be a quasi-distribution and let $\boldsymbol{r}$ be a probability distribution with strictly non-zero components. Then, $L_{a\boldsymbol{w}+b\boldsymbol{r}|\boldsymbol{r}}(x) = aL_{\boldsymbol{w}|\boldsymbol{r}}(x) + bx$ for any constants $a > 0$ and $b \in \mathbb{R}$.*

Proof. The Lorenz curve of $a\boldsymbol{w} + b\boldsymbol{r}$ relative to $\boldsymbol{r}$ passes through $(0, 0)$ and the points $(\sum_{i=1}^k r_{\pi(i)}, \sum_{i=1}^k (a\boldsymbol{w} + b\boldsymbol{r})_{\pi(i)})$ where π is the permutation that puts $(aw_i/r_i + b)$ in non-increasing order. Since $a > 0$, the permutation π puts (w_i/r_i) in non-increasing order too. We thus have

$$\left(\sum_{i=1}^k r_{\pi(i)}, \sum_{i=1}^k (a\boldsymbol{w} + b\boldsymbol{r})_{\pi(i)}\right) = \left(\sum_{i=1}^k r_{\pi(i)}, a\sum_{i=1}^k w_{\pi(i)} + b\sum_{i=1}^k r_{\pi(i)}\right),$$

so the value of the Lorenz curve at each point $x_k = \sum_{i=1}^k r_{\pi(i)}$ is given by

$$L_{a\boldsymbol{w}+b\boldsymbol{r}|\boldsymbol{r}}(x_k) = aL_{\boldsymbol{w}|\boldsymbol{r}}(x_k) + bL_{\boldsymbol{r}|\boldsymbol{r}}(x_k) = aL_{\boldsymbol{w}|\boldsymbol{r}}(x_k) + bx_k, \tag{3.24}$$

so we have $L_{a\boldsymbol{w}+b\boldsymbol{r}|\boldsymbol{r}}(x) = aL_{\boldsymbol{w}|\boldsymbol{r}}(x) + bx$ for any $x \in [0, 1]$ due to linearity. □

We can then reduce to the problem of genuine probability distributions by first masking the negativity in the quasi-distribution $\boldsymbol{w}$, i.e. mixing it with the reference distribution $\boldsymbol{r}$ until all components of $\boldsymbol{w}$ are non-negative, and then applying the conditions for relative majorization. This negativity masking gives the Lorenz curve condition on relative majorization.

Claim 3.5. *Let $\boldsymbol{w} \in \mathbb{R}^n, \boldsymbol{w}' \in \mathbb{R}^m$ be quasi-distributions, and let $\boldsymbol{r} \in \mathbb{R}^n, \boldsymbol{r}' \in \mathbb{R}^m$ be probability distributions with non-zero components. Then, $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ if and only if $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all $x \in [0, 1]$.*

Proof. Since the components of $\boldsymbol{r}$ are strictly positive, there always exists an $\epsilon > 0$ such that $\boldsymbol{w}_\epsilon := \epsilon\boldsymbol{w} + (1-\epsilon)\boldsymbol{r}$ is a genuine probability distribution. A similar result holds for $\boldsymbol{w}'$ and $\boldsymbol{r}'$ and we choose ϵ sufficiently small so that both $\boldsymbol{w}_\epsilon$ and $\boldsymbol{w}'_\epsilon$ are probability distributions. We now have that $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ if and only if $(\boldsymbol{w}_\epsilon, \boldsymbol{r}) \succ (\boldsymbol{w}'_\epsilon, \boldsymbol{r}')$. This equivalence holds because there exists a stochastic map A such that $A\boldsymbol{w} = \boldsymbol{w}'$ and $A\boldsymbol{r} = \boldsymbol{r}'$ if and only if

$$A[\epsilon\boldsymbol{w} + (1-\epsilon)\boldsymbol{r}] = \epsilon\boldsymbol{w}' + (1-\epsilon)\boldsymbol{r}' \text{ and } A\boldsymbol{r} = \boldsymbol{r}'. \tag{3.25}$$

In terms of a Lorenz curve condition we have that $(\boldsymbol{w}_\epsilon, \boldsymbol{r}) \succ (\boldsymbol{w}'_\epsilon, \boldsymbol{r}')$ if and only if $L_{\boldsymbol{w}_\epsilon|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'_\epsilon|\boldsymbol{r}'}(x)$ for all $x \in [0,1]$. Additionally, we use the result of Claim 3.4 that the Lorenz curve for any quasi-distribution obeys the relation

$$L_{a\boldsymbol{w}+b\boldsymbol{r}|\boldsymbol{r}}(x) = aL_{\boldsymbol{w}|\boldsymbol{r}}(x) + bx, \tag{3.26}$$

for any $a > 0$ and $b \in \mathbb{R}$. This relation implies that $L_{\boldsymbol{w}_\epsilon|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'_\epsilon|\boldsymbol{r}'}(x)$ if and only if $\epsilon L_{\boldsymbol{w}|\boldsymbol{r}}(x) + (1-\epsilon)x \geq \epsilon L_{\boldsymbol{w}'|\boldsymbol{r}'}(x) + (1-\epsilon)x$. Finally, the $(1-\epsilon)x$ terms cancel on both sides and we get the required relative majorization conditions that $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ if and only if $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all $x \in [0,1]$, as required. □

We frame the result of Claim 3.5 in order to highlight the link between quasi-distributions and probability distributions. This result therefore assumes that the Lorenz curve condition holds for proper probability distributions, a known result which we replicate later in Claim 3.7. Setting $\boldsymbol{r}' = \boldsymbol{r}$ in Claim 3.5 retrieves the special case of $\boldsymbol{d}$–majorization, which we discuss in more detail now.

3.2.3 Fixed-point majorization in magic theories

We now demonstrate the special case of $\boldsymbol{d}$–majorization in the context of the Wigner representation. Our aim here is to motivate the use of the Lorenz curve to visualize magic and how one can compare states in terms of their magic.

We consider some $\sigma \in \mathcal{F}$, and the resource theory $\mathcal{R}_\sigma$ which leaves its Wigner representation unchanged. Specifically, given a theory of magic $\mathcal{R} = (\mathcal{F}, \mathcal{O})$, we define the sub-theory $\mathcal{R}_\sigma = (\mathcal{F}, \mathcal{O}_\sigma)$, by restricting the free operations

$$\mathcal{O}_\sigma := \{\mathcal{E} \in \mathcal{O} : \mathcal{E}(\sigma) = \sigma\}, \tag{3.27}$$

to the subset that leave the distinguished state $\sigma \in \mathcal{F}$ invariant. This gives a simple way to break up any theory into smaller, more manageable parts where $\boldsymbol{d}$–majorization can be exploited to analyze each part, while the union over all sub-theories still returns the parent theory $\mathcal{R}$.

We are interested in investigating the ability to transform many copies of a magic state ρ towards a more pure form of magic. The state ρ is assumed to have negativity in the Wigner representation, so $W_\rho(\boldsymbol{z}) < 0$ for some regions of $\boldsymbol{z} \in \mathcal{P}_d$. The state σ is assumed to have a Wigner representation with $W_\sigma(\boldsymbol{z}) > 0$ for all $\boldsymbol{z}$ in the phase space. This is justified because a non-full rank state σ can be handled as a limiting case in which we first add an infinitesimal fraction of depolarizing noise $\epsilon(\mathbb{1}/d)$ and then take $\epsilon \to 0$.

The free operations within the magic theory $\mathcal{R}$ are represented by stochastic maps, and within $\mathcal{R}_\sigma$ by stochastic maps that leave W_σ invariant. Therefore, a necessary condition for magic state transformations $\rho_1 \to \rho_2$ within $\mathcal{R}_\sigma$ will be that

$$(W_{\rho_1}, W_\sigma) \succ (W_{\rho_2}, W_\sigma), \tag{3.28}$$

meaning that W_{ρ_1} is more ordered than the quasi-distribution W_{ρ_2} relative to W_σ. To simplify notation, we denote by $L_{\rho|\sigma}(x)$ the Lorenz curve $L_{W_\rho|W_\sigma}(x)$, and therefore have that

$$\rho_1 \to \rho_2 \text{ within } \mathcal{R}_\sigma \text{ implies } L_{\rho_1|\sigma}(x) \geq L_{\rho_2|\sigma}(x), \tag{3.29}$$

for all $x \in [0, 1]$, which restricts the transformations that are possible. Note that this is not a single numerical constraint, but a family of constraints. For n copies of a qudit system of dimension d the number of terms in W_ρ is d^{2n}, so imposing the Lorenz curve condition corresponds to exponentially many constraints.

We are now in a position to provide examples of how to construct non-monotonic Lorenz curves for Wigner representations, and more generally quasi-distributions, by computing the Lorenz curves corresponding to a family of qutrit magic states. Specifically, we can define the ϵ–noisy Strange state as

$$\rho_S(\epsilon) \coloneqq (1-\epsilon) \left|S\right\rangle \left\langle S\right| + \epsilon \frac{1}{3}\mathbb{1}, \tag{3.30}$$

where ϵ is the depolarizing error parameter. This magic state is canonical in the sense that any magic state ρ can be processed via Clifford operations [83, 134] and put into this form for some $\epsilon \geq 0$. We will therefore use it as a prominent example in chapter 5 to demonstrate our results. For now, it suffices to state that the Wigner representation of the single-copy, ϵ–noisy Strange state is given by

$$W_{\rho_S(\epsilon)}(\boldsymbol{z}) = (1-\epsilon) W_{|S\rangle\langle S|}(\boldsymbol{z}) + \epsilon W_{\frac{1}{3}\mathbb{1}}(\boldsymbol{z}), \tag{3.31}$$

due to property 1 in Claim 2.7. The representation contains a single negative component

$$-v(\epsilon) \coloneqq -\left(\frac{1}{3} - \frac{4}{9}\epsilon\right) \tag{3.32}$$

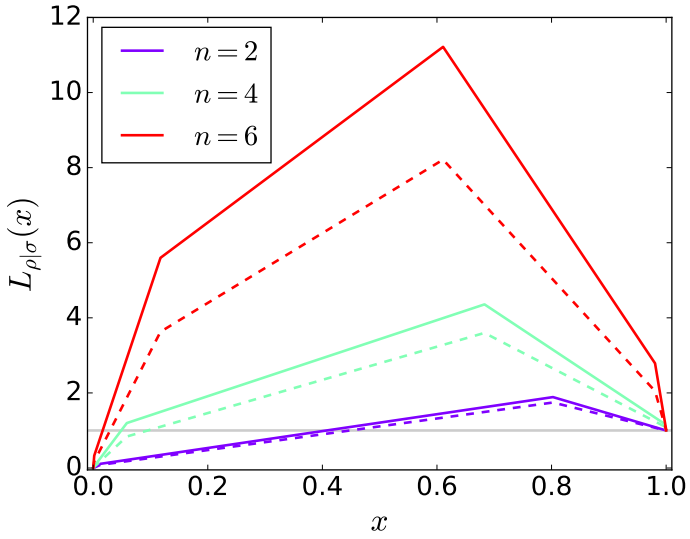

Figure 3.1: **"These Lorenz curves go to 11".** Traditionally, Lorenz curves for distributions are monotone increasing cumulative functions that reach a maximum value of 1. In contrast, Lorenz curves for quasi-distributions generally achieve values higher than $L(x) = 1$ and can, in fact, reach arbitrarily high values due to the presence of negativity in their components. The above family of curves corresponds to multiple copies of noisy Strange states $\rho := \rho_S(\epsilon)^{\otimes n}$ (Eq. (3.30)) for $n = 2, 4, 6$ within $\mathcal{R}_\sigma$, where $\sigma = \mathbb{1}/3$. Solid lines represent pure Strange states, while dashed lines represent ϵ–noisy Strange states with depolarizing error $\epsilon = 0.1$.

located at the origin $\boldsymbol{z} = \mathbf{0}$ and positive components

$$u(\epsilon) := \frac{1}{6} - \frac{1}{18}\epsilon \tag{3.33}$$

at the 8 phase space points $\boldsymbol{z} \neq \mathbf{0}$. We assume that $\epsilon < 3/4$ to ensure the presence of negativity in the Wigner representation.

Due to negativity, the ordering of the components of the rescaled Wigner representation $W_{\rho_n|\sigma_n}(\boldsymbol{z}) := W_{\rho_n}(\boldsymbol{z})/W_{\sigma_n}(\boldsymbol{z})$, for $\rho_n := \rho_S(\epsilon)^{\otimes n}$, and $\sigma_n := \sigma^{\otimes n} = (\mathbb{1}/3)^{\otimes n}$, depends on whether $-v(\epsilon)$ is raised to an even or odd power. Treating the Wigner representations as vectors, the component values w_i and associated multiplicities m_i in the n–copy case are found to be

$$m_i = 8^i \binom{n}{i}, \tag{3.34}$$

$$w_i = u^i(-v)^{n-i}, \tag{3.35}$$

where index i runs through $0, \ldots, n$, and we assume for simplicity that $\epsilon \leq 3/7$, which implies that $v \geq u$.

Therefore, assuming further that n is even, we find the order in which we should add up components, according to the definition of the Lorenz curve in Eq. (3.23). In particular, all components with even i are positive, while the smaller i is, the larger the contribution of v over u is, hence the larger the value of the component. Adding up all the positive components, we see that the Lorenz curve $L_{\rho_n|\sigma_n}(x)$ reaches a maximum value of

$$L_\star := L_{\rho_n|\sigma_n}(x_\star) = \sum_{i=0}^{n/2} m_{2i} w_{2i} = \frac{1}{2} + \frac{1}{2}\left(\frac{15-8\epsilon}{9}\right)^n > 1, \tag{3.36}$$

which occurs at $x = x_\star$ given by

$$x_\star = \frac{1}{2} + \frac{1}{2}\left(\frac{7}{9}\right)^n. \tag{3.37}$$

We plot the Lorenz curves for different numbers of copies n and error rate ϵ in Fig. (3.1) for the Strange state (Eq. (3.30)) in sub-theory $\mathcal{R}_{\mathbb{1}/3}$, where the free channels are unital. We see that all Lorenz curves achieve values higher than 1, with curves corresponding to higher numbers of copies n' lying completely above curves with fewer copies n, for $\epsilon = 0$. It is therefore implied by Claim 3.5 that one can transform many copies of the Strange state to few via unital channels. Introducing non-zero noise lowers the Lorenz curves. If enough noise is introduced, a curve representing a n'–copy state may intersect a curve corresponding to a noiseless n–copy state with $n < n'$, thus indicating that the n'–copy state is too noisy to be transformed to the noiseless n–copy state.

3.3 Relative majorization for quasi-distributions

In this section, we first prove a link between simple majorization and relative majorization in the context of quasi-distributions. This is well-known for probability distributions (e.g. [131]), but entirely novel in the context of quasi-distributions. We then use this result to complete the Lorenz curve condition in Claim 3.5 by proving the statement for proper probability distributions. Finally, we provide equivalent formulations of relative majorization and prove some useful properties of the Lorenz curves.

3.3.1 Relative majorization is still majorization

The following result is well-known for probability distributions [127, 133, 135] and we extend it here to quasi-distributions. It is a satisfying connection between relative majorization and simple majorization.

Claim 3.6. *Let $\boldsymbol{w} \in \mathbb{R}^N, \boldsymbol{w}' \in \mathbb{R}^{N'}$ be quasi-distributions and $\boldsymbol{r} \in \mathbb{R}^N, \boldsymbol{r}' \in \mathbb{R}^{N'}$ probability distributions with positive rational entries given by $r_i = a_i/K$ and $r'_i = a'_i/K$ for positive integers a_i, a'_i and $K = \sum_{i=1}^N a_i = \sum_{i=1}^{N'} a'_i$. Then,*

$$(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}') \text{ if and only if } \Gamma_{\boldsymbol{a}}(\boldsymbol{w}) \succ \Gamma_{\boldsymbol{a}'}(\boldsymbol{w}'), \tag{3.38}$$

where the embedding map $\Gamma_{\boldsymbol{a}} : \mathbb{R}^N \to \mathbb{R}^K$ is given by

$$\Gamma_{\boldsymbol{a}}(\boldsymbol{z}) \coloneqq \bigoplus_{i=1}^N z_i \boldsymbol{\eta}_{a_i} = \left(\frac{z_1}{a_1}, \frac{z_1}{a_1}, \ldots, \frac{z_1}{a_1}, \frac{z_2}{a_2}, \frac{z_2}{a_2}, \ldots, \frac{z_2}{a_2}, \ldots, \frac{z_N}{a_N}, \frac{z_N}{a_N}, \ldots, \frac{z_N}{a_N} \right), \tag{3.39}$$

with $\boldsymbol{\eta}_{a_i} = (1/a_i, 1/a_i, \ldots, 1/a_i)$ the uniform distribution on a_i elements.

Proof. We first note that the map $\Gamma_{\boldsymbol{a}}$ is stochastic and has a well-defined left-inverse $\Gamma_{\boldsymbol{a}}^{-1} : \mathbb{R}^K \to \mathbb{R}^N$ given by

$$\Gamma_{\boldsymbol{a}}^{-1}(\boldsymbol{x}) = \left(\sum_{i=1}^{a_1} x_i, \sum_{i=a_1+1}^{a_1+a_2} x_i, \ldots, \sum_{i=a_1+\cdots+a_{N-1}+1}^{K} x_i \right), \tag{3.40}$$

that obeys $(\Gamma_{\boldsymbol{a}}^{-1} \circ \Gamma_{\boldsymbol{a}})(\boldsymbol{z}) = \boldsymbol{z}$ for all $\boldsymbol{z} \in \mathbb{R}^N$. To obtain this inverse, we sum up all components within each independent block $z_i \boldsymbol{\eta}_{a_i}$ that appears in $\Gamma_{\boldsymbol{a}}$ for all $i = 1, \ldots, N$. It is not a right-inverse as not all vectors in $\mathbb{R}^K$ is of the form of Eq. (3.39).

The Claim is equivalent to the statement that there exists a bistochastic map B sending $\Gamma_{\boldsymbol{a}}(\boldsymbol{w})$ to $\Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')$ if and only if there exists a stochastic map A sending $\boldsymbol{w}$ to $\boldsymbol{w}'$ and $\boldsymbol{r}$ to $\boldsymbol{r}'$.

Suppose there is a stochastic map A such that $A\boldsymbol{w} = \boldsymbol{w}'$ and $A\boldsymbol{r} = \boldsymbol{r}'$. We define $B : \mathbb{R}^K \mapsto \mathbb{R}^K$ by $B \coloneqq \Gamma_{\boldsymbol{a}'} \circ A \circ \Gamma_{\boldsymbol{a}}^{-1}$ so that it is stochastic as a composition of stochastic maps and it preserves the uniform distribution in $\mathbb{R}^K$, since

$$B(1/K, \ldots, 1/K) = \left(\Gamma_{\boldsymbol{a}'} \circ A \circ \Gamma_{\boldsymbol{a}}^{-1}\right)\left(\Gamma_{\boldsymbol{a}'}(\boldsymbol{r})\right) = \Gamma_{\boldsymbol{a}'}(\boldsymbol{r}') = (1/K, \ldots, 1/K), \tag{3.41}$$

therefore B is bistochastic. Finally, B maps the embedded distributions as follows,

$$B\Gamma_{\boldsymbol{a}}(\boldsymbol{w}) = \left(\Gamma_{\boldsymbol{a}'} \circ A \circ \Gamma_{\boldsymbol{a}}^{-1}\right)\left(\Gamma_{\boldsymbol{a}}(\boldsymbol{w})\right) = \Gamma_{\boldsymbol{a}'}(\boldsymbol{w}'). \tag{3.42}$$

Conversely, suppose a bistochastic map B exists sending $\Gamma_{\boldsymbol{a}}(\boldsymbol{w})$ to $\Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')$. Again, define $A : \mathbb{R}^N \mapsto \mathbb{R}^{N'}$ by $A \coloneqq \Gamma_{\boldsymbol{a}'}^{-1} \circ B \circ \Gamma_{\boldsymbol{a}}$ so that it is stochastic as a composition of stochastic maps and $A\boldsymbol{w} = \left(\Gamma_{\boldsymbol{a}'}^{-1} \circ B \circ \Gamma_{\boldsymbol{a}}\right)\boldsymbol{w} = \boldsymbol{w}'$, as well as $A\boldsymbol{r} = \boldsymbol{r}'$. □

We note that the vectors $\boldsymbol{r}, \boldsymbol{r}'$ with rational components form a dense subset of the positive probability distributions, and do not consider further technicalities that have no impact on actual physical measurements, which always have a finite resolution.

Restricting the above result in the context of proper probability distributions, we can replicate the Lorenz curve condition for probability distributions (see e.g. [130]).

Claim 3.7. *Given probability distributions $\boldsymbol{w}, \boldsymbol{r} \in \mathbb{R}^N$ and $\boldsymbol{w}', \boldsymbol{r}' \in \mathbb{R}^{N'}$ with $\boldsymbol{r}, \boldsymbol{r}'$ having positive components, then*

$$(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}') \text{ if and only if } L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x),$$

for all $x \in [0, 1]$.

Proof. In the case of $N = N'$ and $\boldsymbol{r} = \boldsymbol{r}' = \boldsymbol{\eta}$, where $\boldsymbol{\eta} = (1/N, 1/N, \ldots, 1/N)$ is the uniform distribution on N elements, the statement reduces to the Lorenz curve condition $L_{\boldsymbol{w}}(x) \geq L_{\boldsymbol{w}'}(x)$ for all x, for simple majorization which follows immediately from the defining set of inequalities for majorization [115]. Namely, the Lorenz curve $L_{\boldsymbol{w}}(x)$ for $\boldsymbol{w}$ is obtained from the partial sums of $\boldsymbol{w}$ sorted in non-increasing order. It is also clear from the definition that the Lorenz curve of $\boldsymbol{w}$ is given by $L_{\boldsymbol{w}}(x) = L_{\boldsymbol{w}|\boldsymbol{\eta}}(x)$.

To prove the general statement we reduce the relative majorization Lorenz curve condition to standard majorization. Using the notation and assumptions of Claim 3.6 for distributions with rational components, we can define $\Gamma_{\boldsymbol{a}}(\boldsymbol{w}), \Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')$ and the uniform distribution $\boldsymbol{\eta} = (1/K, 1/K, \ldots, 1/K)$.

The key ingredient in the proof is that the Lorenz curve of $\boldsymbol{w}$ relative to $\boldsymbol{r}$ coincides with the Lorenz curve of $\Gamma_{\boldsymbol{a}}(\boldsymbol{w})$, namely

$$L_{\boldsymbol{w}|\boldsymbol{r}}(x) = L_{\Gamma_{\boldsymbol{a}}(\boldsymbol{w})|\boldsymbol{\eta}}(x) = L_{\Gamma_{\boldsymbol{a}}(\boldsymbol{w})}(x) \text{ for all } x \in [0, 1]. \tag{3.43}$$

To see this we consider the elbows of $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$, as defined under Eq. (3.23)),

$$(x_k, L_{\boldsymbol{w}|\boldsymbol{r}}(x_k)) = \left(\sum_{i=1}^{k} r_{\pi(i)}, \sum_{i=1}^{k} w_{\pi(i)} \right), \tag{3.44}$$

where the permutation π sorts (w_i/r_i) in non-increasing order. Expressing $\Gamma_{\boldsymbol{a}}(\boldsymbol{w})$ as

$$\Gamma_{\boldsymbol{a}}(\boldsymbol{w}) = \frac{1}{K} \bigoplus_{i=1}^{N} \left(\frac{w_i}{r_i}, \ldots, \frac{w_i}{r_i} \right), \tag{3.45}$$

where $(w_i/r_i, \ldots, w_i/r_i)$ has a_i elements, it is clear that permutation π sorts $\Gamma_{\boldsymbol{a}}(\boldsymbol{w})$ in non-increasing order too. The Lorenz curve elbows $(y_k, L_{\Gamma_{\boldsymbol{a}}(\boldsymbol{w})|\boldsymbol{\eta}}(y_k))$ occur at

$$y_k = \sum_{j=1}^{a_{\pi(1)}+\cdots+a_{\pi(k)}} \frac{1}{K} = \sum_{i=1}^{k} \frac{a_{\pi(i)}}{K} = x_k \tag{3.46}$$

and take values

$$L_{\Gamma_{\boldsymbol{a}}(\boldsymbol{w})|\boldsymbol{\eta}}(y_k) = \sum_{i=1}^{k} a_{\pi(i)} \frac{1}{K} \frac{w_{\pi(i)}}{r_{\pi(i)}} = L_{\boldsymbol{w}|\boldsymbol{r}}(x_k). \tag{3.47}$$

Therefore, the Lorenz curve of the embedded distribution coincides with the Lorenz curve in the relative majorization setting.

Finally, we have $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ if and only if $\Gamma_{\boldsymbol{a}}(\boldsymbol{w}) \succ \Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')$, which holds if and only if $L_{\Gamma_{\boldsymbol{a}}(\boldsymbol{w})|\boldsymbol{\eta}}(x) \geq L_{\Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')|\boldsymbol{\eta}}(x)$ for all $x \in [0,1]$, which in turn holds if and only if $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x),\ x \in [0,1]$, which concludes the proof. □

Having shown the Lorenz curve condition on probability distributions in Claim 3.7, we have completed the proof of the Lorenz curve condition on quasi-distributions in Claim 3.5. We now list a concise summary of useful equivalent formulations of relative majorization on quasi-distributions.

Claim 3.8. *Given quasi-distributions $\boldsymbol{w}, \boldsymbol{w}'$, $\boldsymbol{r}, \boldsymbol{r}'$, such that the components of $\boldsymbol{r}$ and $\boldsymbol{r}'$ are positive, the following statements are equivalent:*

(i) $\boldsymbol{w}' = A\boldsymbol{w}$ *and* $\boldsymbol{r}' = A\boldsymbol{r}$ *for a stochastic map* A*;*

(ii) $L_{\boldsymbol{w}|\boldsymbol{r}}(t) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(t)$ *for* $t \in [0,1)$ *and* $L_{\boldsymbol{w}|\boldsymbol{r}}(1) = L_{\boldsymbol{w}'|\boldsymbol{r}'}(1)$*;*

(iii) $\sum_{i=1}^{n} |w_i - r_i t| \geq \sum_{i=1}^{n} |w'_i - r'_i t|$ *for all* $t \in \mathbb{R}$.

Proof. The equivalence between (i) and (ii) on quasi-distributions was proven in Theorem 1 in the main text.

The equivalence between (i) and (iii) follows from a similar argument where we mask negativity with a probability distribution. Namely, let $\epsilon > 0$ be such that $\boldsymbol{w}(\epsilon) \coloneqq \epsilon\boldsymbol{w} + (1-\epsilon)\boldsymbol{r}$ and $\boldsymbol{w}'(\epsilon) \coloneqq \epsilon\boldsymbol{w} + (1-\epsilon)\boldsymbol{r}'$ are genuine probability distributions. This is guaranteed by picking a sufficiently small $\epsilon > 0$, since $\boldsymbol{r}$ has positive components. We now have that $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ if and only if $(\boldsymbol{w}(\epsilon), \boldsymbol{r}) \succ (\boldsymbol{w}'(\epsilon), \boldsymbol{r}')$. Moreover, we have that for all $c \in \mathbb{R}$, $\sum_{i=1}^{n} |w(\epsilon)_i - r_i c| \geq \sum_{i=1}^{n} |w'(\epsilon)_i - r'_i c|$, which leads to

$$\sum_{i=1}^{n} \left| w_i - r_i \frac{\epsilon + c - 1}{\epsilon} \right| \geq \sum_{i=1}^{n} \left| w'_i - r'_i \frac{\epsilon + c - 1}{\epsilon} \right|$$

Replacing $t = (\epsilon + c - 1)/\epsilon$, we see that t can attain any real value for $c \in \mathbb{R}$, so we deduce the required L_1–norm condition on quasi-distributions $\boldsymbol{w}, \boldsymbol{w}'$. □

The first statement in Claim 3.8 is simply the definition of majorization, the second statement is the Lorenz curve condition and the third statement is one that we use in showing that majorization is more powerful than the mana monotone (Eq. (2.33)) later on, in Claim 4.2.

3.3.2 Lorenz curve properties

Aiming to obtain a firmer grip on the Lorenz curve formulation, we verify useful properties in the context of quasi-distributions, starting from the simple property that the Lorenz curve is concave.

Claim 3.9 (Concavity)**.** *Let $\boldsymbol{w}$ be a quasi-distribution and let $\boldsymbol{r}$ be a probability distribution with non-zero components. Then, $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ is a concave function on $[0,1]$.*

Proof. Let $x_\star$ be the point where $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ first attains it maximum. Therefore, on $[0, x_\star]$ the function rises monotonically to $L_{\boldsymbol{w}|\boldsymbol{r}}(x_\star)$, via the sum of all positive entries of $\boldsymbol{w}$, taken in decreasing order. Likewise on $[x_\star, 1]$, the function decreases monotonically from its maximum via the partial sums of the negative entries of $\boldsymbol{w}$ in decreasing order. Let $f_1(x)$ be equal to $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ on $[0, x_\star]$ and 0 otherwise. Also let $f_2(x)$ be equal to $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ on $[x_\star, 1]$ and and 0 otherwise. By inspection both f_1 and f_2 are concave functions, and $f_1(x) + f_2(x) = L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ for all $x \in [0,1]$. However, the sum of two concave functions is also concave which concludes the proof. □

This result is known for the case of probability distributions, where the Lorenz curve is in fact non-decreasing. It is also clear for quasi-probability distributions, as we can inspect in Fig. (3.1), with the difference that now the concave Lorenz curve contains segments with negative slope.

The following result reduces the Lorenz curve condition on the interval $[0,1]$ to a the smallest set of independent inequalities one can condier, thus simplifying the computation of relative majorization constraints in practical scenarios.

Claim 3.10 (Lorenz curve inequalities)**.** *Let $\boldsymbol{w} \in \mathbb{R}^N, \boldsymbol{w}' \in \mathbb{R}^{N'}$ be quasi-distributions and $\boldsymbol{r} \in \mathbb{R}^N, \boldsymbol{r}' \in \mathbb{R}^{N'}$ probability distributions with positive entries. Assume that $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ has t elbows at locations $x_1, \ldots, x_t$. Then, $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all $x \in [0,1]$ iff $L_{\boldsymbol{w}|\boldsymbol{r}}(x_i) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x_i)$ for all $i = 1, \ldots, t$.*

Proof. If $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all $x \in [0,1]$, then the condition on the elbows follows trivially.

Conversely, assume $L_{\boldsymbol{w}|\boldsymbol{r}}(x_i) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x_i)$ for all $i = 1, \ldots t$. Suppose on the contrary that $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ dips below $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ at some point $x = y$ where $x_i < y < x_{i+1}$ for some i. Then this implies

$$\begin{aligned} L_{\boldsymbol{w}|\boldsymbol{r}}(x_i) &\geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x_i) \\ L_{\boldsymbol{w}|\boldsymbol{r}}(y) &< L_{\boldsymbol{w}'|\boldsymbol{r}'}(y) \\ L_{\boldsymbol{w}|\boldsymbol{r}}(x_{i+1}) &\geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x_{i+1}). \end{aligned}$$

However, since $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ is linear between x_i and x_{i+1}, the above conditions imply that $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ is not concave, which contradicts Claim 3.9. Therefore, $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all x in $[0,1]$. □

In fact, we reduce the result of Claim 3.10 to the first elbow, hence considering only the part of the Lorenz curves between the origin $(0,0)$ and the first elbow (Eq. (3.23)).

Claim 3.11 (First elbow constraint)**.** *Let $\boldsymbol{w} \in \mathbb{R}^N, \boldsymbol{w}' \in \mathbb{R}^{N'}$ be quasi-distributions and $\boldsymbol{r} \in \mathbb{R}^N, \boldsymbol{r}' \in \mathbb{R}^{N'}$ probability distributions with positive entries. Denote the coordinates of the first elbow of $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ by (X_0, L_0) and the coordinates of the first elbow of $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ by (X_0', L_0').*

Then, given any coordinates (x, L) and (x', L') on Lorenz curves $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ and $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ respectively, where $0 < x \leq X_0$ and $0 < x' \leq X_0'$, the process is possible only if

$$\frac{L}{x} \geq \frac{L'}{x'}. \tag{3.48}$$

Proof. If the transformation is possible, then $L_{\boldsymbol{w}|\boldsymbol{r}}(x) \geq L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ for all x in $[0,1]$. Restricting to the initial line segment of $L_{\boldsymbol{w}|\boldsymbol{r}}(x)$ joining $(0,0)$ to (X_0, L_0), we then require that it be above the initial linear segment of $L_{\boldsymbol{w}'|\boldsymbol{r}'}(x)$ joining $(0,0)$ to (X_0', L_0'). Since both these linear segments start at the origin, this condition is equivalent to the slope of the initial segment of the input Lorenz curve being greater than or equal to the slope of the initial segment of the output Lorenz curve. This slope constraint is then equivalent to Eq. (3.48). □

This simple condition ignores the rest of the Lorenz curves, but it can still be sufficient to build, for example, magic bounds that beat bounds from certain monotones. We call it the *first elbow constraint* and we use it in deriving some explicit upper bounds on magic distillation rates in section 4.1.2 and section 4.2.2.

Moving on from Lorenz curves, we now turn our attention to the emergence of single-shot entropies from relative majorization in the context of quasi-distributions.

3.4 Single-shot entropies on quasi-distributions

The obstacle in linking Wigner negativity on the phase space with macroscopic thermodynamic notions, such as equilibrium free energy is that the Wigner representation is generally a quasi-distribution, so the standard Boltzmann entropy $H(\boldsymbol{w}) = -\sum_i w_i \log w_i$ is not well-defined for representations $\boldsymbol{w}$ of states that admit

negativities, let alone have a physically meaningful interpretation. However, in this section we provide the tools to overcome this annoying hiccup. We do so by deriving in section 3.4.1 a wide family of entropic families that are well-defined on the discrete phase space, and, more broadly, on the space of quasi-distributions of any dimension. We then show in section 3.4.2 that these well-defined entropies allow us to further define a notion of relative entropy on quasi-distributions that is physically meaningful in terms of data processing. These functions form single-shot measures of disorder between quasi-distributions and we derive them based on our work on relative majorization of quasi-distributions that we have described thus far. The narrative of these two sections takes the form of Claims that extend results known for proper probabilities to the entire quasi-probability space.

3.4.1 Rényi entropies on quasi-distributions

To extend our majorization framework to a wide family of entropic functions, we make use of known results from probability majorization theory to establish the Schur-concavity of the Rényi entropy H_α [125], as given in Eq. (3.53), on quasi-distributions, for a subset of α values. We consider the set of all quasi-distributions

$$\mathcal{Q}_N = \left\{ \boldsymbol{w} \in \mathbb{R}^N : \sum_{i=1}^{N} w_i = 1 \right\}, \tag{3.49}$$

which is a hyper-plane in $\mathbb{R}^N$.

We first need to define the notion of Schur-convex and Schur-convave functions on a subset $\mathcal{D} \subseteq \mathbb{R}^N$. A continuous, real-valued function f defined on $\mathcal{D} \subseteq \mathbb{R}^N$ is called *Schur-convex* on the domain $\mathcal{D}$ if $\boldsymbol{p} \prec \boldsymbol{q}$ on $\mathcal{D}$ implies that $f(\boldsymbol{p}) \leq f(\boldsymbol{q})$. A continuous function $f : \mathcal{D} \to \mathbb{R}$ is called *Schur-concave* if $(-f)$ is Schur-convex.

For any subset $\mathcal{D} \subseteq \mathbb{R}^N$ we further define

$$\mathcal{D}^{\downarrow} := \{\boldsymbol{w} \in \mathcal{D} : w_1 \geq w_2 \geq \cdots \geq w_N\}. \tag{3.50}$$

We now have the following fundamental result in majorization theory.

Claim 3.12 (Schur-Ostrowski criterion [115, 136, 137])**.** *Let $\mathcal{D} \subseteq \mathbb{R}^N$ be a convex set with non-empty interior, and invariant under permutations of vector components. Let $f : \mathcal{D} \to \mathbb{R}$ be a continuously differentiable function on the interior of $\mathcal{D}$ and continuous on the whole of $\mathcal{D}$. Then f is Schur-convex on $\mathcal{D}$ if and only if f is a symmetric function on $\mathcal{D}$ and*

$$\partial_1 f(\boldsymbol{w}) \geq \partial_2 f(\boldsymbol{w}) \geq \cdots \geq \partial_N f(\boldsymbol{w}), \tag{3.51}$$

for all $\boldsymbol{w}$ in the interior of $\mathcal{D}^{\downarrow}$.

The Schur-Ostrowski criterion is used to derive the monotonicity of the Rényi entropy H_α on probability distributions. Since the set $\mathcal{Q}_N$ of quasi-distributions obey the necessary conditions of the Schur-Ostrowski criterion, we can use it to establish that the Rényi entropy H_α is well-defined and monotonic on the entire set $\mathcal{Q}_N$. To this end, we first state the following preliminary result, which follows directly from Claim 3.12.

Claim 3.13. *Let f be a real-valued continuous function defined on $\mathcal{Q}_N$ that is continuously differentiable on the interior of $\mathcal{Q}_N$. Then, f is Schur-convex on $\mathcal{Q}_N$ if and only if it is a symmetric function and*

$$\partial_1 f(\boldsymbol{w}) \geq \partial_2 f(\boldsymbol{w}) \geq \cdots \geq \partial_N f(\boldsymbol{w}), \tag{3.52}$$

for all quasi-distributions $\boldsymbol{w}$ in the interior of $\mathcal{Q}_N^{\downarrow}$.

The classical α–Rényi entropy [125] on a probability distribution $\boldsymbol{p} = (p_i)$ is given by

$$H_\alpha(\boldsymbol{p}) \coloneqq \frac{1}{1-\alpha} \log \sum_i p_i^\alpha, \tag{3.53}$$

where $\alpha \geq 0$. This means the function is well-defined on $\mathcal{Q}_N^+$, namely the set of quasi-distributions with non-negative components. We wish to extend the domain to the whole of $\mathcal{Q}_N$, but this requires that w_i^α be a real number for any $w_i \in \mathbb{R}$.

The exponent function $x \mapsto x^\alpha$ is well-defined and real-valued if α is a rational number of the form $\alpha = r/s$ where s is an odd integer and $r \in \mathbb{N}$. However, we also require that $\sum_i w_i^\alpha > 0$ in order for the logarithm to return a real value. To ensure this, we restrict to $\alpha = r/s$ with $r = 2a$ being an even integer and $s = 2b-1$ being an odd integer, for some integers $a, b \in \mathbb{N}$, which implies that the following are all well-defined, real-valued expressions,

$$0 \leq w_i^\alpha = w_i^{\frac{2a}{2b-1}} = \left(w_i^{\frac{1}{2b-1}}\right)^{2a} = \left(w_i^{2a}\right)^{\frac{1}{2b-1}} = |w_i|^{\frac{2a}{2b-1}}.$$

We note that previous work [138] has considered the $\alpha = 2$ entropy of Wigner representations, and other work exists that uses the Wehrl entropy, based on the Hussimi function of a quantum state [139].

We are now in a position to show that $H_\alpha(\boldsymbol{w})$ for $\alpha = 2a/(2b-1)$ is a Schur-concave function on $\mathcal{Q}_N$ provided that $a \geq b$.

Claim 3.14. *If $\alpha = \frac{2a}{2b-1}$ for positive integers a, b with $a \geq b$, then $H_\alpha(\boldsymbol{w})$ is well-defined on the set of quasi-distributions $\mathcal{Q}_N$, and moreover if $\boldsymbol{w} \succ \boldsymbol{w}'$ for two quasi-distributions $\boldsymbol{w}, \boldsymbol{w}' \in \mathcal{Q}_N$ then $H_\alpha(\boldsymbol{w}) \leq H_\alpha(\boldsymbol{w}')$.*

Proof. We have $\alpha = \frac{2a}{2b-1} \geq \frac{2b}{2b-1} > 1$ and α is a rational with even numerator and odd denominator, so $\sum_i w_i^\alpha > 0$ and $H_\alpha(\boldsymbol{w})$ is well-defined for all quasi-distributions $\boldsymbol{w}$.

We also have that $H_\alpha(\boldsymbol{w})$ is well-defined, continuous, differentiable on the interior of $\mathcal{Q}_N$ and symmetric in the components of $\boldsymbol{w}$.

Consider the partial derivatives

$$\frac{\partial H_\alpha}{\partial w_i} = \frac{\alpha}{(\alpha-1)\sum_i w_i^\alpha}\left(-w_i^{\alpha-1}\right). \tag{3.54}$$

For $\alpha = \frac{2a}{2b-1}$, with $a \geq b$, we have that

$$\frac{\alpha}{(\alpha-1)\sum_i w_i^\alpha} > 0. \tag{3.55}$$

The first derivative of the function $g(w) \coloneqq -w^{\alpha-1}$ is given by

$$g'(w) = -(\alpha-1)w^{\alpha-2} = -(\alpha-1)w^{\frac{2a-4b+2}{2b-1}} \leq 0, \tag{3.56}$$

because $\alpha > 1$ and

$$w^{\frac{2a-4b+2}{2b-1}} = \left(w^{\frac{a-2b+1}{2b-1}}\right)^2 \geq 0, \tag{3.57}$$

so $g(w)$ is non-increasing in w.

Therefore, whenever $w_i \geq w_j$, we have that

$$-w_i^{\alpha-1} \leq -w_j^{\alpha-1}, \tag{3.58}$$

which implies that

$$\frac{\partial H_\alpha}{\partial w_i}(\boldsymbol{w}) \leq \frac{\partial H_\alpha}{\partial w_j}(\boldsymbol{w}), \tag{3.59}$$

for any $\boldsymbol{w}$ in the interior of $\mathcal{Q}_N^\downarrow$. Therefore, H_α is Schur-concave on $\mathcal{Q}_N$ and

$$H_\alpha(\boldsymbol{w}) \leq H_\alpha(\boldsymbol{w}'), \tag{3.60}$$

for any quasi-distributions $\boldsymbol{w}, \boldsymbol{w}'$ that obey $\boldsymbol{w} \succ \boldsymbol{w}'$. □

While we have integers a, b such that $0 < \alpha = \frac{2a}{2b-1} < 1$ and $H_\alpha(\boldsymbol{w})$ is well-defined, it turns out that monotonicity does not hold if we drop the condition $a \geq b$. If $\alpha = 2a/(2b-1)$ with $\alpha < 1$, then $g(w) \coloneqq \frac{-w^{\alpha-1}}{\alpha-1}$ is no longer monotonic for all $w \in \mathbb{R}$, and the problem occurs when comparing $w_i < 0$ and $w_j > 0$. As an example of this dependence of monotonicity on the domain of the function, consider the function $g(x) = \frac{1}{x^3}$ which is monotone decreasing on both $x < 0$ and $x > 0$ separately, however it is not monotone on the full real-line.

We note that if $\alpha = r/s$ with both r and s odd, then H_α is not well-defined for all quasi-distributions, although if the actual set of quasi-distributions has a sufficiently bounded negativity, then $\log \sum_i w_i^\alpha$ can still be obtained for r odd, provided the total sum is never negative. For Wigner representations, this may require a stricter bound than the generic bound $|W_\rho(\boldsymbol{z})| \leq 1/d$ for a d–dimensional quantum system.

We also note that the set $F \coloneqq \{2a/(2b-1) : a, b \in \mathbb{N} \text{ and } a \geq b\}$ is dense in the reals $\mathbb{R}_{>1}$, as any rational c/d with $c > d$ can be approximated by $c2^n/(d2^n - 1) \in F$ for sufficiently large n.

We have that H_α is additive on products of quasi-distributions, which we state for completeness.

Claim 3.15. *For any $\boldsymbol{w} \in \mathcal{Q}_N$ and $\boldsymbol{w}' \in \mathcal{Q}_{N'}$, we have*

$$H_\alpha(\boldsymbol{w} \otimes \boldsymbol{w}') = H_\alpha(\boldsymbol{w}) + H_\alpha(\boldsymbol{w}'), \tag{3.61}$$

where $\alpha = 2a/(2b-1)$ with positive integers $a \geq b$.

Proof.

$$\begin{aligned} H_\alpha(\boldsymbol{w} \otimes \boldsymbol{w}') &= \frac{1}{1-\alpha} \log \sum_{i,j} \left[\boldsymbol{w}_i \boldsymbol{w}'_j\right]^\alpha = \frac{1}{1-\alpha} \log \left[\sum_i \boldsymbol{w}_i^\alpha \sum_j \boldsymbol{w}_j'^\alpha\right] \\ &= \frac{1}{1-\alpha} \log \left[\sum_i \boldsymbol{w}_i^\alpha\right] + \frac{1}{1-\alpha} \log \left[\sum_j \boldsymbol{w}_j'^\alpha\right] \\ &= H_\alpha(\boldsymbol{w}) + H_\alpha(\boldsymbol{w}'). \end{aligned} \tag{3.62}$$

□

The additivity property in representing many-copy magic states.

Applied to Wigner representations $W_\rho(\boldsymbol{z})$ for a quantum state ρ we have

$$H_\alpha(W_\rho) \coloneqq \frac{1}{1-\alpha} \log \sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z})^\alpha, \tag{3.63}$$

where α can take values of the form $2a/(2b-1)$, for non-negative integers a, b. For the noisy Strange state $\rho = \rho_S(\epsilon)$ (Eq. (3.30)), this becomes

$$H_\alpha\left(W_{\rho_S(\epsilon)}\right) = \frac{1}{1-\alpha} \log \left[8\left(\frac{1}{6} - \frac{1}{18}\epsilon\right)^\alpha + \left(-\frac{1}{3} + \frac{4}{9}\epsilon\right)^\alpha\right] \tag{3.64}$$

We plot this entropy as a function of the index α and the error ϵ in Fig. (3.2). We see that $H_\alpha\left(W_{\rho_S(\epsilon)}\right)$ attains a negative value for α in the neighborhood of 1, at any value

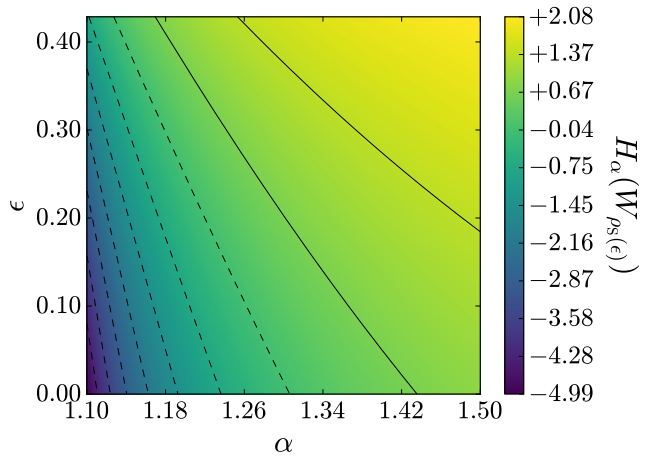

Figure 3.2: **Rényi entropy H_α of the ϵ–noisy Strange state.** H_α is increasing in α and ϵ for the ϵ–noisy Strange state of Eq. (3.30). The rightmost dashed contour line corresponds to $H_\alpha(W_{\rho_S(\epsilon)}) = 0$. In particular, $H_\alpha(W_{\rho_S(0)}) = 0$ occurs at $\alpha \approx 1.31$.

of error ϵ for which $W_{\rho_S(\epsilon)}(\boldsymbol{z})$ also contains a negativity. This demonstrates Claim 4.9 that we prove later on, which establishes an equivalence between a negative Wigner representation and a negative-valued entropy. We see in Fig. (3.2) that the more noisy the state is, the closer it lies to the set of free states, therefore one needs to restrict to a smaller neighborhood of α to find entropic negativity.

3.4.2 Rényi divergences on quasi-distributions

We now define the α–Rényi divergence, or relative entropy, for a quasi-distribution $\boldsymbol{w} \in \mathcal{Q}_N$ relative to a full-rank probability distribution $\boldsymbol{r} \in \mathcal{Q}_N^+$ as

$$D_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r}) \coloneqq \frac{1}{\alpha - 1} \log \sum_i w_i^\alpha r_i^{1-\alpha}, \tag{3.65}$$

where $\alpha = 2a/(2b-1)$ for positive integers $a \geq b$. The min-relative divergence is the limit of the α–Rényi divergence, when $\alpha \to \infty$,

$$D_\infty(\boldsymbol{w} \,||\, \boldsymbol{r}) \coloneqq \frac{1}{\alpha - 1} \log \sum_i w_i^\alpha r_i^{1-\alpha}. \tag{3.66}$$

The physical motivation for defining these divergences is to quantify the entropic distance between a magic state and a reference stabilizer state. Our magic bounds later on will depend on how this distance varies before and after the application of a magic-preserving channel. We therefore need to show that the divergence remains meaningful under data processing, which is the goal of the rest of this section.

It is clear that $D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r})$ and $D_\infty(\boldsymbol{w} \mid\mid \boldsymbol{r})$ remain well-defined when $\boldsymbol{w}$ is a quasi-distribution. We can further prove the following result that relates the Rényi divergence to the Rényi entropy on a dense subset of probability distributions.

Claim 3.16. *Let $\boldsymbol{w} \in \mathcal{Q}_N$ be a quasi-distribution and $\boldsymbol{r} \in \mathcal{Q}_N^+$ a probability distribution with positive rational entries given by $r_i = a_i/K$ for positive integers a_i and $K = \sum_{i=1}^N a_i$. Then,*

$$H_\alpha(\Gamma_{\boldsymbol{a}}(\boldsymbol{w})) = K - D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r}), \tag{3.67}$$

where $\alpha = 2a/(2b-1)$ for positive integers $a \geq b$.

Proof. For the given form of α and positive $r_i,\ i = 1, \ldots, N$, $D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r})$ is well-defined for all $\boldsymbol{w}$. From the definition of $\Gamma_{\boldsymbol{a}}$ we have

$$\Gamma_{\boldsymbol{a}}(\boldsymbol{w}) = \bigoplus_{i=1}^N w_i(1/a_i, 1/a_i, \ldots, 1/a_i). \tag{3.68}$$

This leads to

$$\begin{aligned} H_\alpha(\Gamma_{\boldsymbol{a}}(\boldsymbol{w})) &= \frac{1}{1-\alpha} \log \sum_{i=1}^N w_i^\alpha a_i^{1-\alpha} = K + \frac{1}{1-\alpha} \log \sum_{i=1}^N w_i^\alpha r_i^{1-\alpha} \\ &= K - D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r}), \end{aligned} \tag{3.69}$$

where in the first equality we used the definition of H_α (Eq. (3.53)), in the second equality we substituted $\alpha_i = r_i/K$, and in the last equality we used the definition of D_α (Eq. (3.65)). □

With this we now establish monotonicity for α–Rényi relative divergences.

Claim 3.17. *Let $\alpha = \frac{2a}{2b-1}$ for any positive integers a, b with $a \geq b$. Let $\boldsymbol{w} \in \mathcal{Q}_N$, $\boldsymbol{w}' \in \mathcal{Q}_{N'}$ and $\boldsymbol{r} \in \mathcal{Q}_N^+$, $\boldsymbol{r}' \in \mathcal{Q}_{N'}^+$ with positive rational components a_i/K and a_i'/K respectively. Then $D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r}) \geq D_\alpha(\boldsymbol{w}' \mid\mid \boldsymbol{r}')$, whenever $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$.*

Proof. The statement $(\boldsymbol{w}, \boldsymbol{r}) \succ (\boldsymbol{w}', \boldsymbol{r}')$ is equivalent to $\Gamma_{\boldsymbol{a}}(\boldsymbol{w}) \succ \Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')$. Therefore, due to Claim 3.14 we have

$$H_\alpha(\Gamma_{\boldsymbol{a}}(\boldsymbol{w})) \leq H_\alpha(\Gamma_{\boldsymbol{a}'}(\boldsymbol{w}')) \tag{3.70}$$

which, due to Claim 3.16, is equivalent to

$$D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r}) \geq D_\alpha(\boldsymbol{w}' \mid\mid \boldsymbol{r}'). \tag{3.71}$$

□

Since the rationals are dense in the real numbers we can assume that any Wigner representation considered in the following chapters has rational components without affecting results.

In this chapter we have rigorously extended relative majorization to the entire space of quasi-distributions of any finite dimension. Given the numerous applications of majorization on probability distributions in the context of quantum information, we expect that our mathematical extension to quasi-distributions can lead to new applications in tasks where characterizing non-classicalities is essential. Our extension is summarized in Claim 3.8, where all the equivalent statements listed constitute different ways of thinking about majorization of general quasi-distributions.

We make the first step in finding applications for this framework within the context of magic state transformations. In chapter 4 and chapter 5, majorization on Wigner representation of magic states will give rise to fundamental constraints on the resource cost of magic. For example, both $D_\alpha(\boldsymbol{w} \mid\mid \boldsymbol{r})$ (Eq. (3.65)) and $D_\infty(\boldsymbol{w} \mid\mid \boldsymbol{r})$ (Eq. (3.66)) will be shown to result in monotonicity conditions that lead to bounds on general magic distillation rates. In order to achieve this, the data-processing inequality we have derived in Claim 3.17 will be crucial.

Chapter 4

Entropic constraints on odd-dimensional magic distillation

In this chapter, we develop a framework to analyze magic state distillation protocols in odd dimensions, where explicit physics of the system is incorporated in the distillation bounds. This is achieved by considering how a given magic state protocol transforms a pair of quantum states – one being a noisy magic state, and the other a stabilizer state that is distinguished by the physics of the system, such as a state at a characteristic temperature, and acts as a reference state for the protocol. The magic distillation bounds can then be expressed in terms of the physics of the reference state (e.g. free energy changes in the case of temperature). Such bounds are of potential interest in assessing how physical features such as temperature, noise biases or fixed-point structure associated with restricted gate-sets constrain distillation protocols [140–146].

The approach we take is most closely aligned with resource theories of magic, although it differs in key ways. We obtain distillation upper bounds without employing monotones, as defined in Eq. (1.1), that quantify magic as a resource. We instead employ the statistical mechanical tools and recent work in single-shot resource theories [120, 127, 133, 135, 147–149], detailed in chapter 3. Our analysis relies on applying these tools on the discrete Wigner representation of quantum systems, which is detailed in chapter 2. Hence, our approach is more ambitious than previous attempts that develop monotones to characterize magic. Monotones are only capable of capturing properties of the input and output magic states. Instead, our approach incorporates all the characteristics of the magic states as well as reference stabilizer states which probe the action of the magic protocol itself on the phase space. We find that explicit properties of the protocol are thus reflected in our bounds.

Crucially, we focus on magic states with negativity in their Wigner representation, which is known to be a necessary condition for universality in the state-injection model [41, 42, 103, 150, 151]. However, taking a statistical mechanical perspective in this context raises a problem: the standard Boltzmann entropy is not defined for quasi-distributions. We circumvent this obstacle by making use of the majorization theory discussed in chapter 3, which constitutes a more fundamental tool in statistical mechanics and leads to the emergence of entropies.

We consider families of magic protocols with increasing complexity and analyze how increasingly more complex variants of majorization lead to bounds on magic distillation rates. The simplest case we consider is for magic protocols that lead to unital channels. These are a subset of protocols that leave invariant some distinguished state or have some equilibrium fixed-point structure. We then consider protocols that are non-equilibrium processes, but only generate sub-linear correlations that enable a simple description in the thermodynamic limit before giving entropic analysis that can be applied generally.

The relations between the sets of protocols considered are as follows:

$$\text{Unital} \subset \text{ local fixed-points} \subset \text{ sub-linear correlations} \subset \text{ general protocols.}$$

A general schematic of the protocols we consider is provided in Fig. (4.2).

We provide explicit bounds for magic protocols that generate unital channels in Eq. (4.5), as well as bounds that incorporate the temperature and Hamiltonian of the system in Claim 4.5. We also compare our bounds with monotone bounds, as shown in Fig. (4.3), and discuss extensions to more general scenarios.

We find that the analysis in the presence of negativity displays a range of features that do not appear in classical statistical mechanics, and leads to a picture of Wigner negativity in a quantum circuit being described as non-classical free energy that is processed under stochastic dynamics illustrated in Fig. (4.1). This is demonstrated by non-monotonic Lorenz curves as already displayed in Fig. (3.1), and by single-shot entropies for general Wigner representations that are well-defined, obey a data-processing inequality under operations that preserve Wigner positivity, such as Clifford operations, and can take on negative values, introduced in Eq. (2.45).

4.1 Constraints on fixed-point protocols

We begin with fixed-point protocols as they require the simpler form of d–majorization, instead of the full power of relative majorization, so they are simpler to visualize. Recall from section 3.2.3 that, given a theory of magic $\mathcal{R} = (\mathcal{F}, \mathcal{O})$, we can restrict

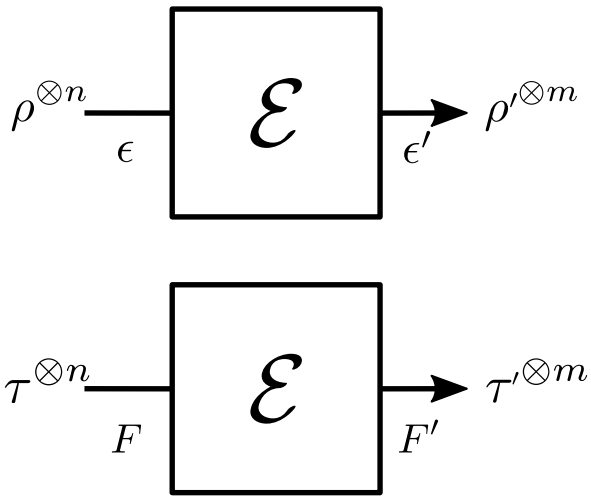

Figure 4.1: **Probing magic protocols with reference states.** A magic protocol (top figure) converts n copies of some noisy magic state ρ with error parameter ϵ to m copies of some less noisy state ρ' with $\epsilon' < \epsilon$. The physics of the protocol can be analyzed by considering how it would transform some distinguished reference stabilizer state $\tau^{\otimes n}$ (bottom figure). For the case where this reference state is a thermal state at some temperature T our magic distillation bounds are functions of T, the error parameters ϵ, ϵ', and free energies F, F'. In addition, a quantity ϕ appearing in the bound of Eq. (4.14) relates the computational and energy bases, and corresponds to the degree to which the Wigner negativity of ρ can be given a sharp energy via the system Hamiltonian.

to a sub-theory $\mathcal{R}_\sigma = (\mathcal{F}, \mathcal{O}_\sigma)$, where the free operations $\mathcal{O}_\sigma$ leave the state $\sigma \in \mathcal{F}$ invariant. This is in fact always possible as we now show.

Claim 4.1. *Let $\mathcal{R} = (\mathcal{F}, \mathcal{O})$ be a theory of magic. Then a transformation $\rho_1 \to \rho_2$ is possible in $\mathcal{R}$ if and only if the transformation $\rho_1 \to \rho_2$ is possible in at least one sub-theory $\mathcal{R}_\sigma$.*

Proof. Suppose the transformation is possible in a $\mathcal{R}_\sigma$ via some $\mathcal{E} \in \mathcal{O}_\sigma$. But since $\mathcal{O}_\sigma \subseteq \mathcal{O}$ it is also possible in $\mathcal{R}$. Conversely, suppose the transformation is possible in $\mathcal{R}$ via some $\mathcal{E}$ in $\mathcal{O}$. The free states $\mathcal{F}$ are a closed, bounded set and moreover the image of $\mathcal{F}$ under the map $\mathcal{E}$ is in $\mathcal{F}$. By Brouwer's fixed-point theorem [152], this mapping must therefore have a fixed point $\sigma \in \mathcal{F}$, so $\mathcal{E} \in \mathcal{O}_\sigma$ and the transformation is possible in $\mathcal{R}_\sigma$. □

We can compare this approach with resource monotones. A complete set of monotones $\{\mathcal{M}_i\}_i$ is defined so that $\rho \to \sigma$ if and only if $\mathcal{M}_i(\rho) \geq \mathcal{M}_i(\sigma)$ for all i. In our case the set of sub-theories $\{\mathcal{R}_\sigma\}_{\sigma \in \mathcal{F}}$ can therefore be viewed as a complete set of

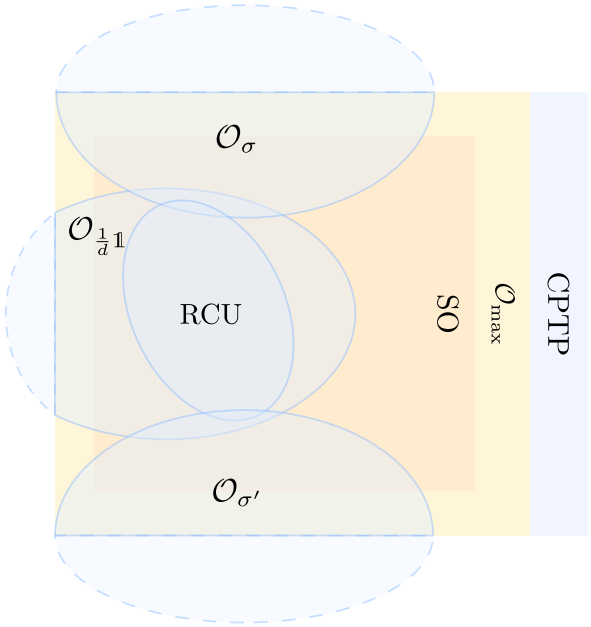

Figure 4.2: **Schematic of magic protocols in odd dimensions**. The set of the quantum channels which preserve Wigner positivity $\mathcal{O}_{\max}$ is the largest set of free operations for any magic protocol. It includes the stabilizer operations SO, which consist of stabilizer state preparations, random mixtures of Clifford unitaries (RCU) and Pauli measurements. We study sub-theories with operations in $\mathcal{O}_\sigma$, which cover all magic protocols according to Claim 4.1, but also include stochastic maps on the phase space outside the image of CPTP operations. $\mathcal{O}_{\mathbb{1}/d}$ includes the set of d-dimensional unital protocols.

of sub-theories fully characterizing the parent resource theory $\mathcal{R}$, where the complicated pre-order in $\mathcal{R}$ of states is now mapped to a simpler pre-order for a particular $\mathcal{R}_\sigma$. We proceed to show that a given $\mathcal{R}_\sigma$ can be approximated by a majorization pre-order, which is later exploited in section 4.1 to compute magic distillation bounds under distillation protocols with a fixed point.

4.1.1 General properties of fixed points

Here, we state some generic aspects of Lorenz curves for magic states, which allow us to interpret previous magic monotones as features of the curves. The first result gives a simple way to see that the mana (Eq. (2.33)) and sum-negativity (Eq. (2.34)) of a magic state is a monotone [104]. It is then evident that these two monotones naturally arise due to more holistic majorization constraints.

Claim 4.2. *Majorization in any $\mathcal{R}_\sigma$ implies the monotonicity of sum-negativity/mana.*

Proof. The sum-negativity of a magic state ρ can be written as $sn(\rho) = \frac{1}{2}(\sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})| - 1)$. We make use of the L_1-norm form of relative majorization presented in Claim 3.8, which states that $\boldsymbol{w} \succ_{\boldsymbol{r}} \boldsymbol{w}'$ if and only if

$$\sum_i |w_i - r_i t| \geq \sum_i |q_i - r_i t|, \tag{4.1}$$

for all $t \in \mathbb{R}$. Choosing $t = 0$ we get the single condition that $\sum_i |p_i| \geq |\sum_i |q_i|$, independent of the choice of $\boldsymbol{r}$. Applying this to the Wigner representations of two quantum states immediately gives the result. □

If we have a magic state ρ that has negativity in its Wigner representation, then, as discussed, its Lorenz curve attains values larger than 1, reaching a maximum $L_\star > 1$ that depends on the particular state. There is a simple relation between $L_\star$ and sum-negativity/mana, which is provided by the following result.

Claim 4.3. *Given a magic state ρ, the maximum $L_\star$ of its Lorenz curve $L_{\rho|\sigma}(x)$ is independent of the $\mathcal{R}_\sigma$ and equal to $1 + \mathrm{sn}(\rho)$. Moreover, the majorization constraint is stronger than mana in every $\mathcal{R}_\sigma$.*

Proof. We denote the Wigner representations of the states single-component vectors $\boldsymbol{w}(\rho) = W_\rho(\boldsymbol{z})$ and $\boldsymbol{w}(\sigma) = W_\sigma(\boldsymbol{z})$. Likewise, we write $\boldsymbol{w}(\rho|\sigma) = W_{\rho|\tau}(\boldsymbol{z})$. We choose the component indexing so that $\boldsymbol{w}(\rho|\sigma)^\downarrow = \boldsymbol{w}(\rho|\sigma)$, so the components are sorted in non-increasing order.

Note that all components of $\boldsymbol{w}(\sigma)$ are positive, so $w(\rho|\sigma)_k \geq 0$ if and only if $w(\rho)_k \geq 0$ for any $k = 1, \ldots, d^2$.

Let $k_\star$ be the index of the smallest non-negative component of $\boldsymbol{w}(\rho|\sigma)^\downarrow$. Then, $w(\rho)_k < 0$ if and only if $k > k_\star$, so the maximum of Lorenz curve $L_{\rho|\sigma}(x)$ takes the value

$$L_{\rho|\sigma}(x_{k_\star}) = \sum_{k=1}^{k_\star} w(\rho)_k, \tag{4.2}$$

and is achieved at $x_{k_\star} \coloneqq \sum_{k=1}^{k_\star} w(\sigma)_k$. The location of this maximum varies with $\mathcal{R}_\sigma$, but its value is independent of σ,

$$\begin{aligned} L_\star \coloneqq L_{\rho|\sigma}(x_{k_\star}) &= \sum_{\boldsymbol{z}: W_\rho(\boldsymbol{z}) \geq 0} W_\rho(\boldsymbol{z}) \\ &= 1 + \mathrm{sn}(\rho). \end{aligned} \tag{4.3}$$

Since mana is a monotonic function of sum-negativity, $\mathcal{M}(\rho) = \log(2\mathrm{sn}(\rho) + 1)$, we see that mana determines the peak of the Lorenz curve $L_{\rho|\sigma}(x)$. However, mana is one of d^{2n} constraints, so majorization is strictly a stronger constraint in any $\mathcal{R}_\sigma$. □

Therefore, the mana of a state can be thought of as the maximum of its Lorenz curve, independent of the reource theory $\mathcal{R}_\sigma$. Using the Lorenz curve perspective, it is also simple to construct new magic monotones. For example, within a given $\mathcal{R}_\sigma$, the area above the horizontal line $y = 1$ is a magic monotone.

Claim 4.4. *Given a magic state ρ and a free state σ, let $\mathcal{A}_\sigma(\rho)$ be the area of the region $\{(x, y) : 1 \leq y \leq L_{\rho|\sigma}(x)\}$. Then $\mathcal{A}_\sigma(\rho)$ is a magic monotone for $\mathcal{R}_\sigma$.*

Proof. Consider the transformation $\rho_1 \to \rho_2$ within $\mathcal{R}_\sigma$. We have that $L_{\rho_1|\sigma}(x)$ is never below the curve $L_{\rho_2|\sigma}(x)$, and therefore the region above $L = 1$ for ρ_2 is a subset of the corresponding region for ρ_1. Thus, $\mathcal{A}_\sigma(\rho_1) \geq \mathcal{A}_\sigma(\rho_2)$, so $\mathcal{A}_\sigma$ is a magic monotone in $\mathcal{R}_\sigma$. □

Note though, in contrast to mana, the area monotone is specific to $\mathcal{R}_\sigma$, and its value will vary as we change σ. Therefore, its monotonicity depends on the physics of the fixed point and provides a means to analyze magic distillation only under free operations that leave σ invariant. We can see an example of this monotone within the resource theory $\mathcal{R}_{\mathbb{1}/3}$ in Fig. (3.1). For two noisy Strange states of different number of copies n and $n' > n$, the Lorenz curve of the n'–copy state lies above the Lorenz curve of the n–copy state, therefore $\mathcal{A}_{\mathbb{1}/3}(\rho_S(\epsilon)^{\otimes n'}) > \mathcal{A}_{\mathbb{1}/3}(\rho_S(\epsilon)^{\otimes n})$ accordingly, for a given error ϵ. One can also define an area monotone in the more general setting of relative majorization, where one drops the fixed point emphasis.

4.1.2 Unital protocols

We now construct magic state distillation bounds for protocols that generate unital channels, meaning that $\mathcal{E}(\mathbb{1}/d) = \mathbb{1}/d$, represented by the theory $\mathcal{R}_{\mathbb{1}/d}$. This is a broad family of channels and includes any noiseless unitary or Clifford dynamics on systems of dimension d. Our approach works for any odd d, but for simplicity we consider qutrit magic states ($d = 3$) and in particular the problem of purifying n copies of a noisy Strange state $\rho_S(\epsilon)^{\otimes n}$ into a smaller number of copies n' of a less noisy Strange state $\rho_S(\epsilon')^{\otimes n'}$, with $\epsilon' < \epsilon$ and $n' \leq n$. To this end, we consider the transformation

$$\rho_S(\epsilon)^{\otimes n} \longrightarrow \rho_S(\epsilon')^{\otimes n'}, \tag{4.4}$$

and use the Lorenz curves computed in section 3.2.3 in order to bound the magic distillation rate $R(\epsilon, \epsilon') \coloneqq n'/n$.

Distillation bounds can be obtained from any part of the Lorenz curve by imposing the constraint that the Lorenz curve of the input state never dips below the curve of the output. One simple constraint that proves sufficient for meaningful distillation

bounds can be obtained by considering the initial slopes of the two Lorenz curves. This is the first elbow constraint of Claim 3.11.

The coordinates of the first point after the origin for $\rho_S(\epsilon)^{\otimes n}$ as calculated in section 3.2.3 are given by $(1/9^n, v(\epsilon)^n)$, while for $\rho_S(\epsilon')^{\otimes n'}$ by $(1/9^{n'}, v(\epsilon')^{n'})$, and by requiring that the initial slope for the input state Lorenz curve is larger than that of the output state (Claim 3.11) we find that

$$R(\epsilon, \epsilon') \leq R_\infty := \frac{\log(3-4\epsilon)}{\log(3-4\epsilon')}. \tag{4.5}$$

The choice of denoting the bound by R_∞ is explained in the derivation of Eq. (4.77) in the context of single-shot entropies. For the limiting case of pure magic states on the output ($\epsilon' = 0$), this simplifies to

$$R \leq 1 + \frac{\log\left(1 - \frac{4}{3}\epsilon\right)}{\log 3}. \tag{4.6}$$

We compare this to two known distillation bounds based on magic monotones. The bound based on mana (Eq. (2.33)) can be directly calculated as

$$R \leq \frac{\mathcal{M}(\rho_S(\epsilon))}{\mathcal{M}(\rho_S(0))} = 1 + \frac{\log\left(1 - \frac{8}{15}\epsilon\right)}{\log \frac{5}{3}}. \tag{4.7}$$

The max–thauma [153] is defined as

$$\theta_{\max}(\rho) := \log \min\{2\,\mathrm{sn}(V) + 1 : V \geq \rho\}, \tag{4.8}$$

and can be calculated numerically via a semi-definite program. For the noisy Strange state, the max–thauma bound coincides with the mana bound, and we find that they are both looser than the majorization bound R_∞ as shown in Fig. (4.3).

This figure includes numerical estimates of the optimal majorization bounds R_{num} obtained by considering the full Lorenz curve, which show that the bound R_∞ can be further improved. While this is an improvement on prior results, all known distillation protocols have rates much lower than these upper bounds. It remains a major open question to determine what are the best possible rates that can be achieved.

4.2 Magic free energy

In this section, we generalize our analysis from fixed-point theories to relative majorization, where it is no longer necessary to identify the, commonly complicated, fixed points of the magic protocols. As this forms the subject of quantum thermodynamics in the case of proper probability distributions, we take the approach of drawing rigorous analogies between the fields of thermodynamics and magic, leading, for

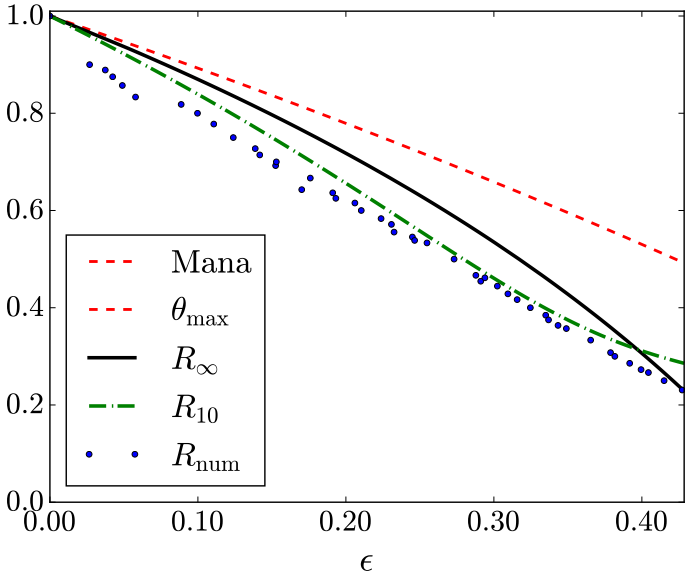

Figure 4.3: **Magic distillation bounds from majorization of Wigner representations.** The above plot shows distillation upper bounds for qutrits obtained from majorization on unital protocols. The constraints are plotted as a function of the depolarizing error ϵ for a noisy Strange state (Eq. (3.30)). Numerical bounds R_{num} are obtained by using the entirety of the Lorenz curves, namely not only the first elbow analysis, but the full set of majorization constraints. The bound R_∞ follows from a Lorenz curve analysis culminating in Eq. (4.5), while the bound R_{10} comes from single-shot entropies on the Wigner representations of the states, as discussed later in the derivation of Eq. (4.77). Bounds $R_\infty, R_{\text{num}}, R_{10}$ all provide stricter constraints than mana [104] and max–thauma $\theta_{\max}$ [153] bounds for noisy Strange states.

example, to a computational analog of free energy, which captures the Wigner negativity present in a magic state. Therefore, we see that magic constitutes the potential for quantum computational advantage, much like free energy constitute the potential for useful thermodynamic work. The way we proceed is by considering how the protocol would disturb a reference equilibrium state τ to some different state τ', if the protocol is applied to this reference state instead of the desired magic state.

We consider a magic distillation protocol on multiple identical qudits in a noisy magic state $\rho(\epsilon)$, with noise parameter ϵ, sending

$$\rho(\epsilon)^{\otimes n} \longrightarrow \mathcal{E}(\rho(\epsilon)^{\otimes n}) = \rho(\epsilon')^{\otimes m} \tag{4.9}$$

with $n \geq 1$, $\epsilon' < \epsilon$ and $\mathcal{E}$ denoting the quantum channel induced by the protocol. We also assume each qudit has a Hamiltonian H and neglect interaction terms. We choose some temperature $T = (k\beta)^{-1}$ where k is Boltzmann's constant and β the

inverse temperature. The reference equilibrium state of the n qudits is given by

$$\tau^{\otimes n} = \left(\frac{e^{-\beta H}}{\mathcal{Z}}\right)^{\otimes n}. \tag{4.10}$$

We may also assume that the reference state τ is not a magic state, and has a strictly positive Wigner representation $W_\tau(\boldsymbol{z})$.

4.2.1 Sub-linear correlations in the thermodynamic limit

We firstly provide motivation drawn from thermodynamics as to why we expect the output reference state to be in product form. Our general results in Claim 4.11 do not require this property, but it is helpful for further simplification of expressions.

A given magic protocol on the n qudits will correspond to a quantum channel $\mathcal{E}$. We also assume for simplicity that $U_\pi \mathcal{E}(X) U_\pi^\dagger = \mathcal{E}(X)$ for all X and any permutation U_π of the m output subsystems. This is justified because the protocol on the input magic state results in state $\rho(\epsilon')^{\otimes m}$, which is invariant under permutations. Therefore, we are always allowed to symmetrize the output $\mathcal{E}(X)$ by performing a group average over the permutation group for the output systems without changing the performance of the distillation protocol on $\rho(\epsilon)^{\otimes n}$. Thus, we can assume that $\mathcal{E}$ always outputs a symmetric state in general. This means that $\mathcal{E}(\tau^{\otimes n})$ is a symmetric state on m subsystems, so by the quantum de Finetti theorem [154, 155] we have that

$$\mathcal{E}(\tau) \approx \int d\mu(x) \tau_x'^{\otimes m}, \tag{4.11}$$

for $m \geq 1$, where $d\mu(x)$ is a probability measure over a set $\{\tau_x'\}$ of single qudit states.

To keep our analysis simple we make the following physical assumption. We assume that in the asymptotic/thermodynamic limit $n, m \to \infty$ the correlations generated on the reference equilibrium state are negligible. This implies that $d\mu(x)$ is peaked on a particular state τ', and $\mathcal{E}(\tau^{\otimes n}) \approx \tau'^{\otimes m}$. This scenario occurs in the context of traditional thermodynamics, and states that the output system is well-described by intrinsic variables that do not scale in m, and correlations are sub-linear in m. In particular, this allows us to compute a free energy per qudit of the output state.

Non-trivial correlations in the thermodynamic limit can also be considered, but leads to a more complex analysis within our majorization framework of Wigner representations. In this direction, we highlight recent work in majorization in which stochastic independence and correlations are analyzed. It has been shown [156] that stochastic independence (no correlations) of independent distributions can be

viewed as a resource in an extension of catalytic majorization, and leads to a single-shot operational interpretation of the Shannon entropy [157, 158].

Our bound depends on the von Neumann entropy $S(\rho)$ of a state, and the free energy at a particular temperature. For a state $\tau = e^{-\beta H}/\mathcal{Z}$, the Helmholtz free energy F is given by

$$F \coloneqq \mathrm{tr}[H\tau] - \beta^{-1} S(\tau) = -\beta^{-1} \log \mathcal{Z}, \tag{4.12}$$

which is obtained from the internal energy via a Legendre transform [159].

The protocol transforms the equilibrium as $\tau^{\otimes n} \to \mathcal{E}(\tau^{\otimes n}) = \tau'^{\otimes m}$. The protocol does not generate magic on its own, so we assume that the output state τ' is also a Wigner-positive state. However, this is generally a non-equilibrium state for the system. Despite this, it is useful to associate an effective Hamiltonian H' to the output state by considering the change $H \to H'$ such that equilibrium is restored at the reference temperature T. This Hamiltonian is defined by the expression $\tau' = e^{-\beta H'}/\mathcal{Z}'$, and has free energy $F' = -\beta^{-1} \log \mathcal{Z}'$.

The magic state protocol is now considered by how it transforms the pair of quantum states $(\rho^{\otimes n}, \tau^{\otimes n})$ and is then constrained by the relative majorization condition $(W_\rho, W_\tau) \succ (W_{\rho'}, W_{\tau'})$ that holds due to the protocol being a stochastic map in the Wigner representation.

4.2.2 Temperature-dependent bounds

Here, we provide an analysis of the pair of transformations $(\rho^{\otimes n}, \tau^{\otimes n}) \to (\rho'^{\otimes n}, \tau'^{\otimes n})$ based on the Lorenz curve condition, exploiting the fact that the output state is in product form in the presence of sub-linear correlations. In Claim 4.5, we obtain magic distillation bounds that combine computational measures ϵ, ϵ', with terms that depend on the Hamiltonian H and reference temperature T of the physical system. We state the result for the case of qutrits, where the magic state is taken as the noisy Strange state $\rho_S(\epsilon)$ (Eq. (3.30)), but a similar analysis works for general odd dimensional qudits.

Claim 4.5. *Consider a magic distillation protocol on qutrits that transforms n copies of an ϵ–noisy Strange state into m copies of an ϵ'–noisy Strange state, with depolarizing errors $\epsilon' \leq \epsilon \leq 3/7$. We also allow pre/post-processing by local Clifford unitaries.*

Let $T = (k\beta)^{-1}$ be any finite temperature for the physical system and let $H = \sum_{i \in \mathbb{Z}_3} E_i \, |E_i\rangle\langle E_i|$ be the Hamiltonian of each qutrit subsystem in its eigendecomposition. Assume that in the thermodynamic limit ($n, m \geq 1$), the protocol applied to the equilibrium state $\tau^{\otimes n} = (e^{-\beta H}/\mathcal{Z})^{\otimes n}$ maps $\tau^{\otimes n} \longrightarrow \tau'^{\otimes m}$, where we write $\tau' = e^{-\beta H'}/\mathcal{Z}'$ for some Hermitian H'.

Then the asymptotic magic distillation rate $R = m/n$ is bounded as

$$R \leq \frac{\log\left(1 - \frac{4}{3}\epsilon\right) + \beta(\phi - F)}{\log\left(1 - \frac{4}{3}\epsilon'\right) + \beta(\phi' - F')}, \tag{4.13}$$

where F is the free energy of τ, and

$$\phi \coloneqq -\beta^{-1} \log \zeta \tag{4.14}$$

with ζ given by the expressions

$$\zeta \coloneqq \sum_{i \in \mathbb{Z}_3} \alpha_i e^{-\beta E_i}, \tag{4.15}$$

$$\alpha_k \coloneqq \langle E_k | \, A_{\boldsymbol{z}_\star} \, | E_k \rangle \,, \tag{4.16}$$

and $W_\tau(\boldsymbol{z})$ attaining a minimum at $\boldsymbol{z} = \boldsymbol{z}_\star$. The primed variables are defined similarly for the output system.

Proof. For the sake of clarity, we write $\rho_n \coloneqq \rho_S(\epsilon)^{\otimes n}$, $\rho'_m \coloneqq \rho_S(\epsilon')^{\otimes m}$, $\tau_n \coloneqq \tau^{\otimes m}$ and $\tau'_m \coloneqq \tau'^{\otimes m}$. We also assume, without loss of generality on the asymptotic distillation rate, that n and m are both even.

To establish the distillation bound we consider the distillation protocol that gives $\mathcal{E}(\rho_n) = \rho'_m$ for the magic states. We then consider that the protocol transforms the reference equilibrium state as $\mathcal{E}(\tau_n) = \tau'_m$. Since we have a finite temperature we have that τ and τ' are full rank stabilizer states and have a strictly positive Wigner representation. The input and output magic states generally have Wigner representations that are quasi-distributions. For any such protocol we thus have

$$(W_{\rho_n}(\boldsymbol{z}), W_{\tau_n}(\boldsymbol{z})) \succ (W_{\rho'_m}(\boldsymbol{z}), W_{\tau'_m}(\boldsymbol{z})), \tag{4.17}$$

or, equivalently, in terms of the relevant Lorenz curves,

$$L_{\rho_n|\tau_n}(x) \geq L_{\rho'_m|\tau'_m}(x) \text{ for all } x \in [0, 1]. \tag{4.18}$$

We now consider the rescaled Wigner representation $W_{\rho|\tau}(\boldsymbol{z}) \coloneqq W_\rho(\boldsymbol{z})/W_\tau(\boldsymbol{z})$, which is well-defined since τ is full-rank. Due to the multiplicative property of the Wigner representation, the rescaled distribution is also multiplicative in the sense that

$$W_{\rho\otimes\rho'|\tau\otimes\tau'}(\boldsymbol{z}_1 \oplus \boldsymbol{z}_2) = W_{\rho|\tau}(\boldsymbol{z}_1) W_{\rho'|\tau'}(\boldsymbol{z}_2), \tag{4.19}$$

for any states ρ, ρ' and any full-rank stabilizer states τ, τ'. Therefore, we have that

$$W_{\rho_n|\tau_n}(\boldsymbol{z}) = \prod_{i=1}^{n} W_{\rho|\tau}(\boldsymbol{z}_i) \tag{4.20}$$

$$W_{\rho_n}(\boldsymbol{z}) = \prod_{i=1}^{n} W_{\rho}(\boldsymbol{z}_i) \tag{4.21}$$

where $\boldsymbol{z} = \oplus_{i=1}^{n} \boldsymbol{z}_i \in \mathbb{Z}_3^{2n}$ is the phase space point for the full system in terms of those of the individual subsystems.

The points defining the Lorenz curve $L_{\rho_n|\tau_n}(x)$ are obtained from sorting the components of $W_{\rho_n|\tau_n}(\boldsymbol{z})$ in non-increasing order and then computing the partial sums of $W_{\rho_n|\tau_n}(\pi(\boldsymbol{z}))$ where π is the permutation that realizes the sorting. However, similar to the unital protocol analysis, we use the slope constraint (Claim 3.11) obtained by considering the line segments connecting the origin to the first elbow of both Lorenz curves.

The Wigner representation of a single noisy Strange state consists of 1 negative component $W_\rho(\mathbf{0}) = -v(\epsilon)$ and 8 positive components $W_\rho(\boldsymbol{z}) = u(\epsilon)$ for $\boldsymbol{z} \neq \mathbf{0}$. The Wigner representation of the full-rank, stabilizer equilibrium state τ is $W_\tau(\boldsymbol{z}) > 0$ for all $\boldsymbol{z} \in \mathbb{Z}_3^2$.

Assume that the smallest component of the distribution $W_\tau(\boldsymbol{z})$ is at $\boldsymbol{z} = \boldsymbol{z}_\star$. Since the magic content of the state is unchanged under a Clifford unitary C, we can instead consider the state

$$\rho_S(\epsilon) \to C\rho_S(\epsilon)C^\dagger = D_{\boldsymbol{z}_\star}\rho_S(\epsilon)D_{\boldsymbol{z}_\star}^\dagger, \tag{4.22}$$

which has its single negative Wigner component $-v(\epsilon)$ at the point $\boldsymbol{z}_\star$.

The components of the rescaled distribution $W_{\rho_n|\tau_n}$ for this transformed state are given by

$$\left(\frac{-v}{W_\tau(\boldsymbol{z}_\star)}\right)^{i_{\boldsymbol{z}_\star}} \prod_{\boldsymbol{z}\neq\boldsymbol{z}_\star} \left(\frac{u}{W_\tau(\boldsymbol{z})}\right)^{i_{\boldsymbol{z}}}, \tag{4.23}$$

where the integer indices obey the following conditions:

$$0 \leq i_{\boldsymbol{z}} \leq n \text{ for all } \boldsymbol{z} \in \mathbb{Z}_3^2, \text{ and } \sum_{\boldsymbol{z}\in\mathbb{Z}_3^2} i_{\boldsymbol{z}} = n. \tag{4.24}$$

We now compute the largest rescaled component. Firstly, note that n is even, so we require that $i_{\boldsymbol{z}_\star} \in \{0, 2, \ldots, n\}$ for the component to be positive. Then, we have that $v \geq u$ because $\epsilon \leq 3/7$, and we have already ensured that $W_\tau(\boldsymbol{z}_\star) \leq W_\tau(\boldsymbol{z})$ for all $\boldsymbol{z} \in \mathbb{Z}_3^2$. Therefore, the largest rescaled component occurs when $i_{\boldsymbol{z}_\star} = n$ and $i_{\boldsymbol{z}} = 0$ for $\boldsymbol{z} \neq \boldsymbol{z}_\star$ and is equal to $(v/W_\tau(\boldsymbol{z}_\star))^n$. The coordinates of the first Lorenz curve point after the origin are given by

$$(x_0, L_0) = ((W_\tau(\boldsymbol{z}_\star))^n, v^n) \tag{4.25}$$

Given a Hamiltonian decomposition,

$$H = \sum_{i\in\mathbb{Z}_3} E_i \left|E_i\right\rangle\!\left\langle E_i\right|, \tag{4.26}$$

we now express the coordinates of the first point in terms of free energy quantities. We expand as follows

$$W_\tau(\boldsymbol{z}_\star) = \frac{1}{3\mathcal{Z}}\mathrm{tr}\left[A_{\boldsymbol{z}_\star}e^{-\beta H}\right] = \frac{e^{\beta F}}{3}\mathrm{tr}\left[A_{\boldsymbol{z}_\star}\sum_{i\in\mathbb{Z}_3}e^{-\beta E_i}\left|E_i\right\rangle\!\left\langle E_i\right|\right]$$
$$= \frac{e^{\beta F}}{3}\sum_{i\in\mathbb{Z}_3}\alpha_i e^{-\beta E_i} = \frac{e^{\beta F}}{3}\zeta = \frac{e^{\beta(F-\phi)}}{3}, \tag{4.27}$$

where in the first equality we use the definition of the Wigner function, the second equality follows from the decomposition in Eq. (4.26), the third from the definition of α_k in Eq. (4.16), the fourth from the definition of ζ in Eq. (4.15) and, finally, the last equality follows from the definition of ϕ in Eq. (4.14). We can view ϕ as a *magic free energy* in the sense that its definition is analogous to the classical free energy. We can now express the coordinates of the input state using Eq. (4.25) as

$$(x_0, L_0) = \left(\frac{e^{n\beta(F-\phi)}}{3^n}, v(\epsilon)^n\right). \tag{4.28}$$

The output state obeys the same conditions, so we can express the coordinates of the first elbow of the output Lorenz curve as

$$(x_0', L_0') = \left(\frac{e^{m\beta(F'-\phi')}}{3^m}, v(\epsilon')^m\right). \tag{4.29}$$

If the largest rescaled component of a state is distinct with no multiplicities, then these coordinates correspond to the first elbow of the corresponding Lorenz curve, whereas if it appears multiple times, then the coordinates derived correspond to a point on the interior of the line segment connecting the origin to the first elbow. In both cases, the distillation bound remains the same, as is clear by its derivation in Claim 3.11, given by $L_0/x_0 \geq L_0'/x_0'$, leading to

$$\left(3v(\epsilon)e^{-\beta(F-\phi)}\right)^n = \frac{L_0}{x_0} \geq \frac{L_0'}{x_0'} = \left(3v(\epsilon')e^{-\beta(F'-\phi')}\right)^m. \tag{4.30}$$

We note that $3v(\epsilon')e^{-\beta(F'-\phi')} > 1$ always, as $\rho_S(\epsilon')$ is not a free state, so the initial slope of its Lorenz curve exceeds 1. Taking the natural logarithm on both sides and rearranging gives the bound

$$\frac{m}{n} \leq \frac{\log\left(1-\frac{4}{3}\epsilon\right)+\beta(\phi-F)}{\log\left(1-\frac{4}{3}\epsilon'\right)+\beta(\phi'-F')}, \tag{4.31}$$

which completes the proof. □

The following result shows that if the energy eigenbasis is a stabilizer basis then the bound in Eq. (4.13) simplify further.

Claim 4.6. *Let H be a Hamiltonian H with spectrum $\{E_k\}_{k\in\mathbb{Z}_d}$. If H has a stabilizer eigenbasis, then $\phi = E_{k_\star}$ for some $k_\star \in \mathbb{Z}_d$.*

Proof. We first write α_k as

$$\alpha_k = \langle E_k| \, |A_{\boldsymbol{z}_\star}| \, |E_k\rangle = dW_{|E_k\rangle\langle E_k|}(\boldsymbol{z}_\star) \tag{4.32}$$

Since each $|E_k\rangle\langle E_k|$ is a stabilizer state, their Wigner representations have orthogonal supports in the phase space, so only one of them, say $k = k_\star$ contains point $\boldsymbol{z}_\star$ in its support. Additionally, its distribution is uniform on its support, so $W_{|E_{k_\star}\rangle\langle E_{k_\star}|}(\boldsymbol{z}_\star) = 1/d$.

Therefore, only the coefficient $\alpha_{k_\star}$ is non-zero and

$$\phi = -\beta^{-1}\log\left(\sum_i \alpha_i e^{-\beta E_i}\right) = E_{k_\star} - kT\log\big(dW_{|E_{k_\star}\rangle\langle E_{k_\star}|}(\boldsymbol{z}_\star)\big) = E_{k_\star}. \tag{4.33}$$

□

In order to isolate the effect of majorization on the magic distillation process, we drop the freedom to pre/post-process the magic states via Clifford unitaries in Claim 4.7. We illustrate explicitly that this comes at the expense of a more complex final expression for the bounds due to the dependence of the largest component in the rescaled distribution on a number of factors.

Claim 4.7. *Consider a magic distillation protocol on qutrits that transforms*

$$\rho_S(\epsilon)^{\otimes n} \longrightarrow \mathcal{E}(\rho_S(\epsilon)^{\otimes n}) = \rho_S(\epsilon')^{\otimes m}$$

with $n, m \gg 1$.

Let each qutrit have a Hamiltonian H with energies E_0, E_1, E_2 and stabilizer eigenstates, and define $E_{\max} = \max\{E_0, E_1, E_2\}$ and E_s as the eigenvalue of the Hamiltonian eigenstate whose Wigner representation overlaps the negative component of $\rho_S(\epsilon)$ in the phase space. Let $T = (k\beta)^{-1}$ be any finite temperature and assume that in the thermodynamic limit ($n, m \gg 1$), the protocol applied to the equilibrium state $\tau^{\otimes n} = (e^{-\beta H}/\mathcal{Z})^{\otimes n}$ maps $\tau^{\otimes n} \longrightarrow \tau'^{\otimes m}$, where we express state τ' as $\tau' = e^{-\beta H'}/\mathcal{Z}'$ for some Hermitian H'.

Define $\beta_\star = (kT_\star)^{-1}$ through the relation

$$E_{\max} - E_s =: kT_\star \ln 2, \tag{4.34}$$

and define a threshold error,

$$\epsilon_\star(\beta) := \begin{cases} 3 - \dfrac{9}{4 - 2^{\beta/\beta_\star - 1}}, & \text{for } \beta \leq \beta_\star \\ 0, & \text{for } \beta > \beta_\star. \end{cases} \tag{4.35}$$

Then, the distillation rate $R = m/n$ of the magic protocol is bounded as:

$$R \leq \begin{cases} \dfrac{\ln\left(1 - \frac{4}{3}\epsilon\right) + \beta(E_s - F)}{\ln\left(1 - \frac{4}{3}\epsilon'\right) + \beta(E'_s - F')}, & \epsilon \leq \epsilon_\star, \epsilon' \leq \epsilon'_\star, \\ \dfrac{\ln\left(1 - \frac{4}{3}\epsilon\right) + \beta(E_s - F)}{\ln\left(\frac{1}{2} - \frac{1}{6}\epsilon'\right) + \beta(E'_{\max} - F')}, & \epsilon \leq \epsilon_\star, \epsilon' > \epsilon'_\star, \\ \dfrac{\ln\left(\frac{1}{2} - \frac{1}{6}\epsilon\right) + \beta(E_{\max} - F)}{\ln\left(1 - \frac{4}{3}\epsilon'\right) + \beta(E'_s - F')}, & \epsilon > \epsilon_\star, \epsilon' \leq \epsilon'_\star, \\ \dfrac{\ln\left(\frac{1}{2} - \frac{1}{6}\epsilon\right) + \beta(E_{\max} - F)}{\ln\left(\frac{1}{2} - \frac{1}{6}\epsilon'\right) + \beta(E'_{\max} - F')}, & \epsilon > \epsilon_\star, \epsilon' > \epsilon'_\star, \end{cases} \tag{4.36}$$

where F is the free energy of state τ and all primed quantities are similarly defined for the ouptut system.

Proof. The proof proceeds in the same manner as in Claim 4.5 up to the point where we need to evaluate $W_{\rho_n|\tau_n}(\boldsymbol{z})$ and $W_{\rho_n}(\boldsymbol{z})$ explicitly, beyond which we do not have the freedom of additional Clifford processing.

The equilibrium state at inverse temperature β on a single qutrit is given by $\tau = e^{-\beta H}/\mathcal{Z}$. Moreover, we have that τ is a full-rank stabilizer state, where $\beta \geq 0$ and $H = E_0 \left|\phi_0\right\rangle\!\left\langle\phi_0\right| + E_1 \left|\phi_1\right\rangle\!\left\langle\phi_1\right| + E_2 \left|\phi_2\right\rangle\!\left\langle\phi_2\right|$ is an eigendecomposition of H. The state τ can now be written as

$$\tau = \frac{e^{-\beta E_0}}{\mathcal{Z}} \left|\phi_0\right\rangle\!\left\langle\phi_0\right| + \frac{e^{-\beta E_1}}{\mathcal{Z}} \left|\phi_1\right\rangle\!\left\langle\phi_1\right| + \frac{e^{-\beta E_2}}{\mathcal{Z}} \left|\phi_2\right\rangle\!\left\langle\phi_2\right|, \tag{4.37}$$

where the eigenstates $\{\left|\phi_k\right\rangle\!\left\langle\phi_k\right|\}$ are pure, orthonormal stabilizer states, which can be represented in terms of generalized Pauli operators. We are free to redefine the computational basis so it coincides with the basis of H. Abstractly, let C be the unitary transforming each $\left|\phi_k\right\rangle\!\left\langle\phi_k\right|$ to $\left|k\right\rangle\!\left\langle k\right|$. Since the Clifford group is the normalizer of the Heisenberg-Weyl group, C is a Clifford unitary. Therefore, C maps τ to another stabilizer state that is diagonal in the computational basis, and we can assume without loss of generality that τ is diagonal in $\{\left|0\right\rangle, \left|1\right\rangle, \left|2\right\rangle\}$. This re-definition of the computational basis means that the coordinates on the discrete phase space are also permuted, so the negative Wigner component $-v(\epsilon)$ of the Strange state is in a new position. We denote by E_s the eigenvalue of H where the associated eigenvector has

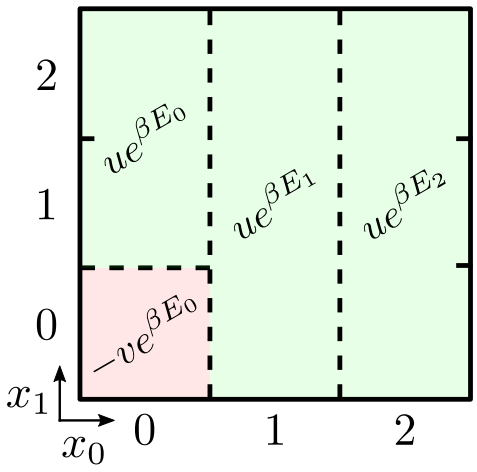

Figure 4.4: **Qutrit phase space regions for $W_{\rho|\tau}(\boldsymbol{z})$.** Here, the negative component of the magic state overlaps the Wigner representation of $|0\rangle$. The rescaled distribution attains a single value in each of the four regions, proportional to the value depicted in the region, see Eq. (4.42).

Wigner representation overlapping the negative component of the state $C\rho_S(\epsilon)C^\dagger$ in the original phase space, hence also of the state Strange state $\rho_S(\epsilon)$ in the re-defined phase space. The eigenvalue E_s is unique, since the eigenstates form an orthonormal basis.

The Wigner representation of state τ is then given by

$$W_\tau(\boldsymbol{z}) = \sum_{k=0}^{2} \frac{e^{-\beta E_k}}{\mathcal{Z}} W_{|k\rangle\langle k|}(q,p) = \sum_{k=0}^{2} \frac{e^{-\beta E_k}}{\mathcal{Z}} \delta_{q,k} = \frac{e^{-\beta E_q}}{3\mathcal{Z}}, \tag{4.38}$$

where x labels one of the three vertical lines in the phase space. The rescaled Wigner representation $W_{\rho|\tau}(\boldsymbol{z})$ can now be computed. It has 9 components, but these come with multiplicities. In total, there are four distinct values on the phase space, as illustrated in Fig. (4.4).

We define vectors $\boldsymbol{w}(\rho), \boldsymbol{w}(\tau), \boldsymbol{w}(\rho|\tau)$ based on the values occurring in $W_\rho, W_\tau, W_{\rho|\tau}$ respectively and $\boldsymbol{m}$ as the vector of associated multiplicities of each value in $W_\rho(\boldsymbol{z})$. Specifically, the component values and multiplicities of the relevant distributions in the four distinct regions are given by

$$\boldsymbol{m} \coloneqq (1,2,3,3), \tag{4.39}$$

$$\boldsymbol{w}(\rho) \coloneqq (-v,u,u,u), \tag{4.40}$$

$$\boldsymbol{w}(\tau) \coloneqq \frac{1}{3\mathcal{Z}} \left(e^{-\beta E_0}, e^{-\beta E_0}, e^{-\beta E_1}, e^{-\beta E_2}\right), \tag{4.41}$$

$$\boldsymbol{w}(\rho|\tau) \coloneqq 3\mathcal{Z} \left(-ve^{\beta E_0}, ue^{\beta E_0}, ue^{\beta E_1}, ue^{\beta E_2}\right). \tag{4.42}$$

Using this notation, the values and multiplicities of the n–copy distribution $\boldsymbol{w}(\rho_n|\tau_n)$ are computed according to the component-multiplicity pair analysis of section 3.1.3. The values are given by

$$[\boldsymbol{w}(\rho_n|\tau_n)]_{ijk} = (3\mathcal{Z})^n(-v)^{n-\alpha}u^\alpha e^{\beta(n-\alpha)E_s}e^{\beta(iE_0+jE_1+kE_2)}, \tag{4.43}$$

where the indices i, j, k are non-negative integers that obey the constraint $\alpha \coloneqq i + j + k \leq n$. The multiplicity of this above value is m_{ijk} with

$$m_{ijk} = \frac{n!}{i!j!k!(n-\alpha)!}2^i 3^j 3^k. \tag{4.44}$$

The associated components of $\boldsymbol{w}(\rho_n)$ are given by

$$[\boldsymbol{w}(\rho_n)]_{ijk} = (-v)^{n-\alpha}u^\alpha, \tag{4.45}$$

$$[\boldsymbol{w}(\tau)]_{ijk} = (3\mathcal{Z})^{-n}e^{-\beta(n-\alpha)E_s}e^{-\beta(iE_0+jE_1+kE_2)}. \tag{4.46}$$

In order to construct the n–copy Lorenz curve $L_{\rho_n|\tau_n}(x)$ we need to order the components of the distribution, $w(\rho|\tau)_{ijk}$ in decreasing order, and identify the sequence of indices that give us $W_{\rho_n}(\pi(\boldsymbol{z}))$.

However, in order to obtain a non-trivial bound, it again suffices to use the constraint from the first elbow (x_0, L_0) of $L_{\rho_n|\tau_n}(\boldsymbol{z})$. In order to calculate the coordinates of the first elbow, we compute the largest component

$$w_{\max} \coloneqq \max_{i,j,k}[\boldsymbol{w}(\rho_n|\tau_n)]_{ijk}, \tag{4.47}$$

and determine the indices at which this occurs. Putting in the values from Eq. (4.45) and Eq. (4.46), we obtain

$$(3\mathcal{Z})^{-n}w_{\max} = \max_{i,j,k}\left\{(-v)^{n-\alpha}u^\alpha e^{\beta(n-\alpha)E_s}e^{\beta(iE_0+jE_1+kE_2)}\right\}, \tag{4.48}$$

where $0 \leq i, j, k \leq n$ and $\alpha \coloneqq i + j + k \leq n$. Now for $0 \leq \epsilon \leq 3/7$, we have $v \geq u$. Since we assume that n is even, we need the sum $\alpha = i + j + k$ to be even too, so that the expression is positive.

Given an even value for α, the term $v^{n-\alpha}u^\alpha e^{-\beta(n-\alpha)E_s}$ is fixed, so the expression is maximized by setting the coefficient of the highest energy $E_{\max}$ equal to α. Hence, we have

$$w_{\max} = (3\mathcal{Z})^n v^n e^{n\beta E_s} \max_{\substack{\alpha=0,2,\\ \ldots,n-2,n}}\left\{\left(\frac{u}{v}e^{\beta(E_{\max}-E_s)}\right)^\alpha\right\}. \tag{4.49}$$

If the expression $\frac{u(\epsilon)}{v(\epsilon)}e^{\beta(E_{\max}-E_s)}$ is less than 1 then the maximum occurs at $\alpha = 0$, otherwise it occurs at $\alpha = n$. For a fixed state τ, this transition is determined by the

value of the depolarizing error parameter ϵ of the noisy magic state. The transition occurs at $\epsilon = \epsilon_\star$ where

$$\frac{u(\epsilon_\star)}{v(\epsilon_\star)} e^{\beta(E_{\max}-E_s)} = \frac{3-\epsilon_\star}{6-8\epsilon_\star} e^{\beta(E_{\max}-E_s)} = 1. \tag{4.50}$$

If $E_{\max} = E_s$, namely if the state negativity lies in the same phase space region as the highest energy, this threshold is constant in temperature and given by $\epsilon_\star = 3/7$. However, the condition that $\epsilon_\star \geq 0$ also implies a constraint on the effective temperature of the stabilizer state. Specifically, there is a threshold temperature value $\beta_\star$ given by

$$\beta_\star := \frac{1}{E_{\max} - E_s} \ln 2, \tag{4.51}$$

such that for the regime $0 \leq \beta \leq \beta_\star$ a threshold error $\epsilon_\star$ exists, and for $\beta > \beta_\star$ no such transition exists, so we choose $\epsilon_\star = 0$. Therefore, the transition value for the error is given by

$$\epsilon_\star(\beta) := \begin{cases} 3 - \dfrac{9}{4 - 2^{\beta/\beta_\star - 1}}, & \text{for } \beta \leq \beta_\star \\ 0, & \text{for } \beta > \beta_\star. \end{cases} \tag{4.52}$$

The quantity $w(\rho_S|\sigma)_{\max}$ is now given by

$$w_{\max} = \begin{cases} (3\mathcal{Z})^n v^n e^{n\beta E_s}, & \text{if } \epsilon \leq \epsilon_\star, \ \ (\mathrm{C1}) \\ (3\mathcal{Z})^n u^n e^{n\beta E_{\max}}, & \text{if } \epsilon > \epsilon_\star. \ \ (\mathrm{C2}) \end{cases}$$

Case $(\mathrm{C1})$ corresponds to $(i,j,k) = (0,0,0)$, so the multiplicity is $m_{000} = 1$, while Case $(\mathrm{C2})$ corresponds to

$$(i,j,k) = \begin{cases} (0,n,0), & \text{if } E_{\max} = E_1, \\ (0,0,n), & \text{if } E_{\max} = E_2, \end{cases} \tag{4.53}$$

so the multiplicity in both cases is 3^n.

Using that $F = -\beta^{-1} \log \mathcal{Z}$, the first elbow coordinates in the two cases are now given by

$$(x_0, L_0) = \begin{cases} \left(\frac{1}{3^n} e^{-n\beta(E_s - F)}, v^n\right), & \epsilon \leq \epsilon_\star \\ \left(e^{-n\beta(E_{\max} - F)}, (3u)^n\right). & \epsilon > \epsilon_\star \end{cases} \tag{4.54}$$

Similarly, considering the output magic state with respect to state σ', the image of equilibrium state σ under the magic protocol, we get output Lorenz curve coordinates,

$$(x_0', L_0') = \begin{cases} \left(\frac{1}{3^{n'}} e^{-n\beta(E_s' - F')}, v(\epsilon')^{n'}\right), & \epsilon' \leq \epsilon_\star' \\ \left(e^{-n'\beta(E_{\max}' - F')}, (3u(\epsilon'))^{n'}\right), & \epsilon' > \epsilon_\star' \end{cases} \tag{4.55}$$

There are four combinations of coordinates, depending on the error parameters ϵ, ϵ' for the input and output states. In each of these combinations, we simply use the first elbow constraint, as described in Claim 3.11, and manipulate the coordinates as in the proof of Claim 4.5, leading to the bounds in the statement of this theorem. □

We devote the next section to discussing the results of Claim 4.5 and Claim 4.7.

4.2.3 Remarks on parameter-dependence of bounds

The bounds provided in Claim 4.5 and Claim 4.7 depend on:

- Quantum computational measures ϵ, ϵ',
- Thermodynamic quantities β, F, F',
- Intermediate terms ϕ, ϕ'.

The intermediate terms specify how the energy eigenbasis of the system $\{|E_k\rangle\}$ relates to the computational stabilizer basis $\{|k\rangle\}$. In particular, the term ϕ quantifies to what degree a sharp energy value can be associated to the negativity in the Wigner representation. Its form is similar to F, which is why we say that ϕ can be viewed as a *magic free energy*.

The coefficients α_k may be negative for some k, when the Hamiltonian has non-stabilizer eigenstates, but the quantity ϕ is always well-defined, since the function ζ is always positive for τ in the interior of $\mathcal{F}$. The quantity ϕ can diverge if the state τ acquired zero Wigner components, which occurs for τ on the boundary of the set of Wigner-positive states, and is not defined outside of $\mathcal{F}$. Finally, if H has a stabilizer eigenbasis then $\phi = E_i$ for some i, independent of the temperature T, as seen in Claim 4.6.

Similar to the unital protocol bounds, the above result is based on only part of the Lorenz curves and so can certainly be tightened with further analysis. The primary role of the pre/post-processing by Clifford unitaries in Claim 4.5 is to simplify the form of the bound, and allow it to be expressed in terms of the free energy per particle in a way that does not depend on the parameters $\epsilon, \epsilon', \beta$ in a complex form. In Claim 4.7, we give bounds in which one does not include these Clifford changes of basis, thus having a more non-trivial dependency on the parameters.

The analysis makes other simplifying assumptions that could easily be dropped, at the price of more complex expressions. We could perform similar analysis for general qudits, and different choices of magic states, for example. It might also be

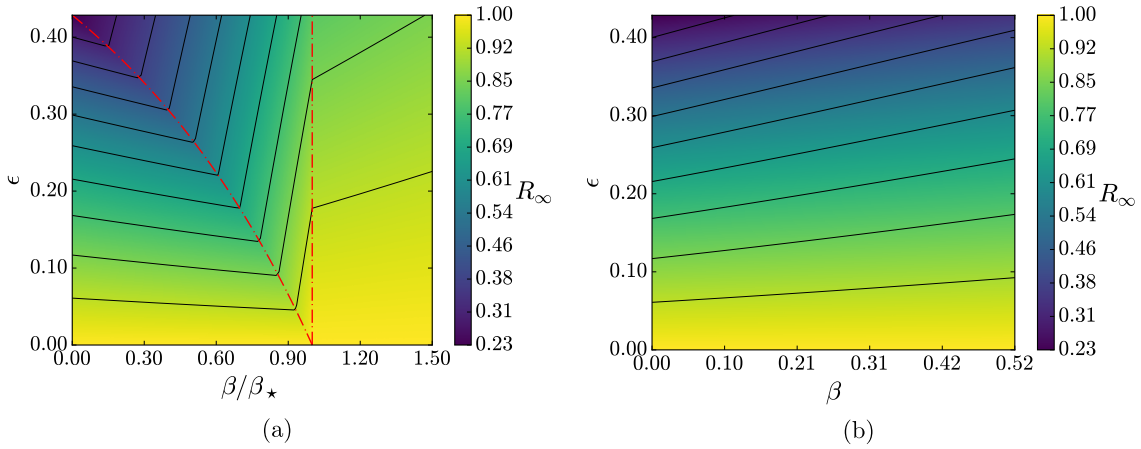

Figure 4.5: **Temperature-dependent bounds for magic distillation.** Shown are two contour plots of the bound on $R(\epsilon, \beta)$ for the case $H = H' = \sum_{k\in\mathbb{Z}_3} k\,|k\rangle\langle k|$ and $\epsilon' = 0$, where β is the inverse temperature and ϵ is the depolarizing error of the input magic state. Figure (a) does not use any changes of Clifford basis, and the form of the bound depends on both the error parameter and temperature. The curved dashed line is $\epsilon_\star(\beta)$ and given by Eq. (4.36) and $\beta_\star = (kT_\star)^{-1}$ is given by $E_2 - \phi = kT_\star \ln 2$. Figure (b) assumes Clifford processing is used resulting in a smoother bound. In both figures, the $\beta = 0$ line correspond to the unital bounds.

of interest to consider other choices of reference states τ that are more appropriate to the hardware physics, for example for photonic set-ups [160].

The assumption which is non-trivial is that we neglect correlations in the reference state in the thermodynamic limit. However, for more general scenarios one could make use of variational tools such as the Bogoliubov inequality [161] for approximating the free energy of a system via product states, to obtain similar bounds.

The simplest special case to consider is where the Hamiltonian H is diagonal in the computational basis, and $H = H'$, implying that the protocol leaves the reference equilibrium state unchanged, hence it corresponds to a Gibbs-preserving map [162]. For the limiting case of $\epsilon' \to 0$ we obtain Fig. (4.5) which is a contour plot of the bound on R as a function of inverse temperature β and the depolarizing error ϵ for the noisy input magic states. In this figure we show both the bounds without the power of Clifford pre/post-processing (in (a) – Eq. (4.36)) and with Clifford processing (in (b) – Eq. (4.13)). As a pedagogical explanation, we can consider a temperature limitation that allows a device to operate with thermal operations only below a certain β. Then, we would read the vertical line on the plot to decide what is the upper bound on the input noise level to achieve a certain distillation bound. In the more general case, $\Delta F := F' - F \neq 0$, so the protocol, when applied to the

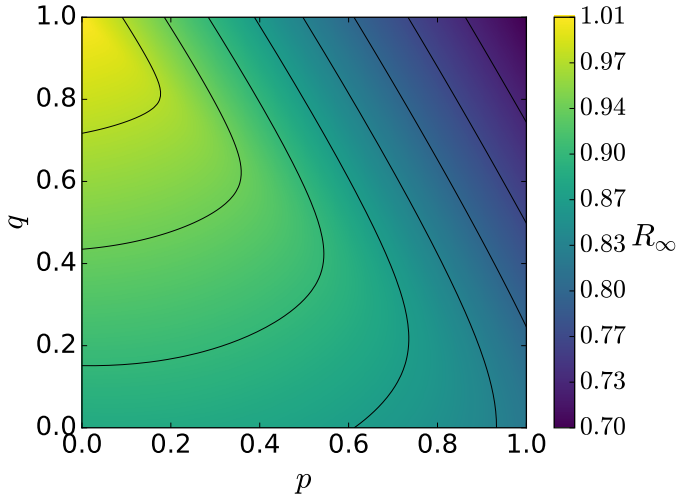

Figure 4.6: **Variation of distillation bounds with Hamiltonian.** We illustrate variation of the bound in Eq. (4.13) with respect to the system Hamiltonian. We fix $(\epsilon, \epsilon', \beta) = (0.1, 0.0, 0.2)$ and look at a family of Hamiltonians for the qutrit systems, given by $H = A_{\mathbf{0}}$ and $H' = (1-p-q)A_{\mathbf{0}} + pA_{(1,2)} + q\mathrm{diag}(0,1,2)$, with varying parameters p, q, and where $A_{\mathbf{0}}$ and $A_{(1,2)}$ are phase-point operators.

reference state, adds or extracts free energy from the system.

We demonstrate change of the bounds with respect to a varying output Hamiltonian example in Fig. (4.6). In particular, we create a parameterized family of Hamiltonians and explore how our bounds vary in terms of the parameters. The distillation bound contours depend on the specifics of the Hamiltonian, and they allow us to identify regimes where the bound attains low values and where it saturates to 1 for a fixed choice of error rates ϵ, ϵ'.

4.3 General entropic constraints

4.3.1 Shifting from Lorenz curves to entropies

We have considered Wigner representations of magic states within a statistical mechanical setting in which it was argued that magic can be viewed as a non-classical form of free energy. This was done at the level of majorization constraints that appear due to Clifford operations being described by stochastic maps in the Wigner representation. In this representation, non-equilibrium states with substantial free energy have Lorenz curves that deviate noticeably from the line $y = x$. We also found that when considered in a distillation protocol, magic could also be linked to

physical free energies in a non-trivial way. We therefore ask:

Is it possible to make a more direct link between magic and free energy or entropy?

One obstacle to linking with macroscopic, equilibrium free energy is that the Wigner representation is generally a quasi-distribution, so the Boltzmann entropy

$$H(\boldsymbol{w}) = -\sum_i w_i \log w_i \tag{4.56}$$

of the Wigner representation is not defined outside the set $\mathcal{F}$, let alone have a physically meaningful interpretation.

However, it turns out that the analysis for the temperature-dependent bound given by Eq. (4.13) implicitly made use of the single-shot Rényi divergence [125] $D_\infty(\boldsymbol{p} \,||\, \boldsymbol{r})$.

Claim 4.8. *Let τ be in the interior of $\mathcal{F}$. Then $D_\infty(W_\rho \,||\, W_\tau)$ is well-defined for all ρ, and the following hold:*

1. *$D_\infty(W_\rho \,||\, W_\tau) \geq 0$ for all quantum states ρ.*
2. *$D_\infty(W_\rho \,||\, W_\tau) = 0$ if and only if $\rho = \tau$.*
3. *$D_\infty(W_{\rho^{\otimes 2n}} \,||\, W_{\tau^{\otimes 2n}}) = nD_\infty(W_{\rho^{\otimes 2}} \,||\, W_{\tau^{\otimes 2}})$ for any $n \in \mathbb{N}$.*
4. *$D_\infty(W_\rho \,||\, W_\tau) \geq D_\infty(W_{\mathcal{E}(\rho)} \,||\, W_{\mathcal{E}(\tau)})$ for any free operation $\mathcal{E}$ such that $\mathcal{E}(\tau)$ is in the interior of $\mathcal{F}$.*

Proof. Since τ is in the interior of $\mathcal{F}$, its Wigner representation obeys $W_\tau(\boldsymbol{z}) > 0$ for all points $\boldsymbol{z}$ in the phase space. In general, $W_\rho(\boldsymbol{z})$ is a quasi-distribution, but the given form of α ensures that $W_\rho(\boldsymbol{z})^\alpha \geq 0$. Therefore $D_\infty(W_\rho \,||\, W_\tau)$ is always well-defined.

1. We have that

$$D_\infty(W_\rho \,||\, W_\tau) = \log \max_{\boldsymbol{z}} \frac{W_\rho(\boldsymbol{z})}{W_\tau(\boldsymbol{z})}, \tag{4.57}$$

so $2^{D_\infty(W_\rho \,||\, W_\tau)}$ equals the slope of the Lorenz curve $L_{\rho|\tau}(x)$ at $x = 0$. Since $L_{\rho|\tau}(x)$ is a concave function passing through $(0, 0)$ and $(1, 1)$ this implies that $L_{\rho|\tau}(x) \geq x$ for all $x \in [0, 1]$, and in particular its slope at $x = 0$ is always greater than or equal to 1, which implies $D_\infty(W_\rho \,||\, W_\tau) \geq 0$ for all ρ.

2. $D_\infty(W_\rho \,||\, W_\tau) = 0$ implies that $\max_{\boldsymbol{z}} \frac{W_\rho(\boldsymbol{z})}{W_\tau(\boldsymbol{z})} = 1$ and the initial slope of $L_{\rho|\tau}(x)$ is 1. From the concavity of the function and the fact that $L_{\rho|\tau}(1) = 1$, this implies that the slope of $L_{\rho|\tau}(x)$ must equal 1 throughout the interval $[0, 1]$ and this, together with the definition of the Lorenz curve, implies that $W_\rho(\boldsymbol{z})/W_\tau(\boldsymbol{z}) = 1$ for all $\boldsymbol{z}$.

Since the Wigner representation $\rho \mapsto W_\rho(\boldsymbol{z})$ is bijective this implies that $\rho = \tau$ only. The converse holds by inspection.

3. Given a vector $\boldsymbol{w} \in \mathbb{R}^N$, it is generally the case that $\max_k w_k \neq \max_k |w_k|$. However, we do have that

$$\max_{k_1,k_2}(w_{k_1}w_{k_2}) = \max_{k_1,k_2}|w_{k_1}w_{k_2}|, \tag{4.58}$$

and if additionally $w_k \geq 0$ for all k, then we also have for $2n$ copies that

$$\max_{k_1,k_2,\ldots,k_{2n}}(w_{k_1}w_{k_2}\cdots w_{k_{2n}}) = (\max_k |w_k|)^{2n}, \tag{4.59}$$

and also,

$$\max_{k_1,k_2,\ldots,k_{2n}}(w_{k_1}w_{k_2}\cdots w_{k_{2n}}) = \max_{k_1,k_2,\ldots,k_{2n}}\left(|w_{k_1}w_{k_2}|\cdots|w_{k_{2n-1}}w_{k_{2n}}|\right) = \left(\max_{k_1,k_2}(w_{k_1}w_{k_2})\right)^n. \tag{4.60}$$

If we now let $\boldsymbol{w} = W_{\rho|\tau}(\boldsymbol{z}) \coloneqq W_\rho(\boldsymbol{z})/W_\tau(\boldsymbol{z})$ we have

$$\max_{\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_{2n}}\prod_{k=1}^{2n} W_{\rho|\tau}(\boldsymbol{z}_k) = \left(\max_{\boldsymbol{z}_1,\boldsymbol{z}_2} W_{\rho|\tau}(\boldsymbol{z}_1)W_{\rho|\tau}(\boldsymbol{z}_2)\right)^n, \tag{4.61}$$

and therefore taking logarithms we have that $D_\infty(W_{\rho^{\otimes 2n}} \,||\, W_{\tau^{\otimes 2n}}) = nD_\infty(W_{\rho^{\otimes 2}} \,||\, W_{\tau^{\otimes 2}})$ as required.

4. The free operation $\mathcal{E}$ in the Wigner representation corresponds to a stochastic map, which sends $W_\tau(\boldsymbol{z})$ to $W_{\mathcal{E}(\tau)}(\boldsymbol{z})$, which is another strictly positive probability distribution on the phase space, due to $\mathcal{E}(\tau)$ being in the interior of $\mathcal{F}$. Moreover, $(W_\rho, W_\tau) \succ (W_{\mathcal{E}(\rho)}, W_{\mathcal{E}(\tau)})$. As shown in the main text this condition holds if and only if $L_{\rho|\tau}(x) \geq L_{\mathcal{E}(\rho)|\mathcal{E}(\tau)}(x)$ for all x. In particular, this implies the slope at the origin of the input curve is never less than the slope at the origin of the output curve, and hence the result follows. □

The reason for the power of 2 in the third property is that the maximization is sensitive to the presence of negativity in the Wigner representation. An important remark is that the state dimension does not enter the proof of the theorem, so this result is directly transferable to our qubit analysis in chapter 5, once we have defined a well-behaved Wigner representation.

Considering the unital magic protocol $\mathcal{E}$ that transforms n copies of a magic state ρ to m copies of ρ', and combining property 3 and property 4, we obtain

$$\frac{m}{n} \leq \frac{D_\infty(W_\rho \,||\, W_{\mathbb{1}/d})}{D_\infty(W_{\rho'} \,||\, W_{\mathbb{1}/d})}. \tag{4.62}$$

By substituting $\rho = \rho_S(\epsilon)$ and $\rho' = \rho_S(\epsilon')$ for $d = 3$, we retrieve the bound R_∞ in Eq. (4.5), justifying its index. This is because this bound is derived using the first elbow constraint (Claim 3.11) on the input and output Lorenz curves, which is equivalent to the formulation of the Rényi divergence D_∞ given by Eq. (4.57).

4.3.2 Entropic measure of magic disorder

We can go further, and show that in fact a range of Rényi entropies remain both well-defined and meaningful in the Wigner representation. The α–Rényi entropy was defined in section 3.4 as

$$H_\alpha(\boldsymbol{w}) \coloneqq \frac{1}{1-\alpha} \log \sum_i w_i^\alpha, \tag{4.63}$$

for quasi-distribution $\boldsymbol{w}$.

We can prove the following result, which establishes the equivalence between Wigner negativity and the existence of negative Rényi entropy.

Claim 4.9. *A quantum state ρ has Wigner negativity if and only if $H_\alpha(W_\rho) < 0$ for some $\alpha = \frac{2a}{2b-1}$, with positive integers $a \geq b$.*

Proof. If we have $H_\alpha(W_\rho) < 0$ for $\alpha > 1$, then it follows that $\log \sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z})^\alpha > 1$. However, it is known that H_α is always non-negative on probability distributions, so $W_\rho(\boldsymbol{z})$ must be a strict quasi-distribution with negativity.

Conversely, suppose ρ has negativity in its Wigner representation. This in particular implies that $\sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})| > 1$. We now consider $\sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z})^{\frac{2a}{2b-1}}$ for positive integers $a \geq b$. We have that

$$\sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z})^{\frac{2a}{2b-1}} = \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|^{\frac{2a}{2b-1}} = \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|^{1+\epsilon}, \tag{4.64}$$

where $\epsilon = \alpha - 1 = \frac{2a}{2b-1} - 1 > 0$. By choosing the positive integers a and b sufficiently large we can make ϵ arbitrarily close to 0. This implies that there exists a sequence $\epsilon_n = \alpha_n - 1 = \frac{2a_n}{2b_n-1} - 1$ with integer pairs a_n, b_n such that

$$\sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|^{1+\epsilon_n} \to \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})| > 1, \tag{4.65}$$

as n increases. Since $\frac{1}{1-\alpha_n} < 0$ for any n it follows that

$$H_{\alpha_n}(W_\rho) = \frac{1}{1-\alpha_n} \log \sum_{\boldsymbol{z}} W_\rho(\boldsymbol{z})^{\alpha_n} < 0 \tag{4.66}$$

at some point in the sequence, which completes the proof. □

Therefore, the statistical mechanical description of Clifford operations provides a setting in which negative entropy is fully meaningful, and quantifies the fact that magic states are more ordered than a perfectly sharp, deterministic classical state $\boldsymbol{w} = (1, 0, \ldots, 0)$ with zero entropy. We note that prior work has shown that negative conditional entropy [163] arises in the context of quantum correlations, but is unrelated to the present negative entropy.

The $\alpha \to 1$ limit of the Rényi entropy diverges if negativity is present, however we can assign a meaningful value to the residue at $\alpha = 1$. We write the entropy as

$$H_{1+\epsilon}(W_\rho) = -\epsilon^{-1} \log \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|^{1+\epsilon} \tag{4.67}$$

with $\epsilon \to 0^+$ through a sequence of rational values. We recall that mana can be written as a sum of the Wigner components in absolute vale, $\mathcal{M}(\rho) = \log \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z})|$, and now we express mana as an entropic limit,

$$\mathcal{M}(\rho) = -\lim_{\epsilon \to 0^+} \epsilon H_{1+\epsilon}(W_\rho). \tag{4.68}$$

Therefore, the mana of a state ρ is minus the residue of the pole at $\alpha = 1$ for $H_\alpha(W_\rho)$, and quantifies the divergence of the Rényi-entropy as we approach the limiting Shannon entropy $H(W_\rho)$.

We also note that the Rényi entropy can be described as a q–deformation of the Shannon entropy [164], and that $H_\alpha(\boldsymbol{p})$ can be related to the q–derivative of free energy. For this, the Rényi parameter becomes a temperature term $\alpha = \frac{T_0}{T}$ where T_0 is a reference temperature and derivatives are considered via the limit $T \to T_0$. Therefore, if we try to push this to the present setting, the presence of Wigner negativity would correspond to a divergence in the first derivative of an effective free energy as $T \to T_0$. It is interesting to speculate whether this non-classicality could be interpreted in terms of a phase transition [165].

4.3.3 General entropic distillation bounds

We now extend to Rényi divergences of finite α for Wigner representations with the aim of deriving an infinite family of entropic bounds for magic transformations, parameterized by parameter α. In section 3.4, we defined the α–Rényi divergence $D_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r})$ as

$$D_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r}) \coloneqq \frac{1}{\alpha - 1} \log \sum_i w_i^\alpha r_i^{1-\alpha},$$

where $\alpha = 2a/(2b-1)$ for positive integers $a \geq b$. This is well-defined for quasi-distribution $\boldsymbol{w}$ and probability distribution $\boldsymbol{p}$ with positive components, which corresponds to a state lying in the interior of $\mathcal{F}$, in the context of Wigner representations.

Claim 4.10. *Let τ be in the interior of $\mathcal{F}$. If $\alpha = \frac{2a}{2b-1}$ for positive integers a, b with $a \geq b$, then the α-Rényi divergence $D_\alpha(W_\rho \,||\, W_\tau)$ is well-defined for all states ρ, and the following hold:*

1. *$D_\alpha(W_\rho \,||\, W_\tau) \geq 0$ for all quantum states ρ.*
2. *$D_\alpha(W_\rho \,||\, W_\tau) = 0$ if and only if $\rho = \tau$.*
3. *$D_\alpha(W_{\rho^{\otimes n}} \,||\, W_{\tau^{\otimes n}}) = nD_\alpha(W_\rho \,||\, W_\tau)$ for any $n \in \mathbb{N}$.*
4. *$D_\alpha(W_\rho \,||\, W_\tau) \geq D_\alpha(W_{\mathcal{E}(\rho)} \,||\, W_{\mathcal{E}(\tau)})$ for any free operation $\mathcal{E}$ such that $\mathcal{E}(\tau)$ is in the interior of $\mathcal{F}$.*

Proof. Since τ is in the interior of $\mathcal{F}$, its Wigner representation obeys $W_\tau(\boldsymbol{z}) > 0$ for all points $\boldsymbol{z}$ in the phase space. In general, $W_\rho(\boldsymbol{z})$ is a quasi-distribution, but for $\alpha = 2a/2b - 1$ we that $W_\rho(\boldsymbol{z})^\alpha \geq 0$. Therefore $D_\alpha(W_\rho \,||\, W_\tau)$ is always well-defined and real-valued.

1. D_α is Schur-convex and every pair (W_ρ, W_τ) majorizes the pair $(W_{\mathbb{1}/d}, W_{\mathbb{1}/d})$, so

$$D_\alpha(W_\rho \,||\, W_\tau) \geq D_\alpha(W_{\mathbb{1}/d} \,||\, W_{\mathbb{1}/d}) = \frac{1}{\alpha - 1} \log \sum_{\boldsymbol{z}} W_{\mathbb{1}/d}(\boldsymbol{z}) = 0. \tag{4.69}$$

2. In the inequality of property 1, equality holds iff $L_{\rho|\tau}(x) = L_{\mathbb{1}/d|\mathbb{1}/d}(x) = L_{\tau|\tau}(x)$ for all $x \in [0, 1]$ due to Claim 3.8 which in turn holds iff $\rho = \tau$.

3. This property follows from the multiplicativity of the Wigner representation. In particular,

$$\begin{aligned} D_\alpha(W_{\rho^{\otimes n}} \,||\, W_{\tau^{\otimes n}}) &= \frac{1}{\alpha - 1} \log \sum_{\boldsymbol{z}} W_{\rho^{\otimes n}}(\boldsymbol{z})^\alpha W_{\tau^{\otimes n}}(\boldsymbol{z})^{1-\alpha} \\ &= \frac{1}{\alpha - 1} \log \sum_{\boldsymbol{z}} \prod_{i=1}^{n} W_\rho(\boldsymbol{z}_i)^\alpha W_\tau(\boldsymbol{z}_i)^{1-\alpha} \\ &= \frac{1}{\alpha - 1} \log \prod_{i=1}^{n} \sum_{\boldsymbol{z}_i} W_\rho(\boldsymbol{z}_i)^\alpha W_\tau(\boldsymbol{z}_i)^{1-\alpha} \\ &= \sum_{i=1}^{n} \frac{1}{\alpha - 1} \log \sum_{\boldsymbol{z}'} W_\rho(\boldsymbol{z}')^\alpha W_\tau(\boldsymbol{z}')^{1-\alpha} \\ &= nD_\alpha(W_\rho \,||\, W_\tau). \end{aligned} \tag{4.70}$$

4. This follows immediately from the fact that $(W_\rho, W_\tau) \succ (W_{\mathcal{E}(\rho)}, W_{\mathcal{E}(\tau)})$ for any free quantum channel $\mathcal{E}$ that sends τ into the interior of $\mathcal{F}$, and the Schur-convexity of D_α. □

Note again that, crucially, the state dimension does not enter the proof of the theorem, so α-Rényi entropies will be used in our qubit analysis in chapter 5.

The properties of the Rényi divergence immediately yield a family of magic distillation bounds.

Claim 4.11. *Consider a general magic state distillation protocol on odd prime dimension qudits, that converts a magic state $\rho^{\otimes n} \longrightarrow \mathcal{E}(\rho^{\otimes n}) = \rho'^{\otimes m}$ and let τ be any full-rank stabilizer reference state on a qudit. Then, the distillation rate $R \coloneqq m/n$ is upper bounded as*

$$R \leq \frac{D_\alpha(W_\rho \,||\, W_\tau)}{\widetilde{D}_\alpha(\rho', \tau')}, \tag{4.71}$$

where $\alpha = \frac{2a}{2b-1}$ for any positive integers a, b with $a \geq b$ and the average divergence per qudit

$$\widetilde{D}_\alpha(\rho', \tau') \coloneqq \frac{1}{m} D_\alpha(W_{\rho'^{\otimes m}} \,||\, W_{\tau'_m}), \tag{4.72}$$

between the output magic state $\rho'^{\otimes m}$ and $\tau'_m = \mathcal{E}(\tau^{\otimes n})$.

Proof. The bound is a direct consequence of the properties of the α–Rényi divergence in Claim 4.10.

Due to the action of the magic protocol channel, we get

$$D_\alpha(W_{\rho^{\otimes n}} \,||\, W_{\tau^{\otimes n}}) \geq D_\alpha(W_{\rho'^{\otimes m}} \,||\, W_{\tau'_m}). \tag{4.73}$$

We can use the additivity to rewrite this as

$$n D_\alpha(W_\rho \,||\, W_\tau) \geq m \frac{1}{m} D_\alpha(W_{\rho'^{\otimes m}} \,||\, W_{\tau'_m}). \tag{4.74}$$

Since $\rho'^{\otimes m} \neq \tau'_m$, we have $D_\alpha(W_{\rho'^{\otimes m}} \,||\, W_{\tau'_m}) > 0$, which directly leads to the bound

$$\frac{m}{n} \leq \frac{D_\alpha(W_\rho \,||\, W_\tau)}{\widetilde{D}_\alpha(\rho', \tau')}. \tag{4.75}$$

□

While this bound is abstract in its present form, it does suggest future applications by viewing $\widetilde{D}_\alpha(\rho', \tau')$ within the context of hypothesis testing [166], as a measure of distinguishability between $\rho'^{\otimes m}$ and the m qudit state τ'_m. However, in order to properly link with hypothesis-testing one would have to first extend such results to quasi-distributions.

If correlations in τ'_m between subsystems can be neglected, then $\tau'_m = \tau'^{\otimes m}$ for some qudit state τ' and

$$\widetilde{D}_\alpha(\rho', \tau') = D_\alpha(W_{\rho'} \,||\, W_{\tau'}), \tag{4.76}$$

which gives a generalized version of Claim 4.5, with a term such as $D_\alpha(W_\rho \,||\, W_\tau)$ that behaves essentially as an α–free energy difference [135] in the Wigner representation.

We can illustrate a basic application of the general bound in Eq. (4.71) for the case of unital magic protocols. In this case it is easy to see that the above bound reduces to the simple form

$$R \leq R_\alpha \coloneqq \frac{2\log d - H_\alpha(W_\rho)}{2\log d - H_\alpha(W_{\rho'})}. \tag{4.77}$$

In particular, for the case of noisy Strange states (Eq. (3.30)) on qutrits and $\epsilon' = 0$, this bound becomes

$$R \leq R_\alpha = \frac{2(1-\alpha)\log 3 - \log\left[8(\frac{1}{6} - \frac{\epsilon}{18})^\alpha + (-\frac{1}{3} + \frac{4\epsilon}{9})^\alpha\right]}{2(1-\alpha)\log 3 + \alpha\log 6 - \log(8+2^\alpha)}. \tag{4.78}$$

This upper bound can be numerically minimized over α, for any ϵ. The cases R_{10} and R_∞ are shown in Fig. (4.3) and numerical evidence shows that index 10 is an integer which lies close to the index of the optimal bound. It is clear that different values of α allow R_α to capture different aspects of the full set of majorization conditions, or equivalently, of the Lorenz curve constraint. For example, R_∞ captures exactly the first elbow constraint (Claim 3.11), while R_{10} captures more aspects of the Lorenz curve than just the first elbow up to $\epsilon \approx 0.40$ as Fig. (4.3) suggests for the noisy Strange state.

In summary, this chapter has presented a methodology for deriving bounds on magic distillation rates of odd dimensional systems based on the statistical mechanics of the Wigner representation. We start off with providing a bound (Eq. (4.5)) for unital protocols, which outperforms previous bounds that are based on monotones. Then, we demonstrate how majorization bounds reflect properties of the magic protocols and physical properties of the device performing the distillation by deriving bounds inspired from thermodynamics in Claim 4.5 and Claim 4.7. Finally, we generalize these results in establishing entnropic bounds that apply in the context of any odd dimensional magic state transformation in Eq. (4.62) and Claim 4.11. We extend this generalized family of magic distillation bounds to qubits in chapter 5. To this end, we identify a subset of universal qubit computation in which the crucial entropic properties of Claim 4.8 and Claim 4.10 naturally extend to qubits.

Chapter 5

Entropic constraints on qubit magic distillation

Most quantum algorithms are formulated for systems of qubits, so an important question remains whether a similar description to the odd dimensional framework discussed in chapter 4 exists for qubits. In this case, well-known obstacles exist, as was apparent in the qubit constructions of chapter 2 due to qubits having non-trivial Wigner representations [167]. For instance, recall that the odd dimensional Wigner representation [97] cannot provide a positive representation for all qubit stabilizer states and operations, and the link between Wigner negativity and quantum computational power breaks down [106, 168]. Although substantial work has been done to develop representations in which all qubit stabilizer operations are non-negatively represented [105, 107], these representations no longer factorize over tensor product composition, a key property needed in the odd dimensional case to obtain bounds on distillation rates. This is no surprise, since it is known to be impossible to construct a multiplicative representation in which every Clifford operation is represented stochastically [167]. We therefore exert an effort in justifying that the CSS subclass of stabilizer operations, introduced in section 2.5.4, is computationally important and admits a powerful Wigner representation that allows us to transfer results from chapter 4 to qubits.

Almost all protocols [72, 75–77, 169] to-date for qubit magic distillation make use of a sub-class of stabilizer codes, known as Calderbank-Shor-Steane (CSS) codes [70, 71]. CSS codes can be constructed from two classical linear codes and therefore the CSS construction allows one to draw on a plethora of results from classical coding theory to construct quantum codes with desirable properties. For instance, it has been shown that CSS codes are optimal when it comes to constructing quantum error correcting codes that support a transversal T–gate [170], a key feature in

many of the aforementioned distillation protocols.

We first discuss CSS protocols, then distinguish an important sub-class of protocols, which we call CSS code projections, before finally applying majorization to the protocols in order to obtain magic distillation bounds. Our first result is to engineer a set of channels $\mathcal{O}$, including state preparations, gates and measurements, which is useful in terms of universality and practical due to the fact that it admits a stochastic description on the phase space, allowing for majorization constraints. We then show how our framework can exploit core structures in CSS code distillation protocols and arrive at upper and lower bounds which act as trade-off relations between important performance metrics of CSS code distillation protocols such as the desired output accuracy and the acceptance probability.

Throughout this chapter we will make use of some simplifying notation and terminology, which we lay down here. Given a 1–qubit operator O, and a binary vector $\boldsymbol{u} \coloneqq (u_1, \ldots, u_n) \in \mathbb{Z}_2^n$ let us denote

$$O(\boldsymbol{u}) \coloneqq O^{u_1} \otimes \cdots \otimes O^{u_n}. \tag{5.1}$$

Furthermore, we use $\langle \cdot \, , \, \cdot \rangle$ to denote the symplectic inner product, where the arguments are phase space vectors, while we also borrow the notation $\langle \ldots \rangle$ from section 2.5.4 to denote the set of stabilizer generators for a given state. The partial trace $\mathrm{tr}_{[a,b]}[\cdot]$, denotes tracing on subsystems a through b inclusive. A Pauli observable is called X–type or Z–type if it is a sequence of 1–qubit Pauli X operators or 1–qubit Pauli Z operators respectively. A n–k distillation protocol is a magic distillation protocol that converts n copies of a specified magic state to k copies of a purer magic state. Finally, the set of free operations will be specified in each section, but the set of free states $\mathcal{F}$ will commonly be the set of convex combinations of CSS states, unless specified otherwise.

5.1 CSS magic protocols

In section 2.5.4, we discussed the CSS formalism, which forms the basis for the magic distillation protocols we will be discussing in this chapter. We first discuss the stochasticity of completely CSS–preserving operations, starting from the unitary operations (section 5.1.1) and moving to the measurements (section 5.1.2), before combining them (section 5.1.3) into one set that constitutes the general form of CSS protocols.

5.1.1 Completely CSS–preserving unitaries

We recall from section 2.5.4 that the set of completely CSS–preserving unitaries is

$$\mathcal{G}(n) \coloneqq \langle \{\mathrm{CNOT}_{i,j}, Z_i, X_i\}_{i,j=1,\ldots,n,i\neq j} \rangle. \tag{5.2}$$

The n–qubit CSS–preserving unitaries are [98]

$$\mathcal{G}_+(n) \coloneqq \langle \{H(n), \mathrm{CNOT}_{i,j}, X_i, Z_i\}_{i,j=1,\ldots,n,i\neq j} \rangle, \tag{5.3}$$

where $H(n) \coloneqq H^{\otimes n}$ is the collective Hadamard gate on n–qubits.

Any unitary $U_+ \in \mathcal{G}_+(n)$ can be written as $U_+ = [H(n)]^a U$ for some $a \in \{0,1\}$ and $U \in \mathcal{G}(n)$. This follows from the commutation relations satisfied by the generators of $\mathcal{G}(n)$ and the fact that $H(n)$ is self-inverse. In particular, we have that $H(n)\mathrm{CNOT}_{i,j} = \mathrm{CNOT}_{j,i}H(n)$ and $H(n)X_i = Z_i H(n)$.

Since every generator of $\mathcal{G}(n)$ is in turn a generator of $\mathcal{G}_+(n')$ for all $n' \geq n$, every member of $\mathcal{G}(n)$ is completely CSS–preserving. Therefore, the only members of $\mathcal{G}_+(n)$ that are not also completely CSS–preserving are those of the form $H(n)U$ for some $U \in \mathcal{G}(n)$, since $H(n)$ is not a member of $\mathcal{G}(n')$ for any $n' > n$. We conclude that $\mathcal{G}(n)$ must be the group of completely CSS–preserving unitaries on n–qubits.

5.1.2 Completely CSS–preserving measurements

The projective measurement of any n–qubit Pauli observable S is carried out using two projectors $P_\pm(S) \coloneqq \frac{1}{2}(\mathbb{1}_n \pm S)$ corresponding to the outcomes ± 1. We can then denote the post-selection channel for the ± 1 outcome as $\mathcal{P}_\pm(S) \coloneqq P_\pm(S)(\cdot)P_\pm(S)$.

We now establish which Pauli observables can be projectively measured in CSS–preserving and completely CSS–preserving ways. The latter form elements of the CSS formalism that we can describe stochastically with the Wigner representation that we developed in section 2.5.

Claim 5.1. *Let S be a Pauli observable on n qubits, and ρ be a* CSS *state on $(n+m)$ qubits for any $m \geq 0$. Then the state $\sigma_\pm$ obtained by projectively measuring S on the final n qubits of ρ and post-selecting on the ± 1 outcome,*

$$\sigma_\pm \coloneqq \frac{[\mathbb{1}_m \otimes \mathcal{P}_\pm(S)](\rho)}{p_\pm}, \quad p_\pm \coloneqq \mathrm{tr}\big[[\mathbb{1}_m \otimes P_\pm(S)]\rho\big] \tag{5.4}$$

is a CSS *state if and only if S is X–type or Z–type. Moreover, if S is neither X–type nor Z–type, then there exists a* CSS *state ρ such that $\sigma_\pm$ are not* CSS *for any m.*

Proof. We first establish that if $S = X(\boldsymbol{a})$ and ρ is a pure state $|\psi\rangle\langle\psi|$, then $\sigma_\pm$ are CSS. Let the stabilizer group of $|\psi\rangle$ be

$$\langle (-1)^{b_1} X(\boldsymbol{a}_1), \ldots, (-1)^{b_r} X(\boldsymbol{a}_r), (-1)^{b_{r+1}} Z(\boldsymbol{a}_{r+1}), \ldots, -(1)^{b_{n+m}} Z(\boldsymbol{a}_{n+m}) \rangle, \tag{5.5}$$

where each $\boldsymbol{a}_i$ is a non-zero $(n+m)$–bit string and each b_i is a binary digit. Then, the (un-normalized) state obtained by projectively measuring S on the final n qubits of $|\psi\rangle$ and post-selecting on the ± 1 outcome is

$$[\mathbb{1}_m \otimes P_\pm(X(\boldsymbol{a}))] |\psi\rangle = \left[\frac{1}{2}(\mathbb{1}_{n+m} \pm \mathbb{1}_m \otimes X(\boldsymbol{a}))\right] |\psi\rangle = P_\pm(X(\boldsymbol{a}')) |\psi\rangle \tag{5.6}$$

where $\boldsymbol{a}' \coloneqq \boldsymbol{0}_m \oplus \boldsymbol{a}$ and $\boldsymbol{0}_m$ is the m–dimensional zero vector.

There are now two possibilities:

1. $X(\boldsymbol{a}')$ commutes with $Z(\boldsymbol{a}_i)$ for all i in the range $r+1 \leq i \leq n+m$. Therefore, either $X(\boldsymbol{a}')$ or $-X(\boldsymbol{a}')$ stabilizes $|\psi\rangle$, so either $P_+(X(\boldsymbol{a}')) |\psi\rangle = |\psi\rangle$ and $P_-(X(\boldsymbol{a}')) |\psi\rangle = 0$, or $P_-(X(\boldsymbol{a}')) |\psi\rangle = |\psi\rangle$ and $P_+(X(\boldsymbol{a}')) |\psi\rangle = 0$.

2. $X(\boldsymbol{a}')$ does not commute with $Z(\boldsymbol{a}_i)$ for at least one i. Without loss of generality, choose i in the range $r+1 \leq i \leq r+c$ for some $c \geq 1$. We then have that $P_\pm(X(\boldsymbol{a}')) |\psi\rangle = \frac{1}{\sqrt{2}} |\phi_\pm\rangle$, where $|\phi_\pm\rangle$ are respectively stabilized by

$$\begin{aligned}&\langle (-1)^{b_1} X(\boldsymbol{a}_1), \ldots, (-1)^{b_r} X(\boldsymbol{a}_r), \pm X(\boldsymbol{a}'), (-1)^{(b_{r+1} \oplus b_{r+2})} Z(\boldsymbol{a}_{r+1} \oplus \boldsymbol{a}_{r+2}), \ldots, \\ &(-1)^{(b_{r+1} \oplus b_{r+c})} Z(\boldsymbol{a}_{r+1} \oplus \boldsymbol{a}_{r+c}), (-1)^{b_{r+c+1}} Z(\boldsymbol{a}_{r+c+1}), \ldots, (-1)^{b_n} Z(\boldsymbol{a}_n) \rangle,\end{aligned} \tag{5.7}$$

   where $\oplus$ denotes additional modulo 2, from which we see that $|\phi_\pm\rangle$ are both CSS states.

Summarizing these possibilities, we conclude that, given any pure CSS state $|\psi\rangle$ on n qubits, the state obtained from projectively measuring $X(\boldsymbol{a})$ on the last n qubits and post-selecting on the ± 1 outcome,

$$|\phi_\pm\rangle \coloneqq \frac{1}{\sqrt{p_\pm}} [P_\pm(X(\boldsymbol{a}')) |\psi\rangle] = \frac{1}{\sqrt{p_\pm}} [\mathbb{1} \otimes P_\pm(X(\boldsymbol{a})) |\psi\rangle], \tag{5.8}$$

where $p_\pm \coloneqq \mathrm{tr}[\mathbb{1}_m \otimes P_\pm(X(\boldsymbol{a})) |\psi\rangle\langle\psi|]$, is always CSS.

Since the projective measurement of $-X(\boldsymbol{a})$ requires the same projectors as that of $X(\boldsymbol{a})$, the above argument carries over directly for $S = -X(\boldsymbol{a})$ and $\rho = |\psi\rangle\langle\psi|$. By reversing the roles of X and Z, the argument also carries over to $S = \pm Z(\boldsymbol{a})$ and $\rho = |\psi\rangle\langle\psi|$. Therefore, by decomposing an arbitrary CSS state ρ on $n+m$ qubits into a statistical mixture of pure CSS states, we can show that the state $\sigma_\pm$ obtained from

projectively measuring a Pauli observable S on the final n qubits and post-selecting on the ± 1 outcome is CSS if $S = \pm X(\boldsymbol{a})$ or $S = \pm Z(\boldsymbol{a})$.

We next prove that, if $S \neq \pm X(\boldsymbol{a})$ and $S \neq \pm Z(\boldsymbol{a})$ for all n–bit strings $\boldsymbol{a}$, then it is always possible to find a pure CSS state ρ on n qubits alone such that $\sigma_{\pm}$ are not CSS. This argument will only be made for positive Pauli observables, since the projective measurements for $\pm S$ involve the same projectors.

Recall that every positive n–qubit Pauli observable S can be represented as

$$S = \bigotimes_{i=1}^{n} Q_i, \text{ for } {}_i \in \{\mathbb{1}, X, Y, Z\}. \tag{5.9}$$

When $S \neq X(\boldsymbol{a})$ and $S \neq Z(\boldsymbol{a})$ for all possible $\boldsymbol{a}$, there are three possibilities:

1. $Q_i \neq Y$ for all i. In this case, $Q_i = \mathbb{1}, X, Z$, and there must exist j and k for which $Q_j = X$ while $Q_k = Z$.

   Let us define the sets $\mathcal{X} \coloneqq \{i \neq j | Q_i = X\}$, $\mathcal{Z} \coloneqq \{i \neq k | Q_i = Z\}$ and $\mathcal{N} \coloneqq \{i | Q_i = \mathbb{1}\}$. Consider the CSS state $|\psi\rangle$ defined by the stabilizer group

   $$\mathrm{STAB}(|\psi\rangle) \coloneqq \langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, Z_j, Z_j X_k \rangle. \tag{5.10}$$

   By construction, S commutes with every generators of $\mathrm{STAB}(|\psi\rangle)$ except Z_j. Therefore, $P_{\pm}(S)\,|\psi\rangle = \frac{1}{\sqrt{2}}\,|\phi_{\pm}\rangle$, where $|\phi_{\pm}\rangle$ are stabilizer states defined by stabilizer groups

   $$\begin{aligned} \mathrm{STAB}(|\phi_{\pm}\rangle) &= \langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, \pm S, Z_j X_k \rangle \\ &= \langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_j Z_k, Z_j X_k \rangle. \end{aligned} \tag{5.11}$$

   where we obtained the second equality by multiplying $\pm S$ by every othere generator except for $Z_j X_k$. This shows explicitly that $|\phi_{\pm}\rangle$ are not CSS states.

2. There is an odd number w of values for i at which $Q_i = Y$.

   Let us define the sets $\mathcal{X} \coloneqq \{i | Q_i = X\}, \mathcal{Y} \coloneqq \{i | Q_i = Y\}, \mathcal{Z} \coloneqq \{i | Q_i = Z\}$ and $\mathcal{N} \coloneqq \{i | Q_i = \mathbb{1}\}$. Let j be a member of $\mathcal{Y}$. Where $\mathcal{Y}$ contains more than one member, let k be another member of $\mathcal{Y}$ besides j.

   Consider the CSS state $|\psi\rangle$ defined by the stabilizer group

   $$\begin{aligned} \mathrm{STAB}(|\psi\rangle) = \big\langle &\{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\{Z_i Z_k | i \in \mathcal{Y}, i \neq j, k\}, Z_j \big\rangle, \end{aligned} \tag{5.12}$$

   where $X_\star \coloneqq \Big((-1)^{(w-1)/2} \prod_{i \in \mathcal{Y}, i \neq j} X_i \Big)$.

By construction, S commutes with every generator of $\mathrm{STAB}(|\psi\rangle)$ except Z_j. Therefore, $P_{\pm}(S)\,|\psi\rangle = \frac{1}{\sqrt{2}}\,|\phi_{\pm}\rangle$, where $|\phi_{\pm}\rangle$ are stabilizer states defined by the stabilizer groups

$$\begin{aligned}\mathrm{STAB}(|\phi_{\pm}\rangle) &= \big\langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\qquad \{Z_i Z_k | i \in \mathcal{Y}, i \neq j, k\}, \pm S \big\rangle \\ &= \big\langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\qquad \{Z_i Z_k | i \in \mathcal{Y}, i \neq j, k\}, \pm Y_j \big\rangle, \end{aligned} \tag{5.13}$$

where we obtained the second equality by multiplying $\pm S$ by every other generator. This explicitly shows that $|\phi_{\pm}\rangle$ are not CSS states.

3. There is an even number w of values for i at which $Q_i = Y$.

   Let us define the sets $\mathcal{X} \coloneqq \{i | Q_i = X\}, \mathcal{Y} \coloneqq \{i | Q_i = Y\}, \mathcal{Z} \coloneqq \{i | Q_i = Z\}$ and $\mathcal{N} \coloneqq \{i | Q_i = \mathbb{1}\}$. Let j and k be two distinct members of $\mathcal{Y}$. Where $\mathcal{Y}$ contains more than two members, let l be another member of $\mathcal{Y}$ besides j and k.

   Consider the CSS state $|\psi\rangle$ defined by the stabilizer group

$$\begin{aligned}\mathrm{STAB}(|\psi\rangle) &= \big\langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\qquad \{Z_i Z_l | i \in \mathcal{Y}, i \neq j, k, l\}, Z_j X_k, X_k \big\rangle, \end{aligned} \tag{5.14}$$

   where now $X_\star \coloneqq \Big((-1)^{\left(\frac{w-2}{2}\right)} \prod_{i \in \mathcal{Y}, i \neq j,k} X_i \Big)$.

   By construction, S commutes with every generator of $\mathrm{STAB}(|\psi\rangle)$ except for X_k, which means $P_{\pm}(S)\,|\psi\rangle = \frac{1}{\sqrt{2}}\,|\phi_{\pm}\rangle$, where $|\phi_{\pm}\rangle$ are defined by the stabilizer groups

$$\begin{aligned}\mathrm{STAB}(|\phi_{\pm}\rangle) &= \big\langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\qquad \{Z_i Z_l | i \in \mathcal{Y}, i \neq j, k, l\}, Z_j X_k, \pm S \big\rangle \\ &= \big\langle \{Z_i | i \in \mathcal{N}\}, \{Z_i | i \in \mathcal{Z}\}, \{X_i | i \in \mathcal{X}\}, X_\star, \\ &\qquad \{Z_i Z_l | i \in \mathcal{Y}, i \neq j, k, l\}, Z_j X_k, \pm X_j Z_k \big\rangle, \end{aligned} \tag{5.15}$$

   where the second equality was obtained by multiplying $\pm S$ by every other generator. This explicitly shows that $|\phi_{\pm}\rangle$ are not CSS states.

□

We therefore group Pauli observables of the forms $\pm X(\boldsymbol{a})$ or $\pm Z(\boldsymbol{a})$ together as CSS observables. Furthermore, we highlight the fact that Claim 5.1 shows CSS observables to be the only Pauli observables whose projective measurement can be carried out in a completely CSS–preserving way ($m > 0$).

5.1.3 General CSS protocols

Essential to the application of majorization to magic state protocols is identifying a sufficiently large subset of stabilizer states and Clifford operations that can be stochastically represented. In section 2.5.4, we mentioned that all *completely* CSS*–preserving channels* are stochastically represented and listed a substantial subset of Clifford operations that fall under this description. We now provide a proof of this by utilizing the rigorous definitions of the previous sections.

Claim 5.2. *Any sequential composition of the following operations:*

1. *Preparation of a* CSS *state;*
2. *Any gate* $G \in \mathcal{G}(n)$*;*
3. *Projective measurement of any* X*– or* Z*–type Pauli observable, followed by a completely* CSS*–preserving operation* $\mathcal{E}_\pm$ *conditioned on the outcome* ± 1*;*
4. *Tracing out;*

as well as statistical mixtures of such processes, is completely CSS*–preserving.*

Proof. Let $\{\mathcal{E}_i\}$ be a set of completely CSS–preserving operations from n to m qubits, and let ρ be a state on $(a+n)$ qubits. Given any probability distribution, we then have that

$$\sigma \coloneqq \left(\sum_i p_i \mathbb{1}_a \otimes \mathcal{E}_i\right) \rho = \sum_i p_i [(\mathbb{1}_a \otimes \mathcal{E}_i)\rho], \tag{5.16}$$

is a CSS state, since $(\mathbb{1}_a \otimes \mathcal{E}_i)\rho$ is a CSS state on $(a+m)$ qubits due to $\mathcal{E}_i$ being completely CSS–preserving. Therefore, any statistical mixture of completely CSS–preserving channels (between the same input and output systems) are also completely CSS–preserving.

Furthermore, let $\mathcal{E}$ be a completely CSS–preserving channel from n to m qubits, and let $\mathcal{F}$ be a completely CSS–preserving channel from m to l qubits. Letting ρ once again be a CSS state on $(a+n)$ qubits, we have that $(\mathbb{1}_a \otimes \mathcal{F}) \circ (\mathbb{1}_a \otimes \mathcal{E})\rho$ is another CSS state, since $(\mathbb{1}_a \otimes \mathcal{E})\rho$ must be a CSS state due to $\mathcal{E}$ being completely CSS–preserving, which $(\mathbb{1}_a \otimes \mathcal{F})$ would map onto another CSS state since $\mathcal{F}$ is also completely CSS–preserving. Therefore, any sequential composition of completely CSS–preserving channels (where the input of one channel matches the output of the other) is also completely CSS–preserving.

All that remains is to prove that the listed stabilizer operations are completely CSS–preserving.

1. Preparation is CSS preserving as the tensor product of two CSS states is also CSSstate.

2. Gates $G \in \mathcal{G}(n)$ are completely CSS–preserving since their Choi matrices are CSS.

3. Let $\{\mathcal{P}_\pm(S) \coloneqq P_\pm(S)(\cdot)P_\pm(S)\}$ be the projective channels carrying out a measurement of the CSS observable S, and ρ be a CSS state on $(n+m)$ qubits. From Claim 5.1, we see that $(\mathbb{1}_m \otimes \mathcal{P}_\pm(C)))\rho = p_\pm \sigma_\pm$, where $p_\pm \coloneqq \mathrm{tr}[(\mathbb{1}_m \otimes P_\pm(S))\rho]$ and $\sigma_\pm$ are CSS states. Therefore, since $\mathcal{E}_\pm$ are completely CSS–preserving, we have that $\sigma' \coloneqq \left[\sum_\pm \mathbb{1}_m \otimes \mathcal{E}_\pm \circ \mathbb{1}_m \otimes P_\pm(S)\right](\rho) = \sum_\pm p_\pm[\mathbb{1}_m \otimes \mathcal{E}_\pm](\sigma_\pm)$ is also CSS.

4. Consider tracing out m qubits from n. Since we can freely relabel subsystems, we can, without loss of generality, only consider tracing out the last m qubits of n. Since tracing out is unaffected by first performing a computational basis measurement on the last m qubits. Let $|\psi\rangle$ be a pure CSS state on $(a+n)$ qubits, we have that

$$\mathbb{1}_a \otimes \mathrm{tr}_{[n-m,n]}\left[|\psi\rangle\langle\psi|\right] = \sum_{\boldsymbol{k}\in\{0,1\}^m} \mathrm{tr}_{[a+n-m,a+n]}\Big[\mathbb{1}_{a+n-m} \otimes |\boldsymbol{k}\rangle\langle\boldsymbol{k}|\, |\psi\rangle\langle\psi| \\ \mathbb{1}_{a+n-m} \otimes |\boldsymbol{k}\rangle\langle\boldsymbol{k}|\,\Big]. \tag{5.17}$$

We then observe that

$$\mathbb{1}_{a+n-m} \otimes |\boldsymbol{k}\rangle\langle\boldsymbol{k}| = \mathcal{P}_+((-1)^{k_1} Z_{a+n-m+1}) \circ \cdots \circ \mathcal{P}_+((-1)^{k_m} Z_{a+n}), \tag{5.18}$$

and so by Claim 5.1, $(\mathbb{1}_{a+n-m} \otimes |\boldsymbol{k}\rangle\langle\boldsymbol{k}|)\,|\psi\rangle$ is a (sub-normalized) pure CSS state of the form $\sqrt{p_{\boldsymbol{k}}}\,|\phi_{\boldsymbol{k}}\rangle \otimes |\boldsymbol{k}\rangle$, where $|\phi_{\boldsymbol{k}}\rangle$ must be a CSS state on the first $(a+n-m)$ qubits in order to keep the complete state CSS, and $p_{\boldsymbol{k}}$ is the probability of getting the $|\boldsymbol{k}\rangle$ outcome in the computational basis measurement. Therefore,

$$\begin{aligned}\mathbb{1}_a \otimes \mathrm{tr}_{[n-m,n]}[|\psi\rangle\langle\psi|] &= \sum_{\boldsymbol{k}\in\{0,1\}^m} \mathrm{tr}_{[a+n-m,a+n]}[p_{\boldsymbol{k}}\,|\phi_{\boldsymbol{k}}\rangle\langle\phi_{\boldsymbol{k}}| \otimes |\boldsymbol{k}\rangle\langle\boldsymbol{k}|] \\ &= \sum_{\boldsymbol{k}\in\{0,1\}^m} p_{\boldsymbol{k}}\,|\phi_{\boldsymbol{k}}\rangle\langle\phi_{\boldsymbol{k}}|\,,\end{aligned} \tag{5.19}$$

which is a CSS state on $(a+n-m)$ qubits. Therefore, by decomposing an arbitrary CSS state ρ on $(a+n)$ qubits as a statistical mixture of CSS states, we can show that $\mathbb{1}_{a+n-m} \otimes \mathrm{tr}_{[n-m,n]}[\rho]$ must be CSS, which implies tracing out subsystems is completely CSS–preserving.

□

CSS *protocols* can therefore be viewed as the maximal class $\mathcal{O}$ of completely CSS–preserving operations.

Consider now any quantum circuit $\mathcal{E}$ formed from preparing CSS states, performing gates from the larger group of CSS–preserving unitaries, discarding qubits and projectively measuring CSS observables. Let $\mathcal{U}_H(n) \coloneqq H(n)(\cdot)H(n)$. We make the following observations.

- Given any n–bit string $\boldsymbol{a}$, because $U_H(n)X(\boldsymbol{a})U_H(n) = Z(\boldsymbol{a})$, and every CSS observable S on n qubits is of the form $\pm X(\boldsymbol{a})$ or $\pm Z(\boldsymbol{a})$, we have the conjugation relation $P_\pm(S)U_H(n) = U_H(n)P_\pm(S')$, where $S' \coloneqq H(n)SH(n)$ is another CSS observable. Therefore, letting $\mathcal{E}$ be the projective measurement of a CSS observable S, we conclude that $\mathcal{E}\circ\mathcal{U}_H(n) = \mathcal{U}_H(n)\circ\mathcal{E}'$, where $\mathcal{E}'$ is the projective measurement of the CSS observable S'.

- Given any n–qubit state ρ, we have that $H(n)\rho H(n) \otimes \sigma = H(n)\rho H(n) \otimes H(m)\sigma' H(m)$, where $\sigma' \coloneqq H(m)\sigma H(m)$ is also CSS state since $H(m)$ is CSS–preserving on m qubits. Therefore, letting $\mathcal{E}$ be the channel introducing a CSS state σ, we conclude that $\mathcal{U}_H(n) \circ \mathcal{E} = \mathcal{E}' \circ \mathcal{U}_H(n+m)$, where $\mathcal{E}'$ is the channel introducing another CSS state σ'.

- Let $\mathcal{R}$ be any subset of m qubits out of n. We then have $\mathrm{tr}_\mathcal{R} \circ \mathcal{U}_H(n) = \mathcal{U}_H(n-m) \circ \mathrm{tr}_\mathcal{R}$.

- As we have already seen in section 5.1.1, given any completely CSS–preserving n–qubit unitary channel $\mathcal{U}$, we have that $\mathcal{U}_H(n) \circ \mathcal{U} = \mathcal{U}' \circ \mathcal{U}_H(n)$, where $\mathcal{U}'$ is another completely CSS–preserving unitary channel.

We also saw in section 5.1.1 that any CSS–preserving unitary channel can be written as $[\mathcal{U}_H(n)]^b \circ \mathcal{U}_+$, where $\mathcal{U}_+$ is a completely CSS–preserving unitary channel, and b is a binary digit. After decomposing every CSS–preserving unitary in $\mathcal{E}$ into this form, we can conjugate any collective Hadamard gate to the end of the circuit as described above. Therefore, $\mathcal{E}$ is operationally equivalent to a CSS protocol described by a subset of $\mathcal{O}$, followed by a collective Hadamard gate conditioned upon certain some measurement outcomes obtained during the protocol. We conclude that $\mathcal{E}$ can be converted reversibly into a CSS protocol using Clifford post-processing, which implies CSS–preserving and completely CSS–preserving unitaries generate equally powerful distillation protocols.

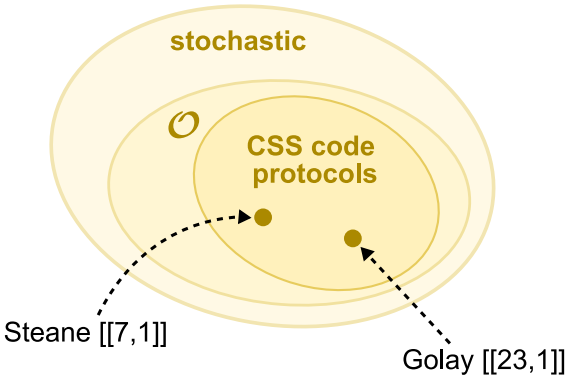

Figure 5.1: **Schematic of qubit magic protocols**. The set of CSS protocols $\mathcal{O}$, which include the family of CSS code projections as a subset according to Claim 5.5, are a subset of the stochastically represented CPTP maps. Elements of $\mathcal{O}$ include important distillation codes, such as the 7–1 and 23–1 CSS distillation protocols based on the Steane $[[7,1]]$ and Golay $[[23,1]]$ codes, respectively [169].

5.2 CSS code projections

We now discuss an elementary protocol for distilling magic, which was proposed in the seminal work [72] and forms a suclass of CSS protocols. It is performed via projecting onto the code-space of a quantum error correcting code. The basic method involves taking n copies of your noisy magic state $\rho^{\otimes n}$, measuring the stabilizer generators of an $[[n,k]]$ code $\mathcal{C}$, and post-selecting on the no-error syndrome. The net effect is to project onto $\mathcal{C}$. Conditional on no errors being detected in the syndrome measurements, one then decodes onto k output qubits, whereas when an error is detected the output state is simply discarded.

In general, the protocol will only succeed probabilistically with some acceptance probability p. The core idea is analogous to stabilizer code projections [171] in that if the likelihood of an undetectable error occurring is less than the input error rate ϵ, then the post-selected output state will have a higher per-copy fidelity with the target magic state of choice. Many other examples are based on CSS codes, for instance the 15–1 protocol [72] based on the $[[15,1]]$ punctured Reed-Muller code [172,173], and the protocols based on Steane $[[7,1]]$ and Golay $[[23,1]]$ CSS codes analyzed in [169]. A general schematic of the protocols we considered is provided in Fig. (5.1).

5.2.1 Definition of CSS code projections

We can write any CSS code projection as a completely positive trace non-increasing map $\mathcal{E}_{\mathcal{P}}$ for the given CSS code $\mathcal{C}$ as

$$\mathcal{E}_{\mathcal{P}}(\cdot) \coloneqq \mathrm{tr}_{[1,n-k]}[\mathcal{U} \circ \mathcal{P}(\cdot)], \tag{5.20}$$

where we have defined $\mathcal{U} \coloneqq U(\cdot)U^{\dagger}$ and $\mathcal{P} \coloneqq P(\cdot)P$ for the decoding unitary U and code-space projector P of $\mathcal{C}$ respectively. Given n copies of a noise magic state ρ, $\mathcal{E}_{\mathcal{P}}$ acts as

$$\mathcal{E}_{\mathcal{P}}(\rho^{\otimes n}) = p\rho', \tag{5.21}$$

where $\rho' \in \mathcal{B}(\mathcal{H}_2^k)$ is the output magic state on k qubits and we have defined the acceptance probability $p \coloneqq \mathrm{tr}[P\rho^{\otimes n}]$ for a single successful run of $\mathcal{E}_{\mathcal{P}}$. Distillation of the magic state $\psi \coloneqq |\psi\rangle\langle\psi|$ is successful if the output ρ' from a successful run has a greater per-copy fidelity with respect to a target (pure) magic state of choice than ρ.

The majorization tools that we want to apply to the protocols do not immediately apply to the projection $\mathcal{E}_{\mathcal{P}}$ as we require trace-preservation. However, this can be fixed by simply recording the success (labeled '0') or failure (labeled '1') of $\mathcal{E}_{\mathcal{P}}$ in an ancillary qubit, and without loss of generality we can assume an arbitrary CSS state σ on k qubits is outputted in case of failure. We can therefore extend $\mathcal{E}_{\mathcal{P}}$ into the following completely positive trace preserving (CPTP) map which describes the CSS code projection:

$$\mathcal{E}(\cdot) \coloneqq \mathcal{E}_{\mathcal{P}}(\cdot) \otimes |0\rangle\langle 0| + \mathrm{tr}[\overline{\mathcal{P}}(\cdot)]\sigma \otimes |1\rangle\langle 1|\,, \tag{5.22}$$

where $\overline{\mathcal{P}} \coloneqq \overline{P}(\cdot)\overline{P} \coloneqq (\mathbb{1}_n - P)(\cdot)(\mathbb{1}_n - P)$ performs a projection onto the orthogonal complement of $\mathcal{C}$. Under such a channel, n copies of ρ are mapped onto

$$\mathcal{E}[\rho^{\otimes n}] = p\rho' \otimes |0\rangle\langle 0| + (1-p)\sigma \otimes |1\rangle\langle 1| \coloneqq \rho_p, \tag{5.23}$$

which captures the structure of the probabilistic distillation protocol.

5.2.2 Stochasticity of CSS code projections

It is not obvious from the definition in Eq. (5.22) that CSS code projections form a subset of CSS protocols, therefore we prove this fact here, establishing the stochasticity of such protocols. Our strategy is as follows: we first establish in Claim 5.3 that the decoding unitary appearing in $\mathcal{E}_{\mathcal{P}}$ is completely CSS–preserving. We then prove in Claim 5.4 that a convex combination of terms in the form of a projective measurement $\mathcal{P}_k$ followed by some completely CSS–preserving post-processing $\mathcal{E}_k$ is

stochastically represented. We then bring these two results together in Claim 5.5 by identifying that the channel $\mathcal{E}$ in Eq. (5.22) constitutes a sum of two terms in the form $\mathcal{E}_k \circ \mathcal{P}_k$, therefore it is a stochastically represented CSS protocol.

We first demonstrate that any CSS code can be decoded using a completely CSS–preserving unitary. This result allows for the use of simple explicit expressions to describe very general CSS code projections.

Claim 5.3. *Let* $\mathcal{S} \coloneqq \langle \{(-1)^{b_i} S_i\}_{i=1,\dots,n-k} \rangle$ *be the stabilizer group defining an* $[[n,k]]$ CSS *code, where*

$$S_i = \begin{cases} X(\boldsymbol{u}_i) & \text{for } 1 \leq i \leq r \\ Z(\boldsymbol{v}_i) & \text{for } r+1 \leq i \leq n-k \end{cases} \tag{5.24}$$

in which $\boldsymbol{u}_i, \boldsymbol{v}_i$ *are non-zero* n*–dimensional binary vectors and* $b_i \in \{0,1\}$ *is a binary digit for all* i*.*

Then, there exists a completely CSS*–preserving unitary* U *such that*

$$U((-1)^{b_i} S_i) U^\dagger = \begin{cases} X_i & \text{for } 1 \leq i \leq r \\ Z_i & \text{otherwise.} \end{cases} \tag{5.25}$$

Proof. The proof proceeds by construction.

Let us first consider the X–type generators of $\mathcal{S}$ without their signs, i.e. $\mathcal{X} \coloneqq \{X(\boldsymbol{u}_i)\}_{i=1,\dots,r}$. We will prove that there exists a sequence of CNOT operations that transforms $X(\boldsymbol{u}_i)$ to X_i for all $1 \leq i \leq r$.

Let G be the group formed by the set of positive X–type n–qubit CSS observables under matrix multiplication, i.e. $G \coloneqq (\{X(\boldsymbol{a}) | \boldsymbol{a} \in \{0,1\}^n\}, \cdot)$, and let G' be the group formed by the set of n–dimensional binary strings under binary addition of corresponding entries, i.e. $G' \coloneqq (\{0,1\}^n, \oplus)$. Then $G \cong G'$ under the intuitive isomorphism $X(\boldsymbol{a}) \leftrightarrow \boldsymbol{a}$.

We can represent an m–tuple $\mathcal{A}$ of m positive X–type n–qubit CSS observables, $\mathcal{A} \coloneqq (X(\boldsymbol{a}_1), \dots, X(\boldsymbol{a}_m))$ where $\boldsymbol{a}_i \in \{0,1\}^n$ for $1 \leq i \leq m$, as the columns of an $n \times m$ matrix

$$M_{\mathcal{A}} \coloneqq \begin{bmatrix} | & \dots & | \\ \boldsymbol{a}_1 & \dots & \boldsymbol{a}_m \\ | & \dots & | \end{bmatrix}. \tag{5.26}$$

Because no element of $\mathcal{X}$ can be formed by multiplying other elements together, their image in G', $\{\boldsymbol{u}_i\}_{i=1,\dots,r}$, is a set of r linearly independent elements in V_n, the n–dimensional vector space over F_2. Therefore, $M_{\mathcal{X}}$ has rank r.

We now demonstrate that the following two operations on $M_{\mathcal{A}}$ can be accomplished by performing sequences of CNOT operations on all the members of $\mathcal{A}$ simultaneously:

1. *(Swapping rows j and k)*. The unitary operation $\mathcal{U}_{SWAP}(\cdot) \coloneqq U_{SWAP}(\cdot)U_{SWAP}^{\dagger}$, where $U_{SWAP} \coloneqq \mathrm{CNOT}_{j,k}\mathrm{CNOT}_{k,j}\mathrm{CNOT}_{j,k}$, swaps qubits j and k. Therefore, the matrix $M_{\mathcal{A}'}$ representing the m–tuple

$$\mathcal{A}' \coloneqq (\mathcal{U}_{SWAP}(X(\boldsymbol{a}_1)), \ldots, \mathcal{U}_{SWAP}(X(\boldsymbol{a}_m)) \tag{5.27}$$

   can be obtained from $M_{\mathcal{A}}$ by swapping rows j and k.

2. *(Adding row j to row k)*. The action of $\mathrm{CNOT}_{j,k}$ on X_k is

$$\mathrm{CNOT}_{j,k}X_i\mathrm{CNOT}_{j,k} = \begin{cases} X_jX_k & \text{for } i = j \\ X_i & \text{otherwise.} \end{cases} \tag{5.28}$$

   Therefore,

$$\mathrm{CNOT}_{j,k}X(\boldsymbol{a})\mathrm{CNOT}_{j,k} = X(\boldsymbol{a} \oplus a_j\boldsymbol{\eta}_k) \coloneqq X(\boldsymbol{a}'), \tag{5.29}$$

   where $\boldsymbol{\eta}_k$ is an n–bit string with 1 in the k–th entry and 0 everywhere else. In words, $\boldsymbol{a}'$ is formed from adding the j–th entry of $\boldsymbol{a}$ to its k–th entry. Therefore, the matrix $M_{\mathcal{A}'}$ representing the m–tuple

$$\mathcal{A}' \coloneqq (\mathrm{CNOT}_{j,k}X(\boldsymbol{a}_1)\mathrm{CNOT}_{j,k}, \ldots, \mathrm{CNOT}_{j,k}X(\boldsymbol{a}_1)\mathrm{CNOT}_{j,k}) \tag{5.30}$$

   can be obtained from $M_{\mathcal{A}}$ by entry-wise binary addition of row j to row k.

We further demonstrate that Gauss-Jordan elimination reduces to a sequence of such row swaps and additions on V_n. For a vector space over a general field F, Gauss-Jordan elimination is a sequence of three moves:

1. Swap the positions of two rows.

2. Add to one row a non-zero scalar multiple of another.

3. Multiply any row by a non-zero scalar.

Since the only scalars available in F_2 are 0 and 1, the third move has no effect when $F = F_2$ and can be neglected, while the second move reduces to the entry-wise binary addition of rows.

We now note that we can perform Gauss-Jordan elimination on $M_{\mathcal{X}}$ and convert it into its unique reduced row echelon form,

$$D_{\mathcal{X}} \coloneqq \begin{bmatrix} \mathbb{1}_{r,r} \\ \mathbb{O}_{n-r,r} \end{bmatrix}, \tag{5.31}$$

where $\mathbb{1}_{r,r}$ is an $r \times r$ identity matrix while $\mathbb{O}_{n-r,r}$ is a $(n-r) \times r$ null matrix.

The CNOT sequence corresponding to this Gauss-Jordan elimination, which we can denote by the unitary operation $\mathcal{U}_{\mathcal{X}}(\cdot) \coloneqq U_{\mathcal{X}}(\cdot)U_{\mathcal{X}}^{\dagger}$, accomplishes $\mathcal{U}_{\mathcal{X}}(X(\boldsymbol{u}_i)) = X_i$.

We next consider the Z–type generators of $\mathcal{S}$ without their signs, i.e. $\mathcal{Z} \coloneqq \{Z(\boldsymbol{v}_{r+1}), \ldots, Z(\boldsymbol{v}_{n-k})\}$. The action of $\mathrm{CNOT}_{j,k}$, on Z_i is

$$\mathrm{CNOT}_{j,k} Z_i \mathrm{CNOT}_{j,k} = \begin{cases} Z_j Z_k & \text{for } i = k \\ Z_i & \text{otherwise,} \end{cases} \tag{5.32}$$

so $\mathcal{U}_{\mathcal{X}}$ only transforms positive Z–type CSS observables into other positive Z–type CSS observables, i.e. we can find n–bit binary strings $\{\boldsymbol{v}_i'\}_{i=r+1,\ldots,n-k}$ such that $\mathcal{U}_{\mathcal{X}}(Z(\boldsymbol{v}_i)) = Z(\boldsymbol{v}_i')$ for all i in the range $r+1 \leq i \leq n-k$.

However, since the X–type generators of $\mathcal{S}$ commute with the Z–type generators, $Z(\boldsymbol{v}_i')$ must commute with X_1 through X_r for all i in the range $r+1 \leq i \leq n-k$. Therefore, $Z(\boldsymbol{v}_i')$ acts non-trivially on qubits $r+1$ through n only.

Everything that we have done for the X–type generators can then be repeated for the Z–type generators. The only thing that needs to be checked is that, when we represent the m–tuple $\mathcal{B} \coloneqq (Z(\boldsymbol{a}_1), \ldots, Z(\boldsymbol{a}_m))$ of m positive Z–type n–qubit CSS observables as the columns of a matrix $M_{\mathcal{B}}$, row addition on $M_{\mathcal{B}}$ can be performed by executing a CNOT operation on all elements of $\mathcal{B}$ simultaneously, just as in the case of X–type CSS observables. This is confirmed via Eq. (5.32), which implies

$$\mathrm{CNOT}_{j,k} Z(\boldsymbol{a}) \mathrm{CNOT}_{j,k} = Z(\boldsymbol{a} \oplus a_k \boldsymbol{\eta}_j) \coloneqq Z(\boldsymbol{a}') \tag{5.33}$$

where we see that $\boldsymbol{a}'$ is formed by adding entry k in $\boldsymbol{a}$ to entry j. Therefore, the matrix $M_{\mathcal{B}'}$ representing the m–tuple

$$\mathcal{B}' \coloneqq (\mathrm{CNOT}_{j,k} Z(\boldsymbol{a}_1) \mathrm{CNOT}_{j,k}, \ldots, \mathrm{CNOT}_{j,k} Z(\boldsymbol{a}_m) \mathrm{CNOT}_{j,k}) \tag{5.34}$$

can be obtained from $M_{\mathcal{B}}$ by adding row k to row j.

We conclude that there also exists a sequence of CNOT operations $\mathcal{U}_{\mathcal{Z}} \coloneqq U_{\mathcal{Z}}(\cdot)U_{\mathcal{Z}}^{\dagger}$ such that $\mathcal{U}_{\mathcal{Z}}(Z(\boldsymbol{v}_i')) = Z_i$ for all i in the range $r+1 \leq i \leq n-k$. Since $Z(\boldsymbol{v}_i')$ acts non-trivially on qubits $r+1$ through n only for all i in the range $r+1 \leq i \leq n-k$,

we can choose $\mathcal{U}_{\mathcal{Z}}$ to act only on qubits $r+1$ through n, and so $\mathcal{U}_Z(X_i) = X_i$ for all i in the range $1 \leq i \leq r$.

We now have

$$(\mathcal{U}_{\mathcal{Z}} \circ \mathcal{U}_{\mathcal{X}})[(-1)^{b_i} S_i] = \begin{cases} (-1)^{b_i} X_i & \text{for } 1 \leq i \leq r \\ (-1)^{b_i} Z_i & \text{for } r+1 \leq i \leq n-k. \end{cases} \tag{5.35}$$

Defining $\mathcal{U}_C(\cdot) \coloneqq U_C(\cdot)U_C^\dagger$, where $U_C \coloneqq \left[\prod_{i=r+1}^{n-k} X_i^{b_i}\right] \left[\prod_{i=1}^{r} Z_i^{b_i}\right]$, we see that

$$(\mathcal{U}_C \circ \mathcal{U}_Z \circ \mathcal{U}_{\mathcal{X}})[(-1)^{b_i} S_i] = \begin{cases} X_i & \text{for } 1 \leq i \leq r \\ Z_i & \text{for } r+1 \leq i \leq n-k. \end{cases} \tag{5.36}$$

Since $U_{\mathcal{X}}, U_{\mathcal{Z}}, U_C$ are completely CSS–preserving, as they are sequences of CNOT gates or local qubit X and Z gates, we have constructed a completely CSS–preserving unitary $U \coloneqq U_C U_{\mathcal{Z}} U_{\mathcal{X}}$ that completes the Claim. □

We next identify a collection of primitive channels that are stochastically represented and which can be arbitrarily combined to construct magic protocols.

Claim 5.4. *Let $\{\mathcal{P}_k\}$ be a set of projective channels on n qubits defined as $\mathcal{P}_k(X) \coloneqq p_k P_k X P_k$, where p_k is a probability and P_k is a product of commuting projectors onto the eigenspaces of* CSS *observables. Furthermore, let $\{\mathcal{E}_k\}$ be a set of completely* CSS*–preserving channels from n to m qubits.*

Then, $\sum_k \mathcal{E}_k \circ \mathcal{P}_k$ is stochastically represented whenever $\sum_k p_k P_k = \mathbb{1}_n$.

Proof. Let ρ be a CSS state on $(n+a)$ qubits. By repeated applications of Claim 5.1 to each projection forming the product P_k, we obtain

$$(\mathbb{1}_a \otimes \mathcal{P}_k)\rho = p_k \mathrm{tr}[(\mathbb{1}_a \otimes P_k)\rho]\sigma_k, \tag{5.37}$$

where σ_k is a CSS state on $(n+m)$ qubits. Therefore,

$$\sigma \coloneqq \mathbb{1}_a \otimes \left(\sum_k \mathcal{E}_k \circ \mathcal{P}_k\right) \rho = \sum_k p_k \mathrm{tr}[(\mathbb{1}_a \otimes P_k)\rho](\mathbb{1}_a \otimes \mathcal{E}_k)(\sigma_k). \tag{5.38}$$

Since $\mathcal{E}_k$ is completely CSS–preserving, $(\mathbb{1}_a \otimes \mathcal{E}_k)(\sigma_k)$ is a CSS state. When $\sum p_k P_k = \mathbb{1}_n$, we have that $\sum_k p_k \mathrm{tr}[(\mathbb{1}_a \otimes P_k)\rho] = 1$, which means σ is a CSS state. We conclude that, if $\sum p_k P_k = \mathbb{1}_n$, then $\sum_k \mathcal{E}_k \circ \mathcal{P}_k$ is completely CSS–preserving and thus stochastically represented. □

We are now in a position to prove that CSS code projections are stochastic.

Claim 5.5. *The quantum channel for any n–k CSS code projection is a completely CSS–preserving operation, and hence it is stochastically represented.*

Proof. Let $\mathcal{C}$ be an $[[n,k]]$ CSS code. The trace-preserving version of its code projection can be represented as

$$\mathcal{E}(\rho) \coloneqq \mathrm{tr}_{[1,n-k]}[\mathcal{U}\circ\mathcal{P}(\rho)]\otimes|0\rangle\langle 0| + \mathrm{tr}[\overline{\mathcal{P}}(\rho)]\sigma\otimes|1\rangle\langle 1|\,, \tag{5.39}$$

where $\mathcal{U} \coloneqq U(\cdot)U$ and $\mathcal{P} \coloneqq P_{\mathbf{0}}(\cdot)P_{\mathbf{0}}$ for the decoding unitary U and codespace projector $P_{\mathbf{0}}$ of $\mathcal{C}$, and $\overline{\mathcal{P}} \coloneqq \overline{P}_{\mathbf{0}}(\cdot)\overline{P}_{\mathbf{0}}$ for $\overline{P}_{\mathbf{0}} \coloneqq \mathbb{1}_n - P_{\mathbf{0}}$.

Let the code-space of $\mathcal{C}$ be stabilized by $\mathcal{S} \coloneqq \langle\{S_i\}_{i=1,\ldots,n-k}\rangle$, where $\{S_i\}_{i=1,\ldots,n-k}$ is a set of $(n-k)$ commuting and independent n–qubit CSS observables, of which the first $r \leq n-k$ are X–type while the rest are Z–type.

Let us define channels $\mathcal{E}_0, \mathcal{E}_1$ as

$$\mathcal{E}_0(\cdot) \coloneqq (\mathrm{tr}_{[1,n-k]}\circ\mathcal{U})(\cdot)\otimes|0\rangle\langle 0|\,, \text{ and} \tag{5.40}$$

$$\mathcal{E}_1(\cdot) \coloneqq \sigma\otimes|1\rangle\langle 1|\,\mathrm{tr}(\cdot) \tag{5.41}$$

We can then express $\mathcal{E}$ as $\mathcal{E} = \mathcal{E}_0\circ\mathcal{P} + \mathcal{E}_1\circ\overline{\mathcal{P}}$.

By Claim 5.3, $\mathcal{C}$ can be decoded by a unitary $U = [H^{\otimes r}\otimes\mathbb{1}_{n-k-r}]\,V$ for some $V\in\mathcal{G}(n)$. Since the Hadamards before V only take place on qubits 1 through $n-k$, we have by the cyclic property of the trace that

$$\mathcal{E}_0(\cdot) = (\mathrm{tr}_{[1,n-k]}\circ\mathcal{V})(\cdot)\otimes|0\rangle\langle 0|\,, \tag{5.42}$$

where $\mathcal{V}(\cdot) \coloneqq V(\cdot)V^\dagger$. We thereby see from Claim 5.2 that $\mathcal{E}_0$ and $\mathcal{E}_1$ are both completely CSS–preserving.

Let $\boldsymbol{s}$ be an n–bit string denoting the outcome of the syndrome measurement for $\mathcal{C}$, where s_i is the outcome of measuring the observable S_i. By definition, $P_{\mathbf{0}}$ projects onto the subspace corresponding to the zero syndrome outcome, and is therefore the product of commuting CSS projectors that successively project onto the $(+1)$–eigenspaces of each S_i, i.e.

$$P_{\mathbf{0}} = \prod_{i=1}^{n-k} P_0(S_i), \quad P_0(S_i) \coloneqq \frac{1}{2}(\mathbb{1}_n + S_i). \tag{5.43}$$

Furthermore, we have that

$$\mathcal{E}_1\circ\overline{\mathcal{P}}(X) = \sigma\otimes|1\rangle\langle 1|\,\mathrm{tr}\left[\overline{P}_{\mathbf{0}}X\overline{P}_{\mathbf{0}}\right] = \sigma\otimes|1\rangle\langle 1|\,\mathrm{tr}\left[\overline{P}_{\mathbf{0}}X\right] \tag{5.44}$$

Now $\overline{P}_{\mathbf{0}} = \sum_{\boldsymbol{s}\neq\mathbf{0}} P_{\boldsymbol{s}}$, where each $P_{\boldsymbol{s}}$ is a product of commuting projectors onto the $(-1)^{s_i}$ eigenspace of each S_i, i.e.

$$P_{\boldsymbol{s}} = \prod_{i=1}^{n-k} P_{s_i}(S_i), \quad P_{s_i}(S_i) \coloneqq \frac{1}{2}(\mathbb{1}_n + (-1)^{s_i}S_i). \tag{5.45}$$

Substituting into Eq. (5.44), and defining $\mathcal{P}_{\boldsymbol{s}} \coloneqq P_{\boldsymbol{s}}(\cdot)P_{\boldsymbol{s}}$, we obtain

$$\begin{aligned}\mathcal{E}_1 \circ \overline{\mathcal{P}}(X) &= \sigma \otimes |1\rangle\langle 1| \operatorname{tr}\left[\left(\sum_{\boldsymbol{s}\neq\boldsymbol{0}} P_{\boldsymbol{s}}\right) X\right] \\ &= \sum_{\boldsymbol{s}\neq\boldsymbol{0}} \sigma \otimes |1\rangle\langle 1| \operatorname{tr}\left[P_{\boldsymbol{s}} X P_{\boldsymbol{s}}\right] \\ &= \sum_{\boldsymbol{s}\neq\boldsymbol{0}} \mathcal{E}_1 \circ \mathcal{P}_{\boldsymbol{s}}. \end{aligned} \tag{5.46}$$

We have thereby shown that $\mathcal{E} = \mathcal{E}_0 \otimes \mathcal{P}_{\boldsymbol{0}} + \sum_{\boldsymbol{s}\neq\boldsymbol{0}} \mathcal{E}_1 \otimes \mathcal{P}_{\boldsymbol{s}}$, in which $P_{\boldsymbol{s}}$ is a product of commuting projectors onto the eigenspaces of CSS observables such that $\sum_{\boldsymbol{s}} P_{\boldsymbol{s}} = \mathbb{1}_n$, and $\mathcal{E}_0$ and $\mathcal{E}_1$ are both completely CSS–preserving channels. Therefore, by Claim 5.4, we conclude that $\mathcal{E}$ can be stochastically represented. □

The stochasticity of CSS code projections allows us to describe induced state transformations via relative majorization, thereby the entropic results of chapter 4 of odd dimensions are transferable to qubits, as discussed in the following section.

5.3 Entropic bounds for CSS protocols

CSS circuits become capable of universal quantum computation with the injection of rebit magic states [98, 108], which are real and therefore represented in Eq. (2.65) by valid quasi-distributions with negative probabilities. So far we have established the stochasticity of CSS protocols. In this section, we therefore turn our attention to deriving constraints on magic distillation performed by CSS protocols based on the statistical mechanics of the Wigner representation.

It turns out that it is natural to study the quantity ΔD_α that expresses the difference of the output from the input divergences between magic and reference states, which we define in Eq. (5.48). In section 5.3.1, we provide abstract bounds on generic CSS protocols in terms of ΔD_α, given in Claim 5.6. In section 5.3.2, we specialize in CSS code projections and prove several technical aspects of ΔD_α which we use to demonstrate the existence of upper and lower bounds on the number of input copies n in Claim 5.9. The upper bounds appear conditionally on the α–Rényi entropy of the input magic state. We then derive explicit expressions for these bounds in Claim 5.13 that are applicable in the context of any CSS code projection. In section 5.3.3, we discuss an example application of our bounds on Hadamard state (Eq. (2.88)) distillation and finally, as a sanity check, we show in section 5.3.4 that our bounds are non-trivial, by comparing them with a simpler data-processing inequality applicable directly on the quantum density matrices of the magic states.

5.3.1 Bounds on generic CSS protocols

Since every CSS protocol is stochastically represented, we can use the α–Rényi divergence as defined in Eq. (3.65) with properties outlined in Claim 4.10, to obtain the following family of constraints on rebit magic distillation.

Claim 5.6. *Let ρ be a noisy rebit magic state, and τ be a* CSS *state in the interior of $\mathcal{F}$. If there exists a* CSS *distillation protocol $\mathcal{E}$ such that $\rho' = \mathcal{E}(\rho^{\otimes n})$ and $\tau' \coloneqq \mathcal{E}(\tau^{\otimes n})$ is also in the interior of $\mathcal{F}$, then*

$$\Delta D_\alpha \geq 0, \tag{5.47}$$

for $\alpha = \frac{2a}{2b-1}$ with positive integers a, b such that $a \geq b$, where

$$\Delta D_\alpha \coloneqq n D_\alpha(W_\rho \,||\, W_\tau) - D_\alpha(W_{\rho'} \,||\, W_{\tau'}). \tag{5.48}$$

Proof. The result follows by direct substitution of ρ', τ' and use of the properties in Claim 4.10. □

Similarly to the odd dimensional scenario, the reference process $\tau^{\otimes n} \to \tau'$ can be viewed in three different ways: (1) as a variational parameter, (2) as encoding physics of the quantum device, or (3) as a way to exploit structure in a family of protocols, which we now elaborate on in turn.

Firstly, it can simply be treated as a variational parameter, which can be optimized over to obtain the following set of monotones on CSS protocols

$$\Lambda_\alpha(\rho) \coloneqq \inf_{\tau \in \mathcal{F}} D_\alpha(W_\rho \,||\, W_\tau), \tag{5.49}$$

for $\alpha = 2a/(2b-1)$ with positive integers a, b such that $a \geq b$, and where the infimum is taken over the convex set of CSS states $\mathcal{F}$. For odd dimensional systems, we can obtain the same family of monotones, for example by letting $\mathcal{F}$ denote the set of stabilizer states STAB. To see that Eq. (5.49) constitutes a monotone family for CSS protocols, we recall $D_\alpha(W_\rho \,||\, W_\tau) \geq 0$ for all ρ, τ, with equality if and only if $\rho = \tau$ according to Claim 4.10. Given any rebit state ρ, let τ_ρ be a solution to the optimization problem in Eq. (5.49). Then if there exists a CSS protocol $\mathcal{E}$ such that $\mathcal{E}(\rho) = \rho'$, we obtain

$$\Lambda_\alpha(\rho) = D_\alpha(W_\rho \,||\, W_{\tau_\rho}) \geq D_\alpha(W_{\rho'} \,||\, W_{\mathcal{E}(\tau_\rho)}) \geq \Lambda_\alpha(\rho'), \tag{5.50}$$

where the first inequality follows from generalized relative majorization and the second inequality follows by the definition in Eq. (5.49). Therefore, the set of all well-defined $\{\Lambda_\alpha\}$ form an infinite set of monotones on the class of CSS protocols. It

is straightforward to see that Λ_α are sub-additive, i.e. $\Lambda_\alpha(\rho^{\otimes n}) \geq n\Lambda_\alpha(\rho)$ (which follows from the additivity of the generalized α–Rényi divergences). Therefore, these Λ_α monotones allow us to set global bounds on any CSS protocol. More precisely, if there exists a completely CSS–preserving distillation protocol $\mathcal{E} \in \mathcal{O}$ such that $\mathcal{E}(\rho^{\otimes n}) = \rho'$, then the overhead n is lower bounded as

$$n \geq \frac{\Lambda_\alpha(\rho')}{\Lambda_\alpha(\rho)}. \tag{5.51}$$

Secondly, the reference process can be chosen appropriately to encode the particular physics of a protocol (or family of protocols) of interest. For instance, in section 4.2.2, the Gibbs state defined at reference temperature T is assumed to be preserved in order to encode some background temperature or free energy production that takes place during the distillation protocol.

The third way of interpreting the reference process is by choosing it appropriately in order to exploit some structural symmetry of CSS protocols. We demonstrate this in the next section with CSS code projection protocols.

5.3.2 Bounds on CSS code projection protocols

We will now see that there is a natural choice of CSS states τ and τ' such that $\tau^{\otimes n} \to \tau'$ common to all n–k CSS code projections, based on the intuition that the code projection component $\mathcal{E}_\mathcal{P}$ of the channel in Eq. (5.22),

$$\mathcal{E}(\cdot) \coloneqq \mathcal{E}_\mathcal{P}(\cdot) \otimes |0\rangle\langle 0| + \mathrm{tr}[\overline{\mathcal{P}}(\cdot)]\sigma \otimes |1\rangle\langle 1|\,,$$

is always sub-unital. To see this we first note that the identity operator on n qubits can be decomposed as $\mathbb{1}_n = P + \overline{P}$ for the code-space projector P of any $[[n,k]]$ CSS code. Therefore, $\mathcal{E}_{\mathbb{1}_n}$ always acts as

$$\mathcal{E}_{\mathbb{1}_n} = \mathcal{E}_{P+\overline{P}} = \mathcal{E}_P. \tag{5.52}$$

Since P is the logical identity on k logical qubits, i.e.,

$$P = \sum_{\boldsymbol{k}\in\{0,1\}^k} |\boldsymbol{k}_L\rangle\langle\boldsymbol{k}_L| \equiv \mathbb{1}_L, \tag{5.53}$$

the decoding of P in Eq. (5.52) must give an output state that is proportional to the maximally mixed state on k physical qubits, and so (omitting normalization constants) we obtain $\mathcal{E}(\mathbb{1}_n) \propto \mathbb{1}_k$. Therefore, the code projection component of the distillation channel is indeed sub-unital. Finally we note that since P is a rank-2^k projector, the acceptance probability associated with this protocol is $p = \mathrm{tr}\left[P\frac{\mathbb{1}_n}{2^n}\right] =$

2^{k-n}. Putting this all together, we find that under any n–k CSS code projection the maximally mixed state on n qubits gets mapped to

$$\mathcal{E}\left[\left(\frac{\mathbb{1}}{2}\right)^{\otimes n}\right] = \tau_{n,k}, \tag{5.54}$$

where we have defined the output state

$$\tau_{n,k} \coloneqq 2^{k-n}\frac{\mathbb{1}_k}{2^k} \otimes |0\rangle\langle 0| + (1-2^{k-n})\sigma \otimes |1\rangle\langle 1|\,, \tag{5.55}$$

which only depends on the number of stabilizer generators $n-k$ of the CSS code.

We can therefore conclude that, if there exists an n–k CSS code projection such that $\rho^{\otimes n} \mapsto p\rho'$, then $\Delta D_\alpha \geq 0$, where

$$\Delta D_\alpha \coloneqq nD_\alpha\left(W_\rho \middle|\middle| W_{\frac{\mathbb{1}}{2}}\right) - D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right), \tag{5.56}$$

which we define over the restricted domain $n \in [k, \infty]$, so that the number of logical qubits cannot exceed the number of physical qubits.

For any given p and rebit state ρ, we have the following continuous function of n defined on the interval $n \in [k, \infty)$,

$$\begin{aligned}\Delta D_\alpha \coloneqq& D_\alpha(W_{\rho^{\otimes n}} \,||\, W_{\left[\frac{\mathbb{1}}{2}\right]^{\otimes n}}) - D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right)\\ =& nD_\alpha(W_\rho \,||\, W_{\frac{\mathbb{1}}{2}}) - D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right),\end{aligned} \tag{5.57}$$

which gives change in distance, as measured by the α–Rényi divergence, between two quasi-distributions and their reference probability distributions. This allows us to express the necessary condition in Claim 5.6 specialized to the transition $\rho^{\otimes n} \to p\rho'$ under a n–k CSS code projection via $\Delta D_\alpha \geq 0$.

It will also be quite useful to introduce the following general mean $Q_\alpha(\cdot \,||\, \cdot)$ on the quasi-distributions $\boldsymbol{w} \coloneqq (w_1, \ldots w_N)^T$, $\boldsymbol{r} \coloneqq (r_1, \ldots r_N)^T$, which we define via

$$Q_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r}) \coloneqq 2^{(\alpha-1)D_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r})} = \sum_{i=1}^{N} w_i^\alpha r_i^{1-\alpha}. \tag{5.58}$$

We now have the following technical result, which will allow us to simplify the expression of our constraint functions ΔD_α.

Claim 5.7. *Consider the rebit quantum states $\rho_0, \rho_1, \tau_0, \tau_1$, where τ_i for $i \in \{0,1\}$ lie in the interior of $\mathcal{F}$. Moreover, let ψ_0, ψ_1 be two perfectly distinguishable register states in $\mathcal{F}$ such that $\mathrm{tr}[\psi_0\psi_1] = 0$. Then, the following identity holds,*

$$\begin{aligned}& Q_\alpha(W_{p_0\rho_0\otimes\psi_0 + p_1\rho_1\otimes\psi_1} \,||\, W_{q_0\tau_0\otimes\psi_0 + q_1\tau_1\otimes\psi_1})\\ =\;& p_0^\alpha q_0^{1-\alpha} Q_\alpha(W_{\rho_0} \,||\, W_{\tau_0}) + p_1^\alpha q_1^{1-\alpha} Q_\alpha(W_{\rho_1} \,||\, W_{\tau_1}),\end{aligned} \tag{5.59}$$

which in turn implies the inequality

$$Q_\alpha(W_{p_0\rho_0\otimes\psi_0+p_1\rho_1\otimes\psi_1} \,||\, W_{q_0\tau_0\otimes\psi_0+q_1\tau_1\otimes\psi_1}) \geq p_0^\alpha q_0^{1-\alpha} Q_\alpha(W_{\rho_0} \,||\, W_{\tau_0}). \tag{5.60}$$

Proof. By assumption the register states ψ_0, ψ_1 have zero overlap:

$$\mathrm{tr}[\psi_0\psi_1] = 2\sum_{\boldsymbol{u}} W_{\psi_0}(\boldsymbol{u}) W_{\psi_1}(\boldsymbol{u}) = 0. \tag{5.61}$$

Now since $\psi_i \in \mathcal{F}$ we must have $W_{\psi_i}(\boldsymbol{u}) \geq 0$ for all $\boldsymbol{u} \in \mathcal{P}$ and for each $i \in \{0,1\}$. We can thus conclude from Eq. (5.61) that

$$V_0 \coloneqq \mathrm{supp}(W_{\psi_0}) \subseteq \ker(W_{\psi_1}); \tag{5.62}$$

$$V_1 \coloneqq \mathrm{supp}(W_{\psi_1}) \subseteq \ker(W_{\psi_0}). \tag{5.63}$$

With this in hand, we can explicitly evaluate:

$$\begin{aligned}
&Q_\alpha(W_{p_0\rho_0\otimes\psi_0+p_1\rho_1\otimes\psi_1} \,||\, W_{q_0\tau_0\otimes\psi_0+q_1\tau_1\otimes\psi_1}) \\
=&\sum_{\boldsymbol{u}\in\mathcal{P}} \Bigg[\sum_{\boldsymbol{v}\in V_0} \left(p_0 W_{\rho_0}(\boldsymbol{u}) W_{\psi_0}(\boldsymbol{v})\right)^\alpha \left(q_0 W_{\tau_0}(\boldsymbol{u}) W_{\psi_0}(\boldsymbol{v})\right)^{1-\alpha} + \\
&\qquad \sum_{\boldsymbol{v}\in V_1} \left(p_1 W_{\rho_1}(\boldsymbol{u}) W_{\psi_1}(\boldsymbol{v})\right)^\alpha \left(q_1 W_{\tau_1}(\boldsymbol{u}) W_{\psi_1}(\boldsymbol{v})\right)^{1-\alpha} \Bigg] \\
=&\sum_{\boldsymbol{u}\in\mathcal{P}} \left[p_0^\alpha q_0^{1-\alpha} W_{\rho_0}(\boldsymbol{u})^\alpha W_{\tau_0}(\boldsymbol{u})^{1-\alpha} \sum_{\boldsymbol{v}\in V_0} W_{\psi_0}(\boldsymbol{v}) + p_1^\alpha q_1^{1-\alpha} W_{\rho_1}(\boldsymbol{u})^\alpha W_{\tau_1}(\boldsymbol{u})^{1-\alpha} \sum_{\boldsymbol{v}\in V_1} W_{\psi_1}(\boldsymbol{v}) \right] \\
=&p_0^\alpha q_0^{1-\alpha} \sum_{\boldsymbol{u}\in\mathcal{P}} W_{\rho_0}(\boldsymbol{u})^\alpha W_{\tau_0}(\boldsymbol{u})^{1-\alpha} + p_1^\alpha q_1^{1-\alpha} \sum_{\boldsymbol{u}'\in\mathcal{P}} W_{\rho_1}(\boldsymbol{u}')^\alpha W_{\tau_1}(\boldsymbol{u}')^{1-\alpha} \\
=&p_0^\alpha q_0^{1-\alpha} Q_\alpha(W_{\rho_0} \,||\, W_{\tau_0}) + p_1^\alpha q_1^{1-\alpha} Q_\alpha(W_{\rho_1} \,||\, W_{\tau_1}),
\end{aligned} \tag{5.64}$$

where in the third equality we used the normalization of the representation. The inequality in the statement of the Claim then follows from the fact that both terms on the right hand side of Eq. (5.59) must be non-negative, completing the proof. □

With this property in hand we can make the non-trivial n–dependence in ΔD_α more explicit and moreover we highlight in Claim 5.8 the fact that, in case of a failed run, the choice of CSS state the system is left in is arbitrary.

Claim 5.8. *Let us define the function $Q_\alpha(\boldsymbol{w} \,||\, \boldsymbol{r}) \coloneqq 2^{(\alpha-1)D_\alpha(\boldsymbol{w}\,||\,\boldsymbol{r})}$, for quasi-distribution $\boldsymbol{w}$ and probability distribution $\boldsymbol{r}$ with positive components. Let the maximally mixed state on k qubits be written as $\mathbb{1}_k/2^k$,. Then, we have:*

$$\Delta D_\alpha = n\left(1 - H_\alpha[W_\rho]\right) + k - \frac{1}{\alpha-1}\log\left[p^\alpha Q_\alpha\left(W_{\rho'} \middle|\middle| W_{\mathbb{1}_k/2^k}\right) + (1-p)^\alpha \left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1} \right]. \tag{5.65}$$

It follows that the function ΔD_α is independent of the choice of state σ.

Proof. By Claim 5.7, we have the following expansion

$$\begin{aligned} Q_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right) &= p^\alpha \left(2^{k-n}\right)^{1-\alpha} Q_\alpha\left(W_{\rho'} \,||\, W_{\frac{\mathbb{1}_k}{2^k}}\right) + (1-p)^\alpha \left(1-2^{k-n}\right)^{1-\alpha} Q_\alpha\left(W_\sigma \,||\, W_\sigma\right) \\ &= \left(2^{n-k}\right)^{\alpha-1} \left[p^\alpha Q_\alpha\left(W_{\rho'} \,||\, W_{\frac{\mathbb{1}_k}{2^k}}\right) + (1-p)^\alpha \left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1}\right], \end{aligned} \tag{5.66}$$

where in the last equality we have used $Q_\alpha(\boldsymbol{r} \,||\, \boldsymbol{r}) = 1$ for all probability distributions $\boldsymbol{r}$. Therefore, since $D_\alpha(\cdot \,||\, \cdot) = \frac{1}{\alpha-1}\log Q_\alpha(\cdot \,||\, \cdot)$ we have

$$D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right) = n - k + \frac{1}{\alpha-1}\log\left[p^\alpha Q_\alpha\left(W_{\rho'} \,||\, W_{\frac{\mathbb{1}_k}{2^k}}\right) + (1-p)^\alpha \left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1}\right]. \tag{5.67}$$

Substituting Eq. (5.67) and $D_\alpha\left(W_\rho \middle|\middle| W_{\frac{\mathbb{1}}{2}}\right) = 2 - H_\alpha[W_\rho]$ into Eq. (5.57) gives the result as claimed. □

By inspection, the form for ΔD_α given in Claim 5.8 has no dependence on σ. This also follows from a resource-theoretic argument. In particular, we observe that the following quantum channel is straightforwardly stochastically represented for any state $\omega \in \mathcal{F}$

$$\mathcal{E}'(\cdot) := \mathbb{1} \otimes P_0(\cdot) + \sigma' \mathrm{tr} \otimes P_1(\cdot), \tag{5.68}$$

where $P_k(\cdot) := |k\rangle\langle k|\,(\cdot)\,|k\rangle\langle k|$ for $k = 0, 1$, since it is a composite channel formed from elements in $\mathcal{O}$. It then follows that for any σ' in the interior of $\mathcal{F}$, we have

$$\begin{aligned} D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right) &\geq D_\alpha\left(W_{\mathcal{E}'(\rho_p)} \,||\, W_{\mathcal{E}'(\tau_{n,k})}\right) \geq D_\alpha\left(W_{(\mathcal{E}\circ\mathcal{E}')(\rho_p)} \,||\, W_{(\mathcal{E}\circ\mathcal{E}')(\tau_{n,k}}\right) \\ &= D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right). \end{aligned} \tag{5.69}$$

We therefore conclude that

$$D_\alpha\left(\mathcal{E}'(W_{\rho_p}) \,||\, \mathcal{E}'(W_{\tau_{n,k}})\right) = D_\alpha\left(W_{\rho_p} \,||\, W_{\tau_{n,k}}\right), \tag{5.70}$$

and thus the corresponding bounds are unaffected by the choice of the reference state σ.

We now highlight some properties of the relative entropy difference ΔD_α stemming from basic properties of α–Rényi entropies and divergences.

Claim 5.9. *The relative entropy difference ΔD_α satisfies the following properties:*

1. *ΔD_α is concave in n on the domain $n \in [k, \infty]$;*

2. *ΔD_α is negative in the limit where $n = k$:*

$$\lim_{n\to k^+} \Delta D_\alpha < 0; \tag{5.71}$$

3. *If $H_\alpha[W_\rho] > 1$, then ΔD_α is also negative in the asymptotic limit*

$$\lim_{n\to\infty} \Delta D_\alpha < 0, \tag{5.72}$$

for $p < 1$, any rebit states ρ, ρ', and $\alpha = \frac{2a}{2b-1}$ with positive integers a, b such that $a \geq b$.

Proof.

1. Let us define the function

$$g(n) \coloneqq \left[c_1 + c_2 \left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1}\right]. \tag{5.73}$$

We can now re-express ΔD_α from Claim 5.8 as

$$\Delta D_\alpha = n\left(1 - H_\alpha[W_\rho]\right) + k - \frac{1}{\alpha-1}\log g(n), \tag{5.74}$$

and since the first term is linear we need only check the second derivative of the second term to establish that ΔD_α is concave. We have

$$\begin{aligned}\partial_n^2 \Delta D_\alpha &= -\frac{1}{\alpha-1}\partial_n^2 \log g(n) \\ &= -\left[\frac{\log 2\, c_2 2^{k+n}\left(c_1\left(2^k + (\alpha-1)2^n\right)\left(2^{n-k}-1\right)^\alpha + c_2\left(2^n - 2^k\right)\right)}{\left(2^n - 2^k\right)\left(c_1 2^k\left(2^{n-k}-1\right)^\alpha + c_2\left(2^n - 2^k\right)\right)^2}\right].\end{aligned} \tag{5.75}$$

Since $c_1, c_2 \geq 0$ for all ρ' and p, the term in square brackets is non-negative for all $n > k, \alpha > 1, \rho'$ and p (strictly positive for $p < 1$), which implies $\partial_n^2 \Delta D_\alpha$ is non-positive everywhere on our restricted domain. Therefore ΔD_α is concave, as claimed.

2. Recalling that $\alpha > 1$, we have from Claim 5.8 that

$$\begin{aligned}\lim_{n\to k^+} \Delta D_\alpha &= \lim_{n\to k^+}\left\{n\left(1 - H_\alpha[W_\rho]\right) + k - \frac{1}{\alpha-1}\log\left[c_1 + c_2\left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1}\right]\right\} \\ &= -kH_\alpha[W_{\rho'}] - \frac{1}{\alpha-1}\lim_{n\to k^+}\left\{\log\left[c_1 + c_2\left(\frac{1}{2^{n-k}-1}\right)^{\alpha-1}\right]\right\} \\ &= -\infty < 0,\end{aligned} \tag{5.76}$$

so long as $c_2 > 1$, which is true if and only if $p < 1$.

3. We have:

$$\begin{aligned}
\lim_{n\to\infty} \Delta D_\alpha &= \lim_{n\to\infty} \left\{ n\left(1 - H_\alpha[W_\rho]\right) + k - \frac{1}{\alpha - 1} \log\left[c_1 + c_2 \left(\frac{1}{2^{n-k} - 1} \right)^{\alpha - 1} \right] \right\} \\
&= k - \frac{1}{\alpha - 1} \log[c_1] + \lim_{n\to\infty} \left\{ n\left(1 - H_\alpha[W_\rho]\right) \right\} \\
&= k - D_\alpha \left(W_{p\rho'} \,||\, W_{\frac{\mathbb{1}_k}{2^k}} \right) + \lim_{n\to\infty} \left\{ n\left(1 - H_\alpha[W_\rho]\right) \right\} \\
&= H_\alpha[W_{p\rho'}] - k + \lim_{n\to\infty} \left\{ n\left(1 - H_\alpha[W_\rho]\right) \right\} \\
&= \begin{cases} -\infty, & H_\alpha[W_\rho] > 1, \\ H_\alpha[W_{p\rho'}] - k, & H_\alpha[W_\rho] = 1, \\ +\infty, & H_\alpha[W_\rho] < 1. \end{cases}
\end{aligned} \tag{5.77}$$

Therefore, if $H_\alpha[W_\rho] > 1$ then $\lim_{n\to\infty} \Delta D_\alpha < 0$, as claimed.

□

An immediate consequence of Claim 5.9 is that if ΔD_α is non-negative anywhere on its well-defined domain, i.e., an n–k CSS code projection protocol is not completely ruled out for any number of input copies n, then it has one or two roots located at n_L^α and n_U^α. These roots correspond to lower and upper bounds on the permissible code length n, respectively. More formally, we arrive at the following statement.

Claim 5.10. *Let ρ be a noisy magic state on $\mathcal{H}_2$. If $\rho^{\otimes n} \mapsto p\rho'$ under a n–k CSS code projection with acceptance probability p, then we have the following lower and upper bounds on n, respectively*

$$n \geq n_L^\alpha \coloneqq \inf_n \{ n : \Delta D_\alpha \geq 0 \}, \tag{5.78}$$

$$n \leq n_U^\alpha \coloneqq \sup_n \{ n : \Delta D_\alpha \geq 0 \}, \tag{5.79}$$

for $\alpha = \frac{2a}{2b-1}$ with positive integers a, b such that $a \geq b$. Moreover, if there exists an α such that $H_\alpha[W_\rho] > 1$, we obtain a finite upper bound on n from this latter expression.

We highlight that n in Claim 5.10 refers to the code length of a single run of a distillation protocol, as opposed to the the asymptotic overhead. However, single-run n still constitutes a useful metric for analyzing the actual resource cost of a given stage of a protocol. Moreover, distillation costs are typically dominated by the final round of a multi-stage distillation protocol, see for example [43] and references

contained therein, so we expect the above bounds to be particularly informative in this context.

For sufficiently low k, these bounds can be computed numerically using basic root-finding methods. However, it turns out that by upper bounding n_U^α and lower bounding n_L^α, we can find useful closed analytic forms for upper and lower bounds on n. To this end, let us first define the trace norm, also known as the Schatten-1 norm of an operator X,

$$\|X\|_1 \coloneqq \mathrm{tr}|X| = \mathrm{tr}\sqrt{X^\dagger X}. \tag{5.80}$$

For a vector $\boldsymbol{w} \in \mathbb{R}^d$, this reduces to the ℓ_1–norm,

$$\|\boldsymbol{w}\|_1 \coloneqq \sum_{i=1}^{d} |w_i|. \tag{5.81}$$

For a Hermitian matrix H, $\|H\|_1 = \|\boldsymbol{\lambda}\|_1$, where $\boldsymbol{\lambda}$ denotes the vector of the eigenvalues of H. Using these standard norms, we can reproduce a basic result stating that two states being close is equivalent to their Wigner representations being close.

Claim 5.11. *If $\|\rho - \sigma\|_1 \leq \delta$ for states $\rho, \sigma \in \mathcal{B}(\mathcal{H}_2^{\otimes n}$, then $\|W_\rho - W_\sigma\|_1 \leq \delta$, where the Wigner representations W_ρ, W_τ are defined according to Eq. (2.65).*

Proof. The operator $\rho - \sigma$ is Hermitian, so we can write out its eigendecomposition as

$$\rho - \sigma = \sum_i \lambda_i \left|\lambda_i\right\rangle\!\left\langle\lambda_i\right|. \tag{5.82}$$

Then, we can proceed as follows,

$$\begin{aligned}
\|W_\rho - W_\sigma\|_1 &= \sum_{\boldsymbol{z}} |W_\rho(\boldsymbol{z}) - W_\sigma(\boldsymbol{z})| = \frac{1}{2^n} \sum_{\boldsymbol{z}} \left|\mathrm{tr}[A_{\boldsymbol{z}}^\dagger(\rho - \sigma)]\right| \\
&= \frac{1}{2^n} \sum_{\boldsymbol{z}} \left| \sum_i \lambda_i \left\langle\lambda_i\right| A_{\boldsymbol{z}}^\dagger \left|\lambda_i\right\rangle \right| \leq \frac{1}{2^n} \sum_{\boldsymbol{z},i} |\lambda_i| \left|\left\langle\lambda_i\right| A_{\boldsymbol{z}}^\dagger \left|\lambda_i\right\rangle\right| \\
&= \frac{1}{2^n} \sum_{\boldsymbol{z},i} |\lambda_i| \left\langle\lambda_i\right| A_{\boldsymbol{z}}^\dagger \left|\lambda_i\right\rangle = \sum_i |\lambda_i| \left\langle\lambda_i\right| \left[\frac{1}{2^n} \sum_{\boldsymbol{z}} A_{\boldsymbol{z}}^\dagger\right] \left|\lambda_i\right\rangle \\
&= \sum_i |\lambda_i| = \|\rho - \sigma\|_1.
\end{aligned} \tag{5.83}$$

The inequality is an application of the Cauchy-Schwarz inequality, and it is followed by employing the properties of positivity and completeness for the qubit phase-point operators as laid out in Claim 2.12. □

We can further bound the entropic distance between representations when the states are δ–close.

Claim 5.12. *Let ρ and σ be two quantum states on $\mathcal{H}_2^{\otimes n}$ such that $\|\rho - \sigma\|_1 \leq \delta$. Then,*

$$H_\alpha(W_\rho) - H_\alpha(W_\sigma) \leq \frac{\alpha}{\alpha - 1} \log\left(1 + 4^n \delta\right). \tag{5.84}$$

Proof. We restate Theorem 7 of [174] for two 4^n–dimensional probability distributions $\boldsymbol{w}, \boldsymbol{w}'$,

$$|H_\alpha(\boldsymbol{w}) - H_\alpha(\boldsymbol{w}')| \leq \frac{\alpha}{\alpha - 1} \log\left(1 + 4^n \|\boldsymbol{w} - \boldsymbol{w}'\|_1\right). \tag{5.85}$$

We can show that this continuity statement on the α–Rényi entropies applies to quasi-distributions as well. The proof of this is essentially reliant on the monotonicity of p–norms,

$$\|\boldsymbol{w}\|_p \geq \|\boldsymbol{w}\|_q, \text{ for } 1 \leq p < q \leq \infty, \tag{5.86}$$

which also holds for quasi-distributions $\boldsymbol{w}$ of dimension 4^n, where

$$\|\boldsymbol{w}\|_p \coloneqq \left(\sum_{i=1}^{4^n} |w_i|^p\right)^{1/p}. \tag{5.87}$$

The result then follows immediately from Claim 5.11. □

We are now in a position to derive analytic constraints on n–k CSS code projection protocols.

Claim 5.13. *Let ρ be a rebit magic state and assume that $\rho^{\otimes n} \to p\rho'$, under a CSS code projection, where $\left\|\rho' - \psi^{\otimes k}\right\|_1 \leq \delta$ and ψ is a pure rebit magic state.*

Then, for all $\alpha = 2a/(2b-1)$ with positive integers a, b such that $a \geq b$, we have the following family of upper bounds on n,

$$n \leq \frac{k\left[H_\alpha(W_\psi) - 1\right] + \frac{\alpha}{1-\alpha} \log \frac{p}{1+4\delta}}{H_\alpha(W_\rho) - 1}, \tag{5.88}$$

when $H_\alpha(W_\rho) > 1$ and the following family of lower bounds on n,

$$n \geq \frac{k[1 - H_\alpha(W_\psi)] - \frac{\alpha}{1-\alpha} \log \frac{p}{1+4\delta}}{1 - H_\alpha(W_\rho)}, \tag{5.89}$$

$H_\alpha(W_\rho) < 1$.

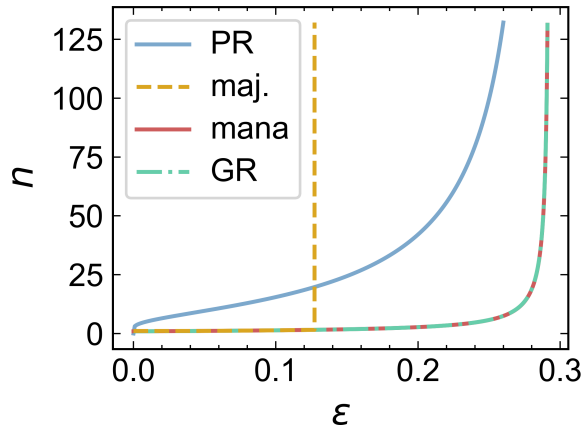

Figure 5.2: **(Lower bound comparison).** We plot lower bounds on the number of copies n of the noisy H–state $(1-\epsilon)\left|H\right\rangle\!\left\langle H\right| + \epsilon\frac{\mathbb{1}}{2}$ required to distill a single output qubit $|H\rangle$ with output noise $\epsilon' = 10^{-9}$ and acceptance probability $p = 0.9$ under a CSS code projection protocol, plotted as a function on input noise ϵ. Our lower bound from majorization (maj.) is shown to be tighter those from mana [104] and generalized robustness (GR) in [89]. However, it only outperforms the lower bound from projective robustness (PR) in [175] in the high p, high ϵ regime.

Proof. From Claim 5.7, we have

$$Q_\alpha(W_{\rho_p} \,||\, W_{\tau_{n,k}}) \geq p^\alpha \left(\frac{1}{2^{n-k}}\right)^{1-\alpha} Q_\alpha\left(W_{\rho'} \middle|\middle| W_{\mathbb{1}/2^k}\right). \tag{5.90}$$

Since for $\alpha > 1$, $\log x/(\alpha-1)$ is a monotonically increasing function in $x \in \mathbb{R}$, it follows that

$$\begin{aligned} D_\alpha(W_{\rho_p} \,||\, W_{\tau_{n,k}}) &\geq \frac{1}{\alpha-1}\log\left[p^\alpha \left(\frac{1}{2^{n-k}}\right)^{1-\alpha} Q_\alpha\left(W_{\rho'} \middle|\middle| W_{\mathbb{1}/2^k}\right)\right] && (5.91)\\ &= D_\alpha(W_{\rho'} \,||\, W_{\mathbb{1}/2^k}) + \frac{\alpha}{\alpha-1}\log p + (n-k) && (5.92)\\ &= k - H_\alpha(W_{\rho'}) + \frac{\alpha}{\alpha-1}\log p + n, && (5.93) \end{aligned}$$

where in the final equality we used the identity $D_\alpha(W_\rho \,||\, W_{\mathbb{1}_d/d}) = 2\log d - H_\alpha(W_\rho)$.

We can make use of the continuity of the α–Rényi entropy from Claim 5.12 to further lower bound this divergence as

$$\begin{aligned} D_\alpha(W_{\rho_p} \,||\, W_{\tau_{n,k}}) &\geq k - H_\alpha(W_{\psi^{\otimes k}}) - \frac{\alpha}{\alpha-1}\log(1+4\delta) + \frac{\alpha}{\alpha-1}\log p + n \\ &= k\left[1 - H_\alpha(W_\psi)\right] - \frac{\alpha}{1-\alpha}\log\frac{p}{1+4\delta} + n, \end{aligned} \tag{5.94}$$

where the equality follows from the multiplicativity of the α–Rényi entropy.

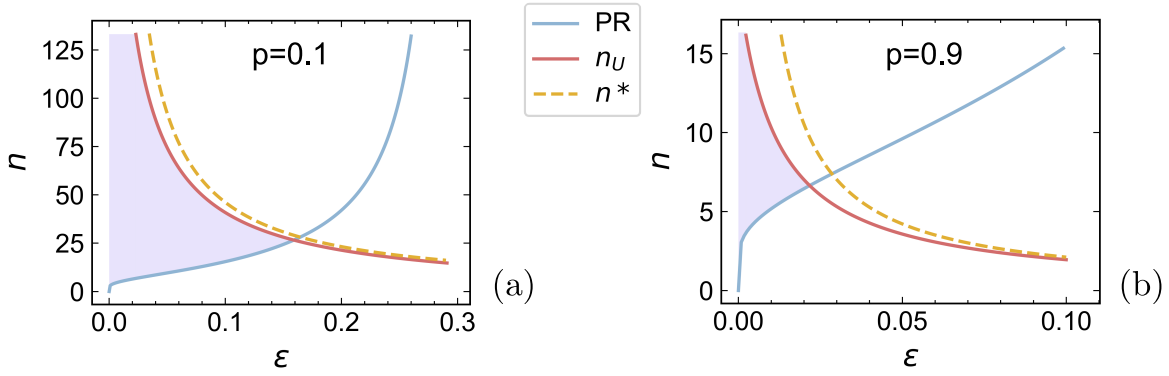

Figure 5.3: **Finite bounds on CSS code lengths for magic state distillation protocols.** We plot upper and lower bounds on the number of copies n of the noisy H–state $(1-\epsilon)\,|H\rangle\langle H| + \epsilon\frac{\mathbb{1}}{2}$ required to distill a single output qubit $|H\rangle$ with output noise $\epsilon' = 10^{-9}$ via projection onto the code-space of an $[[n,1]]$ CSS code. The shaded purple region shows the accessible region of parameter space prescribed by the intersection of our numeric upper bound n_U (red curve) defined in Claim 5.10 and the lower bound from projective robustness (PR) introduced in [175] (blue curve). The analytic upper bound n^* (dashed yellow curve) defined inEq. (5.100) is shown to form a good approximation to the numeric bound n_U. (a) Bounds for $p = 0.1$. (b) Increasing p to $p = 0.9$ the upper bounds become considerably tighter.

This gives rise to the following upper bound on the relative entropy difference $\Delta D_\alpha \coloneqq nD_\alpha(W_\rho \,||\, W_{\mathbb{1}/d}) - D_\alpha(W_{\rho_p} \,||\, W_{\tau_{n,k}})$ as follows

$$0 \leq \Delta D_\alpha \leq n\big(1 - H_\alpha(W_\rho)\big) + k\big(H_\alpha(W_\psi) - 1\big) + \frac{\alpha}{1-\alpha}\log\frac{p}{1+4\delta}. \tag{5.95}$$

This gives a weaker but still necessary constraint on the transformation $\rho^{\otimes n} \mapsto \rho_p$ and $(\mathbb{1}/2)^{\otimes n} \mapsto \tau_{n,k}$, which we rearrange to write as

$$n(H_\alpha(W_\rho) - \log d) \leq k(H_\alpha(W_\psi) - 1) + \frac{\alpha}{1-\alpha}\log\frac{p}{1+4\delta}. \tag{5.96}$$

Therefore, for $H_\alpha(W_\rho) > 1$, we can rearrange Eq. (5.96) to obtain Eq. (5.88), whereas for $H_\alpha(W_\rho) < 1$ we obtain Eq. (5.89), which completes the proof. □

5.3.3 Hadamard state distillation

As an explicit example illustrating the results of Claim 5.13, we discuss here the n–1 distillation of the Hadamard state which is given in Eq. (2.88). It is sufficient to consider input magic states of the form

$$\rho_H(\epsilon) \coloneqq (1-\epsilon)\,|H\rangle\langle H| + \epsilon\frac{\mathbb{1}}{2}, \tag{5.97}$$

with $0 \leq \epsilon \leq 1$ depolarization noise, since any other input magic state ρ can be converted into this form via the pre-processing channel

$$\mathcal{E}_{\text{prep}}(\cdot) \coloneqq \frac{1}{2}\mathbb{1}(\cdot)\mathbb{1} + \frac{1}{2}H(\cdot)H \tag{5.98}$$

without altering its fidelity with respect to the Hadamard state $|H\rangle$.

In Fig. (5.2) we plot the performance of our lower bounds applied to any CSS code projection protocol for a noisy H–state. In all parameter regimes, our lower bounds are observed to be tighter than mana [104] and the generalized robustness bound of [89]. Our lower bound gives tighter constraints than the p–independent projective robustness bound [175] in the high p, high ϵ regime. Outside of this regime, however, our upper bound is still able to give additional constraints over the projective robustness bound. In particular, in Fig. (5.3) we plot the combined information of our upper bound with the lower bound from projective robustness. As illustrated, combining our upper bound with a lower bound significantly restricts the parameter regime in which distillation is allowed.

Taking CSS protocols to be our free operations, the appearance of upper bounds on n might first seem to contradict a resource theory perspective, where, since discarding subsystems is a CSS operation, we would anticipate $n+1$ copies of a noisy magic state to be at least as good as n copies at distilling magic. However, in specializing to code projection protocols, we are in fact considering a resource theoretic approach to magic subject to the additional constraint $(\mathbb{1}/2)^{\otimes n} \to \tau_{n,k}$. Crucially, the output state depends non-trivially on n.

For n–1 CSS code projection protocols of Hadamard distillation, one can also show that there always exists a valid set of indices α such that

$$H_\alpha\left(W_{\rho_H(\epsilon)}\right) \geq H_\alpha\left(W_{|H\rangle\langle H|}\right) > 1, \tag{5.99}$$

for all ϵ for which $W_{\rho_H(\epsilon)}$ contains a negative component. This is illustrated in Fig. (5.4), where we see that even the $\epsilon = 0$ condition gives a finite range of α such that $H_\alpha(W_{|H\rangle\langle H|}) > 1$. We can therefore conclude that n–1 CSS code projection protocols for H–state distillation are ruled out in the limit $n \to \infty$ as there is always valid upper bound on n.

Moreover, we find that for the n–1 Hadamard-state distillation protocols, the upper bound in Eq. (5.88) takes on a particularly simple form. By evaluating the $\alpha = 2$ condition explicitly, we find that given a code projection $\mathcal{E}$ such that $\mathcal{E}[\rho_H(\epsilon)^{\otimes n}] = p\rho'$, where $\|\rho' - |H\rangle\langle H|\|_1 \leq \epsilon'$, then

$$n \leq n^* = \frac{2\log\left(\frac{1+4\delta}{p}\right)}{1 - \log[1 + (1-\epsilon)^2]}. \tag{5.100}$$

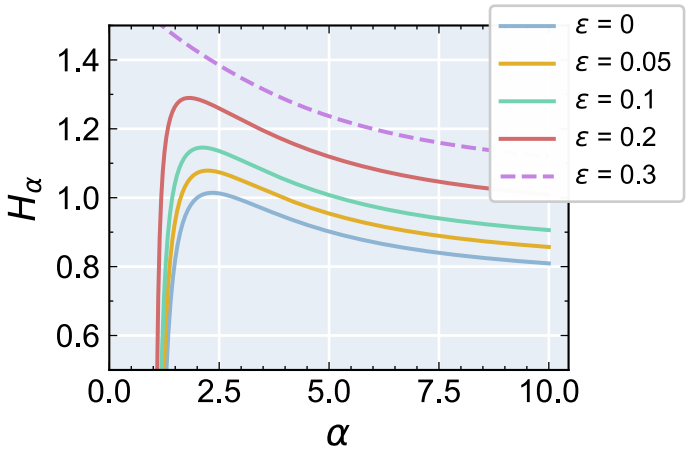

Figure 5.4: **α–Rényi entropies & magic distillation** We plot the condition in Claim 5.10 for the existence of finite upper bounds on n in an n–1 CSS distillation for the qubit H–state distillation, signified by region where $H_\alpha[W_{\rho(\epsilon)}] > 1$. Even in the limit of zero input noise $\epsilon = 0$ we obtain a valid set of permissible α, which implies that H–state distillation under n–1 CSS code projection is ruled out in the asymptotic limit $n \to \infty$. We further highlight that the noise level $\epsilon = 0.3$ (dashed curve) is outside of the region where $\rho(\epsilon)$ is magic ($0 \leq \epsilon < 1 - \frac{1}{\sqrt{2}}$), and therefore $W_{\rho(\epsilon)}$ is a proper probability distribution at $\epsilon = 0.3$, which is why H_α is only seen to satisfy standard monotonicity properties at this input error.

This expression captures the fact that under a CSS code projection protocol, there is a fundamental trade-off between acceptance probability and output fidelity. In particular, for any fixed code length n and input noise $\epsilon > 0$, we cannot simultaneously obtain zero output noise $\epsilon' = 0$ and unit acceptance probability $p = 1$. To further investigate this trade-off, in Fig. (5.5)(b) we plot the *maximum achievable fidelity* with respect to the Hadamard state,

$$F_{\max}(\rho) = \max_{\mathcal{E}} \left\{ \langle H | \rho' | H \rangle \, : \, \mathcal{E}(\rho^{\otimes n}) \mapsto p\rho' \right\}, \tag{5.101}$$

under an n–1 CSS code projection. The maximization is performed over the acceptance probability that defines the set of all CSS code projection protocols.

5.3.4 Majorization bounds and data processing

We have seen that stochastically represented $[[n, k]]$ code projection protocols give rise to a set of no-go results according to Claim 5.13, in the form of a set of upper bounds on n. By comparing to the data-processing inequality (DPI) on quantum states [114], we see that although the existence of upper bounds is a generic feature

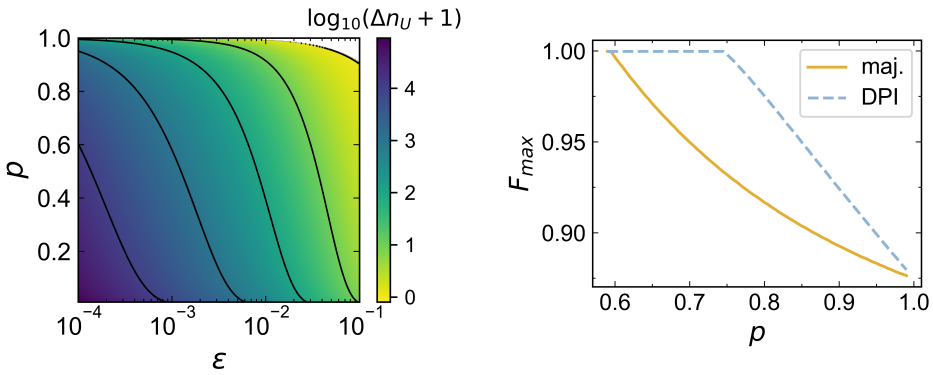

Figure 5.5: **Majorization gives independent constraints over DPI.** (a) Shown is how (scaled) $\Delta n_U := n_U^{\mathrm{DPI}} - n_U^{\mathrm{maj}}$ varies over all possible values of acceptance probability p and a realistic range of input noise ϵ, with fixed $\epsilon' = 10^{-9}$. Whenever we have $\log_{10}(\Delta n_U + 1) > 0$ means that upper bounds from majorization give tighter constraints than the DPI, reaching $\Delta n_U = O(10^4)$ in the low p, low ϵ regime. (b) We show the trade-off relation given by bounds on the maximum achievable fidelity $F_{\max}(\rho)$ vs. target acceptance probability p, under an n–1 CSS code projection, where $\rho = \frac{3}{4}|H\rangle\langle H| + \frac{1}{8}\mathbb{1}$. For p above a given threshold (≈ 0.6) no perfect distillation is theoretically possible, even for $n \to \infty$ copies of the input state. Majorization (maj.) is shown to give stronger constraints than that of DPI.

of code projection protocols, exploiting the stochasticity of the protocol representations gives strictly stronger bounds.

To begin, we note that if there exists a (stochastic or otherwise) code projection channel $\mathcal{E}$ that maps $\rho^{\otimes n} \to p\rho'$, then the DPI for quantum channels states that

$$\Delta\tilde{D}_\alpha := n\tilde{D}_\alpha\left(\rho\middle\|\frac{\mathbb{1}}{2}\right) - \tilde{D}_\alpha\left(\rho_p \,||\, \tau_{n,k}\right) \geq 0, \tag{5.102}$$

for all $\alpha \in (1,\infty)$ [176], where $\tilde{D}_\alpha(\rho \,||\, \tau)$ is the sandwiched α–Rényi divergence [177,178] on the normalized quantum states ρ and τ, which is defined for $\alpha \in (1,\infty)$ as

$$\tilde{D}_\alpha(\rho \,||\, \tau) := \frac{1}{\alpha - 1}\log \mathrm{tr}\left[\left(\tau^{\frac{1-\alpha}{2\alpha}}\rho\tau^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]. \tag{5.103}$$

Now if $\rho^{\otimes n} \to p\rho'$ under any such channel, then we have the following family of upper bounds on n,

$$n \leq \tilde{n}_U := \min_\alpha \max_n \{n : \Delta\tilde{D}_\alpha \geq 0\}. \tag{5.104}$$

which turns out to be finite whenever α is such that $H_\alpha(\rho) > 1$.

Since one can also obtain upper bounds from simple data processing of quantum states, we now ask whether majorization gives genuinely new constraints on magic state distillation using CSS protocols that go beyond DPI constraints. Since our majorization conditions are a consequence of the stochastic representation of all CSS operations, while the DPI arises from the fact that all quantum channels are CPTP, this question may be loosely rephrased as asking whether stochasticity imposes additional constraints beyond those imposed by CPTP on CSS magic state distillation.

Fig. (5.5) shows that we can answer this question in the affirmative, since our majorization upper bound is observed to impose much stronger restrictions than that given by the DPI over a wide range of parameter regimes. For example, in Fig. (5.5)(a) the low acceptance probability p and input noise ϵ regime, we find the difference in upper bounds $\Delta n_U \coloneqq n_U^{\mathrm{DPI}} - n_U^{\mathrm{maj}}$ is of the order 10^4. We thus conclude that the upper bounds on CSS code projections stemming from majorization on quasi-distributions go beyond those of DPI.

Overall, we have overcome known obstacles in the qubit Wigner representation, to develop a meaningful statistical mechanical framework that allows us to derive bounds on qubit magic distillation in the form of relative entropy changes between the magic states and reference "free" states throughout the distillation process. To this end, we first establish in Claim 5.5 the stochasticity of a particular, useful subclass of stabilizer protocols, called the n–k CSS code projections. We then proceed to demonstrate how our statistical mechanical framework on quasi-distributions from chapter 3 leads to the existence of upper and lower bounds on the number of input copies n in Claim 5.9, before deriving explicit expressions for these bounds in Claim 5.13.

We can view the derived bounds as non-trivial trade-off relations between protocol parameters, including the input and output number of copies n, k, errors ϵ, ϵ' as well as the protocol success probability p or fidelity of the output state δ, i.e. the distance δ of the output state from a pure magic state. Therefore, they provide significant constraints on the parameter regimes that magic distillation is allowed in the context of qubit CSS code projections. Our lower bounds complement previous work on lower bounds from qubit magic monotones, while the appearance of upper bounds is striking as they constitute the first set of trade-off relations that act as fundamental upper bounds on the resource cost for a family of distillation protocols. Upper bounds are therefore particularly promising in bounding allowed parameter regimes.

Chapter 6

Faster Born probability estimation

In chapter 4 and chapter 5, we have utilized the Wigner representation of quantum states and channels to characterize magic state transformations in qudit systems of odd and even dimension d respectively. The work of the preceding two chapters thus sets bounds on the potential of quantum processes to provide quantum computational advantage. In this chapter, we explore the boundary between classical and quantum computational capabilities, via a different approach: we enhance the ability of classical estimation methods to approximate quantum processes. For this purpose, we use a family of representations, called frames, that generalize the Wigner representations used in the previous chapters. The negativity present in these representations is still the central figure of merit, but it now manifests itself as the cost of the classical sampling algorithms we use to sample quantum circuit probabilities. We therefore develop methods to parameterize the frames that represent a given quantum circuit, allowing us to minimize the total negativity present, by varying the parameters.

This chapter is organized as follows. In section 6.1, we review known methods of classical estimation of quantum probabilities and outline our original contributions. In section 6.2, we provide parameterizations applicable to quasi-probability representations for qudits of any dimension, as well as the main known sampling algorithm that uses quasi-probability representations of the components of a given quantum circuit. In section 6.3, we outline our results within the context of the current state of quasi-probability simulators, and then we proceed to describe our two novel sub-routines in detail in section 6.4 and section 6.5.

6.1 Techniques for quantum probability estimation

6.1.1 Review of probability estimation methods

Probability estimation methods are varied and aim to explore the efficiency of circuit sampling or simulation beyond the regime of quantum circuits that admit tractable classical representations. The estimation methods we mention throughout this chapter can be classified under one of two leading approaches [89,179]. The first involves stabilizer rank-based simulators [113, 180–185], which rely on approximating the circuit components by stabilizer operators. Every state or operation is assigned an exact or approximate stabilizer rank [181] indicating the number of stabilizer (Pauli) operators required to perform an exact or approximate decomposition of that component. If a circuit component is non-classical, its stabilizer rank grows large thus inducing an exponential run-time cost for the estimation of the outcome probability. Algorithms based on stabilizer decompositions have been very successful in estimating outcome probabilities of circuits dominated by Clifford gates and supplemented by a few types of magic states [113,179,180]. These Clifford simulators are generalized from pure to noisy settings by the recently proposed density-operator stabilizer-rank simulator [89]. Furthermore, computing the stabilizer rank of arbitrary gates appears to be an intractable problem in the general case, so recent improvements on computing stabilizer rank bounds for specific non-Clifford states enhance run-times significantly [186].

The other family of estimation methods relies on quasi-probabilistic representations of circuit components [40, 89, 92, 187–189] which are more general forms of Wigner representations, as we discuss in section 6.2. Such methods are in principle directly applicable to any quantum circuit without the need for state decompositions, in particular circuits with induced noise. They are based on the notion of a *frame representation* for the components of the circuit [95, 96]. Specifically, all components are represented by quasi-distributions in a certain frame and sampling on these distributions can be performed. Since any state or gate admits such a representation, quasi-probability simulators naturally apply to arbitrary circuits with noise. Many such frames have been studied [95–97, 189–192], and the run-time depends on the total negativity that is present in the circuit representation [187]. A notable frame simulator is the dyadic frame simulator [89] which relies on operator decompositions into stabilizer dyads $|L\rangle\langle R|$, where $|L\rangle$ and $|R\rangle$ are pure stabilizer states. This method assigns dyadic negativity to non-classical elements, which quantifies the extent to which the operator's optimal linear decomposition into stabilizer dyads departs from a convex combination. The dyadic simulator is a state-of-the-art quasi-

probability frame simulator for qubits, as demonstrated by its low run-time scaling $O(4^{0.228t})$ with t non-Clifford gates [89]. However, optimizing the decomposition of an operator in dyads is computationally challenging.

Stabilizer rank simulators generally offer two advantages over estimation methods based on frame representations. Firstly, they can be used for sampling the circuit output probabilities, which can be viewed as a stronger notion of simulation than probability estimation. Frame representation methods produce probability estimates with additive precision, which does not suffice for sampling [21, 180]. Secondly, the stabilizer-rank algorithms developed in [89, 113, 180] are quadratically faster as they achieve a scaling of $O(2^{0.228t})$ in the asymptotic limit. However, specialized simulators (e.g. [113, 180]) suffer from additional polynomial run-time factors, which tend to be more significant compared to the exponential run-time for the experimentally relevant case of big circuits with a low number of non-Clifford elements. Recently, an algorithm of additive precision [179] has also been shown to asymptotically outperform the methods of [113, 180], at least in certain parameter regimes.

6.1.2 Novel contributions to improved estimation

In this chapter, we focus on quasi-probability estimation methods based on frame representations and look for a way to improve the performance of outcome probability estimation. Recently, there has been a proposal of a Monte Carlo sampling algorithm which allows for quasi-probability estimation of circuits that contain a bounded amount of negativity in their representation [187]. For classes of circuits in which negativity grows only polynomially in the number of input states, this estimation algorithm is efficient. The negativity of the circuit therefore indicates the hardness of the probability sampling problem. Although the negativity scales exponentially with the number of non-Clifford gates, the scaling factors hugely depend on the frame choice. Until now, however, the same fixed representation has been applied on every circuit component and the flexibility on reducing negativity has been limited.

Our aim is to explore the extent to which varying the frame representations of the components in a given circuit can lead to a reduction in the total circuit negativity. To this end, we propose a pre-processing routine for any general quantum circuit, which aims at reducing the negativity overhead required for probability estimation. Our proposed routine consists of two distinct sub-routines:

1. **Circuit gate merging:** We introduce the idea of merging gates together into new n–qudit gates for fixed n in the context of reducing sampling overhead.

This sub-routine reduces the negativity of the entire circuit and is independent of the estimation method used.

We demonstrate numerically that the average negativity reduction over a random ensemble of circuits is greater as the number of non-Clifford elements, e.g., T gates, increases and is comparable to recent asymptotic negativity bounds [89, 193, 194]. Our routine does not depend on the specifics of the circuit gate set and can therefore be used in cases of gates which are hard to decompose, e.g., Haar-random gates.

2. **Frame optimization:** We introduce the idea of using different frames to represent the input and output phase spaces of the gates in the circuit. This is inspired by work in continuous variables [195], but our approach is novel in the context of discrete quasi-probability sampling methods.

   We argue that this sub-routine compliments gate merging as an additional source of negativity reduction when merging is no longer efficient. We then demonstrate numerically that instances of Clifford+T circuits and circuits with Haar-random gates admit significant negativity reductions by introducing additional frames in the circuit representation.

We note that a polynomial run-time for these classical sub-routines with respect to the circuit size should be guaranteed to effectively reduce the overall run-time of the sampling method. As proof of principle, we provide explicit algorithms in the main text that ensure this condition for each sub-routine.

6.2 Preliminaries

6.2.1 Frame representation of quantum circuits

We first give a brief overview on classical circuit sampling based on the method of frame representation. This is a generalization of the Wigner representation presented in chapter 2. Suppose that an N–qudit quantum circuit C is composed of the initial state preparation ρ, sequential quantum gates $U_1, U_2, \ldots, U_L$ and the measurement effect E. The outcome probability of the quantum circuit

$$p_C = \mathrm{tr}[U_L \ldots U_2 U_1 \rho U_1^\dagger U_2^\dagger \ldots U_L^\dagger E] \tag{6.1}$$

can be estimated by describing the quantum state ρ as quasi-distributions over phase space points $\boldsymbol{\lambda} \in \mathbb{Z}_d^{2N}$ and the quantum operations U_i as the transition matrices of the distributions. More specifically, a phase space can be constructed from a frame

defined as a set of operators $\mathcal{F} \coloneqq \{F(\boldsymbol{\lambda})\}$ and its dual $\mathcal{G} \coloneqq \{G(\boldsymbol{\lambda})\}$ [95, 96], such that any operator O is expressed as

$$O = \sum_{\boldsymbol{\lambda}} \mathrm{tr}[F(\boldsymbol{\lambda})O]G(\boldsymbol{\lambda}). \tag{6.2}$$

For a given frame, the outcome probability can be expressed in terms of the representation as

$$p_C = \sum_{\boldsymbol{\lambda}_0,\ldots,\boldsymbol{\lambda}_L} W_E(\boldsymbol{\lambda}_L) \left[\prod_{l=1}^{L} W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})\right] W_\rho(\boldsymbol{\lambda}_0), \tag{6.3}$$

where we define

$$W_\rho(\boldsymbol{\lambda}) = \mathrm{tr}[F(\boldsymbol{\lambda})\rho], \tag{6.4}$$

$$W_U(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) = \mathrm{tr}[F(\boldsymbol{\lambda}')UG(\boldsymbol{\lambda})U^\dagger], \text{ and} \tag{6.5}$$

$$W_E(\boldsymbol{\lambda}) = \mathrm{tr}[EG(\boldsymbol{\lambda})]. \tag{6.6}$$

The notation is suggestive as the frame representation is a generalized version of the Wigner representation discussed in chapter 2. In the case where 1) ρ and E are products of local initial states and measurement effects, 2) $W_\rho(\boldsymbol{\lambda}_0)$ and $W_E(\boldsymbol{\lambda}_L)$ are classical probability distributions, and 3) $W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})$, for $\ell = 1, \ldots, L$, are classical conditional probability distributions for all l, efficient classical simulation is possible, where the sampling run-time scales polynomially with N and L [103]. The simulation is performed by sampling the trajectories of $(\boldsymbol{\lambda}_0, \ldots, \boldsymbol{\lambda}_L)$ from the initial distribution $P(\boldsymbol{\lambda}_0) = W_\rho(\boldsymbol{\lambda}_0)$ and the transition matrix at each step $P_l(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1}) = W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})$, which leads to the probability estimate, $\hat{p}_C = W_E(\boldsymbol{\lambda}_L)$. Taking an average over M probability estimates converges to the Born probability as M increases.

6.2.2 Overhead of classical simulation

Non-classicality in the quantum process is represented by negativities in $W_\rho(\boldsymbol{\lambda})$ or $W_U(\boldsymbol{\lambda}'|\boldsymbol{\lambda})$, which gives rise to quasi-probabilities. In general, $W_\rho(\boldsymbol{\lambda})$ and $W_U(\boldsymbol{\lambda}'|\boldsymbol{\lambda})$ consist of real components that can attain negative values, while satisfying the normalisation conditions,

$$\sum_{\boldsymbol{\lambda}\in\mathbb{Z}_d^{2N}} W_\rho(\boldsymbol{\lambda}) = 1 \text{ and} \tag{6.7}$$

$$\sum_{\boldsymbol{\lambda}'\in\mathbb{Z}_d^{2N}} W_U(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) = 1 \text{ for all } \boldsymbol{\lambda} \in \mathbb{Z}_d^{2N}. \tag{6.8}$$

Despite the presence of negativities in the distributions and update matrices, Monte Carlo methods can still be used with adjustments as introduced by Pashayan *et*

al. [187] in order to perform probability sampling. This can be done by sampling over $P(\boldsymbol{\lambda}_0) = |W_\rho(\boldsymbol{\lambda}_0)|/\sum_{\boldsymbol{\lambda}_0}|W_\rho(\boldsymbol{\lambda}_0)|$ for the initial state preparation and taking the transition matrix of $P_l(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1}) = |W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})|/\sum_{\boldsymbol{\lambda}_l}|W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})|$ for the quantum gate, while keep track of the signs. In this case, the probability estimate is modified to

$$\hat{p}_C = \mathrm{sign}\left(W_\rho(\boldsymbol{\lambda}_0)\prod_{l=1}^{L}W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})\right)\times N_\rho\left(\prod_{l=1}^{L}N_{U_l}(\boldsymbol{\lambda}_{l-1})\right)W_E(\boldsymbol{\lambda}_L), \tag{6.9}$$

where we have defined

$$N_\rho \coloneqq \sum_{\boldsymbol{\lambda}_0}|W_\rho(\boldsymbol{\lambda}_0)| \tag{6.10}$$

$$N_{U_l}(\boldsymbol{\lambda}_{l-1}) \coloneqq \sum_{\boldsymbol{\lambda}_l}|W_{U_l}(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})|\,. \tag{6.11}$$

In order to converge to the Born probability, one can similarly take the average of increasingly many probability estimates sampled over trajectories $(\boldsymbol{\lambda}_0,\ldots,\boldsymbol{\lambda}_L)$ using distributions $P(\boldsymbol{\lambda}_0)$ and $P_l(\boldsymbol{\lambda}_l|\boldsymbol{\lambda}_{l-1})$.

This directly relates the total amount of circuit negativity to the computational overhead: the larger the negativity in the circuit, the more samples required for an accurate estimation.

Claim 6.1 (Pashayan *et al.* [187])**.** *The outcome probability p_C of the quantum circuit C can be estimated by $\hat{p}_C$ from the number of samples*

$$M \geq M(\epsilon,\delta) = \frac{2}{\epsilon^2}N_C^2\ln(2/\delta), \tag{6.12}$$

with at least probability $1-\delta$ of having error less than ϵ. Here,

$$N_C = N_\rho\times\left[\prod_{l=1}^{L}\max_{\boldsymbol{\lambda}_0,\ldots,\boldsymbol{\lambda}_{L-1}}N_{U_l}(\boldsymbol{\lambda}_{l-1})\right]\times\max_{\boldsymbol{\lambda}_L}|W_E(\boldsymbol{\lambda}_L)|\,, \tag{6.13}$$

is the (maximum) circuit negativity.

As is clear by Eq. (6.12)), the negativity of the circuit acts as an overhead for the convergence time of the sampling algorithm, therefore it is desirable to reduce it before executing the sampling by considering different frame choices.

6.3 Main results

In this section, we develop a pre-processing routine to reduce the negativity of the circuit, which in turn reduces the number of samples required to estimate the outcome probability of the circuit. Our routine is applicable to any general circuit consisting of a product input state and product measurement, but independently of the

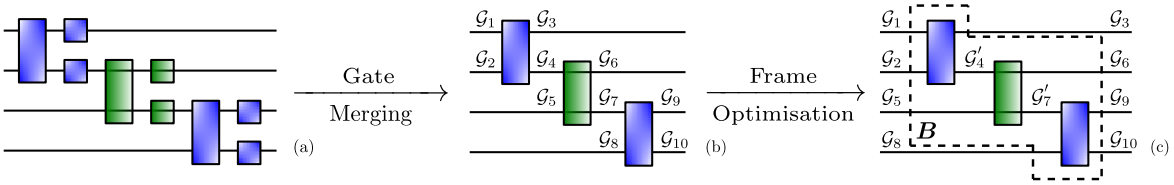

Figure 6.1: **Sketch of routine on a toy circuit**. The first step (a) $\rightarrow$ (b) is gate merging, here implemented with $n = 2$. Gates that share input and output wires merge in the schematic way depicted in the figure. The second step (b) $\rightarrow$ (c) is frame optimization, here implemented with $\ell = 2$ in a block $\mathbf{B}$ comprising the three merged gates. The optimization results into updated frames $\mathcal{G}'_4, \mathcal{G}'_7$, while the remaining frames that connect the block $\mathbf{B}$ to the rest of the circuit components are left unchanged at this optimization cycle.

input state dimension, the gate set (e.g. Clifford unitaries or Haar-random gates) and adaptive operations based on intermediate measurement outcomes.

6.3.1 Frame parameterization

The central focus of our work is to consider frame parameterizations that are allowed to vary across the circuit. It is clear from the definitions that the circuit negativity of a given circuit in Eq. (6.13) depends on the choice of the frame $\mathcal{G}$ and its dual $\mathcal{G}$. Note that $\mathcal{G}$ and $\mathcal{G}$ are uniquely defined by each other for phase space dimension equal to d^2, where d is the qudit dimension. We therefore make the dependence clear by labelling the representation functions by $\mathcal{G}$:

$$W_\rho^{\mathcal{G}}(\boldsymbol{\lambda}) = \mathrm{tr}[F(\boldsymbol{\lambda})\rho], \tag{6.14}$$

$$W_U^{\mathcal{G}'|\mathcal{G}}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) = \mathrm{tr}[F'(\boldsymbol{\lambda}')UG(\boldsymbol{\lambda})U^\dagger], \text{ and} \tag{6.15}$$

$$W_E^{\mathcal{G}}(\boldsymbol{\lambda}) = \mathrm{tr}[EG(\boldsymbol{\lambda})], \tag{6.16}$$

where we used different frames, $\mathcal{G}$ and $\mathcal{G}'$, for the input and output wires respectively in the definition of $W_U^{\mathcal{G}'|\mathcal{G}}$.

In order to ensure that the number of frames does not grow exponentially with the number of qudits N, we restrict to *product frames* that are constructed as tensor products of single qudit frames. This allows us to parameterize each single qudit phase space separately, rather than the entire N–qudit phase space. Therefore, we reserve the label $\mathcal{G}$ for denoting single qudit frames and the boldface symbol $\boldsymbol{\mathcal{G}}$ for denoting a set of single qudit frames. The negativity of each circuit component can

now be expressed as

$$N_\rho^{\boldsymbol{\mathcal{G}}} = \sum_{\boldsymbol{\lambda}} \left| W_\rho^{\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}) \right|, \tag{6.17}$$

$$N_U^{\boldsymbol{\mathcal{G}}'|\boldsymbol{\mathcal{G}}} = \max_{\boldsymbol{\lambda}} \left[\sum_{\boldsymbol{\lambda}'} \left| W_U^{\boldsymbol{\mathcal{G}}'|\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}'|\boldsymbol{\lambda}) \right| \right], \text{ and} \tag{6.18}$$

$$N_E^{\boldsymbol{\mathcal{G}}} = \max_{\boldsymbol{\lambda}} \left| W_E^{\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}) \right|, \tag{6.19}$$

where $\boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{G}}'$ contain elements from the complete set of frames required to represent the circuit. In practice, each circuit component is parameterized only via the frames that correspond to its input and output wires. For example, in Fig. (6.1)(b), when parameterizing the first gate in the sequence, we can simply consider $\boldsymbol{\mathcal{G}}$ as the set $\{\mathcal{G}_1, \mathcal{G}_2\}$ and $\boldsymbol{\mathcal{G}}'$ as the set $\{\mathcal{G}_3, \mathcal{G}_4\}$. If only a unique frame representation $\mathcal{G}$ is used for all circuit components, then $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}}' = \{\mathcal{G}\}$ in the expressions above and the label can be dropped, simplifying to the notation of the previous section. The total negativity of the parameterized circuit can now be expressed as a function of the circuit frame set $\boldsymbol{\mathcal{G}}$:

$$N_C(\boldsymbol{\mathcal{G}}) = N_\rho^{\boldsymbol{\mathcal{G}}} \times \left[\prod_{l=1}^{L} N_{U_l}^{\boldsymbol{\mathcal{G}}'|\boldsymbol{\mathcal{G}}} \right] \times N_E^{\boldsymbol{\mathcal{G}}}. \tag{6.20}$$

We note that Claim 6.1 still holds by replacing N_C with the more general form $N_C(\boldsymbol{\mathcal{G}})$. Our main objective is to study the reduction of this circuit negativity by tuning $\boldsymbol{\mathcal{G}}$.

6.3.2 Examples of frame parameterizations

While our results are general and applicable to any family of parameterized frames, in this chapter we deal with two examples of explicit, product frame parameterizations: (i) parameterized Wigner frames and (ii) rotated Pauli frames.

Parameterized Wigner frames employ the conventional phase space of the discrete Wigner representation [96, 97]. As a reminder, we re-express the discrete displacement operator for a d–dimensional system as

$$D(p, q) = \chi(-2^{-1}pq) Z^p X^q, \tag{6.21}$$

where $\chi(q) = e^{i(2\pi/d)q}$. For a qubit system ($d = 2$), this takes the form $D(p, q) = i^{pq} Z^p X^q$. It can be generalized to an N–qudit system as

$$D(\boldsymbol{\lambda}) = \bigotimes_{i=1}^{N} D(p_i, q_i), \tag{6.22}$$

where $\boldsymbol{\lambda} \coloneqq (p_1, q_1, p_2, q_2, \ldots, p_N, q_N)^T \in \mathbb{Z}_d^{2N}$ denotes a phase space point of the whole system. We then define the frame $\mathcal{F} = \{F(\boldsymbol{\lambda})\} \coloneqq \{D(\boldsymbol{\lambda})F_0 D^\dagger(\boldsymbol{\lambda})\}$ and its dual frame $\mathcal{G} = \{G(\boldsymbol{\lambda})\} \coloneqq \{D(\boldsymbol{\lambda})G_0 D^\dagger(\boldsymbol{\lambda})\}$ using the following reference operators:

$$F_0 = \frac{1}{d}\sum_{\boldsymbol{\lambda}} \left[\frac{1}{g(\boldsymbol{\lambda})}\right] D(\boldsymbol{\lambda}) \tag{6.23}$$

$$G_0 = \frac{1}{d}\sum_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}) D(-\boldsymbol{\lambda})\,, \tag{6.24}$$

where we introduced the parameterization function $g(\boldsymbol{\lambda})$. Note that the following relation holds:

$$g(\boldsymbol{\lambda}) = \mathrm{tr}\left[G_0 D(\boldsymbol{\lambda})\right]\,, \tag{6.25}$$

so the parameterization function $g(\boldsymbol{\lambda}) : \mathbb{Z}_d^{2N} \mapsto \mathbb{C} \setminus \{0\}$ can be fully characterized by the reference operator G_0. In order to impose that $W_\rho^{\mathcal{G}}(\boldsymbol{\lambda})$ is real-valued and that $\sum_{\boldsymbol{\lambda}} W_\rho^{\mathcal{G}}(\boldsymbol{\lambda}) = 1$, we need the additional conditions $g^*(\boldsymbol{\omega}) = g(-\boldsymbol{\omega})$ and $g(\mathbf{0}) = 1$, which are equivalent to $G_0^\dagger = G_0$ and $\mathrm{tr}[G_0] = 1$ respectively. By taking $g(\boldsymbol{\lambda}) = 1$ for all $\boldsymbol{\lambda}$, the conventional discrete Wigner representation [97] is recovered. One can calculate the quasi-distributions of circuit elements via Eq. (6.14)-Eq. (6.16) using the defined frame and dual frame. In odd dimensions, the parameterized Wigner frame is a good choice for Clifford dominated circuits as Clifford gates do not possess any negativity in the conventional Wigner representation. Therefore, $g(\boldsymbol{\lambda}) = 1$ is already optimal for most circuit elements when considered in isolation and constitutes an obvious starting point for frame optimization.

In the qubit case, it is known that the Hadamard and CNOT gates have non-zero negativity even in the conventional Wigner representation [196], which motivates us to introduce the next frame parameterization, valid only for qubits: the rotated Pauli frames.

Rotated Pauli frames are based on the Bloch decomposition of a quantum operator. Consider the set of displacement operators for a single qubit $\{D(\boldsymbol{\lambda})\}$ as defined in Eq. (6.21) for $\boldsymbol{\lambda} \in \mathbb{Z}_2^2 = \{(0,0),(0,1),(1,0),(1,1)\}$. The usual Bloch vector for a single-qubit state ρ can be written as

$$W_\rho(\boldsymbol{\lambda}) = \frac{1}{2}\mathrm{tr}\left[\rho D(\boldsymbol{\lambda})\right]\,, \tag{6.26}$$

and this defines a quasi-distribution with frame $\{\frac{1}{2}D(\boldsymbol{\lambda})\}$. We can define a new frame by applying a rotation to the space of the Bloch vector. Let us consider a rotational angle vector $\boldsymbol{\theta} \coloneqq (\theta_X, \theta_Y, \theta_Z)$ and a corresponding rotation operator

$R(\boldsymbol{\theta}) \coloneqq R(\theta_Z)R(\theta_Y)R(\theta_X)$, where $R(\theta_X) \coloneqq e^{-i\theta_X X/2}$ and similarly for Y, Z. Applying this to the Bloch vector in Eq. (6.26) results in a set of rotated displacement operators, parameterized by $\boldsymbol{\theta}$:

$$D^{\boldsymbol{\theta}}(0,0) \coloneqq \mathbb{1}, \tag{6.27}$$

$$D^{\boldsymbol{\theta}}(0,1) \coloneqq \begin{pmatrix} -\sin\theta_Y & e^{-i\theta_X}\cos\theta_Y \\ e^{+i\theta_X}\cos\theta_Y & \sin\theta_Y \end{pmatrix}, \tag{6.28}$$

$$D^{\boldsymbol{\theta}}(1,0) \coloneqq \begin{pmatrix} \cos\theta_Y\cos\theta_Z & e^{-i\theta_X}(\sin\theta_Y\cos\theta_Z + i\sin\theta_Z) \\ e^{+i\theta_X}(\sin\theta_Y\cos\theta_Z - i\sin\theta_Z) & -\cos\theta_Y\cos\theta_Z \end{pmatrix}, \tag{6.29}$$

$$D^{\boldsymbol{\theta}}(1,1) \coloneqq \begin{pmatrix} \cos\theta_Y\sin\theta_Z & e^{-i\theta_X}(\sin\theta_Y\sin\theta_Z + i\cos\theta_Z) \\ e^{+i\theta_X}(\sin\theta_Y\sin\theta_Z - i\cos\theta_Z) & -\cos\theta_Y\sin\theta_Z \end{pmatrix}. \tag{6.30}$$

Then, we define the frame $\mathcal{F} = \{F(\boldsymbol{\lambda})\}$ and its dual frame $\mathcal{G} = \{G(\boldsymbol{\lambda})\}$ as

$$F(\boldsymbol{\lambda}) \coloneqq \frac{1}{2^N} D^{\boldsymbol{\theta}}(\boldsymbol{\lambda}), \tag{6.31}$$

$$G(\boldsymbol{\lambda}) \coloneqq D^{\boldsymbol{\theta}}(\boldsymbol{\lambda}), \tag{6.32}$$

which provide a parameterized frame representation for a qubit. This can be generalized to an N–qubit system via

$$D^{\boldsymbol{\theta}}(\boldsymbol{\lambda}) = \bigotimes_{i=1}^{N} D^{\boldsymbol{\theta}_i}(\boldsymbol{\lambda}_i) \tag{6.33}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N)$, where $\boldsymbol{\theta}_i$ is the rotational angle vector for the i–th qubit. The rotated Pauli frames possess the desired property that all stabilizer states and Clifford gates have zero negativity in the conventional Bloch frame representation with $\boldsymbol{\theta} = (0,0,0)$. Thus, when a given qubit circuit is dominated by Clifford gates, it can be advantageous to employ the rotated Pauli frame.

6.3.3 Pre-processing routine for negativity reduction

The central idea of our pre-processing routine for negativity reduction can now be expressed by the following lower bounds on gate negativity.

Claim 6.2. *For two consecutive gates U and V, the following bounds on negativity hold:*

$$N_V^{\mathcal{G}|\mathcal{G}} N_U^{\mathcal{G}|\mathcal{G}} \;\geq\; \min_{\mathcal{G}'} \; N_V^{\mathcal{G}|\mathcal{G}'} N_U^{\mathcal{G}'|\mathcal{G}} \;\geq\; N_{VU}^{\mathcal{G}|\mathcal{G}}, \tag{6.34}$$

where $\mathcal{G}$ and $\mathcal{G}'$ are frame sets that represent gates U, V and UV.

Proof. The first inequality holds since $\boldsymbol{\mathcal{G}}$ is one specific choice of the optimization variable set $\boldsymbol{\mathcal{G}}'$. The second inequality is due to Claim 6.3 in the next section. □

Claim 6.2 motivates us to introduce two sub-routines applicable to any quasi-probability estimation algorithm with run-time cost determined by the circuit negativity.

The second inequality in Claim 6.2 suggests that merging two gates into one is generally advantageous in minimizing the total negativity. This leads to the first pre-processing sub-routine, *gate merging*. The inequality is independent of the specific frame parameterization and can be directly extended to an arbitrary number of gates. The trade-off is that the merged gate may be of a larger size. For example, if U and V are 2-qudit gates sharing one wire between them, gate VU will be a 3-qudit gate. The dimension of the merged gate increases exponentially as the number of qudits involved becomes larger, hence one should truncate the maximum number of qudits acted on by the merged gates, which we define as the spatial parameter n.

The first inequality in Claim 6.2 states that, unless the frames between two gates in sequence are already optimal, we can always reduce the total negativity of the two gates by optimizing the frames they share. This leads to the second sub-routine, *frame optimization*. The optimization can be directly generalized to a circuit block $\mathbf{B}$ containing a sequence of ℓ frames $\boldsymbol{\mathcal{G}}$ by simultaneously optimizing all the frames in the block, $\min_{\boldsymbol{\mathcal{G}}} N_{\mathbf{B}}(\boldsymbol{\mathcal{G}})$. The temporal parameter ℓ is the number of frames to be optimized in one optimization cycle. The optimization takes place iteratively in the sense that every optimization cycle optimizes the frames within a block, taking as an initial state the optimized frames obtained from the previous cycle. This ensures that negativity cannot increase above its initial value, no matter how many optimization cycles occur.

Given fixed values for the truncation parameters n, ℓ, we show in the following two sections that the total run-time τ of our routine is polynomial in the number of circuit components,

$$\tau = O(N, L^2). \tag{6.35}$$

In general, larger n or ℓ give larger negativity reduction at the cost of additional classical computation.

We note that gate merging yields lower negativity than any frame optimization between the gates. However, fixing $n < N$ prevents us from merging gates indefinitely, so frame optimization can then be used for further negativity reduction.

We present an algorithm for Born probability estimation, including our complete pre-processing routine and sampling, in Algorithm 1 and illustrate its implementation on a toy circuit in Fig. (6.1). In the following two sections, we discuss in more

Algorithm 1 Outcome probability estimation with merging and optimization

Input: An N–qudit quantum circuit C with a product input state $\rho = \rho_1 \otimes \cdots \otimes \rho_N$, the list of gates $\mathcal{U} = \{U_1, ..., U_L\}$, and the product measurement operator $E = E_1 \otimes \cdots \otimes E_N$; the spatial parameter n; the temporal parameter ℓ; the desired accuracy ϵ.

1: Run gate merging (Sub-routine 1) with the input gate sequence and n and return the merged gate sequence $\{V_1, ..., V_{L'}\}$ with $L' \leq L$ consisting of gates acting on at most n qudits.

2: Run frame optimization (Sub-routine 2) with the merged circuit and ℓ and return the optimized frame sequence $\boldsymbol{\mathcal{G}}_{\text{opt}}$.

3: Run a sampling algorithm to achieve the input accuracy ϵ according to Eq. (6.12) using the quasi-probability representations of the merged circuit obtained with the optimized frame sequence $\boldsymbol{\mathcal{G}}_{\text{opt}}$.

Output: p_{est}, the estimated outcome probability.

detail how the two sub-routines, gate merging and frame optimization, can be implemented. For clarity, we focus on qubit circuits and on the frame parameterizations introduced in the previous section, although our methods are general.

6.4 Gate merging

The central idea of our first sub-routine, gate merging, is that the sampling cost of a merged circuit block consisting of multiple quantum gates is in general lower than sequential sampling of each gate. More precisely, this can be summarized as the following observation:

Claim 6.3. *Let $\{U_1, U_2, \ldots, U_k\}$ be a sequence of quantum gates. The negativity of the merged gate $U = U_k \ldots U_2 U_1$ is always less or equal to the product of the individual negativities, i.e.,*

$$N_U^{\boldsymbol{\mathcal{G}}} \leq \prod_{i=1}^{k} N_{U_i}^{\boldsymbol{\mathcal{G}}}, \tag{6.36}$$

for any frame set $\boldsymbol{\mathcal{G}}$ assigned to the gate sequence.

Proof. It is sufficient to prove the statement for two gates U and V. By noting that the quasi-probability of the merged gate is expressed as

$$W_{VU}^{\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_1) = \sum_{\boldsymbol{\lambda}_2} W_V^{\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_2) W_U^{\boldsymbol{\mathcal{G}}}(\boldsymbol{\lambda}_2|\boldsymbol{\lambda}_1), \tag{6.37}$$

the negativity of the gate can be bounded as

$$\begin{aligned} N_{VU}^{\mathcal{G}} &= \max_{\boldsymbol{\lambda}_1} \sum_{\boldsymbol{\lambda}_3} \left| W_{VU}^{\mathcal{G}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_1) \right| \\ &= \max_{\boldsymbol{\lambda}_1} \sum_{\boldsymbol{\lambda}_3} \left| \sum_{\boldsymbol{\lambda}_2} W_V^{\mathcal{G}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_2) W_U^{\mathcal{G}}(\boldsymbol{\lambda}_2|\boldsymbol{\lambda}_1) \right| \\ &\leq \max_{\boldsymbol{\lambda}_1} \sum_{\boldsymbol{\lambda}_2} \left| W_U^{\mathcal{G}}(\boldsymbol{\lambda}_2|\boldsymbol{\lambda}_1) \right| \sum_{\boldsymbol{\lambda}_3} \left| W_V^{\mathcal{G}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_2) \right| \\ &\leq N_U^{\mathcal{G}} \max_{\boldsymbol{\lambda}_2} \sum_{\boldsymbol{\lambda}_3} \left| W_V^{\mathcal{G}}(\boldsymbol{\lambda}_3|\boldsymbol{\lambda}_2) \right| \\ &= N_V^{\mathcal{G}} N_U^{\mathcal{G}}. \end{aligned} \tag{6.38}$$

We then apply this argument iteratively to any sequence of gates $\{U_1, U_2, \ldots, U_k\}$ to obtain Eq. (6.36), which completes the proof. □

Such a negativity reduction can be exemplified by considering the Toffoli gate, which can be optimally decomposed into four T gates [197] along with Clifford gates and Pauli measurements. We compare the negativity of the Toffoli gate itself and its decomposed gate sequence using the Pauli frame, where the negativity only comes from non-Clifford gates. One can readily observe that the Toffoli gate negativity $N_{\text{Toffoli}}^{\text{Pauli}} = 2$ is lower than the total negativity of the decomposed gate sequence $\left[N_T^{\text{Pauli}}\right]^4 = 4$.

The idea of reducing the negativity of quantum gates by merging (Eq. (6.36)) can be compared to the sub-multiplicativity of magic state negativity characterized by the robustness measure ($\mathcal{R}$), which obeys $\mathcal{R}(\rho_1 \otimes \rho_2) \leq \mathcal{R}(\rho_1)\mathcal{R}(\rho_2)$ [193]. In particular, the robustness of the T state is equivalent to the negativity of the T gate from the sampling cost viewpoint, as one T gate can be constructed using a single T state along with Clifford gates [113]. In [193], the asymptotic negativity per single T gate is $\lim_{t\to\infty} \left[\mathcal{R}\left(|T\rangle^{\otimes t}\right)\right]^{1/t} \approx 2^{0.272}$ which provides a lower bound on their sampling run-time $\Omega(4^{0.272t})$.

In order to compare this with the gate merging method, we consider an n–qubit block consisting of Clifford+T gates (see Fig. (6.2)(f) for an example with $n = 5$). This can be compared to considering n T states in the robustness measure, having the same number of qubits (i.e. the size of Hilbert space) in the block to evaluate the negativity.

Fig. (6.2)(a-d) show the distribution of the negativity of 1000 random n–qubit blocks consisting of 100 Clifford gates and t T gates. We observe that the negativity per T gate after merging the gate sequences in a random n–qubit block can be occasionally lower than the robustness measure of n T states [193]. We also note that

the negativity reduction works efficiently when the number of T gate in the block, t, increases. For example, when $n = 5$ and $t = 15$, 95% of the randomly chosen merged blocks yield negativity per T state lower than the robustness measure. We also plot in Fig. (6.2)(e) the average negativity per T gate versus t, demonstrating that it is decreasing, which implies that our approach can prove efficient when the structure of the gate block considered becomes more complicated.

The main advantage of our approach is that it is not limited to a particular type of gate set, e.g. Clifford+T circuits, but can be directly applied to any types of quantum gates. The aforementioned approaches using stabilizer rank, robustness and generalized robustness rely on the gadgetization of a non-Clifford gate using magic states. Therefore, evaluating the classical overhead should be preceded by finding an optimal Clifford gadget with minimum resource of magic. On the other hand, gate merging does not have such a limitation, so it can be useful when the efficient decomposition of a quantum circuit into non-stabilizer states and Clifford gates is non-trivial. We also highlight that merging gates reduces the negativity independently of the choice of frames.

We now describe the gate merging method for a generic N–qubit quantum circuit with L gates. This can be done by grouping the quantum circuit into n–qubit blocks (see Fig. (6.1)(a)$\rightarrow$(b)), then Claim 6.3 guarantees that the negativity of each block is reduced after merging the gate sequences in it. There are various ways of grouping the circuit into n–qubit blocks, but we introduce the iterative Sub-routine 1 for concreteness. The broad idea of the sub-routine is to iteratively connect any yet unmerged (disjoint) gates. All gates remain in the set $\mathcal{U}_{\text{disj}}$ until they either finally act on n qubits or cannot connect to other gates anymore, when they are move to the output set $\mathcal{U}_{\text{merged}}$.

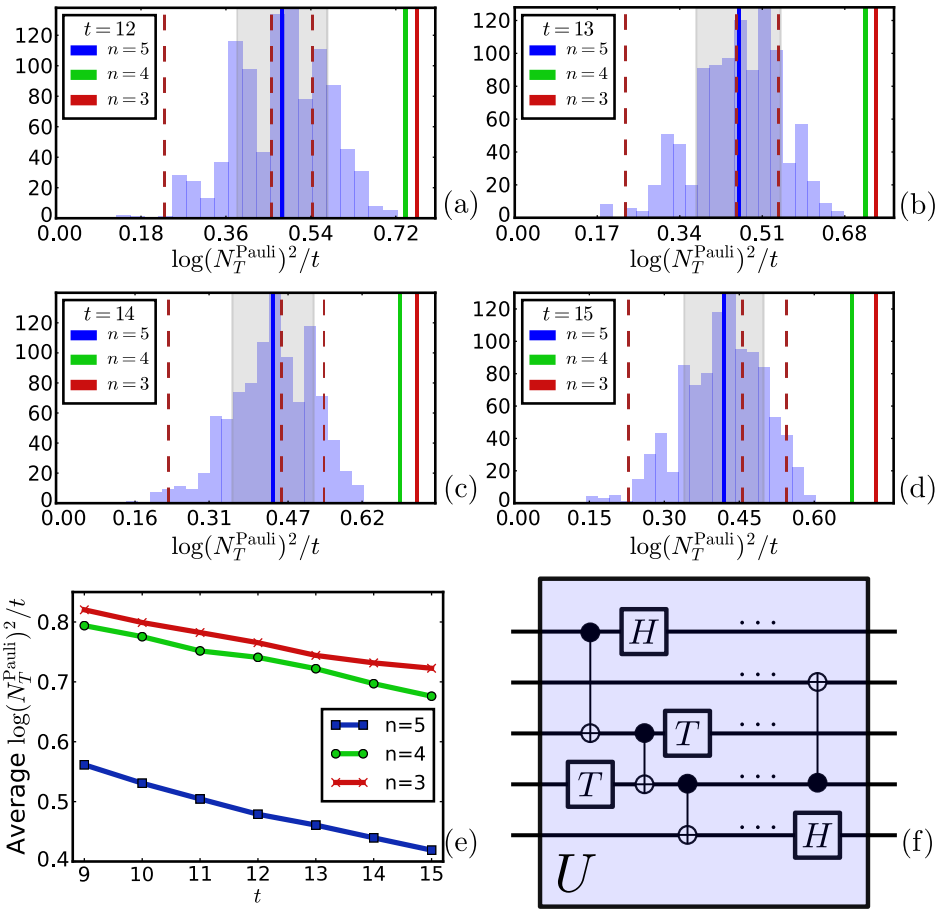

Figure 6.2: **Gate merging analysis.** Histograms of 1000 random Clifford+T circuits with $N = 5$ consisting of 100 1-qubit and 2-qubit Clifford gates, supplemented by t T gates and merged using spatial parameter $n = 5$. The leftmost (blue) solid line with the gray region depict the average and standard deviation of each histogram. The brown and green solid lines (from right to left) represent the higher averages of the corresponding histograms for $n = 3$ and 4 respectively. Vertical dashed lines provide some state-of-the art scalings, more specifically from left to right: $O(2^{0.228t})$ of the Bravyi-Gosset algorithm from [113] based on the stabilizer rank, $O(4^{0.228t})$ of the dyadic frame simulator from [89] and the lower bound $\Omega(4^{0.272t})$ based on the robustness of magic from [193]. As t increases, we observe a higher frequency of circuits with log negativity squared per T gate lower than the robustness lower bound: (a) 71%, (b) 81%, (c) 89%, (d) 95%. (e) Histogram average for $n = 3, 4, 5$ against t. (f) Example 5–qubit merged gate U made up from Clifford gates (CNOT and H) and T gates.

Sub-routine 1 Gate merging

Input: List of gates $\mathcal{U} = \{U_1, ..., U_L\}$ in qudit quantum circuit C and spatial parameter n.

Define list of merged gates $\mathcal{U}_{\text{merged}} \leftarrow \{\}$, and list of disjoint gates $\mathcal{U}_{\text{disj}} \leftarrow \{\}$
for $U_i \in \mathcal{U}$ **do**
    Set target gate $U_{\text{target}} \leftarrow U_i$
    for $V \in \mathcal{U}_{\text{disj}}$ **do**
        if U_{target} shares a wire with V **then**
            Remove V from $\mathcal{U}_{\text{disj}}$.
            if $\text{rank}(U_{\text{target}}V) > d^n$ **then**
                Add V to $\mathcal{U}_{\text{merged}}$.
            else if $\text{rank}(U_{\text{target}}V) \leq d^n$ **then**
                $U_{\text{target}} \leftarrow U_{\text{target}}V$.
    Add U_{target} to $\mathcal{U}_{\text{disj}}$.
for $U_i \in \mathcal{U}_{\text{merged}}$ **do**
    Set target gate $U_{\text{target}} \leftarrow U_i$
    for $V \in \mathcal{U}_{\text{disj}}$ **do**
        if $\text{rank}(U_{\text{target}}V) \leq d^n$ **then**
            $U_{\text{target}} \leftarrow U_{\text{target}}V$.
    Add U_{target} to $\mathcal{U}_{\text{disj}}$.
Append $\mathcal{U}_{\text{disj}}$ to $\mathcal{U}_{\text{merged}}$.

Output: $\mathcal{U}_{\text{merged}}$.

At every step, a target gate U_{target}, the algorithm searches through the disjoint gates to find the next one that is connected to U_{target}. We therefore require to search less than L gates for every U_{target}, while the cost of merging two gates (i.e., multiplying) is $O(2^{2n})$, which is a constant as we fix $n < N$. So the full gate merging sub-routine scales as $O(2^{2n}L^2)$. The computational cost to compute the transition matrix $W_U^{\mathcal{G}}$ for n–qubit unitary U and its negativity also exponentially scales with n as there are $O(2^{2n})$ possible phase space points for a n–qubit system.

As we can observe from the scaling, the limiting factor of gate merging is the spatial parameter n, which stems from the exponential growth of the dimension of Hilbert space by increasing the number of qubits. We find numerically that a practical choice for the spatial parameter n is $n \leq 5$. As this is a fundamental

property of a quantum system, a similar issue arises in the robustness measure as evaluating the robustness of $\mathcal{R}(|T\rangle^{\otimes n})$ and finding its optimal decomposition among $O(2^{n^2})$ stabilizer states is in general a challenging task for a large n [193].

Due to the computational need to truncate the spatial parameter $n < N$, a question arises of whether there exist new methods of manipulating the circuit frames and further reducing the total negativity, after gate merging is completed. We provide a positive answer to this question in the following section, where we describe our second sub-routine, frame optimization.

6.5 Frame optimization

Frame optimization aims to reduce the total circuit negativity by optimally choosing frames for different circuit components. As we discussed in section 6.3.1, we can introduce specific frame parameterizations, such as parameterized Wigner frames or rotated Pauli frames, and iteratively choose the frames throughout the circuit.

Sub-routine 2 Frame optimization

Input: Quantum circuit C and temporal parameter ℓ.

1: Determine the total number of frames, $|\boldsymbol{\mathcal{G}}_{\text{opt}}|$, in the circuit C.
2: Define the set of reference frames, $\boldsymbol{\mathcal{G}}_{\text{opt}} \leftarrow \{\mathcal{G}_1, \ldots, \mathcal{G}_{|\boldsymbol{\mathcal{G}}_{\text{opt}}|}\}$.
3: Fix the number of optimization cycles c.
4: **for** $i = 1, \ldots, c$ **do**
5: Choose a subset $\boldsymbol{\mathcal{G}}^{(i)}_{\text{target}} \subset \boldsymbol{\mathcal{G}}_{\text{opt}}$ with at most ℓ frames.
6: Find a circuit block $\boldsymbol{B}$ containing the frames in $\boldsymbol{\mathcal{G}}^{(i)}_{\text{target}}$.
7: Find $\overline{\boldsymbol{\mathcal{G}}}^{(i)}_{\text{target}} = \text{argmin}_{\boldsymbol{\mathcal{G}}^{(i)}_{\text{target}}} N_{\boldsymbol{B}}\left(\boldsymbol{\mathcal{G}}^{(i)}_{\text{target}}\right)$.
8: Update the corresponding frames in $\boldsymbol{\mathcal{G}}_{\text{opt}}$ with $\overline{\boldsymbol{\mathcal{G}}}^{(i)}_{\text{target}}$.

Output: $\boldsymbol{\mathcal{G}}_{\text{opt}}$.

In principle, the best strategy in terms of achieving the highest negativity reduction would be to carry out global optimization over all circuit frames, requiring that the number of parameters to be optimized should scale with the number of qubits N and

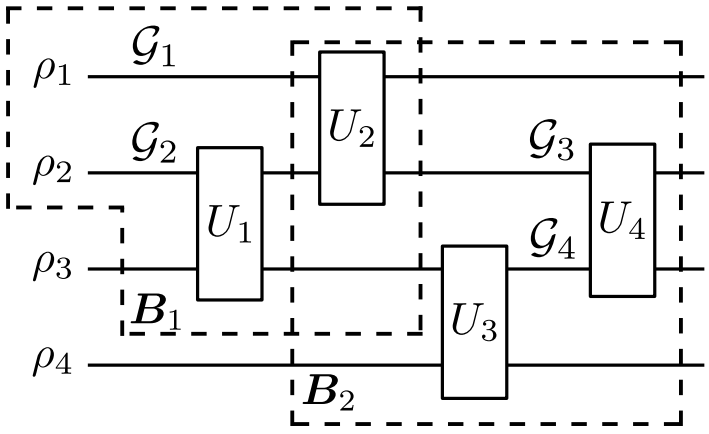

Figure 6.3: **Optimization block toy model.** Example of how to form a block when $\boldsymbol{\mathcal{G}}_{\text{target}}$ is given in the case of $n = 2$ and $\ell = 2$. Only relevant frames are shown. When $\boldsymbol{\mathcal{G}}_{\text{target}} = \{\mathcal{G}_1, \mathcal{G}_2\}$, the corresponding block $\mathbf{B}_1$, which contains all circuit elements connected to the frames in $\boldsymbol{\mathcal{G}}_{\text{target}}$, is $\mathbf{B}_1 = \{\rho_1, \rho_2, U_1, U_2\}$. When $\boldsymbol{\mathcal{G}}_{\text{target}} = \{\mathcal{G}_3, \mathcal{G}_4\}$, then the corresponding block is $\mathbf{B}_2 = \{U_2, U_3, U_4\}$.

circuit length L. In this chapter, we show that local optimization, with only a fixed number of parameters, is sufficient to achieve considerable negativity reduction and scales only linearly in N and L. This optimization sub-routine is implemented by dividing the circuit into blocks containing at most ℓ frames to be optimized, for a fixed temporal parameter ℓ.

To perform the frame optimization on a quantum circuit C consisting of an input state ρ, a gate sequence $\{U_1, ..., U_L\}$ and a measurement effect E, we need to start from an initial frame parameterization. We denote this parameterization as $\boldsymbol{\mathcal{G}}_{\text{opt}} = \{\mathcal{G}_1, \ldots, \mathcal{G}_{|\boldsymbol{\mathcal{G}}_{\text{opt}}|}\}$, where $|\boldsymbol{\mathcal{G}}_{\text{opt}}|$ is the number of frames to be optimized. The procedure is outlined in Sub-routine 2 and explained here. We take a subset $\boldsymbol{\mathcal{G}}^{(1)}_{\text{target}} \subset \boldsymbol{\mathcal{G}}_{\text{opt}}$ with up to ℓ frames (either sequentially or randomly) and create the block $\mathbf{B}$ of circuit components which are attached to those ℓ frames. Keeping all other frames in the block $\mathbf{B}$ fixed with the corresponding frames in $\boldsymbol{\mathcal{G}}_{\text{opt}}$, we want to minimize the total negativity of the block $N_{\mathbf{B}}$ over all possible choices for $\boldsymbol{\mathcal{G}}_{\text{target}}$, so that the minimum

$$\min_{\boldsymbol{\mathcal{G}}_{\text{target}}} N_{\mathbf{B}}(\boldsymbol{\mathcal{G}}_{\text{target}}). \tag{6.39}$$

occurs at $\overline{\boldsymbol{\mathcal{G}}}^{(1)}_{\text{target}}$ allowing us to update the corresponding frames in $\boldsymbol{\mathcal{G}}_{\text{opt}}$, which is the end of the first cycle in our frame optimization. We repeat this process c times by choosing another set of ℓ frames as the new $\boldsymbol{\mathcal{G}}^{(i)}_{\text{target}},\ i = 1, \ldots, c$. The number of optimization cycles c can be chosen arbitrarily, for example it can be chosen as $c \geq$

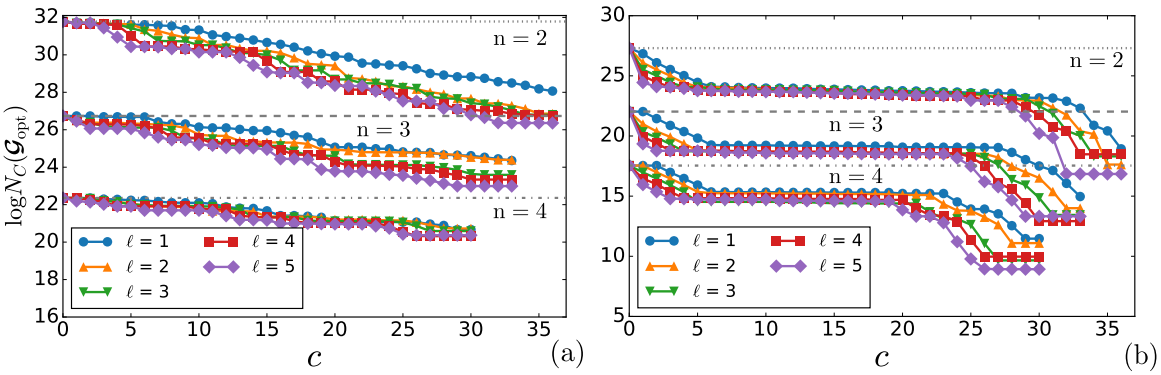

Figure 6.4: **Frame optimization analysis.** Plots showing reduction in the circuit negativity $\log N_C(\mathcal{G}_{\rm opt})$ after c optimization cycles for a circuit consisting of 2-qubit Haar-random gates with $N = 6$ and $L = 15$ and for carious spatial and temporal parameters, n and ℓ. The optimization is carried out sequentially from the first frame to the last frame. optimization is performed via the basin-hopping algorithm as introduced in [198]. (a) Results after frame optimization with rotated Pauli frames. The reference frame is the standard Pauli operators. The total negativity continuously decreases as we optimize more frames. (b) Results after frame optimization with parameterized Wigner frames. The reference frame is the conventional phase space operators for the Wigner representation. Most of the negativity reduction occurs near the initial states and the measurements.

$|\boldsymbol{\mathcal{G}}_{\rm opt}|/\ell$, with the aim of optimizing all frames in the circuit at least once. The order in which frames are optimized can also be chosen arbitrarily and can potentially result in a different overall negativity reduction.

We demonstrate the local frame optimization method with an example. Let us consider the initial part of a simple general circuit depicted in Fig. (6.3) for the case of $n = 2$ and $\ell = 2$. To perform the (i)–th optimization cycle we consider $\boldsymbol{\mathcal{G}}^{(i)}_{\rm target} = \{\mathcal{G}_1, \mathcal{G}_2\}$, we consider the corresponding block $\mathbf{B}_1 = \{\rho_1, \rho_2, U_1, U_2\}$, which is a set of all circuit components connected to the frames in $\boldsymbol{\mathcal{G}}^{(i)}_{\rm target}$. Then, the explicit optimization we perform is

$$\min_{\boldsymbol{\mathcal{G}}^{(i)}_{\rm target}} N_{\mathbf{B}_1}\left(\boldsymbol{\mathcal{G}}^{(i)}_{\rm target}\right) = \min_{\{\mathcal{G}_1,\mathcal{G}_2\}} N_{\rho_1}(\mathcal{G}_1) N_{\rho_2}(\mathcal{G}_2) N_{U_1}(\mathcal{G}_2) N_{U_2}(\mathcal{G}_1), \tag{6.40}$$

where $N_X(\mathcal{G}_X)$ is the negativity of component X as a function of $\mathcal{G}_X$ with all other frames fixed to the corresponding ones in $\boldsymbol{\mathcal{G}}_{\rm opt}$. As an additional example, we could have considered the set $\boldsymbol{\mathcal{G}}^{(i)}_{\rm target} = \{\mathcal{G}_3, \mathcal{G}_4\}$ corresponding to the block $\mathbf{B}_2 = \{U_2, U_3, U_4\}$ in Fig. (6.3). Then, the block negativity we optimize is $N_{\mathbf{B}_2}(\boldsymbol{\mathcal{G}}^{(i)}_{\rm target}) =$

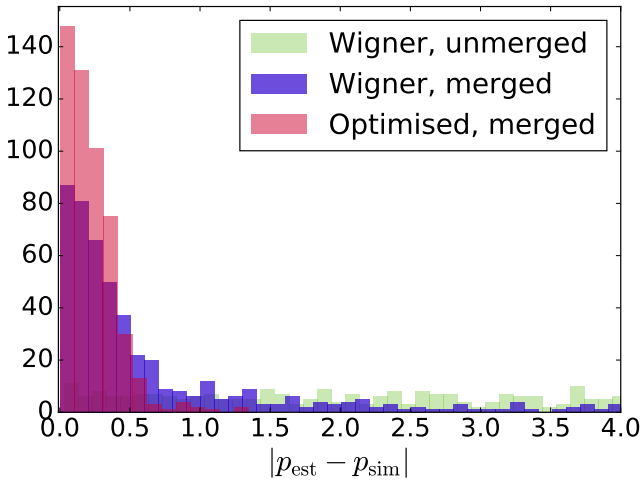

Figure 6.5: **Reduction in sampling overhead.** Histograms of the deviation of estimated probability p_{est} from actual outcome probability p_{sim} (as calculated by Qiskit [199]) for 500 circuits consisting of 2-qubit Haar-random gates with $N = 3$, $L = 8$ and $\ell = 1$. The number of samples taken for each circuit is 10^6, which took around 10 seconds on a standard computer. The plot shown is truncated at $|p_{\text{est}} - p_{\text{sim}}| = 4.0$ to demonstrate the advantage of our routines clearly. The advantage is amplified as N and L increase.

$N_{U_2}(\mathcal{G}_3)N_{U_3}(\mathcal{G}_4)N_{U_4}(\mathcal{G}_3, \mathcal{G}_4)$.

Note that at each optimization step, previously optimized frames in $\boldsymbol{\mathcal{G}}_{\text{opt}}$ are used in the next optimization cycle. This ensures that the negativity never increases compared to the initial frame choice $\{\mathcal{G}_1, \ldots, \mathcal{G}_{|\boldsymbol{\mathcal{G}}_{\text{opt}}|}\}$ between optimization cycles.

The presented local optimization method is efficient in the number of circuit components. Consider an N–qubit circuit of length L where each of L gates acts on at most n qubits. Then, there are at most $N+nL$ different frames to be optimized. Since ℓ is fixed, each optimization cycle takes a constant amount of time $O(1)$. Therefore, the frame optimization of the entire circuit scales as $(N + nL) \times O(1) = O(N, L)$. Note that the exact value depends on truncation parameters n and ℓ as well as the specifics of the circuit and its frame parameterization.

Fig. (6.4) shows the performance of the frame optimization for a circuit with $N = 6$ and $L = 15$ consisting of 2-qubit Haar-random unitaries, which are in general difficult to be simulated with stabilizer-based simulators because they do not admit efficient decompositions. In Fig. (6.4)(a), we use rotated Pauli frames as our frame parameterization and initialized each frame in the circuit to the set of standard qubit Pauli operators. In Fig. (6.4)(b), we choose parameterized Wigner frames

as our frame parameterization and initialize each frame in the circuit to the set of conventional phase space operators corresponding to $g(\boldsymbol{\lambda}) = 1$ (see section 6.3.2). We can observe that the largest negativity reduction comes from gate merging with higher n, but the frame optimization also achieves a significant negativity reduction. In general, larger ℓ results in lower negativity after optimization of all frames with fixed n. In the case of parameterized Wigner frames, together with gate merging, we could considerably decrease the initial log-negativity from ∼27.3 to ∼8.9 with truncation parameters $n = 4$ and $\ell = 5$, which means that we need $\sim 2^{2\times 18}$ times less samples to reach a given accuracy for probability estimation.

We demonstrate the practical significance of our routine, by sampling 500 circuits consisting of Haar-random gates, with the results presented in Fig. (6.5). Unmerged circuits represented entirely by Wigner frames do not show any signs of convergence to the actual probability distribution. Merged circuits clearly converge a lot better, especially when their frame representation is optimized, illustrating the positive effect of both of our sub-routines in reducing the run-time cost of the sampling algorithm.

Chapter 7

Outlook

Short summaries of the work in this thesis are provided in section 1.2, with more involved descriptions at the start/end of each chapter, so, for everyone's sake, I will directly proceed with discussing promising lines for future research.

We have described how relative majorization can be used to establish upper bounds on magic distillation protocols that take into account additional physics of the system. Our bounds exploited relatively simple aspects of the Lorenz curves of the quasi-distributions, so it would be of interest to sharpen these bounds and obtain a better handle on the Lorenz curve structure in the $n \to \infty$ limit. In this vein, it is possible to analyze what features of single-shot entropies can be extended to quasi-distributions in a sensible form [166, 200]. This raises interesting questions as we no longer have a notion of typicality and the central limit theorem does not apply. That said, for special states such as the Strange state, the asymptotic behavior is relatively simple, so exact asymptotics for this are expected to be possible.

Our work on distillation bounds is reminiscent of the second law of bipartite entanglement [117], where majorization is employed to establish entanglement entropy as the unique asymptotic measure of this resource. However, our analysis provides upper and lower bounds that exhibit a considerable gap, so that more than one of the entropies stemming from our magic majorization framework are valid asymptotic measures. The reason is that *stochastic* majorization only approximates the action of magic-preserving channel representations on state distributions. In particular, the Clifford group acts exactly via symplectic affine transformations $G(C) \coloneqq \mathrm{SP}(2, \mathbb{Z}_d) \ltimes \mathbb{Z}_d^2$ on state distributions on the discrete phase space. Analyzing the majorization of group $G(C)$ would provide exact magic distillation rates with a unique entropy.

There are strong results on group majorization in the mathematical literature [201–203], which can be exploited to derive a finite set of conditions in terms of semi-

definite programs, provided that the group has certain properties. A notable structure that could be exploited for G–majorization is when G is a finite reflection group. For this, the G–majorization pre-order is guaranteed to be described by a finite list of conditions [201], just as is the case for the relative majorization ordering. Therefore, one route would be to construct concrete lower bounds on distillation rates by considering reflection subgroups of $G(C)$, for which the majorization relation reduces to a finite set of conditions. An eligible group is $\mathrm{SP}(2, \mathbb{Z}_d)$, the symplectic subgroup of $G(C)$, which can lead to a significant refinement of our stochastic bounds. On the other hand, if the subgroup considered is too simple, the distillation bounds obtained may be trivial. An example of this is when we restrict to the set of Weyl-covariant channels [204], which are represented by a convex mixture of displacement operators on the phase space. Initial work has shown that the majorization conditions for Weyl-covariant channels can be solved exactly in terms of discrete Fourier transforms, however it is found that the resultant distillation rates are trivial.

More generally, majorization of quasi-distributions has not been considered in quantum physics before, and therefore these methods could find application in other studies of non-classicality in quantum systems [205–211]. For example, the topic of G–majorization has been extensively studied, but to our knowledge there has not been work on relative G–majorization. This would correspond to transformations that are not unital. Physically, this regime would correspond to a form of thermo-majorization obtained from looking at the action of the Clifford group at a micro-canonical level, and then reducing to a small subsystem [159]. While this seems like a painful thing to consider, there is motivation for this beyond the aim of magic protocols: in the case of classical statistical mechanics on a phase space this is precisely the situation, albeit in the continuum limit. Statistical mechanics of actual systems obey Hamiltonian dynamics, thus they automatically respect a symplectic form [159, 212]. Therefore, the pre-order of statistical mechanical states with respect to phase space dynamics preserving a Gibbs state must correspond to a symplectic majorization condition.

Of course, technical features arise in the continuum limit when considering distributions on an unbounded phase space. Recently, there has been a comprehensive extension of some majorization tools to continuous phase spaces [213], where quantum states are represented by the original continuous Wigner function. This work raises the straightforward question of extending our results from discrete to continuous quantum mechanics. Some of our results are readily extendable, but in order to define and study analogs of α-Rényi entropies for continuous representations, one would first have to provide additional technical contributions by extending the full

majorization framework we use. In this context, Clifford channels are replaced by Gaussian operations and our results can have impact on quantum optics, such as the characterization of resource theories of non-Gaussianity [214], or the quantum Hall effect [215].

In the particular context of qubits, complex majorization constraints arise naturally when we extend our setup from rebit to all qubit states. The Wigner representations are complex due to the non-Hermiticity of the operator basis $\{A_{\boldsymbol{x}}\}$. We expect such constraints to take the form of a duplet of constraints between the real parts and imaginary parts of the Wigner representation independently. In the context of non-Hermitian quantum mechanics [216], results on complex majorization would also benefit theories that require an ordering between Hamiltonian eigenvalues, such as quantum thermodynamics.

In the interests of expanding the family of operations included within our qubit framework, a natural extension of the qubit framework would be to generalize our results to more sophisticated protocols. For instance, we might ask how the use of m intermediate Clifford measure-and-update operations performed during the distillation may affect these fundamental constraints. One would expect to be able to obtain more refined bounds as a function of m. Moreover, while many protocols are based on CSS codes in part due to their relative ease of construction via tri-orthogonal matrices [75], it would be of interest from an operational perspective to see whether an extension to the full set of Clifford operations on qubit systems is possible.

A clear direction for our work on probability estimation is to improve the classical optimization performed for the frame representation, which currently acts as a black-box. One can investigate the possibility of performing frame optimization analytically, at least for particular classes of quantum circuits. Additional assumptions will likely be required for the circuit structure, but finding optimal frames analytically would eliminate the hidden constant run-time costs of black-box classical algorithms currently employed for the optimization. The difficulty arises when one considers the total negativity of a random sequence of Clifford and magic gates. For example, initial analysis reveals that although a Clifford gate C and a T gate can independently admit representations F_C and F_T with zero negativity respectively, when considered in sequence (C, T), the total negativity is not zero, and in fact it is minimized by a representation sequence different to (F_C, F_T). In order to obtain a theoretical handle, one could therefore start by considering very strict, brick-layered universal architectures that allow for an analytical calculation of optimal representation sequences. An alternative route involves the study of random Clifford+T

circuits via standard group representation theory, such as Clifford t-designs [217]. The technical step would involve translating the analysis into the frame representation, obtaining bounds on negativity and thus proving average-case guarantees for the sampling overhead of quasi-probability simulators.

Bibliography

[1] Alexander, R., Gvirtz-Chen, S., Koukoulekidis, N. & Jennings, D. General entropic constraints on calderbank-shor-steane codes within magic distillation protocols. *PRX Quantum* **4**, 020359 (2023). https://link.aps.org/doi/10.1103/PRXQuantum.4.020359.

[2] Koukoulekidis, N., Kwon, H., Jee, H. H., Jennings, D. & Kim, M. S. Faster Born probability estimation via gate merging and frame optimisation. *Quantum* **6**, 838 (2022). https://doi.org/10.22331/q-2022-10-13-838.

[3] Koukoulekidis, N. & Jennings, D. Constraints on magic state protocols from the statistical mechanics of wigner negativity. *Npj Quantum Inf.* **8**, 42 (2022). https://doi.org/10.1038/s41534-022-00551-1.

[4] Koukoulekidis, N. *et al.* A framework of partial error correction for intermediate-scale quantum computers (2023). https://arxiv.org/abs/2306.15531. 2306.15531.

[5] Koukoulekidis, N., Alexander, R., Hebdige, T. & Jennings, D. The geometry of passivity for quantum systems and a novel elementary derivation of the Gibbs state. *Quantum* **5**, 411 (2021). https://doi.org/10.22331/q-2021-03-15-411.

[6] Dive, B., Koukoulekidis, N., Mousafeiris, S. & Mintert, F. Characterization of multilevel quantum coherence without ideal measurements. *Phys. Rev. Res.* **2**, 013220 (2020). https://link.aps.org/doi/10.1103/PhysRevResearch.2.013220.

[7] Jozsa, R. & Van Den Nest, M. Classical simulation complexity of extended clifford circuits. *Quantum Information & Computation* **14**, 633–648 (2014). https://doi.org/10.48550/arXiv.1305.6190.

[8] Koh, D. E. Further extensions of Clifford circuits and their classical simulation complexities. *Quantum Info. Comput.* **17**, 262–282 (2017). https://doi.org/10.26421/QIC17.3-4-5.

[9] Aaronson, S., Bouland, A., Kuperberg, G. & Mehraban, S. The computational complexity of ball permutations. In *STOC 2017: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, 317–327 (Association for Computing Machinery, New York, NY, USA, 2017). https://doi.org/10.1145/3055399.3055453.

[10] Hebenstreit, M., Jozsa, R., Kraus, B. & Strelchuk, S. Computational power of matchgates with supplementary resources. *Physical Review A* **102**, 052604 (2020). https://link.aps.org/doi/10.1103/PhysRevA.102.052604.

[11] Yoganathan, M., Jozsa, R. & Strelchuk, S. Quantum advantage of unitary clifford circuits with magic state inputs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**, 20180427 (2019). https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0427.

[12] Aaronson, S. & Arkhipov, A. The computational complexity of linear optics (2010). https://arxiv.org/abs/1011.3245. 1011.3245.

[13] Bremner, M. J., Jozsa, R. & Shepherd, D. J. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467**, 459–472 (2011). https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2010.0301.

[14] Morimae, T., Fujii, K. & Fitzsimons, J. F. Hardness of classically simulating the one-clean-qubit model. *Physical Review Letters* **112**, 130502 (2014). https://link.aps.org/doi/10.1103/PhysRevLett.112.130502.

[15] Bremner, M. J., Montanaro, A. & Shepherd, D. J. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical Review Letters* **117**, 080501 (2016). https://link.aps.org/doi/10.1103/PhysRevLett.117.080501.

[16] Gao, X., Wang, S.-T. & Duan, L.-M. Quantum supremacy for simulating a translation-invariant ising spin model. *Physical Review Letters* **118**, 040502 (2017). https://link.aps.org/doi/10.1103/PhysRevLett.118.040502.

[17] Bermejo-Vega, J., Hangleiter, D., Schwarz, M., Raussendorf, R. & Eisert, J. Architectures for quantum simulation showing a quantum speedup. *Physical Review X* **8**, 021010 (2018). https://link.aps.org/doi/10.1103/PhysRevX.8.021010.

[18] Fujii, K. *et al.* Impossibility of classically simulating one-clean-qubit model with multiplicative error. *Physical Review Letters* **120**, 200502 (2018). https://link.aps.org/doi/10.1103/PhysRevLett.120.200502.

[19] Bouland, A., Fitzsimons, J. F. & Koh, D. E. Complexity Classification of Conjugated Clifford Circuits. In Servedio, R. A. (ed.) *33rd Computational Complexity Conference (CCC 2018)*, vol. 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 21:1–21:25 (Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2018). https://drops.dagstuhl.de/opus/volltexte/2018/8867.

[20] Boixo, S. *et al.* Characterizing quantum supremacy in near-term devices. *Nature Physics* **14**, 595–600 (2018). https://doi.org/10.1038/s41567-018-0124-x.

[21] Pashayan, H., Bartlett, S. D. & Gross, D. From estimation of quantum probabilities to simulation of quantum circuits. *Quantum* **4**, 223 (2020). https://doi.org/10.22331/q-2020-01-13-223.

[22] Valiant, L. G. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing* **31**, 1229–1254 (2002). https://doi.org/10.1137/S0097539700377025.

[23] Terhal, B. M. & DiVincenzo, D. P. Classical simulation of noninteracting-fermion quantum circuits. *Physical Review A* **65**, 032325 (2002). https://link.aps.org/doi/10.1103/PhysRevA.65.032325.

[24] Bartlett, S. D., Sanders, B. C., Braunstein, S. L. & Nemoto, K. Efficient classical simulation of continuous variable quantum information processes. *Physical Review Letters* **88**, 097904 (2002). https://link.aps.org/doi/10.1103/PhysRevLett.88.097904.

[25] Aaronson, S. & Gottesman, D. Improved simulation of stabilizer circuits. *Physical Review A* **70**, 052328 (2004). https://link.aps.org/doi/10.1103/PhysRevA.70.052328.

[26] Gross, D., Flammia, S. T. & Eisert, J. Most quantum states are too entangled to be useful as computational resources. *Physical Review Letters* **102**, 190501 (2009). https://link.aps.org/doi/10.1103/PhysRevLett.102.190501.

[27] Brod, D. J. Efficient classical simulation of matchgate circuits with generalized inputs and measurements. *Physical Review A* **93**, 062332 (2016). https://link.aps.org/doi/10.1103/PhysRevA.93.062332.

[28] Haug, T. & Kim, M. Scalable measures of magic resource for quantum computers. *PRX Quantum* **4**, 010301 (2023). https://link.aps.org/doi/10.1103/PRXQuantum.4.010301.

[29] Zhong, H.-S. *et al.* Quantum computational advantage using photons. *Science* **370**, 1460–1463 (2020). http://dx.doi.org/10.1126/science.abe8770.

[30] Arute, F. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019). https://doi.org/10.1038/s41586-019-1666-5.

[31] Bernien, H. *et al.* Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017). http://dx.doi.org/10.1038/nature24622.

[32] Neill, C. *et al.* A blueprint for demonstrating quantum supremacy with superconducting qubits. *Science* **360**, 195–199 (2018). http://dx.doi.org/10.1126/science.aao4309.

[33] Bonilla Ataides, J. P., Tuckett, D. K., Bartlett, S. D., Flammia, S. T. & Brown, B. J. The XZZX surface code. *Nature Communications* **12**, 2172 (2021). https://www.nature.com/articles/s41467-021-22274-1.

[34] Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Physical Review A* **103**, 042605 (2021). https://link.aps.org/doi/10.1103/PhysRevA.103.042605.

[35] Kandala, A. *et al.* Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019). https://www.nature.com/articles/s41586-019-1040-7.

[36] Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Physical Review X* **8**, 031027 (2018). https://link.aps.org/doi/10.1103/PhysRevX.8.031027.

[37] Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Physical Review Letters* **119**, 180509 (2017). https://link.aps.org/doi/10.1103/PhysRevLett.119.180509.

[38] Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018). https://doi.org/10.22331/q-2018-08-06-79.

[39] Harrow, A. W. & Montanaro, A. Quantum computational supremacy. *Nature* **549**, 203–209 (2017). https://doi.org/10.1038/nature23458.

[40] Bennink, R. S. *et al.* Unbiased simulation of near-clifford quantum circuits. *Physical Review A* **95**, 062337 (2017). https://link.aps.org/doi/10.1103/PhysRevA.95.062337.

[41] Veitch, V., Ferrie, C., Gross, D. & Emerson, J. Negative quasi-probability as a resource for quantum computation. *New Journal of Physics* **14**, 113011 (2012). https://doi.org/10.1088/1367-2630/14/11/113011.

[42] Gottesman, D. *Stabilizer codes and quantum error correction*. Ph.D. thesis, California Institute of Technology (1997). https://arxiv.org/abs/quant-ph/9705052.

[43] Campbell, E. T., Terhal, B. M. & Vuillot, C. Roads towards fault-tolerant universal quantum computation. *Nature* **549**, 172–179 (2017). https://www.nature.com/articles/nature23460.

[44] Raussendorf , R. & Briegel, H. J. A one-way quantum computer. *Physical Review Letters* **86**, 5188–5191 (2001). https://link.aps.org/doi/10.1103/PhysRevLett.86.5188.

[45] Raussendorf, R., Browne, D. E. & Briegel, H. J. Measurement-based quantum computation on cluster states. *Physical Review A* **68**, 022312 (2003). https://link.aps.org/doi/10.1103/PhysRevA.68.022312.

[46] Nickerson, N. H., Fitzsimons, J. F. & Benjamin, S. C. Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Physical Review X* **4**, 041041 (2014). https://link.aps.org/doi/10.1103/PhysRevX.4.041041.

[47] Watson, F. H. E., Anwar, H. & Browne, D. E. Fast fault-tolerant decoder for qubit and qudit surface codes. *Phys. Rev. A* **92**, 032309 (2015). https://link.aps.org/doi/10.1103/PhysRevA.92.032309.

[48] Brown, B. J., Nickerson, N. H. & Browne, D. E. Fault-tolerant error correction with the gauge color code. *Nature Communications* **7**, 12302 (2016). https://doi.org/10.1038/ncomms12302.

[49] Nikahd, E., Sedighi, M. & Saheb Zamani, M. Nonuniform code concatenation for universal fault-tolerant quantum computing. *Physical Review A* **96**, 032337 (2017). https://link.aps.org/doi/10.1103/PhysRevA.96.032337.

[50] Auger, J. M., Bergamini, S. & Browne, D. E. Blueprint for fault-tolerant quantum computation with rydberg atoms. *Phys. Rev. A* **96**, 052320 (2017). https://link.aps.org/doi/10.1103/PhysRevA.96.052320.

[51] Auger, J. M., Anwar, H., Gimeno-Segovia, M., Stace, T. M. & Browne, D. E. Fault-tolerant quantum computation with nondeterministic entangling gates. *Phys. Rev. A* **97**, 030301 (2018). https://link.aps.org/doi/10.1103/PhysRevA.97.030301.

[52] Chao, R. & Reichardt, B. W. Fault-tolerant quantum computation with few qubits. *npj Quantum Information* **4**, 1–8 (2018). https://www.nature.com/articles/s41534-018-0085-z.

[53] Vasmer, M. & Browne, D. E. Three-dimensional surface codes: Transversal gates and fault-tolerant architectures. *Phys. Rev. A* **100**, 012312 (2019). https://link.aps.org/doi/10.1103/PhysRevA.100.012312.

[54] Lin, C., Yang, G., Luo, Q. & Li, X. Pieceable fault tolerant conversion between 5-qubit code and 7-CSS code. *Quantum Information Processing* **19**, 243 (2020). https://doi.org/10.1007/s11128-020-02740-3.

[55] Lin, C. & Yang, G. Concatenated pieceable fault-tolerant scheme for universal quantum computation. *Physical Review A* **102**, 052415 (2020). https://link.aps.org/doi/10.1103/PhysRevA.102.052415.

[56] Chamberland, C. *et al.* Building a fault-tolerant quantum computer using concatenated cat codes. *PRX Quantum* **3**, 010329 (2022). https://link.aps.org/doi/10.1103/PRXQuantum.3.010329.

[57] Bourassa, J. E. *et al.* Blueprint for a Scalable Photonic Fault-Tolerant Quantum Computer. *Quantum* **5**, 392 (2021). https://doi.org/10.22331/q-2021-02-04-392.

[58] Jones, N. C. *et al.* Layered architecture for quantum computing. *Physical Review X* **2**, 031007 (2012). https://link.aps.org/doi/10.1103/PhysRevX.2.031007.

[59] Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X* **7**, 021050 (2017). https://link.aps.org/doi/10.1103/PhysRevX.7.021050.

[60] McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. & de Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Physical Review A* **95**, 042308 (2017). https://link.aps.org/doi/10.1103/PhysRevA.95.042308.

[61] Holmes, A. *et al.* Resource optimized quantum architectures for surface code implementations of magic-state distillation. *Microprocessors and Microsystems* **67**, 56–70 (2019). https://www.sciencedirect.com/science/article/pii/S0141933118303314.

[62] Colless, J. I. *et al.* Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Physical Review X* **8**, 011021 (2018). https://link.aps.org/doi/10.1103/PhysRevX.8.011021.

[63] Song, C. *et al.* Quantum computation with universal error mitigation on superconducting quantum processor (2018). https://arxiv.org/abs/2005.07161.

[64] Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Physical Review A* **54**, 3824–3851 (1996). https://link.aps.org/doi/10.1103/PhysRevA.54.3824.

[65] Bravyi, S. B. & Kitaev, A. Y. Quantum codes on a lattice with boundary (1998). https://arxiv.org/abs/quant-ph/9811052.

[66] Freedman, M. H. Quantum computation and the localization of modular functors. *Foundations of Computational Mathematics* **1**, 183–204 (2001). https://doi.org/10.1007/s102080010006.

[67] Dennis, E., Kitaev, A., Landahl, A. & Preskill, J. Topological quantum memory. *Journal of Mathematical Physics* **43**, 4452–4505 (2002). https://aip.scitation.org/doi/10.1063/1.1499754.

[68] Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Physical Review Letters* **98**, 190504 (2007). https://link.aps.org/doi/10.1103/PhysRevLett.98.190504.

[69] Eastin, B. & Knill, E. Restrictions on transversal encoded quantum gate sets. *Physical Review Letters* **102**, 110502 (2009). https://link.aps.org/doi/10.1103/PhysRevLett.102.110502.

[70] Calderbank, A. R. & Shor, P. W. Good quantum error-correcting codes exist. *Physical Review A* **54**, 1098–1105 (1996). https://link.aps.org/doi/10.1103/PhysRevA.54.1098.

[71] Steane, A. Multiple-particle interference and quantum error correction. *Proceedings of the Royal Society of London Series A* **452**, 2551–2577 (1996). http://doi.org/10.1098/rspa.1996.0136.

[72] Bravyi, S. & Kitaev, A. Universal quantum computation with ideal clifford gates and noisy ancillas. *Physical Review A* **71**, 022316 (2005). https://link.aps.org/doi/10.1103/PhysRevA.71.022316.

[73] Jones, C. Multilevel distillation of magic states for quantum computing. *Physical Review A* **87**, 042305 (2013). https://link.aps.org/doi/10.1103/PhysRevA.87.042305.

[74] O'Gorman, J. & Campbell, E. T. Quantum computation with realistic magic-state factories. *Physical Review A* **95**, 032338 (2017). https://link.aps.org/doi/10.1103/PhysRevA.95.032338.

[75] Bravyi, S. & Haah, J. Magic-state distillation with low overhead. *Physical Review A* **86**, 052329 (2012). https://link.aps.org/doi/10.1103/PhysRevA.86.052329.

[76] Haah, J., Hastings, M. B., Poulin, D. & Wecker, D. Magic state distillation with low space overhead and optimal asymptotic input count. *Quantum* **1**, 31 (2017). https://doi.org/10.22331/q-2017-10-03-31.

[77] Hastings, M. B. & Haah, J. Distillation with sublogarithmic overhead. *Physical Review Letters* **120**, 050504 (2018). https://link.aps.org/doi/10.1103/PhysRevLett.120.050504.

[78] Litinski, D. Magic State Distillation: Not as Costly as You Think. *Quantum* **3**, 205 (2019). https://doi.org/10.22331/q-2019-12-02-205.

[79] Campbell, E. T., Anwar, H. & Browne, D. E. Magic-state distillation in all prime dimensions using quantum reed-muller codes. *Physical Review X* **2**, 041021 (2012). https://link.aps.org/doi/10.1103/PhysRevX.2.041021.

[80] Anwar, H., Campbell, E. T. & Browne, D. E. Qutrit magic state distillation. *New Journal of Physics* **14**, 063006 (2012). https://doi.org/10.1088/1367-2630/14/6/063006.

[81] Dawkins, H. & Howard, M. Qutrit magic state distillation tight in some directions. *Physical Review Letters* **115**, 030501 (2015). https://link.aps.org/doi/10.1103/PhysRevLett.115.030501.

[82] Krishna, A. & Tillich, J.-P. Towards low overhead magic state distillation. *Physical Review Letters* **123**, 070507 (2019). https://link.aps.org/doi/10.1103/PhysRevLett.123.070507.

[83] Prakash, S. Magic state distillation with the ternary golay code. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **476**, 20200187 (2020). https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2020.0187.

[84] Chamberland, C. & Noh, K. Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits. *npj Quantum Information* **6**, 91 (2020). https://doi.org/10.1038/s41534-020-00319-5.

[85] Chamberland, C. & Cross, A. W. Fault-tolerant magic state preparation with flag qubits. *Quantum* **3**, 143 (2019). https://doi.org/10.22331/q-2019-05-20-143.

[86] Campbell, E. T. & Howard, M. Unifying gate synthesis and magic state distillation. *Physical Review Letters* **118**, 060501 (2017). https://link.aps.org/doi/10.1103/PhysRevLett.118.060501.

[87] Campbell, E. T. & Howard, M. Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost. *Physical Review A* **95**, 022316 (2017). https://link.aps.org/doi/10.1103/PhysRevA.95.022316.

[88] Prakash, S., Jain, A., Kapur, B. & Seth, S. Normal form for single-qutrit clifford+t operators and synthesis of single-qutrit gates. *Physical Review A* **98**, 032304 (2018). https://link.aps.org/doi/10.1103/PhysRevA.98.032304.

[89] Seddon, J. R., Regula, B., Pashayan, H., Ouyang, Y. & Campbell, E. T. Quantifying quantum speedups: Improved classical simulation from tighter magic monotones. *PRX Quantum* **2**, 010345 (2021). https://link.aps.org/doi/10.1103/PRXQuantum.2.010345.

[90] Chitambar, E. & Gour, G. Quantum resource theories. *Reviews of Modern Physics* **91**, 025001 (2019). https://link.aps.org/doi/10.1103/RevModPhys.91.025001.

[91] Ahmadi, M., Dang, H. B., Gour, G. & Sanders, B. C. Quantification and manipulation of magic states. *Physical Review A* **97**, 062332 (2018). https://link.aps.org/doi/10.1103/PhysRevA.97.062332.

[92] Seddon, J. R. & Campbell, E. T. Quantifying magic for multi-qubit operations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **475**, 20190251 (2019). https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2019.0251.

[93] Wang, X., Wilde, M. M. & Su, Y. Quantifying the magic of quantum channels. *New Journal of Physics* **21**, 103002 (2019). https://doi.org/10.1088/1367-2630/ab451d.

[94] Wootters, W. K. A Wigner-function formulation of finite-state quantum mechanics. *Annals of Physics* **176**, 1–21 (1987). https://doi.org/10.1016/0003-4916(87)90176-X.

[95] Ferrie, C. & Emerson, J. Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations. *Journal of Physics A: Mathematical and Theoretical* **41**, 352001 (2008). https://doi.org/10.1088/1751-8113/41/35/352001.

[96] Ferrie, C. & Emerson, J. Framed hilbert space: hanging the quasi-probability pictures of quantum theory. *New Journal of Physics* **11**, 063040 (2009). https://doi.org/10.1088/1367-2630/11/6/063040.

[97] Gross, D. Hudson's theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics* **47**, 122107 (2006). https://aip.scitation.org/doi/10.1063/1.2393152.

[98] Delfosse, N., Allard Guerin, P., Bian, J. & Raussendorf, R. Wigner function negativity and contextuality in quantum computation on rebits. *Physical Re-*

view X **5**, 021003 (2015). https://link.aps.org/doi/10.1103/PhysRevX.5.021003.

[99] Wigner, E. On the quantum correction for thermodynamic equilibrium. *Physical Review* **40**, 749–759 (1932). https://link.aps.org/doi/10.1103/PhysRev.40.749.

[100] Vourdas, A. Quantum systems with finite hilbert space. *Reports on Progress in Physics* **67**, 267–320 (2004). https://doi.org/10.1088/0034-4885/67/3/r03.

[101] Kenfack, A. & Życzkowski, K. Negativity of the Wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics* **6**, 396–404 (2004). https://doi.org/10.1088/1464-4266/6/10/003.

[102] Catani, L. & Browne, D. E. Spekkens' toy model in all dimensions and its relationship with stabiliser quantum mechanics. *New Journal of Physics* **19**, 073035 (2017). https://doi.org/10.1088/1367-2630/aa781c.

[103] Mari, A. & Eisert, J. Positive wigner functions render classical simulation of quantum computation efficient. *Physical Review Letters* **109**, 230503 (2012). https://link.aps.org/doi/10.1103/PhysRevLett.109.230503.

[104] Veitch, V., Mousavian, S. A. H., Gottesman, D. & Emerson, J. The resource theory of stabilizer quantum computation. *New Journal of Physics* **16**, 013009 (2014). https://doi.org/10.1088/1367-2630/16/1/013009.

[105] Raussendorf, R., Bermejo-Vega, J., Tyhurst, E., Okay, C. & Zurel, M. Phase-space-simulation method for quantum computation with magic states on qubits. *Physical Review A* **101**, 012350 (2020). https://link.aps.org/doi/10.1103/PhysRevA.101.012350.

[106] Howard, M., Wallman, J., Veitch, V. & Emerson, J. Contextuality supplies the 'magic' for quantum computation. *Nature* **510**, 351–355 (2014). https://doi.org/10.1038/nature13460.

[107] Zurel, M., Okay, C. & Raussendorf, R. Hidden variable model for universal quantum computation with magic states on qubits. *Physical Review Letters* **125**, 260404 (2020). https://link.aps.org/doi/10.1103/PhysRevLett.125.260404.

[108] Catani, L. & Browne, D. E. State-injection schemes of quantum computation in spekkens' toy theory. *Physical Review A* **98**, 052108 (2018). https://link.aps.org/doi/10.1103/PhysRevA.98.052108.

[109] Bengtsson, I. & Życzkowski, K. *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, 2006).

[110] Mirsky, L. On the trace of matrix products. *Mathematische Nachrichten* **20**, 171–174 (1959). https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.19590200306.

[111] Appleby, D. M. Symmetric informationally complete–positive operator valued measures and the extended clifford group. *Journal of Mathematical Physics* **46**, 052107 (2005). https://doi.org/10.1063/1.1896384.

[112] Campbell , E. T. & Browne, D. E. Bound states for magic state distillation in fault-tolerant quantum computation. *Physical Review Letters* **104**, 030503 (2010). https://link.aps.org/doi/10.1103/PhysRevLett.104.030503.

[113] Bravyi, S. & Gosset, D. Improved classical simulation of quantum circuits dominated by clifford gates. *Physical Review Letters* **116**, 250501 (2016). https://link.aps.org/doi/10.1103/PhysRevLett.116.250501.

[114] Watrous, J. *The Theory of Quantum Information* (Cambridge University Press, 2018).

[115] Marshall, A. W., Olkin, I. & Arnold, B. C. *Inequalities: Theory of Majorization and Its Applications* (Springer, 2011).

[116] Blackwell, D. Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics* **24**, 265 – 272 (1953). https://doi.org/10.1214/aoms/1177729032.

[117] Nielsen, M. A. Conditions for a class of entanglement transformations. *Physical Review Letters* **83**, 436–439 (1999). https://link.aps.org/doi/10.1103/PhysRevLett.83.436.

[118] Ćwikliński, P., Studziński, M., Horodecki, M. & Oppenheim, J. Limitations on the evolution of quantum coherences: Towards fully quantum second laws of thermodynamics. *Physical Review Letters* **115**, 210403 (2015). https://link.aps.org/doi/10.1103/PhysRevLett.115.210403.

[119] Lostaglio, M., Jennings, D. & Rudolph, T. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature Communications* **6**, 6383 (2015). https://doi.org/10.1038/ncomms7383.

[120] Gour, G., Jennings, D., Buscemi, F., Duan, R. & Marvian, I. Quantum majorization and a complete set of entropic conditions for quantum thermodynamics. *Nature Communications* **9**, 5352 (2018). https://doi.org/10.1038/s41467-018-06261-7.

[121] Gour, G., Müller, M. P., Narasimhachar, V., Spekkens, R. W. & Yunger Halpern, N. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports* **583**, 1–58 (2015). https://doi.org/10.1016/j.physrep.2015.04.003.

[122] Horodecki, M., Horodecki, P. & Oppenheim, J. Reversible transformations from pure to mixed states and the unique measure of information. *Physical Review A* **67**, 062104 (2003). https://link.aps.org/doi/10.1103/PhysRevA.67.062104.

[123] Puchała, Z., Rudnicki, Ł. & Życzkowski, K. Majorization entropic uncertainty relations. *Journal of Physics A: Mathematical and Theoretical* **46**, 272002 (2013). https://doi.org/10.1088/1751-8113/46/27/272002.

[124] Vallejos, R. O., de Melo, F. & Carlo, G. G. Principle of majorization: Application to random quantum circuits. *Physical Review A* **104**, 012602 (2021). https://link.aps.org/doi/10.1103/PhysRevA.104.012602.

[125] Rényi, A. On measures of entropy and information. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability* **1**, 547–561 (1960). https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s4_v1_article-27.pdf.

[126] Veinott, A. F. Least d-majorized network flows with inventory and statistical applications. *Management Science* **17**, 547–567 (1971). https://doi.org/10.1287/mnsc.17.9.547.

[127] Horodecki, M. & Oppenheim, J. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature Communications* **4**, 2059 (2013). https://doi.org/10.1038/ncomms3059.

[128] Ruch, E. & Mead, A. The principle of increasing mixing character and some of its consequences. *Theoretica chimica acta* **41**, 95–117 (1976). https://doi.org/10.1007/BF01178071.

[129] Ruch, E., Schranner, R. & Seligman, T. H. The mixing distance. *The Journal of Chemical Physics* **69**, 386–392 (1978). https://aip.scitation.org/doi/10.1063/1.436364.

[130] Renes, J. M. Relative submajorization and its use in quantum resource theories. *Journal of Mathematical Physics* **57**, 122202 (2016). https://aip.scitation.org/doi/10.1063/1.4972295.

[131] Buscemi, F. & Gour, G. Quantum relative lorenz curves. *Physical Review A* **95**, 012110 (2017). https://link.aps.org/doi/10.1103/PhysRevA.95.012110.

[132] Rethinasamy, S. & Wilde, M. M. Relative entropy and catalytic relative majorization. *Physical Review Research* **2**, 033455 (2020). https://link.aps.org/doi/10.1103/PhysRevResearch.2.033455.

[133] Lostaglio, M. An introductory review of the resource theory approach to thermodynamics. *Reports on Progress in Physics* **82**, 114001 (2019). https://doi.org/10.1088/1361-6633/ab46e5.

[134] Jain, A. & Prakash, S. Qutrit and ququint magic states. *Physical Review A* **102**, 042409 (2020). https://link.aps.org/doi/10.1103/PhysRevA.102.042409.

[135] Brandão, F., Horodecki, M., Ng, N., Oppenheim, J. & Wehner, S. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences* **112**, 3275–3279 (2015). https://www.pnas.org/content/112/11/3275.

[136] Schur, I. Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. *Berliner Mathematische Gesellschaft* **22**, 9–20 (1923).

[137] Ostrowski, A. M. Sur quelques applications des fonctions convexes et concaves au sens de I. Schur. *Journal de Mathématiques Pures et Appliquées* **31**, 253–292 (1952).

[138] Manfredi, G. & Feix, M. R. Entropy and Wigner functions. *Physical Review E* **62**, 4665–4674 (2000). https://link.aps.org/doi/10.1103/PhysRevE.62.4665.

[139] Gnutzmann, S. & Życzkowski, K. Rényi-wehrl entropies as measures of localization in phase space. *Journal of Physics A: Mathematical and General* **34**, 10123–10139 (2001). https://doi.org/10.1088/0305-4470/34/47/317.

[140] Tuckett, D. K. *et al.* Tailoring surface codes for highly biased noise. *Physical Review X* **9**, 041031 (2019). https://link.aps.org/doi/10.1103/PhysRevX.9.041031.

[141] Aliferis, P. & Preskill, J. Fault-tolerant quantum computation against biased noise. *Physical Review A* **78**, 052331 (2008). https://link.aps.org/doi/10.1103/PhysRevA.78.052331.

[142] Stephens, A. M., Munro, W. J. & Nemoto, K. High-threshold topological quantum error correction against biased noise. *Physical Review A* **88**, 060301 (2013). https://link.aps.org/doi/10.1103/PhysRevA.88.060301.

[143] Li, Y. A magic state's fidelity can be superior to the operations that created it. *New Journal of Physics* **17**, 023037 (2015). https://doi.org/10.1088/1367-2630/17/2/023037.

[144] Babbush, R. *et al.* Encoding electronic spectra in quantum circuits with linear t complexity. *Physical Review X* **8**, 041015 (2018). https://link.aps.org/doi/10.1103/PhysRevX.8.041015.

[145] Guillaud, J. & Mirrahimi, M. Repetition cat qubits for fault-tolerant quantum computation. *Physical Review X* **9**, 041053 (2019). https://link.aps.org/doi/10.1103/PhysRevX.9.041053.

[146] Fowler, A. G. & Gidney, C. Low overhead quantum computation using lattice surgery (2019). https://arxiv.org/abs/1808.06709.

[147] Janzing, D., Wocjan, P., Zeier, R., Geiss, R. & Beth, T. Thermodynamic cost of reliability and low temperatures: Tightening landauer's principle and the second law. *International Journal of Theoretical Physics* **39**, 2717–2753 (2000). https://doi.org/10.1023/A:1026422630734.

[148] Vinjanampathy, S. & Anders, J. Quantum thermodynamics. *Contemporary Physics* **57**, 545–579 (2016). https://doi.org/10.1080/00107514.2016.1201896.

[149] Goold, J., Huber, M., Riera, A., del Rio, L. & Skrzypczyk, P. The role of quantum information in thermodynamics—a topical review. *Journal of Physics*

A: Mathematical and Theoretical **49**, 143001 (2016). https://doi.org/10.1088/1751-8113/49/14/143001.

[150] Knill, E. Quantum computing with realistically noisy devices. *Nature* **434**, 39–44 (2005). https://doi.org/10.1038/nature03350.

[151] Campbell, E. T. Catalysis and activation of magic states in fault-tolerant architectures. *Physical Review A* **83**, 032317 (2011). https://link.aps.org/doi/10.1103/PhysRevA.83.032317.

[152] Brouwer, L. E. J. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen* **71** (1911). https://doi.org/10.1007/BF01456931.

[153] Wang, X., Wilde, M. M. & Su, Y. Efficiently computable bounds for magic state distillation. *Physical Review Letters* **124**, 090505 (2020). https://link.aps.org/doi/10.1103/PhysRevLett.124.090505.

[154] Hudson, R. L. & Moody, G. R. Locally normal symmetric states and an analogue of de finetti's theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **33**, 343–351 (1976). https://doi.org/10.1007/BF00534784.

[155] Christandl, M., König, R., Mitchison, G. & Renner, R. One-and-a-half quantum de finetti theorems. *Communications in Mathematical Physics* **273**, 473–498 (2007). https://doi.org/10.1007/s00220-007-0189-3.

[156] Lostaglio, M., Müller, M. P. & Pastena, M. Stochastic independence as a resource in small-scale thermodynamics. *Physical Review Letters* **115**, 150402 (2015). https://link.aps.org/doi/10.1103/PhysRevLett.115.150402.

[157] Müller, M. P. & Pastena, M. A generalization of majorization that characterizes shannon entropy. *IEEE Transactions on Information Theory* **62**, 1711–1720 (2016). https://ieeexplore.ieee.org/document/7404001.

[158] Boes, P., Eisert, J., Gallego, R., Müller, M. P. & Wilming, H. Von neumann entropy from unitarity. *Physical Review Letters* **122**, 210402 (2019). https://link.aps.org/doi/10.1103/PhysRevLett.122.210402.

[159] Pathria, R. K. & Beale, P. D. *Statistical Mechanics* (Elsevier, 2011). https://www.sciencedirect.com/book/9780123821881/statistical-mechanics.

[160] Bombin, H. *et al.* Interleaving: Modular architectures for fault-tolerant photonic quantum computing (2021). https://arxiv.org/abs/2103.08612.

[161] Bogolubov, J., N. N. On model dynamical systems in statistical mechanics. *Physica* **32**, 933–944 (1966). https://doi.org/10.1016/0031-8914(66)90024-3.

[162] Faist, P., Oppenheim, J. & Renner, R. Gibbs-preserving maps outperform thermal operations in the quantum regime. *New Journal of Physics* **17**, 043003 (2015). https://iopscience.iop.org/article/10.1088/1367-2630/17/4/043003.

[163] Rio, L. d., Åberg, J., Renner, R., Dahlsten, O. & Vedral, V. The thermodynamic meaning of negative entropy. *Nature* **474**, 61–63 (2011). https://www.nature.com/articles/nature10123.

[164] Baez, J. C. Rényi entropy and free energy (2011). https://arxiv.org/abs/1102.2098.

[165] Domb, C. *Phase transitions and critical phenomena* (Elsevier, 2000). https://www.elsevier.com/books/phase-transitions-and-critical-phenomena/domb/978-0-12-220318-3.

[166] Tomamichel, M. *A Framework for Non-Asymptotic Quantum Information Theory*. Ph.D. thesis, ETH Zurich (2012). https://arxiv.org/abs/1203.2142.

[167] Schmid, D., Selby, J. H., Wolfe, E., Kunjwal, R. & Spekkens, R. W. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum* **2**, 010331 (2021). https://link.aps.org/doi/10.1103/PRXQuantum.2.010331.

[168] Mermin, N. D. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters* **65**, 3373–3376 (1990). https://link.aps.org/doi/10.1103/PhysRevLett.65.3373.

[169] Reichardt, B. W. Quantum universality from magic states distillation applied to css codes. *Quantum Information Processing* **4**, 251–264 (2005). https://doi.org/10.1007/s11128-005-7654-8.

[170] Rengaswamy, N., Calderbank, R., Newman, M. & Pfister, H. D. On optimality of css codes for transversal t. *IEEE Journal on Selected Areas in Information Theory* **1**, 499–5 (2020).

[171] Campbell, E. T. & Browne, D. E. On the structure of protocols for magic state distillation. In Childs, A. & Mosca, M. (eds.) *Theory of Quantum Computation, Communication, and Cryptography*, 20–32 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).

[172] Knill, E., Laflamme, R. & Żurek, W. Threshold accuracy for quantum computation (1996). https://doi.org/10.48550/arXiv.quant-ph/9610011. 1412.9610011.

[173] Steane, A. M. Quantum reed-muller codes. *IEEE Transactions on Information Theory* **45**, 1701–1703 (1999). https://doi.org/10.1109/18.771249.

[174] Woods, M. P. & Horodecki, M. Autonomous quantum devices: When are they realizable without additional thermodynamic costs? *Phys. Rev. X* **13**, 011016 (2023). https://link.aps.org/doi/10.1103/PhysRevX.13.011016.

[175] Regula, B. Probabilistic transformations of quantum resources. *Physical Review Letters* **128**, 110505 (2022). https://link.aps.org/doi/10.1103/PhysRevLett.128.110505.

[176] Beigi, S. Sandwiched rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics* **54**, 122202 (2013). https://doi.org/10.1063/1.4838855.

[177] Mosonyi, M. Convexity properties of the quantum rényi divergences, with applications to the quantum stein's lemma. In Flammia, S. T. & Harrow, A. W. (eds.) *9th Conference on the Theory of Quantum Computation, Communication and Cryptography, May 21-23, 2014, Singapore*, vol. 27 of *LIPIcs*, 88–98 (Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2014). https://doi.org/10.4230/LIPIcs.TQC.2014.88.

[178] Wilde, M. M., Winter, A. & Yang, D. Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched rényi relative entropy. *Communications in Mathematical Physics* **331**, 593–622 (2014). https://doi.org/10.1007/s00220-014-2122-x.

[179] Pashayan, H., Reardon-Smith, O., Korzekwa, K. & Bartlett, S. D. Fast estimation of outcome probabilities for quantum circuits. *PRX Quantum* **3**, 020361 (2022). https://link.aps.org/doi/10.1103/PRXQuantum.3.020361.

[180] Bravyi, S. *et al.* Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum* **3**, 181 (2019). https://doi.org/10.22331/q-2019-09-02-181.

[181] Bravyi, S., Smith, G. & Smolin, J. A. Trading classical and quantum computational resources. *Physical Review X* **6**, 021043 (2016). https://link.aps.org/doi/10.1103/PhysRevX.6.021043.

[182] Qassim, H., Wallman, J. J. & Emerson, J. Clifford recompilation for faster classical simulation of quantum circuits. *Quantum* **3**, 170 (2019). https://doi.org/10.22331/q-2019-08-05-170.

[183] Huang, Y. & Love, P. Approximate stabilizer rank and improved weak simulation of clifford-dominated circuits for qudits. *Physical Review A* **99**, 052307 (2019). https://link.aps.org/doi/10.1103/PhysRevA.99.052307.

[184] Kocia, L. & Sarovar, M. Classical simulation of quantum circuits using fewer gaussian eliminations. *Physical Review A* **103**, 022603 (2021). https://link.aps.org/doi/10.1103/PhysRevA.103.022603.

[185] Kissinger, A., van de Wetering, J. & Vilmart, R. Classical Simulation of Quantum Circuits with Partial and Graphical Stabiliser Decompositions. In Le Gall, F. & Morimae, T. (eds.) *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, vol. 232 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 5:1–5:13 (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022). https://drops.dagstuhl.de/opus/volltexte/2022/16512.

[186] Qassim, H., Pashayan, H. & Gosset, D. Improved upper bounds on the stabilizer rank of magic states. *Quantum* **5**, 606 (2021). https://doi.org/10.22331/q-2021-12-20-606.

[187] Pashayan, H., Wallman, J. J. & Bartlett, S. D. Estimating outcome probabilities of quantum circuits using quasiprobabilities. *Physical Review Letters* **115**, 070501 (2015). https://link.aps.org/doi/10.1103/PhysRevLett.115.070501.

[188] Stahlke, D. Quantum interference as a resource for quantum speedup. *Physical Review A* **90**, 022302 (2014). https://link.aps.org/doi/10.1103/PhysRevA.90.022302.

[189] Rall, P., Liang, D., Cook, J. & Kretschmer, W. Simulation of qubit quantum circuits via pauli propagation. *Physical Review A* **99**, 062337 (2019). https://link.aps.org/doi/10.1103/PhysRevA.99.062337.

[190] Ruzzi, M., Marchiolli, M. A. & Galetti, D. Extended cahill–glauber formalism for finite-dimensional spaces: I. fundamentals. *Journal of Physics A: Mathematical and General* **38**, 6239–6251 (2005). https://doi.org/10.1088/0305-4470/38/27/010.

[191] Marchiolli, M. A., Ruzzi, M. & Galetti, D. Extended cahill-glauber formalism for finite-dimensional spaces. ii. applications in quantum tomography and quantum teleportation. *Physical Review A* **72**, 042308 (2005). https://link.aps.org/doi/10.1103/PhysRevA.72.042308.

[192] França, D. S., Strelchuk, S. & Studziński, M. Efficient classical simulation and benchmarking of quantum processes in the weyl basis. *Physical Review Letters* **126**, 210502 (2021). https://link.aps.org/doi/10.1103/PhysRevLett.126.210502.

[193] Howard, M. & Campbell, E. Application of a resource theory for magic states to fault-tolerant quantum computing. *Physical Review Letters* **118**, 090501 (2017). https://link.aps.org/doi/10.1103/PhysRevLett.118.090501.

[194] Heinrich, M. & Gross, D. Robustness of Magic and Symmetries of the Stabiliser Polytope. *Quantum* **3**, 132 (2019). https://doi.org/10.22331/q-2019-04-08-132.

[195] Rahimi-Keshari, S., Ralph, T. C. & Caves, C. M. Sufficient conditions for efficient classical simulation of quantum optics. *Physical Review X* **6**, 021039 (2016). https://link.aps.org/doi/10.1103/PhysRevX.6.021039.

[196] Raussendorf, R., Browne, D. E., Delfosse, N., Okay, C. & Bermejo-Vega, J. Contextuality and wigner-function negativity in qubit quantum computation. *Physical Review A* **95**, 052334 (2017).

[197] Jones, C. Low-overhead constructions for the fault-tolerant toffoli gate. *Physical Review A* **87**, 022328 (2013). https://link.aps.org/doi/10.1103/PhysRevA.87.022328.

[198] Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms.

The Journal of Physical Chemistry A **101**, 5111–5116 (1997). http://dx.doi.org/10.1021/jp970984n.

[199] tA v, A. *et al.* Qiskit: An open-source framework for quantum computing (2021).

[200] Renner, R. *Security of Quantum Key Distribution*. Ph.D. thesis, Swiss Federal Institute of Technology Zurich (2005). https://arxiv.org/abs/quant-ph/0512258.

[201] Giovagnoli, A. & Wynn, H. P. G-majorization with applications to matrix orderings. *Linear Algebra and its Applications* **67**, 111–135 (1985). https://www.sciencedirect.com/science/article/pii/0024379585901909.

[202] Eaton, M. L. & Perlman, M. D. Reflection Groups, Generalized Schur Functions, and the Geometry of Majorization. *The Annals of Probability* **5**, 829 – 860 (1977). https://doi.org/10.1214/aop/1176995655.

[203] Steerneman, A. G. M. G-majorization, group-induced cone orderings, and reflection groups. *Linear Algebra and its Applications* **127**, 107–119 (1990). https://www.sciencedirect.com/science/article/pii/002437959090338D.

[204] Fukuda, M. & Holevo, A. S. On weyl-covariant channels (2006). http://arxiv.org/abs/quant-ph/0510148.

[205] Fine, A. Hidden variables, joint probability, and the bell inequalities. *Physical Review Letters* **48**, 291–295 (1982). https://link.aps.org/doi/10.1103/PhysRevLett.48.291.

[206] Allahverdyan, A. E. & Danageozian, A. Excluding joint probabilities from quantum theory. *Physical Review A* **97**, 030102 (2018). https://link.aps.org/doi/10.1103/PhysRevA.97.030102.

[207] Barnett, S. M. & Radmore, P. M. *Methods in Theoretical Quantum Optics*. Oxford science publications (Clarendon Press, 1997). https://books.google.gr/books?id=Hp5z09gPmC8C.

[208] Arvidsson-Shukur, D. R. M. *et al.* Quantum advantage in postselected metrology. *Nature Communications* **11**, 3775 (2020). https://www.nature.com/articles/s41467-020-17559-w.

[209] Yunger Halpern, N., Swingle, B. & Dressel, J. Quasiprobability behind the out-of-time-ordered correlator. *Physical Review A* **97**, 042105 (2018). https://link.aps.org/doi/10.1103/PhysRevA.97.042105.

[210] Lostaglio, M. Quantum fluctuation theorems, contextuality, and work quasiprobabilities. *Physical Review Letters* **120**, 040602 (2018). https://link.aps.org/doi/10.1103/PhysRevLett.120.040602.

[211] Levy, A. & Lostaglio, M. Quasiprobability distribution for heat fluctuations in the quantum regime. *PRX Quantum* **1**, 010309 (2020). https://link.aps.org/doi/10.1103/PRXQuantum.1.010309.

[212] Arnold, V. I. Symplectic geometry and topology. *Journal of Mathematical Physics* **41**, 3307–3343 (2000). https://doi.org/10.1063/1.533315.

[213] Van Herstraeten, Z., Jabbour, M. G. & Cerf, N. J. Continuous majorization in quantum phase space (2021). https://arxiv.org/abs/2108.09167.

[214] Lami, L. *et al.* Gaussian quantum resource theories. *Physical Review A* **98**, 022335 (2018). https://link.aps.org/doi/10.1103/PhysRevA.98.022335.

[215] Klitzing, K. v., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. *Physical Review Letters* **45**, 494–497 (1980). https://link.aps.org/doi/10.1103/PhysRevLett.45.494.

[216] Moiseyev, N. *Non-Hermitian Quantum Mechanics* (Cambridge University Press, 2011).

[217] Gross, D., Nezami, S. & Walter, M. Schur–weyl duality for the clifford group with applications: Property testing, a robust hudson theorem, and de finetti representations. *Communications in Mathematical Physics* **385**, 1325–1393 (2021). https://doi.org/10.1007/s00220-021-04118-7.