Classical restrictions of matrix product states and minimising statistical errors in gate calibration

Inaugural-Dissertation zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

in Theoretischer Physik

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Yaiza Aragonés Soria

aus Vilassar de Mar, Catalunya

14. Februar 2022

Advisor, first reviewer and examiner | Prof. Dr. David Gross
Second reviewer and examiner | Prof. Dr. Ralph Bulla
Chair of the thesis defence committee | Prof. Dr. Peter Schilke

Vols venir a la meva barca?
— Hi ha violetes, a desdir!
anirem lluny sense recança
d'allò que haurem deixat aquí.

Anirem lluny sense recança
— i serem dues, serem tres.
Veniu, veniu, a la nostra barca,
les veles altes, el cel obert.

Maria-Mercè Marçal

Abstract

This thesis consists of two independent parts. In the first part, we investigate the notion of locality in the context of Matrix Product States (MPS), while in the second part, we minimise the statistical error in a calibration protocol for quantum-gate sets.

In the first part of this thesis, we consider an MPS in a one-dimensional lattice and investigate its norm squared amplitudes with respect to a local orthonormal basis. We call this norm the *classical restriction* of the MPS. Concretely, we ask when the classical restriction of an MPS is *quasi-locally Gibbsian*, i.e., it is exponentially well approximated by a Gibbs distribution associated with a local Hamiltonian. We prove that the classical restriction of an injective MPS is quasi-locally Gibbsian if the matrices associated with the MPS satisfy a 'purity' condition, a condition previously established in the theory of random matrix products.

Our result connects two notions of locality: locality of correlations in an MPS and locality of interactions in the Hamiltonian generating the corresponding Gibbs distribution.

The proof of our result consists of two steps. First, given an MPS defined on a lattice for which the purity condition holds, we demonstrate that the classical Conditional Mutual Information (CMI) of any connected tripartition of the lattice is rapidly decaying in the width of the middle region. Then, this decaying of the classical CMI is shown to imply that the probability distribution associated with the classical restriction of the MPS is quasi-locally Gibbsian.

Within this investigation, we present relevant observations around the purity condition. We research how 'typical' the purity condition is by constructing a probabilistic model and showing that, in this model, the purity condition is satisfied in general. Furthermore, we show that violating the purity condition enables a generalised notion of error correction, reinforcing the purity condition's generic nature.

Within the second part of this thesis, we consider a protocol for calibrating quantum-gate sets and implement a statistical analysis to minimise statistical errors. Calibration of quantum gates is a necessary hurdle to overcome on the way to a reliable quantum computer. In a recent paper, a protocol called Gate Set Calibration protocol (GSC) has been introduced and used to extract coherent errors from multi-qubit quantum gates. We build on this study in two ways. First, we take the uncertainty of any measurement in the protocol into account by performing a statistical analysis. Second, we optimise the statistical uncertainty while requiring that the protocol involves only a small number of distinct gates, aiding physical realisability. We numerically demonstrate that, just by adding two more single-qubit gates to GSC, the statistical error produced in the calibration of a CNOT gate is divided by a factor of more than two.

Contents

List of Figures

List of Tables

List of Abbreviations

CLT	Central Limit Theorem
CMI	Conditional Mutual Information
DOE	Design of Experiments
GSC	Gate Set Calibration protocol
GST	Gate Set Tomography
MPS	Matrix Product State
QEC	Quantum Error Correction
QPT	Quantum Process Tomography
RB	Randomized Benchmarking
SPAM	State-Preparation and Measurement
TN	Tensor Network

Introduction

Quantum information is an interdisciplinary field. At the point where quantum physics, computer science, information theory and mathematics meet, quantum information establishes a theoretical framework that studies information processing tasks performed by systems which are governed by the laws of quantum mechanics.

The beginnings of quantum mechanics date back to the turn of the 20th century, acquiring its current form at the end of the 1920s. Once the theory had a proper formalisation, it began to be applied to different fields of physics. This led to important scientific achievements such as establishing the standard model in particle physics or understanding many physical phenomena ranging from polymers to superconductors. Although all these discoveries meant significant advances in understanding natural phenomena, they did not shed much light on the understanding of quantum mechanics.

In the seventies, this started to change. Some scientists began to ask whether the fundamentals of computer science and information theory could be applied to quantum systems. Such a simple question completely transformed our view on quantum systems: instead of seeing them as mere entities of nature, they began to be considered as systems that could be designed, and thus used. As a result, new lines of research into the fundamentals of quantum mechanics opened up and questions combining physics, computer science and information theory emerged.

Since then, the field of quantum information has grown and obtained prominent successes. There are well-known theoretical results such as Shor's algorithm [1], which can factor a prime number in polynomial time, and experimental achievements that take steps towards quantum technologies such as quantum key distribution [2]. Even today, modest applications have started to emerge in industry with the goal of using quantum technology for optimisation problems, chemical simulations, or generating random numbers, among other examples.

There are still many open questions in quantum information. The most intriguing questions are probably about the long-awaited quantum computer. Is a quantum computer experimentally feasible? Which system will it be developed on? Will a quantum computer surpass the capabilities of a classical computer? If so, will the quantum computer be superior in all tasks or only in concrete ones? If the latter is the case, in which tasks would it be advantageous?

As early as 2000, DiVincenzo theorised the requirements for building a large-scale quantum computer, including scalability, qubit initialisation, long decoherence times and the ability to perform measurements on qubits [3]. Many of these requirements have been achieved over the years on different platforms. The preparation of quantum states, as well as the realisation of quantum measurements, has been achieved with high accuracy on a small number of qubits arranged in ion traps [4, 5]. Nuclear magnetic resonance has made it possible to implement dynamics in small molecules [6]. Larger scales, but less accurate implementations, have been achieved by solid-state technologies [7]. However, there is still no unique platform that meets all DiVincenzo's requirements at once.

On the other hand, some researchers have recently claimed to have demonstrated quantum supremacy for the first time [8, 9]. Nevertheless, most experts agree that we are still far away from large-scale quantum computers with practical applications, although it is widely believed that quantum information will bring secure communication and fast computational solutions to, at least, specific problems.

Many achievements in quantum computation are compared to classical computations. Thus, it could be easy to fall into the error of thinking that quantum information is simply a new technology within classical information as, for example, semiconductor transistors have been. This would indeed be a mistake, as quantum information should be understood as a way of processing information of a qualitatively incomparable nature to classical information. Quantum information introduces a new point of view, as well as tools and methodologies that are useful for understanding other fields of science, such as quantum many-body physics.

Quantum many-body physics studies the properties of quantum systems consisting of a large (or infinite) number of degrees of freedom. The traditional approach in quantum many-body physics is to design a Hamiltonian that represents the system's interactions according to some observed physical phenomena. Then, the properties of the system are calculated by numerical methods which, in general, are difficult to control.

Quantum information proposes to determine the state that characterises the system and its correlations instead of the Hamiltonian. However, the state lives in a Hilbert space which is too large to be handled efficiently. Simply the task of characterising the state is already inefficient, since it requires an amount of information that grows exponentially with the dimension of the system. For example, consider a system consisting of N qubits. The dimension of the Hilbert space describing this system is 2^N, so we need 2^N complex numbers to specify the state of the system. This means that if N is of the order of Avogadro's number, i.e., $N \sim \mathcal{O}(10^{23})$, we need more complex numbers than the number of atoms in the observable universe [10].

Fortunately, not all possible states within a Hilbert space are equally relevant in quantum many-body physics. This difference in relevance stems from the fact that physical interactions are local. The locality in interactions has implications for Hamiltonians describing physical

systems and for their fundamental states. A significant consequence is that quantum correlations, known as entanglement, between particles within a region of a system described by local interactions scale with the area of the system [11]. This property is exceptional because a randomly chosen state in a Hilbert space will have an entanglement that grows with the volume of the region, not with the area. By taking advantage of this property, physically relevant systems can be characterised by Tensor Network (TN) states [12, 13]. TN states represent a quantum system using a network of interconnected tensors, where the connections between the tensors encode the correlations of the system. This representation reduces the complexity of the state and allows for a numerically efficient characterisation.

In the first part of this thesis, we will investigate the simplest version of TN states, the Matrix Product States (MPSs). We will review the basic notions of MPSs and present our research on its locality.

One of the main obstacles in constructing a quantum computer is noise. Although significant theoretical results have been achieved assuming closed systems, i.e. systems that do not interact with the environment, the reality is different. Real systems suffer from undesired interactions that introduce noise into quantum information processes. Quantum error correction (QEC) is the branch of quantum information that studies how to protect information from noise of any origin, such as decoherence, faulty measurements, errors in state initialisation, perturbed quantum gates, among others.

All protocols designed in QEC must take into account the intrinsic uncertainty of quantum mechanics. Quantum mechanics is a theory that does not predict future events with certainty but only associates probabilities with them. This feature introduces statistical errors that must be considered to obtain precise results in experimental quantum information.

In the second part of this thesis, we perform a statistical analysis of a quantum-gate calibration protocol and propose a new version of the protocol that achieves smaller statistical errors.

Part I

Classical restrictions of matrix product states

Quantum many-body physics aims to describe the emergent properties of quantum systems made of many interacting particles. The description of many-body systems generally requires an amount of information exponentially large in the number of subsystems. Therefore, this description is feasible for a few particles but not scalable. Physically relevant interactions are, however, local. This property has the far-reaching consequence that ground states of the physically relevant Hamiltonians are not uniformly distributed in the ample Hilbert space. Instead, they live in a subspace that only grows polynomially with the subsystem size [14, 15]. In the first part of this thesis, we will delve into the mathematical representation of this Hilbert-space corner and its applications.

This part is divided into two chapters. Chapter 1 reviews all basic concepts that are necessary to understand the results presented in Chapter 2. Concretely, in the first chapter, we introduce Matrix Product States (MPSs) and Gibbs states.

Chapter 2 is based on our work on classical restrictions of MPSs. We consider finite quantum systems on one-dimensional lattices and study their classical restriction, i.e., their norm squared amplitudes with respect to a local orthonormal basis. The chapter starts establishing the notation and introduces a theoretical framework around the Conditional Mutual Information (CMI) regarding classical restrictions of MPSs. Afterwards, we show that, when a system satisfies that the CMI of any connected tripartition of the lattice decays rapidly in the width of the middle region, then the probability distribution associated with the classical restriction of the system can be exponentially well approximated by a 'local' Gibbs distribution. Furthermore, a condition referred to as 'purity', well established in the context of random matrix products, is shown to be a sufficient condition for the decay of the classical CMI for injective MPSs. We observe later that MPSs corresponding to symmetry-protected topological phases violate the purity condition. Afterwards, a connection is established between the violation of the purity condition and the notion of quantum error correction. For a set of operators, a condition that implies purity is given in terms of the span of the operators. We then investigate how 'often' the purity condition holds by constructing a probabilistic model and demonstrating that purity is a generic property within the model. Before closing the chapter with concrete examples, we compare the convergence rate of the CMI with the rate of convergence of the transfer operator to a fixed point.

Chapter 1

Foundations

In this chapter, we review some theoretical concepts needed to understand the results of this part of the thesis. Particularly, the chapter starts with Section 1.1, that reviews the notion of MPSs and their most relevant properties. We mainly focus on MPSs that satisfy a property called injectivity as it has interesting consequences for us. The section finishes with some examples.

Gibbs states constitute the second relevant concept of the first part of this thesis, and thus they are introduced in Section 1.2. We start reviewing the definition of Gibbs states. Then, the relevance of Gibbs states is pointed out in a fundamental context as well as in the context of quantum computation.

1.1 Matrix product states

Much effort has been put into developing tools to characterize many-body quantum systems, such as mean-field approximation [16], renormalization [17], and tensor networks [18, 19]. The latter has become a powerful tool in recent years, and especially its one-dimensional version, the so-called MPSs [21, 22, 20], which have been shown to be an efficient approximation of ground states of realistic local Hamiltonians [23]. In this section, we present a compact introduction to MPSs with a particular focus on injective MPSs. For formal definitions and proofs, see [24] and references therein.

Consider a one-dimensional lattice of N sites with open boundary conditions and assign to each site a d-dimensional Hilbert space, $\mathcal{H}_i$. An N-particle system can be represented by a *Matrix Product State* (MPS) defined as

$$|\psi\rangle = \sum_{x_1,\ldots,x_N=0}^{d-1} \langle R|A_{x_1}^{(1)} \cdots A_{x_N}^{(N)}|L\rangle |x_1,\ldots,x_N\rangle, \tag{1.1}$$

where $|x_1, \ldots, x_N\rangle = |x_1\rangle \cdots |x_N\rangle$. Here, $\{A_{x_i}^{(i)}\}_{x_i=0}^{d-1}$ is a set of $D_i \times D_{i+1}$ matrices associated to site i that encode the correlations of $|\psi\rangle$, and the vectors $|R\rangle$ and $|L\rangle$ encode the boundary conditions. The maximum matrix dimension, $D := \max_i D_i$, is referred to as the *bond dimension*. If the matrices $\{A_{x_i}^{(i)}\}_{x_i=0}^{d-1}$ are the same for all i, the MPS is called *translationally invariant*.

An MPS can also be defined on a lattice with periodic boundary conditions. In this case, the trace of the product of all A_{x_i} is taken such that

$$|\psi\rangle = \sum_{x_1,\ldots,x_N=0}^{d-1} \mathrm{Tr}\left(A_{x_1}^{(1)} \cdots A_{x_N}^{(N)}\right) |x_1, \ldots, x_N\rangle. \tag{1.2}$$

Eq. (1.1) reveals why MPSs are efficient descriptions of quantum many-body systems. A system of N subsystems is characterised by a state in a d^N-dimensional Hilbert space, $\mathcal{H}$. Therefore, the number of parameters that we need to specify such a state grows exponentially with N. If we restrict the system to be characterised by a state of the form of Eq. (1.1), the number of required parameters grows only linearly with the system size, as an MPS lives in a subspace of $\mathcal{H}$ with dimension at most $d \cdot D^2 \cdot N$.

As mentioned at the beginning of the section, not only are MPSs efficient representations of many-body systems, but they also provide a good approximation of ground states of local Hamiltonians [25]. Moreover, for a given bound dimension, the best MPS can be found efficiently in time, i.e., with a time that grows polynomially with the system size [26, 27]. These properties make MPSs a suitable approach for quantum simulations.

The MPS-representation is not unique. Consider a matrix M_α and its right-inverse M_α^{-1}, then the transformation

$$A_{x_i}^{(\alpha)} \to A_{x_i}^{(\alpha)} M_\alpha, \qquad\qquad A_{x_i}^{(\alpha+1)} \to M_\alpha^{-1} A_{x_i}^{(\alpha+1)}$$

leaves the MPS in Eq. (1.1) (as well as the one in Eq. (1.2)) invariant. This freedom can be practical for both numerical and analytical calculations. Here, we will assume that the MPS in Eq. (1.1) is *left-normalised*, which means that the corresponding matrices satisfy

$$\sum_{x_i=0}^{d-1} A_{x_i}^{(\alpha)\dagger} A_{x_i}^{(\alpha)} = \mathbb{I}.$$

For each MPS, a completely positive map called *transfer operator* is defined as

$$\mathbb{E}_\alpha(\cdot) = \sum_{x_i=0}^{d-1} A_{x_i}^{(\alpha)} \cdot A_{x_i}^{(\alpha)\dagger}.$$

Note that left-normalisation ensures that the transfer operator is trace-preserving. Note further that the transfer operator of a translationally invariant MPS uniquely defines the matrices A_{x_i} up to local unitaries.

The transfer operator of an MPS has become an essential tool to characterise MPSs. Applying the theory of quantum channels to the transfer operator has led to the description and classification of many properties of MPSs [20].

Injectivity of the transfer operator is, for example, one property that has relevant consequences on the MPS. We can define injectivity using its natural physical interpretation on states. Let $\{|\alpha\rangle\}_{\alpha=0}^{D-1}$ be a set of D orthonormal vectors which form an orthonormal basis in the Hilbert space of the MPS. Let also $\{A_{x_1}^{(1)}, \ldots, A_{x_N}^{(N)}\}_{x_i=0}^{d-1}$ be a collection of matrices and define the vector

$$|\psi_{\alpha\beta}\rangle := \sum_{x_1,\ldots,x_N=0}^{d-1} \langle\alpha|A_{x_1}^{(1)} \cdots A_{x_N}^{(N)}|\beta\rangle|x_1, \ldots, x_N\rangle,$$

where $\alpha, \beta = 0, \ldots, D-1$. Then, an MPS with corresponding matrices $\{A_{x_1}^{(1)}, \ldots, A_{x_N}^{(N)}\}_{x_i=0}^{d-1}$ is called *injective* if the set of states $\{|\psi_{\alpha\beta}\rangle\}_{\alpha,\beta=0}^{D-1}$ spans a vector space of dimension D^2, i.e., if $\dim\left[\mathrm{span}\left(\{|\psi_{RL}\rangle\}\right)\right] = D^2$, for a finite N. Note that, if an MPS is injective for N systems, then it is also injective for $N+1$ systems.

A local and frustration-free Hamiltonian with a given MPS as ground state can always be constructed. This Hamiltonian is called the *parent Hamiltonian* of the MPS and was originally defined in Ref. [28].

An injective MPS is the unique ground state of its corresponding parent Hamiltonian [20]. Moreover, injectivity ensures that the Hamiltonian is *gapped*, i.e., that there exists a non-trivial energy gap between the ground state and the first excited state [21]. Exponentially decaying correlations is also a consequence of injectivity of MPSs [29]. The latter is, in turn, equivalent to *primitivity* of the corresponding transfer operator, i.e., that $\mathbb{E}$ has a unique full-rank fixed point [30].

The research presented in Ch. 2 focuses on *classical restrictions* of MPSs. Given a local basis, $\{|x_1, \ldots, x_N\rangle\}_{x_i=0}^{d-1}$, we define the classical restriction as the quantum channel

$$\Phi(\rho) := \sum_{x_1,\ldots,x_N=0}^{d-1} |x_1, \ldots, x_N\rangle\langle x_1, \ldots, x_N|\langle x_1, \ldots, x_N|\rho|x_1, \ldots, x_N\rangle,$$

where ρ is any density matrix. Note that a classical restriction with respect to $\{|x_1, \ldots, x_N\rangle\}_{x_i=0}^{d-1}$ is diagonal in this basis, as Φ deletes all off-diagonal elements of the input state. Therefore, a classical probability distribution, $p_\rho(x_1, \ldots, x_N) = \langle x_1, \ldots, x_N|\rho|x_1, \ldots, x_N\rangle$, can be associated to the normalised density matrix $\Phi(\rho)$ for any basis $\{|x_1, \ldots, x_N\rangle\}_{x_i=0}^{d-1}$.

Consider now a subsystem, which we label by B, made of $M \leq N$ subsystems with associated Hilbert space $\mathcal{H}_B$, and let $\{|b_1, \ldots, b_M\rangle\}_{b_1,\ldots,b_M=0}^{d-1}$ be a local orthonormal basis of $\mathcal{H}_B$. Then, we analogously define a *classical restriction on the subsystem* B as

$$\Phi_B(\rho) := \sum_{b_1,\ldots,b_M=0}^{d-1} |b_1, \ldots, b_M\rangle\langle b_1, \ldots, b_M| \otimes \langle b_1, \ldots, b_M|\rho|b_1, \ldots, b_M\rangle, \tag{1.3}$$

$$= \sum_{b_1,\ldots,b_M=0}^{d-1} p_\rho(b_1, \ldots, b_M)\rho(b_1, \ldots, b_M). \tag{1.4}$$

Here, $p_\rho(b_1, \ldots, b_M) := \langle b_1, \ldots, b_M|\rho_B|b_1, \ldots, b_M\rangle$ is the classical probability associated to $\Phi_B(\rho)$, and $\rho_B = \mathrm{Tr}_{B^c}\,\rho$ is referred to as the reduced state of ρ in the subsystem B, with Tr_{B^c} the partial trace over all subsystems that are not in B. Moreover, we denote

$$\rho(b_1, \ldots, b_M) = \frac{1}{p_\rho(b_1, \ldots, b_M)} |b_1, \ldots, b_M\rangle\langle b_1, \ldots, b_M| \otimes \langle b_1, \ldots, b_M|\rho|b_1, \ldots, b_M\rangle, \tag{1.5}$$

which is the state after measuring the subsystems $1, \ldots, M$ and obtaining the measurement outcomes $b_1, \ldots, b_M$.

Although the input state, ρ, of a classical restriction can be any physical state, in Ch. 2 we investigate the case where ρ is an (injective) MPS. Consider $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle$ of the form of Eq. (1.1), then the probability associated to the classical restriction of $|\psi\rangle$ is

$$p_\psi(x_1, \ldots, x_N) = \left|\langle R|A_{x_1}^{(1)} \cdots A_{x_N}^{(N)}|L\rangle\right|^2. \tag{1.6}$$

Consider now the classical restriction of $|\psi\rangle$, but only applied on the M first subsystems, i.e., $\{b_1, \ldots, b_M\} = \{x_1, \ldots, x_M\} \subset \{x_1, \ldots, x_N\}$. The probability associated to this classical restriction is

$$p_\psi(b_1, \ldots, b_M) = \langle R|A_{b_1}^{(1)} \cdots A_{b_M}^{(M)} \mathbb{E}^{x_{M+1}\cdots x_N}(|L\rangle\langle L|) A_{b_M}^{(M)\dagger} \cdots A_{b_1}^{(1)\dagger}|R\rangle, \tag{1.7}$$

where $\mathbb{E}^{x_{M+1}\cdots x_N}$ is the convolution of the transfer operator, i.e.,

$$\mathbb{E}^{x_{M+1}\cdots x_N}(\sigma) = \sum_{x_{M+1}\cdots x_N=0}^{d-1} A_{x_{M+1}}^{(M+1)} \cdots A_{x_N}^{(N)} \sigma A_{x_N}^{(N)\dagger} \cdots A_{b_{M+1}}^{(M+1)\dagger}.$$

The probabilities in Eqns. (1.6) and (1.7) are the main object of study in the research presented in Ch. 2, but, before delving into that, we present some examples of MPSs in the following section.

1.1.1 Examples

In order to exemplify the concepts introduced previously in this section, we consider three well-known examples of MPSs: the GHZ state, the W state and the AKLT state.

GHZ state

In 1989 Greenberger, Horne and Zeilinger constructed a state with which non-statistical predictions of quantum mechanics were shown to be in contradiction with local realism [31]. Nowadays, this state, called the GHZ state, is one of the most remarkable states in quantum information because of its properties. For a system made of three qubits, for example, there exist two kinds of entanglement, and the GHZ maximises one of them [32].

The GHZ state for N particles can be written in the computational basis as

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}\right).$$

This state also has a translationally invariant MPS-representation with matrices

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

W state

As mentioned in the previous example, for a system of three qubits, there are two kinds of entanglement. One kind is maximised by the GHZ state and the other by the W state.

The W state has been generalised to systems of N qubits as the superposition of all pure N-qubit states which have only one excited state. Mathematically, we write

$$|\mathrm{W}\rangle = \frac{1}{\sqrt{N}}\left(|10\cdots0\rangle + |010\cdots0\rangle + \cdots + |0\cdots01\rangle\right).$$

To express the W state in an MPS-representation, one can take for all $i < N$ the matrices

$$A_0^{(i)} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad A_1^{(i)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and, for $i = N$,

$$A_0^{(N)} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_1^{(N)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

AKLT state

Affleck, Lieb, Kennedy, and Tasaki proposed in 1987 an extension of the one-dimensional quantum Heisenberg spin model [33]. The simplicity of the ground state of this model, known as AKLT state, and its relevance in condensed matter motivated the construction of its MPS representation in Ref. [22].

The AKLT state can be written as a translationally invariant MPS with periodic boundary conditions and corresponding matrices

$$A_+ = \sqrt{\frac{2}{3}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad A_0 = \frac{-1}{\sqrt{3}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_- = -\sqrt{\frac{2}{3}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

1.2 Gibbs states

Statistical thermodynamics is the branch of statistical mechanics focused on describing systems in thermodynamic equilibrium. It applies probability theory to microscopic degrees of freedom within a system to characterise the macroscopic behaviour of the system. In thermal equilibrium and under certain conditions, such a system is characterised by the so-called Gibbs states. In this section, we review the notion of Gibbs states and explain its importance in the context of quantum computation.

Consider a finite-dimensional system at non-zero temperature, T, with a Hamiltonian, H. The system is said to be in a *Gibbs state* if it is characterised by

$$\rho_\beta = \frac{1}{Z} e^{-\beta H}, \tag{1.8}$$

$$= \frac{1}{Z} \sum_i e^{-\beta E_i} |a_i\rangle\langle a_i|, \tag{1.9}$$

where Z is a normalization factor, $\beta = 1/(k_B T)$ with k_B the Boltzmann constant, and $|a_i\rangle$ is an eigenvector of H with corresponding eigenvalue (i.e., energy level) E_i. The normalization factor, which is referred to as *partition function* in statistical mechanics, can be written as

$$Z = \sum_i e^{-\beta E_i}.$$

Note that if no restrictions on the Hamiltonian are imposed, any non-null ρ_β can be written in the form of Eq. (1.8).

Gibbs states are fundamentally relevant in statistical physics because they are the *equilibrium state of the canonical ensemble*. In other words, when a (quantum or classical) system

can freely exchange energy with a bath at a specific temperature, T, the system equilibrates towards a Gibbs state with temperature T and energy-expectation value $\langle E \rangle = \operatorname{Tr} H\rho_\beta$. Furthermore, Gibbs states maximise the entropy of the system for a given energy.

The Hamiltonian characterising a system, H, encodes all system's properties, which will be translated to the corresponding Gibbs state. A Hamiltonian, H, of an N-particle system is called *local* if it can be written as a sum of terms, i.e., $H = \sum_{i=0}^{M-1} H_i$, where every term, H_i, acts non-trivially on a subset of $k < N$ particles. The Gibbs state corresponding to a local Hamiltonian is then referred to as a *local Gibbs state* (or *local Gibbsian distribution*). In Ch. 2, we moreover use the term *quasi-locally Gibbsian* when a probability distribution can be approximated by a local Gibbs distribution and the error of the approximation decays exponentially with k.

In Sec. 1.1, we have seen that MPSs are an efficient approach for simulating quantum many-body systems. Gibbs states turn out to be a relevant tool in quantum simulation as well. Quantum Monte Carlo algorithms, in particular, use a probabilistic representation of Gibbs states to compute the thermal or ground state of a *stoquastic* Hamiltonian. A Hamiltonian is said to be stoquastic on a concrete basis if it is local and all its terms H_i have non-positive off-diagonal matrix elements on that basis. Both simulation approaches rely then on locality, but the notion is slightly different in each case. The research that we present in Ch. 2 connects both types of locality.

Chapter 2

Results

This chapter consists of the article published as

> Y. Aragonés-Soria, J. Åberg, C.-Y. Park, and M. J. Kastoryano. "Classical restrictions of generic matrix product states are quasi-locally Gibbsian". In: *Journal of Mathematical Physics* 62.9 (Oct. 2021), p. 093511

This work is the result of a collaborative effort of all co-authors. While Yaiza Aragonés-Soria is the main author, Johan Åberg has provided an essential mathematical background to the project. Michael J. Kastoryano has supervised the work.

In the research presented here, we consider a quantum system defined on a finite lattice with open boundary conditions, and characterised by a vector, $|\psi\rangle$, on a finite-dimensional Hilbert space. We take a specific local basis, $\{|x\rangle := |x_1 x_2 \ldots\rangle\}$, where each x_i refers to one site of the lattice, and compute the probability distribution associated to the classical restriction of $|\psi\rangle$, i.e., $p_\psi(x) = |\langle x|\psi\rangle|^2$. Within this setting, we ask: when is the probability distribution $p_\psi(x)$ quasi-locally Gibbsian?

Our main result states that, if a state $|\psi\rangle$ is an injective MPS, then the probability $p_\psi(x)$ is quasi-locally Gibbssian if the *purity condition* holds. The matrices associated to an MPS, $\{A_x\}_{x=0}^{d-1}$, satisfy the purity condition if any orthogonal projector, P, that fulfils $PA_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P$ has rank one.

The relevance of this question arises from the notion of locality. Locality is a natural property of physical interactions, implying restrictions on the states that characterise quantum matter. The origin of locality, however, is different depending on the context. In the context of MPSs, locality originates on the description of MPS, which only allows for local correlations (see Eq. (1.1) and the discussion below Eq. (1.2)). On the other hand, in statistical mechanics, a Gibbs state is local if its corresponding Hamiltonian is local, i.e., if it models local interactions. Our result connects these two notions of locality.

During the demonstration of our main result, we make use of the classical CMI of $p_\psi(x)$, which is given by

$$I_{p_\psi}(A:C|B) = H(p_{\psi,AB}) + H(p_{\psi,BC}) - H(p_{\psi,B}) - H(p_{\psi,ABC}),$$

where A, B, C are lattice regions, and $p_{\psi,B} := p_\psi(b_1, \ldots, b_{|B|}) = \langle b_1, \ldots, b_{|B|}|\psi_B|b_1, \ldots, b_{|B|}\rangle$ is the classical probability associated to $\Phi_B(|\psi\rangle\langle\psi|)$ (see Eq. (1.3)), with $\psi_B := \mathrm{Tr}_{B^c}\,|\psi\rangle\langle\psi|$. The probabilities $p_{\psi,AB}$, $p_{\psi,BC}$, and $p_{\psi,ABC}$ are defined analogously.

The proof of our main result consists of two steps:

1. We show that the classical CMI of an injective MPS, $|\psi\rangle$, decays exponentially if the matrices associated with the MPS satisfy the purity condition.

2. We demonstrate that a probability distribution defined on a lattice, $p_\psi(x)$, is quasi-locally Gibbsian if the CMI between any tripartition of the lattice decays rapidly in the width of the middle region.

The second step of the proof is similar to Kozlov's theorem [35], while the first step makes use of the theory of products of random matrices [36], and its application to quantum trajectories [37, 38]. Note that this chapter has its own independent appendix and bibliography.

Classical restrictions of generic matrix product states are quasi-locally Gibbsian

Y. Aragonés-Soria,* J. Åberg, and C-Y. Park
Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, 50937 Köln, Germany.

M. J. Kastoryano
Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, 50937 Köln, Germany.
Amazon Quantum Solutions Lab, Seattle, Washington 98170, USA and
AWS Center for Quantum Computing, Pasadena, California 91125, USA
(Dated: September 20, 2021)

We show that the norm squared amplitudes with respect to a local orthonormal basis (the classical restriction) of finite quantum systems on one-dimensional lattices can be exponentially well approximated by Gibbs states of local Hamiltonians (i.e., are quasi-locally Gibbsian) if the classical conditional mutual information (CMI) of any connected tripartition of the lattice is rapidly decaying in the width of the middle region. For injective matrix product states, we moreover show that the classical CMI decays exponentially, whenever the collection of matrix product operators satisfies a 'purity condition'; a notion previously established in the theory of random matrix products. We furthermore show that violations of the purity condition enables a generalized notion of error correction on the virtual space, thus indicating the non-generic nature of such violations. We make this intuition more concrete by constructing a probabilistic model where purity is a typical property. The proof of our main result makes extensive use of the theory of random matrix products, and may find applications elsewhere.

* yaiza.aragonessoria@gmail.com; Institute for Theoretical Physics, University of Cologne, Zülpicher Str. 77, 50937 Köln, Germany.

I. INTRODUCTION

Considerable effort has been devoted to understanding the entanglement properties of many-body quantum states. For finite one-dimensional-lattice systems, the theory of Matrix Product States (MPSs) provides a complete framework for describing entanglement of gapped many-body systems [1], and allows for efficient high precision simulations via the DMRG algorithm [2–4]. Similarly impressive degrees of numerical precision can be reached in other settings, such as disorder [5], open systems [6], time evolution [7], or critical systems [8]. The success of these simulation methods can be traced back to the accurate parametrization of entanglement in MPSs. There exist extensions to lattices of higher dimensions (projected entangled pair states) but these have been far less useful for simulations, due to their extensive entanglement growth.

In contrast, quantum Monte-Carlo simulations are largely based on heuristic assumptions on the weights and phases of the underlying state. Indeed, if the system under study can be cast in a form with only positive weights, then Monte-Carlo methods often work well, though convergence guarantees are only known in very special cases [9]. This in turn is believed to be due to the local Gibbsian nature of the classical restriction of the state. Classical Monte-Carlo sampling is known to converge rapidly for Ising type problems [10, 11], while quantum variational Monte-Carlo is often successful when using a locally restricted Gibbs Ansatz, such as the Jastrow-Ansatz. Further evidence of the importance of locality in the Ansatz wavefunction has been observed for more expressive Ansätze, such as the complex Restricted Boltzmann machine [12], where the activations naturally preserve locality in many cases. Hence, whereas tensor network states explicitly encode the local entanglement structure in their construction, quantum (variational) Monte Carlo implicitly invokes locality through the pervasive Gibbsian nature of probability distributions.

Here, we connect these two pictures by showing that generic injective MPSs [13] have classical restrictions that are quasi-locally Gibbsian. More precisely, we here refer to a probability distribution as locally Gibbsian if it can be written as the equilibrium distribution of a local Hamiltonian, i.e., as a sum of terms that each spans at most ℓ adjacent sites. Well known examples include the Ising and Potts models. We similarly say that a distribution is quasi-locally Gibbsian, if it can be approximated by Gibbs distributions corresponding to local Hamiltonians, h^ℓ, where the error of the approximation in some sense decays exponentially with increasing ℓ. Such notions appear in various guises in the literature, e.g., Ref. 14, which requires that the coefficients in the cluster expansion of $\log(p)$ are rapidly decaying with the order of the cluster.

As the first step towards proving the generic quasi-local Gibbs property of injective MPSs, we show (in Section IV) that probability distributions on a one-dimensional lattice with open boundary conditions are quasi-locally Gibbsian if the Conditional Mutual Information (CMI) between any tripartition of the lattice is decaying rapidly in the width of the middle region. The stronger the decay of the CMI, the more local the Gibbs distribution. In the case of zero correlation length, the distribution is (strictly) locally Gibbs [15]. A number of recent studies in quantum information theory have revealed connections between the CMI and the Gibbsian nature of density matrices. In Ref. 16, the authors show that the quantum CMI of a full rank density matrix on a one-dimensional lattice is small if and only if the state is Gibbsian. The Gibbsian nature of states has important implications for the nature of edge states of topologically ordered systems [17, 18]. Our results show that similar equivalences hold for classical restrictions of quantum states. Similar conclusions can be reached using perturbative methods for classical restrictions of high temperature quantum Gibbs states of gaped spin chains [19].

The second step towards establishing the quasi-locality is also our main result; that the classical restriction of injective MPSs have an exponentially decaying CMI if the matrices associated to the MPS satisfy a condition referred to as *purity* (see Def. 3). This condition has previously been shown [20, 21] to imply the 'purification' of quantum trajectories resulting from the applications of sequences of random matrices on an initial state. In our setting, the classical CMI can be bounded by a corresponding quantum CMI. The latter can, in turn, be rewritten in terms of the expected entanglement entropy after measurements on the conditional subsystem. A vanishing entanglement entropy is thus equivalent to the purification of the state-trajectory induced by the sequence of measurements on the virtual system. The purification of trajectories implies that the system asymptotically jumps between pure states of a specific stationary measure, irrespective of what (mixed) state the system started in. We are currently not aware of a meaningful operational interpretation of the stationary stochastic process, and believe it to be quite hard to evaluate in practice [22]. Furthermore, and perhaps counter-intuitively, we observe that the rate of decay towards the stationary measure is unrelated to the gap of the transfer operator of the MPS. We moreover do not know of a closed functional form for the decay rate, in terms of the matrices associated to the MPS.

One may note that our setting, which focuses on the degree of conditional post-measurement entanglement, is closely related to the notion of localizable entanglement [23–25]. The latter is obtained by optimizing the measurements over all possible local bases, while we consider a fixed basis. However, to the best of our knowledge, a general proof of the exponential decay of the localizable entanglement has not been shown previously.

As mentioned above, we show that the purity condition is sufficient for the exponential decay of the *classical* CMI. However, it is less clear if it also is a sufficient condition. In order to better pinpoint the significance of the purity condition, we show that it in essence is both necessary and sufficient for the exponential decay of the above mentioned *quantum* CMI.

As a further attempt to gain a better understanding of the purity condition, we moreover investigate the conspicuous similarity between (the violation of) the purity condition (see Def. 3) and the Knill-Laflamme error correction condition [26]. Indeed, we find (in Section VI B) that the purity-condition can be regarded as the non-existence of a non-trivial correctable subspace that

persists indefinitely throughout iterated applications of an error-model, in a somewhat unconventional error correction scenario. One may note that invariant subspaces are special cases of such correctable spaces. As an example, MPSs with Symmetry-Protected Topological (SPT) order are associated to invariant subspaces [27] and would thus violate the purity condition. The above results suggests that violations of the purity condition in some sense are 'fragile'. In order to shed some further light on this question, we construct a probabilistic model (in Sec. VI D), where the purity condition holds, apart for a subset of measure zero.

The proof of our main theorem relies heavily on the theory of random matrix products, and in particular on the work of Benoist et. al. [20] and Maassen and Kümmerer [21]. Since these results involve notions from probability theory that likely are unfamiliar to most of the quantum information community, we reproduce in Appendixes A-D many of the basic results in a language that should be more familiar to the quantum-information reader. We hope that this will facilitate the access to a rich and extensive body of work that should see many more applications in the fields of quantum information and many body physics. For instance, the theory of random matrix products has recently been leveraged in a different setting, to show ergodicity for ensembles of quantum channels [28, 29].

Concerning the structure of the paper, we begin by introducing the notation in Section II, while Section III focuses on the central object in this investigation, namely the CMI with respect to classical restrictions of MPSs. Section IV presents the first result of the paper: an exponentially decaying CMI implies quasi-local Gibbs distributions. Section V is devoted to the main result, namely the exponentially decaying CMI for a broad class of MPSs. Section VI provides examples and observations, where we in Section VI A observe that MPSs corresponding to SPT phases violate the purity condition. In Section VI B we further investigate the purity condition and show that its violation can be regarded as a type or error-correction condition. Section VI C is devoted to give a sufficient condition for purity to hold for a set of operators in terms of the span of the operators. We use this relation in Section VI D, where we construct a model of typicality of purity, to prove that purity is a generic property. Section VI E compares the convergence rate of the CMI with the rate of the converge to the fixed point of the transfer operator. Concrete examples are provided in Section VI F. We finish with an outlook in Section VII.

II. NOTATION

We consider pure states defined on a finite one-dimensional lattice, Λ, and associate a finite dimensional Hilbert space of dimension d to each site. We index the sites of the lattice according to a tripartition of the lattice $\Lambda = ABC$ as follows: we denote sites in region A as $-|A|+1, -|A|+2, \ldots, -1, 0$; sites in region B as $1, \ldots, N$; and sites in region C as $N+1, \ldots, |BC|$ (see Fig. 1). This peculiar indexing of sites will make sense later on when considering the CMI for MPSs.

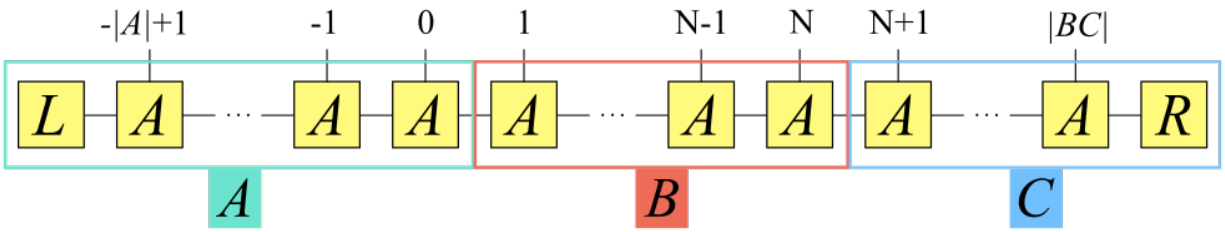

FIG. 1. We consider a MPS on a finite lattice, Λ, which is broken up into three contiguous regions such that $\Lambda = ABC$. We denote sites in region A as $-|A|+1, -|A|+2, \ldots, -1, 0$; sites in region B as $1, \ldots, N$; and sites in region C as $N+1, \ldots, |BC|$.

Let $|x_\Lambda\rangle = |x_{-|A|+1}, \ldots, x_0, x_1, \ldots, x_N, \ldots, x_{|BC|}\rangle$ be a local orthonormal basis, where $\{|x_i\rangle\}_{x_i=0}^{d-1}$ is the local basis at site i. Unless specified otherwise, we will be working with translationally invariant MPSs with open boundary conditions

$$|\Psi\rangle = \frac{1}{K} \sum_{x_{-|A|+1}, \ldots, x_{|BC|}=0}^{d-1} \langle R | A_{x_{|BC|}} \cdots A_{x_{-|A|+1}} | L \rangle | x_{-|A|+1} \cdots x_{|BC|} \rangle, \tag{1}$$

where K is a normalization factor. Here, A_{x_i} are $D \times D$ matrices encoding correlations in the system and $|L\rangle$ and $|R\rangle$ are normalized states on the D-dimensional virtual space specifying the boundary conditions, where D is known as the bond dimension of the MPS. Without loss of generality, we consider (left-)normalized MPSs, which enforces that $\sum_{x_i=0}^{d-1} A_{x_i}^\dagger A_{x_i} = \mathbb{1}$. Left normalization guarantees that the completely positive map

$$\mathbb{E}(\cdot) := \sum_{x_i=0}^{d-1} A_{x_i} \cdot A_{x_i}^\dagger \tag{2}$$

is trace preserving. The map $\mathbb{E}$ is often referred to as the transfer operator and maps density matrices on the virtual space to density matrices from left to right. The adjoint map, $\mathbb{E}^*$, maps operators from right to left along the chain. Our choice of boundary conditions serves mainly for notational simplicity. The results in the paper extend naturally to periodic or mixed boundary conditions. For periodic boundary conditions, the regions ABC need to be chosen differently to ensure that B separates A from C.

The normalization constant can be expressed concisely as $K^2 = \mathrm{Tr}\left[\mathbb{E}^{|\Lambda|}(L)R\right]$, where we use the shorthand notation $R = |R\rangle\langle R|$ and $L = |L\rangle\langle L|$.

a. Classical Restrictions For a given local basis $\{|x_\Lambda\rangle\}$, we define the quantum channel

$$\Phi_\Lambda(\psi) = \sum_{x_\Lambda} |x_\Lambda\rangle\langle x_\Lambda|\langle x_\Lambda|\psi|x_\Lambda\rangle. \tag{3}$$

In other words, Φ_Λ generates a state that is diagonal with respect to the basis $\{|x_\Lambda\rangle\}$, by deleting the off-diagonal elements of the input ψ. We refer to Φ_Λ as the classical restriction (also commonly referred to as a 'dephasing map' or 'pinching'). Since $\Phi_\Lambda(\psi)$ is diagonal, the map Φ_Λ effectively defines a classical probability distribution, $p_\psi(x_\Lambda) = \langle x_\Lambda|\psi|x_\Lambda\rangle$, for any choice of basis $\{|x_\Lambda\rangle\}$.

We also consider the channel that measures a subset of systems $B \subset \Lambda$ and we denote it as

$$\begin{aligned}\Phi_B(\psi) &= \sum_{x_B} |x_B\rangle\langle x_B|\langle x_B|\psi|x_B\rangle, \\ &= \sum_{x_B} p_\psi(x_B)\psi(x_B),\end{aligned} \tag{4}$$

with $|x_B\rangle = \bigotimes_{i\in B}|x_i\rangle$. Here, the channel Φ_B similarly defines a classical probability distribution on the sites in B by $p_\psi(x_B) = \langle x_B|\psi_B|x_B\rangle$, where $\psi_B := \mathrm{Tr}_{\Lambda\setminus AC}\,\psi$ is the reduced state of ψ on B. Note that

$$p_\psi(x_B) = \sum_{x_{AC}} p_\psi(x), \tag{5}$$

where recall that $\Lambda = ABC$. Moreover, we refer to the post-measurement state after obtaining the measurement outcome x_B as

$$\psi(x_B) = \frac{1}{p_\psi(x_B)}|x_B\rangle\langle x_B| \otimes \langle x_B|\psi|x_B\rangle. \tag{6}$$

Consider now the MPS defined in Eq. (1). The probability distribution on B is

$$p_\Psi(x_B) = \frac{1}{K^2}\,\mathrm{Tr}\left[A_{x_N}\cdots A_{x_1}\mathbb{E}^{|A|}(L)A^\dagger_{x_1}\cdots A^\dagger_{x_N}\mathbb{E}^{*|C|}(R)\right], \tag{7}$$

where $\mathbb{E}^n$ is understood as convolution of the map and $x_B := x_1,\ldots,x_N$, with $x_i = 0,\ldots,d-1$.

In this investigation we primarily focus on translationally invariant injective MPSs [13], which equivalently can be defined via the primitivity of the map $\mathbb{E}$ [30]. The latter means that $\mathbb{E}$ has a unique full-rank fixed point, i.e., there exists a unique full-rank density operator ρ such that $\mathbb{E}(\rho) = \rho$. A consequence of the injectivity of the MPS is thus that $\lim_{|A|\to\infty}\mathbb{E}^{|A|}(\chi) = \rho\,\mathrm{Tr}(\chi)$, $\lim_{|C|\to\infty}\mathbb{E}^{*|C|}(Q) = \mathbb{1}\,\mathrm{Tr}(Q\rho)$, and $\lim_{|A|\to\infty,|C|\to\infty}K^2 = \mathrm{Tr}(R\rho) \neq 0$. Hence, if region B is kept fixed, while regions A and C both grow to infinity, the probability distribution (7) on B reduces to

$$p_\Psi(x_B) = \mathrm{Tr}\left[A_{x_N}\cdots A_{x_1}\rho A^\dagger_{x_1}\cdots A^\dagger_{x_N}\right]. \tag{8}$$

III. THE POST-MEASUREMENT CONDITIONAL MUTUAL INFORMATION

Throughout the paper, we use a number of entropic quantities, which we introduce in this section. In particular, we switch back and forth between classical and quantum systems. The quantum von Neumann entropy of a mixed state, χ, is denoted as $S(\chi) = -\mathrm{Tr}\,\chi\log\chi$, while the classical entropy is referred to as $H(p) = -\sum_x p(x)\log p(x)$ for a classical probability distribution $p(x)$. Here, log denotes the natural logarithm. In Sec. IV, we use the classical relative entropy as a measure of distinguishability between probability distributions. The classical relative entropy of $p_1(x)$ with respect to $p_2(x)$ is defined as

$$S(p_1||p_2) = \sum_x p_1(x)\log\left[\frac{p_1(x)}{p_2(x)}\right]. \tag{9}$$

The (quantum) CMI between regions A and C conditioned on region B, is given by

$$I_\chi(A:C|B) = S(\chi_{AB}) + S(\chi_{BC}) - S(\chi_B) - S(\chi_{ABC}). \tag{10}$$

After applying the classical conditioning map, Φ_Λ in Eq. (3), on a quantum state, χ, we get the classical CMI

$$I_{\Phi_\Lambda(\chi)}(A:C|B) = I_{p_\chi}(A:C|B) = H(p_{\chi,AB}) + H(p_{\chi,BC}) - H(p_{\chi,B}) - H(p_{\chi,ABC}), \tag{11}$$

where $p_{\chi,A} := p_\chi(x_A) = \langle x_A|\chi_A|x_A\rangle$.

We now point out an important observation on the CMI [31]. Suppose that we have a pure state, $\psi = |\psi\rangle\langle\psi|$, and we measure all spins in region B. Then, the quantum CMI of the post-measurement state satisfies

$$\begin{aligned} I_{p_\psi}(A:C|B) &\leq I_{\Phi_B(\psi)}(A:C|B), \\ &= \langle S[\psi_A(x_B)]\rangle_{p_\psi(x_B)} + \langle S[\psi_C(x_B)]\rangle_{p_\psi(x_B)}, \\ &= 2\langle S[\psi_C(x_B)]\rangle_{p_\psi(x_B)}, \end{aligned} \tag{12}$$

where the distribution p_ψ is defined by $p_\psi(x_\Lambda) = \langle x_\Lambda|\psi|x_\Lambda\rangle$, and where the state $\psi_X(x_B) = \langle x_B|\rho_{XB}|x_B\rangle$ is the reduced state in region X of the post-measurement state, $\psi(x_B)$, and $\langle S[\psi(x)]\rangle_{p_{\psi(x)}}$ is the average von Neumann entropy of $\psi(x)$ over $p_\psi(x)$, i.e.,

$$\langle S[\psi(x)]\rangle_{p_\psi(x)} := \sum_x p_\psi(x) S[\psi(x)]\,. \tag{13}$$

The inequality in Eq. (12) comes from monotonicity of the relative entropy. Note that $S[\psi_A(x_B)] = S[\psi_C(x_B)]$ since $\langle x_B|\psi|x_B\rangle/p_\psi(x_B)$ is a pure state on the bipartition AC. Eq. (12) allows us to characterise the states that have a small post-measurement CMI by finding the states that have a small average entropy of $\psi_C(x_B)$.

Let us now go back to the MPS described in Sec. II. With the injective MPS in the canonical form of Eq. (1), it can be shown that the reduced state of the post-measurement state, $\Psi_C(x_B)$, is (up to zero eigenvalues) isospectral to

$$\frac{1}{p_\Psi(x_B)K^2}\sqrt{\mathbb{E}^{*|C|}(R)}A_{x_N}\cdots A_{x_1}\mathbb{E}^{|A|}(L)A^\dagger_{x_1}\cdots A^\dagger_{x_N}\sqrt{\mathbb{E}^{*|C|}(R)}. \tag{14}$$

The average von Neumann entropy of the reduced state of a post-measurement translationally invariant injective MPS is then

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} = \sum_{x_B} p_\Psi(x_B) S\left(\frac{1}{p_\Psi(x_B)K^2} F A_{x_N}\cdots A_{x_1}\sigma A^\dagger_{x_1}\cdots A^\dagger_{x_N}F^\dagger\right), \tag{15}$$

where $\sigma := \mathbb{E}^{|A|}(L)$, and

$$F := \sum_{x_{|BC|},\ldots,x_{N+1}} |x_{x_{|BC|},\ldots,x_{N+1}}\rangle\langle R|A_{x_{|BC|},\ldots,x_{N+1}}\,, \tag{16}$$

and thus $F^\dagger F = \mathbb{E}^{*|C|}(R)$. Eq. (15) will be the main object of study throughout this paper. As mentioned earlier, a translationally invariant injective MPS results in a primitive channel $\mathbb{E}$. On a finite-dimensional space, this implies that for sufficiently large $|A|$ and $|C|$, it follows that both σ and $F^\dagger F$ are full-rank operators. We also recall that $\lim_{|A|\to\infty}\mathbb{E}^{|A|}(\chi) = \rho\,\mathrm{Tr}(\chi)$, $\lim_{|C|\to\infty}\mathbb{E}^{*|C|}(Q) = \mathbb{1}\,\mathrm{Tr}(Q\rho)$, and $\lim_{|A|\to\infty,|C|\to\infty}K^2 = \mathrm{Tr}(R\rho) \neq 0$, and consequently (15) reduces to

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} = \sum_{x_B} p_\Psi(x_B) S\left(\frac{A_{x_N}\cdots A_{x_1}\rho A^\dagger_{x_1}\cdots A^\dagger_{x_N}}{p_\Psi(x_B)}\right), \tag{17}$$

for infinite chains.

IV. DISTRIBUTIONS WITH SMALL CMI ARE QUASI-LOCALLY GIBBSIAN

In this section we consider probability distributions, $p_{1,\ldots,|\Lambda|}$, on finite one-dimensional lattices, Λ, and discuss conditions for when these can be well approximated by Gibbs distributions of local Hamiltonians. We say that a Hamiltonian is ℓ-local if it can be written as a sum of terms that each span at most ℓ consecutive sites. A distribution is ℓ-local if it is the Gibbs distribution of some ℓ-local Hamiltonian. In a similar spirit, we say that a distribution is quasi-locally Gibbsian if it can be approximated

by ℓ-local distributions, where the error of this approximation in some sense decays fast with respect to increasing ℓ. In this section we show that, if the CMI $I_{p_{1,\ldots,|\Lambda|}}(A:C|B)$ of the distribution $p_{1,\ldots,|\Lambda|}$ decays sufficiently fast with increasing size $|B|$ of the bridging region in a contiguous tripartition $\Lambda = ABC$ of the lattice, then $p_{1,\ldots,|\Lambda|}$ is quasi-locally Gibbsian. (For convenience we change the notation in this section and enumerate the sites of the entire lattice as $1,\ldots,|\Lambda|$.) This result is similar in spirit to Kozlov's theorem [14] (see also Ref. 31). Although this section exclusively focuses on probability distributions, the application to quantum states becomes apparent in Section V, where we consider classical restrictions of underlying injective MPSs and show that these are quasi-locally Gibbsian under broad conditions.

Let $p_{1,\ldots,|\Lambda|}$ be a probability distribution over a finite sub-chain Λ of a one-dimensional lattice. We let p_j denote the marginal distribution at site j. For $1 \leq j \leq k \leq |\Lambda|$ we let $p_{j,\ldots,k}$ denote the marginal distribution of the chain $j,\ldots,k$. In the following, we assume that

$$p_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) > 0, \quad \forall x_1,\ldots,x_{|\Lambda|}, \tag{18}$$

which consequently leads to $p_{j,\ldots,k}(x_j,\ldots,x_k) > 0$. With these assumptions, we can define

$$h_{j,\ldots,k} := -\log p_{j,\ldots,k}, \quad 1 \leq j \leq k \leq |\Lambda|, \tag{19}$$

and thus $p_{j,\ldots,k} = e^{-h_{j,\ldots,k}}$, where we for notational convenience assume that $h_{j,\ldots,j} := h_j$ and $h_{j,\ldots,j+1} := h_{j,j+1}$.

Hence, we have constructed $h_{j,\ldots,k}$ such that $p_{j,\ldots,k}$ is Gibbs distributed with respect to $h_{j,\ldots,k}$, with $\beta = 1$ in $e^{-\beta h_{j,\ldots,k}}/Z(h_{j,\ldots,k})$, where one may note that $Z(h_{j,\ldots,k}) := \sum_{x_j,\ldots,x_k} e^{-h_{j,\ldots,k}(x_j,\ldots,x_k)} = 1$.

For $1 \leq \ell \leq |\Lambda| - 2$, we define

$$h^{\ell}_{1,\ldots,|\Lambda|} := \sum_{j=1}^{|\Lambda|-\ell} h_{j,\ldots,j+\ell} - \sum_{j=1}^{|\Lambda|-\ell-1} h_{j+1,\ldots,j+\ell}. \tag{20}$$

Hence, $h^{\ell}_{1,\ldots,|\Lambda|}$ only includes the terms for which the range does not exceed ℓ. More precisely, $h^{\ell}_{1,\ldots,|\Lambda|}$ is a $(\ell+1)$-local Hamiltonian. The associated $(\ell+1)$-local Gibbs distribution is

$$p^{\ell}_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) := \frac{e^{-h^{\ell}_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|})}}{Z(h^{\ell}_{1,\ldots,|\Lambda|})}, \tag{21}$$

with

$$Z(h^{\ell}_{1,\ldots,|\Lambda|}) := \sum_{x'_1,\ldots,x'_{|\Lambda|}} e^{-h^{\ell}_{1,\ldots,|\Lambda|}(x'_1,\ldots,x'_{|\Lambda|})}.$$

The following proposition expresses the classical relative entropy (see Eq. (9)) between the Gibbs distribution $p_{1,\ldots,|\Lambda|}$ associated to the full Hamiltonian, $h_{1,\ldots,|\Lambda|}$, and the Gibbs distribution $p^{\ell}_{1,\ldots,|\Lambda|}$ associated to the $(\ell+1)$-local Hamiltonian, $h^{\ell}_{1,\ldots,|\Lambda|}$, in terms of the CMIs between suitable regions of the chain. Hence, if the latter are sufficiently small, then the approximating $(\ell+1)$-local distribution $p^{\ell}_{1,\ldots,|\Lambda|}$ is close to the original distribution $p_{1,\ldots,|\Lambda|}$.

Proposition 1. *For $p_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) > 0$, let $p^{\ell}_{1,\ldots,|\Lambda|}$ be as defined in Eq. (19-21). For $1 \leq \ell \leq |\Lambda| - 2$ it is the case that*

$$S(p_{1,\ldots,|\Lambda|} \| p^{\ell}_{1,\ldots,|\Lambda|}) = \sum_{k=1}^{|\Lambda|-\ell-1} I(1,\ldots,k : k+\ell+1 | k+1,\ldots,k+\ell). \tag{22}$$

Proof. We first note that

$$\langle h_{j,\ldots,k} \rangle_{p_{1,\ldots,|\Lambda|}} = H(p_{j,\ldots,k}). \tag{23}$$

A somewhat lengthy but straightforward calculation moreover yields

$$Z(h^{\ell}_{1,\ldots,|\Lambda|}) = 1. \tag{24}$$

Next we observe that

$$\begin{aligned}
&\sum_{k=1}^{|\Lambda|-\ell-1} I(1,\dots,k:k+\ell+1|k+1,\dots,k+\ell)\\
=&H(p_{1,\dots,\ell+1})-H(p_{1,\dots,|\Lambda|})\\
&+\sum_{k=1}^{|\Lambda|-\ell-1}\Big[H(p_{k+1,\dots,k+\ell+1})-H(p_{k+1,\dots,k+\ell})\Big],\\
&\text{[By (23)]}\\
=&\langle h_{1,\dots,\ell+1}\rangle-\langle h_{1,\dots,|\Lambda|}\rangle\\
&+\sum_{j=1}^{|\Lambda|-\ell-1}\langle h_{j+1,\dots,j+\ell+1}\rangle-\sum_{j=1}^{|\Lambda|-\ell-1}\langle h_{j+1,\dots,j+\ell}\rangle,\\
=&-\langle h_{1,\dots,|\Lambda|}\rangle+\sum_{j=1}^{|\Lambda|-\ell}\langle h_{j,\dots,j+\ell}\rangle-\sum_{j=1}^{|\Lambda|-\ell-1}\langle h_{j+1,\dots,j+\ell}\rangle.
\end{aligned}\tag{25}$$

Next we note that

$$\begin{aligned}
S(p_{1,\dots,|\Lambda|}\|p^{\ell}_{1,\dots,|\Lambda|})=&-\sum_{x_1,\dots,x_{|\Lambda|}}p_{1,\dots,|\Lambda|}(x_1,\dots,x_{|\Lambda|})h_{1,\dots,|\Lambda|}(x_1,\dots,x_{|\Lambda|})\\
&+\sum_{x_1,\dots,x_{|\Lambda|}}p_{1,\dots,|\Lambda|}(x_1,\dots,x_{|\Lambda|})h^{\ell}_{1,\dots,|\Lambda|}(x_1,\dots,x_{|\Lambda|})\\
&+\log Z(h^{\ell}_{1,\dots,|\Lambda|}),\\
=&-\langle h_{1,\dots,|\Lambda|}\rangle+\langle h^{\ell}_{1,\dots,|\Lambda|}\rangle+\log Z(h^{\ell}_{1,\dots,|\Lambda|}),\\
&[\quad\text{By (24)}\quad]\\
=&-\langle h_{1,\dots,|\Lambda|}\rangle+\langle h^{\ell}_{1,\dots,|\Lambda|}\rangle,\\
=&-\langle h_{1,\dots,|\Lambda|}\rangle+\sum_{j=1}^{|\Lambda|-\ell}\langle h_{j,\dots,j+\ell}\rangle-\sum_{j=1}^{|\Lambda|-\ell-1}\langle h_{j+1,\dots,j+\ell}\rangle.
\end{aligned}\tag{26}$$

A comparison with with (25) yields (22). □

Loosely speaking, the above proposition tells us that, if the CMIs $I(1,\dots,k:k+\ell+1|k+1,\dots,k+\ell)$ in some sense decrease sufficiently fast with increasing ℓ, then the $(\ell+1)$-local Gibbs distribution $p^{\ell}_{1,\dots,|\Lambda|}$ approaches the true distribution $p_{1,\dots,|\Lambda|}$. The following lemma formalizes this intuition.

Lemma 2. *Suppose that the probability distribution $p_{1,\dots,|\Lambda|}(x_1,\dots,x_{|\Lambda|})>0$ is such that there exists a function $\xi:\mathbb{N}\to\mathbb{R}$, such that for every contiguous partition $\Lambda=ABC$, it is the case that*

$$I_p(A:C|B)\leq\xi(|B|),\tag{27}$$

where ξ is independent of $|A|$ and $|C|$. Let $p^{\ell}_{1,\dots,|\Lambda|}$ be as defined in (21), via (20) and (19). Then, for $1\leq\ell\leq|\Lambda|-2$, we have

$$S(p_{1,\dots,|\Lambda|}\|p^{\ell}_{1,\dots,|\Lambda|})\leq(|\Lambda|-\ell-1)\xi(\ell)\leq|\Lambda|\xi(\ell).\tag{28}$$

Proof. With the general observation that $I(A:C_1|B)\leq I(A:C_1C_2|B)$, we can use $A=\{1,\dots,k\}$, $B=\{k+1,\dots,k+\ell\}$, $C_1=\{k+\ell+1\}$ and $C_2=\{k+\ell+2,\dots,|\Lambda|\}$ in (22) and assumption (27), which yields

$$S(p_{1,\dots,|\Lambda|}\|p^{\ell}_{1,\dots,|\Lambda|})\leq\sum_{k=1}^{|\Lambda|-\ell-1}\xi(\ell)=(|\Lambda|-\ell-1)\xi(\ell).\tag{29}$$

□

Equation (28) estimates the contribution of the tails of the distribution $p_{1,\dots,|\Lambda|}$: the smaller $|\Lambda|\xi(\ell)$ is, the smaller the contribution of these tails. In suitable joint limits of ℓ and $|\Lambda|$, where $|\Lambda|\xi(\ell)$ vanishes exponentially, we say that $p_{1,\dots,|\Lambda|}$ is a quasi-local Gibbs distribution. At first sight it might not be clear whether there exists an exponentially decreasing bound ξ with the necessary properties. However, in Section V, we establish such a bound, when $p_{1,\dots,|\Lambda|}$ is the classical restriction of a large class of injective MPS, thus showing that those classical restrictions are quasi-locally Gibbsian.

V. THE MAIN THEOREM

In Sec. IV, we showed that probability distributions on finite one-dimensional lattices are quasi-locally Gibbsian if the relevant CMI decays sufficiently rapidly. Here, we apply this result to classical restrictions of injective MPSs, i.e., to the the square amplitudes in a given local basis. We express the relevant CMI in terms of the average post-measurement entropy (as discussed in Section III) and find sufficient conditions for when this average entropy decays exponentially. This approach thus yields sufficient conditions for injective MPSs to have classical restrictions that are quasi-locally Gibbsian.

Our result builds extensively on the theory of products of random matrices [32], and its application to quantum trajectories [20, 21]. We particularly follow the approach of Ref. 20 and formulate the condition for the exponential decay of the average post-measurement entropy in terms of the following condition (referred to as **Pur** in Ref. 20) on the matrices A_x associated to the MPS.

Definition 3 (Purity [20])**.** *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a complex finite-dimensional Hilbert space, $\mathscr{H}$. We say that $\{A_x\}_{x=0}^{d-1}$ satisfies the* purity condition *if the following implication holds:*

$$\begin{aligned}&\textit{If } P \textit{ is an orthogonal projector on } \mathscr{H} \textit{ such that}\\ &PA_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P, \quad \forall N \in \mathbb{N}, \quad \forall (x_1,\ldots,x_N) \in \{0,\ldots,d-1\}^{\times N},\\ &\textit{then } \operatorname{rank}(P) = 1.\end{aligned} \tag{30}$$

Note that the condition $PA_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P$ is trivially true whenever P is a rank-one projector. Hence, the purity condition means that $PA_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P$ *only* holds for rank-one projectors. The purity condition bears some resemblance to the Knill-Laflamme condition [26]. We discuss the relationship between the purity condition and error correction/detection in Section VI B.

An immediate question is if the purity condition is commonly satisfied, or if these cases are rare. One can argue that the error-correction perspective in Section VI B suggests that violations of the purity condition are 'brittle', and thus provides evidence for the purity condition being 'generic' or 'typical'. To shed some further light on this question, we do in Section VI C present a somewhat simpler condition that implies purity, and where the nature of this simplified condition suggests that the purity condition in some sense is 'easily' satisfied. As a concrete application and illustration of this simplified condition, Section VI D considers a specific probabilistic setting, where all $\{A_x\}_{x=0}^{d-1}$ satisfy the purity condition, apart from a subset of measure zero. This construction thus formalizes the notion that purity is a typical or generic property.

Even if the purity condition is generic, another pertinent question is whether it is easy or not to *check* if a given MPS, in terms of the operators $\{A_x\}_{x=0}^{d-1}$, satisfies the purity condition. Although an interesting question, we leave this as an open problem.

The primary focus of this investigation is the classical CMI $I_{p_\Psi}(A:C|B)$ of the distribution $p_\Psi = \langle x_\Lambda|\Psi|x_\Lambda\rangle$. Theorem 4, below, shows that purity is a sufficient condition for an exponential decay of $I_{p_\Psi}(A:C|B)$ with increasing $|B|$. However, in order to facilitate a better understanding of the role of the purity condition, Theorem 4 also includes the closely related quantity $I_{\Phi_B(\Psi)}(A:C|B)$, which we recall is the quantum CMI of the post-measurement state $\Phi_B(\Psi)$ as defined in (4). Theorem 4 in essence shows that purity is both necessary and sufficient condition for the exponential decay of $I_{\Phi_B(\Psi)}(A:C|B)$. Since $I_{\Phi_B(\Psi)}(A:C|B)$ can be viewed as the average entanglement entropy of the post-measurement states $\Psi(x_B)$ (which are pure), this loosely speaking means that the latter typically approach pure product states with respect to the bipartition A and C. Whether purity also is a necessary condition for the exponential decay of $I_{p_\Psi}(A:C|B)$ is less clear, although one may note that one can find pure states for which $I_{p_\Psi}(A:C|B) = 0$, while $I_{\Phi_B(\Psi)}(A:C|B) \neq 0$. (For further details, see the end of Section A 2.) With this observation in mind, it is conceivable that there may exist a weaker condition than purity that would yield an exponential decay of $I_{p_\Psi}(A:C|B)$. Although an interesting question, we leave this as an open problem for future investigations.

Theorem 4. *Let Ψ be an injective MPS on a finite one-dimensional lattice, Λ, with finite bond dimension, D, and open boundary conditions. If the purity condition holds for the matrices associated to the MPS corresponding to a specific local basis, $\{|x\rangle\}$, then there exist constants $1 > \kappa \geq 0$ and $c \geq 0$, such that for any three contiguous regions $\Lambda = ABC$ as in Fig. 1, we have*

$$I_{p_\Psi}(A:C|B) \leq I_{\Phi_B(\Psi)}(A:C|B) \leq c\kappa^{|B|}. \tag{31}$$

The constants c and κ are independent of $|A|$, $|B|$, $|C|$, $|L\rangle$, and $|R\rangle$.

Conversely, suppose that there exist some $|R\rangle$, $|L\rangle$, $|A|$, and $|C|$ such that $\sigma := \mathbb{E}^{|A|}(L)$, and $F^\dagger F = \mathbb{E}^{*|C|}(R)$ are full rank operators. Moreover, suppose that there exist constants $c \geq 0$ and $1 > \kappa \geq 0$, such that

$$I_{\Phi_B(\Psi)}(A:C|B) \leq c\kappa^{|B|}, \tag{32}$$

then $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3.

The following provides an overview of the essential steps of the proof of Theorem 4. For a more detailed account of the first half of Theorem 4, i.e., the purity condition as a sufficient condition for (31), see the proof of Theorem 18 in Appendix A 1. For the second half, with (32) implying the purity condition, see the proof of Theorem 24 in Appendix A 2.

Proof. For the first part of Theorem 4, the first step is to bound $I_{\Phi_B(\Psi)}(A:C|B)$ in terms of the quantity $f(N)$, defined below in Eq. (39). Because of the inequality in (12), we consequently also bound the post-measurement CMI $I_{p_\Psi}(A:C|B)$. The second step is to show that $f(N)$ decays exponentially if the purity condition is satisfied; this step is shown independently in Prop. 5. We relegate much of the technical details of the proof to Appendixes A-D to allow for a clearer presentation of the main ideas.

To start with, we bound the average entropy (Eq. (13)) in terms of a quantity that can be interpreted as the average purity and we get

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq -Q\log Q + Q[1+\log(D-1)], \tag{33}$$

with $Q := 1-\sum_{x_B} p_\Psi(x_B)\|\Psi_C(x_B)\|$. The proof, which is deferred to Lemma 13 in Appendix A, follows from concavity of the entropy functional. It is clear that exponential decay of Q implies exponential decay of $I_{\Phi_B(\Psi)}(A:C|B)$ by Eq. (12).

Next, we show that Q can be bounded above by a function of the ordered singular values of the matrix product defining the classical post-measurement MPS. First, Lemma 14 in Appendix A establishes an upper bound on Q in terms of the average second eigenvalue of the matrix product in Eq. (15) as

$$Q \leq \frac{D-1}{K^2}\sum_{x_B}\lambda_2^{\downarrow}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right), \tag{34}$$

where $\{\lambda_j^{\downarrow}(O)\}$ and $\{\nu_j^{\downarrow}(O)\}$ denote the eigenvalues and singular values of an operator O in decreasing order, i.e., $\lambda_1^{\downarrow}(O)\geq\cdots\geq\lambda_D^{\downarrow}(O)$ and $\nu_1^{\downarrow}(O)\geq\cdots\geq\nu_D^{\downarrow}(O)$, respectively.

Then, recalling that for any operator O, we have $\lambda_j(OO^\dagger)=\nu_j(O)^2$, we get

$$\begin{aligned} Q &\leq \frac{D-1}{K^2}\sum_{x_B}\lambda_2^{\downarrow}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right),\\ &\leq \frac{D-1}{K^2}\sum_{x_B}\sqrt{\lambda_1^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger)\lambda_2^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger)},\\ &= \frac{D-1}{K^2}\sum_{x_B}\nu_1^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})\nu_2^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma}) =: \frac{D-1}{K^2}f(N), \end{aligned} \tag{35}$$

where recall that $|B|=N$.

Next, we need to take into account the fact that K depends on the size of the regions A, B and C, and in principle K could approach zero. However, the assumption that the MPS is injective, implies that $\mathbb{E}(\cdot)=\sum_x A_x\cdot A_x^\dagger$ is primitive, which means that $\mathbb{E}$ has a unique full-rank fixed point. The latter is used in Lemma 17 in Appendix A to show that for all sufficiently large $|B|$ there exists a number $r>0$ such that

$$K^2 = \langle R|\mathbb{E}^{|\Lambda|}(|L\rangle\langle L|)|R\rangle = \langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle \geq r, \tag{36}$$

where r is independent of $|A|$, $|C|$, $|L\rangle$ and $|R\rangle$. We use this to obtain an upper bound on Q that only depends on N via $f(N)$.

Finally, in Proposition 5 below, the function $f(N)$ is shown to decay exponentially if the purity condition holds. Moreover, the constants $\bar{c}$ and γ in the bound (40) can be chosen to be independent of $|A|$, $|B|$, $|C|$, which follows from the fact that $\bar{c}$ and γ are independent of σ and F.

The proof of the first part of Theorem 4, requires us to find an upper bound of the average entropy $\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}$ in terms of the quantity $f(N)$. For the second part of Theorem 4, i.e., that (32) implies the purity condition, we instead need to find an upper bound to $f(N)$ in terms of $\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}$. We obtain this via a chain of inequalities

$$4\log(2)\lambda_2^{\downarrow}(\rho)\lambda_1^{\downarrow}(\rho) \leq 4\log(2)\lambda_1^{\downarrow}(\rho)\big(1-\lambda_1^{\downarrow}(\rho)\big) \leq H_B\big(\lambda_1^{\downarrow}(\rho)\big) \leq S(\rho), \tag{37}$$

where H_B is the binary entropy, i.e., $H_B(\lambda) := -\lambda\log\lambda-(1-\lambda)\log(1-\lambda)$ with $H_B(0):=0$ and $H_B(1):=0$. These observations are utilized to show that

$$f(N) \leq \frac{1}{2\sqrt{\log(2)}}\sqrt{\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}}, \tag{38}$$

with the consequence that an exponential decay of $\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}$ with increasing N, implies an exponential decay of $f(N)$. By Prop. 5, the exponential decay of $f(N)$ implies purity of $\{A_x\}_{x=0}^{d-1}$, if $\sigma := \mathbb{E}^{|A|}(L)$, and $F^\dagger F = \mathbb{E}^{*|C|}(R)$ are full rank operators. □

Note that the bound in Eq. (35) is likely quite sub-optimal. It is an interesting open question whether there exists a more direct bound of the average purity that does not rely on bounding the function $f(N)$. The main reason to work with $f(N)$ rather than the average purity is because $f(N)$ is explicitly submultiplicative.

We now state the key proposition adapted from Ref. 20, and references therein.

Proposition 5 (Ref. 20)**.** *Let $\{A_x\}_{x=0}^{d-1}$ be operators on a finite-dimensional complex Hilbert space, $\mathscr{H}$, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. For an operator σ and an operator $F : \mathscr{H} \to \mathscr{H}'$ for a finite-dimensional complex Hilbert space $\mathscr{H}'$, define*

$$f(N) := \sum_{x_1,\ldots,x_N=0}^{d-1} \nu_1^\downarrow(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})\nu_2^\downarrow(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma}). \tag{39}$$

If $\{A_x\}_{x=1}^{d-1}$ satisfies the purity condition in Definition 3, then there exist real constants, $0 \leq \bar{c}$ and $0 < \gamma < 1$, such that for all density operators σ, and all F such that $F^\dagger F \leq \mathbb{1}$, it is the case that

$$f(N) \leq \bar{c}\gamma^N, \quad \forall N \in \mathbb{N}. \tag{40}$$

Conversely, if there exists constants $0 \leq \bar{c}$ and $0 < \gamma < 1$ such that (40) holds for some σ and $F^\dagger F$ that both are full-rank operators, then $\{A_x\}_{x=1}^{d-1}$ satisfies the purity condition.

In the application of this proposition, the Hilbert space $\mathscr{H}$ is the virtual space, while $\mathscr{H}'$ is the Hilbert space corresponding to sub-chain C, c.f., the definition of F in (16). Theorem 4 provides the necessary bound $\xi(|B|) = c\kappa^{|B|}$ in Lemma 2 for showing the quasi-locality of the classical restriction $p_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) = \langle x_\Lambda|\Psi|x_\Lambda\rangle$. Theorem 4 and Lemma 2 thus yield as a corollary (for a more exact formulation, see Corollary 19 in Appendix A)

$$S(p_{1,\ldots,|\Lambda|}\|p^\ell_{1,\ldots,|\Lambda|}) \leq c|\Lambda|\kappa^\ell, \quad 1 \leq \ell \leq |\Lambda| - 2. \tag{41}$$

A simple example that leads to an exponential decay of the relative entropy is if ℓ is a constant fraction of $|\Lambda|$, i.e.,

$$\ell = \alpha|\Lambda|, \quad 0 < \alpha < 1. \tag{42}$$

The result is that the relative entropy decays exponentially, and the family of ℓ-local distributions thus approaches $p_{1,\ldots,|\Lambda|}$ exponentially fast. The classical restriction of typical injective MPSs is thus in this sense quasi-locally Gibbsian.

A. Proof overview for Proposition 5

Here, we give a brief overview of the general structure and ideas behind the proof of Proposition 5, i.e., that $f(N)$ decays exponentially if $\{A_x\}_x$ satisfies the purity condition. Although we do not always follow the exact same tracks, the essence of the proof is due to Refs. 20 and 21, which we have adapted to our particular setting and cast in a language that is hopefully more accessible to the quantum information theory community. The proof is essentially self-contained, and is presented in Appendices B, C and D, only referencing some standard results from the theory of Martingales (see Appendix B), that can be found in a number of classic textbook on the subject.

As one may note from Eq. (39), the sequence $f(N)$ not only depends on the operators A_x, but also on the operators σ and F. It turns out to be convenient to first focus on the function

$$w(N) = \sum_{x_1,\ldots,x_N=0}^{d-1} \nu_1^\downarrow(A_{x_N}\cdots A_{x_1})\nu_2^\downarrow(A_{x_N}\cdots A_{x_1}). \tag{43}$$

Once we have established the purity condition as a necessary and sufficient condition for exponential decay of $w(N)$, we extend (Proposition 43 in Section D 5) this result to $f(N)$, which thus yields the statement of Proposition 5.

The proof of the exponential convergence of $w(N)$ is essentially done in two steps. First, it is shown that $w(N)$ converges to zero. Next, it is shown that $w(N)$ is submultiplicative, in the sense that $w(N+M) \leq w(N)w(M)$, and thus $\log w(N)$ is subadditive. This observation is used, together with Fekete's subadditive lemma, to show that $w(N)$ goes to zero exponentially fast. These steps are incorporated into the proof of Proposition 42.

The essential approach for proving that $w(N)$ converges to zero is to interpret $w(N)$ as the average over a stochastic process. This process can be viewed as the random measurement outcomes $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ due to a repeated sequential measurement of the positive operator-valued measure (POVM) $\{A_x^\dagger A_x\}_{x=0}^{d-1}$. (This process is described more precisely in Appendix C.) For the proof, it is useful to introduce the operator

$$M_N = \frac{A^\dagger_{\boldsymbol{x}_1}\cdots A^\dagger_{\boldsymbol{x}_N}A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1}}{\mathrm{Tr}(A^\dagger_{\boldsymbol{x}_1}\cdots A^\dagger_{\boldsymbol{x}_N}A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1})}, \tag{44}$$

which thus depends on the sequence of random measurement outcomes $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$. It turns out that one can express $w(N)$ in terms of $\boldsymbol{M}_N$ via the relation $w(N) = E\big(\sqrt{\lambda_1^{\downarrow}(\boldsymbol{M}_N)\lambda_2^{\downarrow}(\boldsymbol{M}_N)}\big)D$, where $\lambda_1^{\downarrow}(\boldsymbol{M}_N)$ and $\lambda_2^{\downarrow}(\boldsymbol{M}_N)$ denote the largest and the second largest eigenvalue of $\boldsymbol{M}_N$, respectively, and D the dimension of the underlying Hilbert space. Moreover, E denotes the expectation value over all possible measurement outcomes. One can realize that $\boldsymbol{M}_N$ is positive semi-definite, has trace 1, and can thus be interpreted as a density operator. The main point is that if $\boldsymbol{M}_N$ would be a rank-one operator, and thus correspond to a pure state, then it follows that $\lambda_2^{\downarrow}(\boldsymbol{M}_N)$ is zero. Intuitively, it thus seems reasonable that $w(N)$ converges to zero if it is 'sufficiently likely' that $\boldsymbol{M}_N$ converges to a rank-one operator.

The starting point for demonstrating that $\boldsymbol{M}_N$ converges to a rank-one operator is to show (Lemma 35) that the sequence $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ is a martingale relative to the sequence of measurement outcomes $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$. This enables us to show (Lemma 36) that $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ almost surely converges to a positive operator $\boldsymbol{M}_\infty$. (All these notions are reviewed in Appendix B.) Once this is established, the bulk of the proof is focused on showing that $\boldsymbol{M}_\infty$ (almost surely) is a rank-one operator if and only if $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition.

The arguable least transparent part of the proof is how to show that the purity condition is sufficient for $\boldsymbol{M}_\infty$ to be a rank-one operator. The first part of the proof (Lemma 37) shows that $\boldsymbol{M}_{N+p}$ and $\boldsymbol{M}_N$ in some sense 'approach' each other, even when conditioned on $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$. The second part (Lemma 39) loosely speaking shows that $\boldsymbol{M}_{N+p}$ gives rise to a term of the form $\sqrt{\boldsymbol{M}_N}\boldsymbol{U}_N^\dagger A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_N\sqrt{\boldsymbol{M}_N}$ for a unitary operator, $\boldsymbol{U}_N$, while $\boldsymbol{M}_N$ gives rise to a term that is proportional to $\boldsymbol{M}_N$. As these operators approach each other when N approaches infinity, one can use this to show that

$$\boldsymbol{M}_\infty\boldsymbol{U}_\infty^\dagger A_{x_1'}^\dagger\cdots A_{x_p'}^\dagger A_{x_p'}\cdots A_{x_1'}\boldsymbol{U}_\infty\boldsymbol{M}_\infty \propto \boldsymbol{M}_\infty\boldsymbol{U}_\infty^\dagger\boldsymbol{U}_\infty\boldsymbol{M}_\infty. \tag{45}$$

In a reformulation (Lemma 38) of the purity condition, the projector, P, is replaced by a general operator, O, again with the conclusion that O must be a rank-one operator. With $O = \boldsymbol{U}_\infty\boldsymbol{M}_\infty$ it follows that $\boldsymbol{M}_\infty$ is a rank-one operator.

To conversely show (Lemma 40) that the purity condition is a necessary condition is somewhat less involved. By assuming that a projector P satisfies the proportionality in (30) while having a rank larger than one, then it follows that the only way in which $\boldsymbol{M}_\infty$ can be a rank-one operator, is if $P\boldsymbol{M}_\infty P = 0$. However, this leads to a contradiction with A_x being such that $\sum_{k=1}^{L} A_x^\dagger A_x = \mathbb{1}$.

Remark. Using the same tools as above, Benoist et. al. show in Ref. 20 that the stochastic process defined in Appendix C equilibrates exponentially. It is worth noting that the average purity can converge to zero much faster than the stochastic process. For instance, if σ is a rank-one operator, then it trivially follows that $f(N)$ is identically zero for all N, irrespective of whether $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition or not.

VI. DISCUSSIONS AND EXAMPLES

In this section, we discuss the purity condition and the decay of the CMI in the context of quantum information theory. In Sec. VI A we specifically study the behaviour of the CMI for SPT phases and obtain that it remains constant. The purity condition is discussed from the point of view of quantum error correction in Sec. VI B. In Sec. VI C we find a simpler condition that implies the purity condition, and based on this simplified condition we discuss the typicality of the purity condition in Sec. VI D. In Sec. VI E we show, by constructing two simple examples, that the decay rate of the CMI is unrelated to the decay of the transfer operator of the corresponding MPS. Finally, in Sec. VI F we discuss some concrete examples.

A. Symmetry-protected phases

Here we briefly discuss systems that do not satisfy the purity condition, and comment on the relation to SPT phases in one dimension.

Consider an MPS, $|\Psi\rangle$, of the form of Eq. (1) with matrices A_{x_i} that have a tensor product decomposition into two subsystems such that

$$A_{x_i} = U_{x_i}\otimes T_{x_i}, \tag{46}$$

where U_{x_i} is a unitary matrix and T_{x_i} is any matrix. Then, the reduced post-measurement state, $\Psi_C(x_B)$, in the infinite chain case (see Eq. (17)) is isospectral to

$$\Psi_C(x_B) \simeq \frac{1}{p_\Psi(x_B)K^2}\left(U_{x_N}\cdots U_{x_1}\otimes T_{x_N}\cdots T_{x_1}\right)\rho\left(U_{x_1}^\dagger\cdots U_{x_N}^\dagger\otimes T_{x_1}^\dagger\cdots T_{x_N}^\dagger\right).$$

For simplicity, let us further consider the case where the unique fixed point of the transfer operator is proportional to the identity, i.e., $\rho = \mathbb{1}/(D_1 D_2)$, where D_1 and D_2 are the dimensions of the two sub-systems respectively. We obtain

$$\Psi_C(x_B) \simeq \frac{1}{D_1 D_2 p_\Psi(x_B) K^2} \left(\mathbb{1} \otimes T_{x_N} \cdots T_{x_1} T_{x_1}^\dagger \cdots T_{x_N}^\dagger\right).$$

The von Neumann entropy of this state has two independent contributions coming from each factor of the tensor product, namely

$$S[\Psi_C(x_B)] = \log D_1 + S\left(\frac{1}{p_\Psi(x_B) D_2 K^2} T_{x_N} \cdots T_{x_1} T_{x_1}^\dagger \cdots T_{x_N}^\dagger\right).$$

Consequently, the average entropy of entanglement of $\Psi_C(x_B)$, and thus the post-measurement CMI, $I_{\Phi_B(\Psi)}(A:C|B)$ (see Eq. (12)), always has a constant contribution independent of the length of the middle region B. More generally, the CMI is non-vanishing for MPSs in a basis where the matrices A_{x_i} can be isometrically mapped to a form as in Eq. (46) [25].

It was shown in Ref. 27 that, for a SPT phase in the MPS framework, there always exists a local basis in which the matrices have the form of Eq. (46), with the additional property that the unitary matrices form a representation of the symmetry group. The AKLT model (see Sec. VI F 1) is such an example.

B. The purity condition: Relation to error correction

In this section we will explore the purity condition (see Def. 3) in more detail. The purity condition states that the only projectors P that satisfy

$$P A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P, \tag{47}$$

for all $N \in \mathbb{N}$, and all $(x_1, \ldots, x_N) \in \{0, \ldots, d-1\}^{\times N}$, are those that have rank one. Here we investigate the relation between this condition (or rather the violation of it) and the Knill-Laflamme error correction condition [26].

Suppose that there exits a projector, P, onto a subspace, $\mathscr{C}$, with $\dim \mathscr{C} \geq 2$ such that

$$P A_x^\dagger A_x P = \lambda_x P, \quad \forall x. \tag{48}$$

This looks suspiciously similar to the Knill-Laflamme error correction condition, which is

$$P A_x^\dagger A_y P = c_{xy} P, \quad \forall x, y. \tag{49}$$

The question is how one can understand the apparent similarity between Eq. (48) and (49). To this end, let us first recall the error correction scenario. If A_x are operators on a Hilbert space, $\mathscr{H}$, with $\sum_x A_x^\dagger A_x = \mathbb{1}$, we define the corresponding noise channel

$$\mathbb{E}(\chi) := \sum_x A_x \chi A_x^\dagger. \tag{50}$$

For any state, χ, with support on the subspace $\mathscr{C} \subseteq \mathscr{H}$, it is the case that $\mathbb{E}(\chi)$ can be restored to χ if and only if (49) is true. More precisely, there exists a recovery operation, $\mathscr{R}$, (that does not depend on χ) such that $\mathscr{R} \circ \mathbb{E}(\chi) = \chi$ for all density operators χ with support on $\mathscr{C}$.

It turns out that Eq. (48) is also a necessary and sufficient condition for error correction, but for a different type of error-model. The channel $\mathbb{E}$, in the standard error-correction scenario, is the effect of a unitary evolution that acts on $\mathscr{H}$ and on an environment, $\mathscr{H}_E$, where the latter is inaccessible to us. In the alternative scenario, we assume that there exists an ancillary system, A, which we do have access to, and which we can use in order to help us restore the initial state on $\mathscr{H}$. More precisely, we assume an error model of the form

$$\tilde{\mathbb{E}}(\chi) = \sum_x |x\rangle_A \langle x| \otimes A_x \chi A_x^\dagger, \tag{51}$$

where $\{|x\rangle_A\}_l$ is an orthonormal basis of the Hilbert space associated to the ancillary system, $\mathscr{H}_A$. We can interpret this as having access to additional classical information about the error in the register, A. We use this additional information in order to restore the state on $\mathscr{C}$. One may note that if we have no access to A, then we are back to the standard scenario, where the channel on $\mathscr{H}$ is $\mathbb{E} = \mathrm{Tr}_A \tilde{\mathbb{E}}$. It turns out that (48) is a necessary and sufficient condition for the existence of a recovery channel $\tilde{\mathscr{R}} : \mathscr{L}(\mathscr{H} \otimes \mathscr{H}_A) \to \mathscr{L}(\mathscr{H})$, such that $\tilde{\mathscr{R}} \circ \tilde{\mathbb{E}}(\chi) = \chi$ for all density operators χ on $\mathscr{C}$. The proof of this statement is nearly identical to that of the original Knill-Laflamme theorem and is omitted here.

If one finds a non-trivial projector P (i.e. if $\mathrm{Tr}(P) = \dim \mathscr{C} \geq 2$) such that Eq. (48) holds, then one can explicitly construct a collection of unitary operators, U_x, such that

$$\sum_x U_x A_x \chi A_x^\dagger U_x^\dagger = \chi, \tag{52}$$

for all density operators χ on $\mathscr{C}$. In other words, the operators U_x perform the error correction on subspace $\mathscr{C}$. More precisely, if we have a set $\{A_x\}$ with $\sum_x A_x^\dagger A_x = \mathbb{1}$, for which there exists a non-trivial projector, P, that satisfies Eq. (48), then we can construct a new 'error-corrected' set, $\{\overline{A}_x\}$, with $\overline{A}_x := U_x A_x$ (and $\sum_x \overline{A}_x^\dagger \overline{A}_x = \mathbb{1}$). For this new set we will thus not get a decay to zero of the average entropy (Eq. (15)), no matter how long a chain $\overline{A}_{x_N} \cdots \overline{A}_{x_1}$ we construct.

Nothing prevents us from repeating the above reasoning for products $\{A_{x_2} A_{x_1}\}_{x_2,x_1}$, i.e., we can try to find the largest subspace $\mathscr{C}_2$ with corresponding projector, P, such that

$$P A_{x_1}^\dagger A_{x_2}^\dagger A_{x_2} A_{x_1} P = \lambda_{x_2,x_1} P. \tag{53}$$

We can similarly ask for the largest subspace $\mathscr{C}_3$ that is correctable for $\{A_{x_3} A_{x_2} A_{x_1}\}_{x_3,x_2,x_1}$. One can realize that we always have $\mathscr{C}_n \subseteq \mathscr{C}_{n-1}$.

The purity condition is violated if and only if there exists a non-trivial projector P such that (47) holds for all N. By the above reasoning we can thus conclude that the purity condition fails if and only if there for all N exists a fixed non-trivial correctable subspace $\mathscr{C}$. Loosely speaking, we can alternatively phrase the purity condition as the non-existence of a non-trivial correctable subspace that persists indefinitely throughout iterated applications of the error channel. Intuitively, this observation suggests that the violation of the purity condition is a rather 'brittle' and non-generic phenomenon.

C. A sufficient condition for purity

It is maybe not entirely clear what is the deeper meaning of the purity-condition, or how easy or difficult it is to satisfy. In order to shed some light on the latter question, we here show that if there exists some N for which the set of operators $A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}$ span the space of linear operators $\mathscr{L}(\mathscr{H})$ on the underlying (finite-dimensional) Hilbert space $\mathscr{H}$, then $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition. Another way of phrasing this is to say that if for some N, the POVM $\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N}$ is informationally complete, then $\{A_x\}_{x=0}^{d-1}$ satisfies purity.

In the general case, it seems intuitively reasonable to expect that the set of products $A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}$ eventually spans the whole of $\mathscr{L}(\mathscr{H})$, for sufficiently large N (assuming linear combinations with complex coefficients). The exception would be if there exists some particular algebraic relation between the operators A_x, which so to speak 'trap' the products $A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}$ within a nontrivial subspace of $\mathscr{L}(\mathscr{H})$. This argument suggests that the purity condition in some sense would be 'easily' satisfied. We investigate this question further in Section VI D.

Let us first note that a set of operators $\{A_x\}_{x=0}^{d-1}$ does *not* satisfy the purity condition if there exists a projector P onto an at least two-dimensional subspace of $\mathscr{H}$, and there exist numbers $r_{x_1,\ldots,x_N}$ such that

$$P A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P = r_{x_1,\ldots,x_N} P, \tag{54}$$

for all $N \in \mathbb{N}$ and all $x_1, \ldots, x_N$.

Proposition 6. *Let $\{A_x\}_{x=0}^{d-1}$ be operators on the finite-dimensional complex Hilbert space $\mathscr{H}$, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. If there exists an $N \in \mathbb{N}$ such that*

$$\mathrm{Sp}\Big(\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N=0}^{d-1}\Big) = \mathscr{L}(\mathscr{H}), \tag{55}$$

then $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition.

Proof. It turns out to be convenient to prove that if $\{A_x\}_{x=0}^{d-1}$ does *not* satisfy the purity condition, then $\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N=0}^{d-1}$ does not span $\mathscr{L}(\mathscr{H})$ for any N. We thus assume that there exists a projector onto an at least two-dimensional subspace, such that (54) is satisfied for all N, and all $x_1, \ldots, x_N$. Since P projects onto an at least two-dimensional subspace, there exists an operator Q such that $PQP = Q$, and where $Q \neq cP$ for all $c \in \mathbb{C}$. Assume that $\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N}$ would span the whole of $\mathscr{L}(\mathscr{H})$. Then, there exist $c_{x_1,\ldots,x_N} \in \mathbb{C}$ such that

$$\sum_{x_1,\ldots,x_N} c_{x_1,\ldots,x_N} A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} = Q. \tag{56}$$

This in turn implies that

$$\sum_{x_1,\ldots,x_N} c_{x_1,\ldots,x_N} P A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P = PQP = Q. \tag{57}$$

However, by (54) we know that

$$\sum_{x_1,\ldots,x_N} c_{x_1,\ldots,x_N} P A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P = \sum_{x_1,\ldots,x_N} c_{x_1,\ldots,x_N} r_{x_1,\ldots,x_N} P. \tag{58}$$

This combined with (57) yields

$$\sum_{x_1,\ldots,x_N} c_{x_1,\ldots,x_N} r_{x_1,\ldots,x_N} P = Q. \tag{59}$$

However, this is in contradiction with $Q \neq cP$ for all $c \in \mathbb{C}$. Hence, we can conclude that $\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N}$ cannot span the whole of $\mathscr{L}(\mathscr{H})$.

To conclude, if $\{A_x\}_{x=0}^{d-1}$ does not satisfy the purity condition, then $\{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N}$ cannot span the whole of $\mathscr{L}(\mathscr{H})$ for any N. This yields the statement of the Lemma. □

D. A model for typicality of purity

In this section, we consider a concrete model for formalizing the notion of typicality of the purity condition. A common method is to assign a probability measure over the set under consideration, and say that a property is typical, or generic, if it holds for all elements in that set, apart from a subset of measure zero. This approach thus requires us to construct a probability measure over the objects $\{A_x\}_{x=0}^{d-1}$. Within this construction, we will use Proposition 6, and a result from the previous literature (Lemma 7 below), to show that the purity condition is satisfied generically. As the reader will note, we here only present a construction for Hilbert spaces of odd dimensions. The reason for why we impose this restriction is to avoid the additional technical complications that arise in the even-dimensional case (briefly explained below). It seems likely that these complications are due to the particular proof-technique that we use, rather than some genuine limitations.

In order to construct a probability measure on the sets $\{A_x\}_{x=0}^{d-1}$, we consider, apart from the Hilbert space $\mathscr{H}$, also an ancillary Hilbert space, $\mathscr{H}_A$, of dimension d. On $\mathscr{H}_A$, we fix an orthonormal basis, $\{|a_x\rangle\}_{x=0}^{d-1}$, and a normalized element, $|a\rangle \in \mathscr{H}_A$. For each unitary operator, U, on $\mathscr{H} \otimes \mathscr{H}_A$, we let

$$A_x := \langle a_x|U|a\rangle, \tag{60}$$

where we note that since U is a mapping on $\mathscr{H} \otimes \mathscr{H}_A$, it follows that A_x is a mapping on $\mathscr{H}$. We can regard this as the result of a procedure where we append an ancillary state, $|a\rangle\langle a|$, to an input state, ρ, evolve the system unitarily with U, and then perform the projective measurement $\{|a_x\rangle\langle a_x|\}_{x=0}^{d-1}$ on the ancillary system. The conditional (unnormalized) post-measurement state resulting from this procedure is $A_x \rho A_x^\dagger$. If we consider the Haar measure over the set of $Dd \times Dd$ unitary matrices U, the construction in (60) thus induces a probability measure on the class of sets $\{A_x\}_{x=0}^{d-1}$. In the following, we shall argue that, with respect to the Haar measure, the set of $\{A_x\}_{x=0}^{d-1}$ that satisfy the purity condition is typical, in the sense that the set that violates the purity condition has measure zero. To reach this conclusion, we make use of the following result, which we have taken from [33]. Consider some polynomial P with real coefficients, over the real and imaginary parts of the elements of complex $K \times K$ matrices. The following lemma says that there are only two possibilities: either P is zero on the whole set of unitary $K \times K$ matrices, $\mathbb{U}(K)$, or P is non-zero on almost all of $\mathbb{U}(K)$.

Lemma 7 (Lemma 4.3 in [33])**.** *Given a polynomial $P \in \mathbb{R}[X_1, \ldots, X_{2K^2}]$, the set*

$$\Big\{[U_{i,j}]_{i,j=1}^K \in \mathbb{U}(K) : P\big(Re(U_{i,j}), Im(U_{i,j})\big) = 0\Big\}, \tag{61}$$

is either equal to the whole of $\mathbb{U}(K)$, or it has Haar measure 0*.*

This means that if we can find a single unitary U for which the polynomial is non-zero, then we know that the polynomial is non-zero for the whole set $\mathbb{U}(K)$, except possibly for a subset of measure zero.

Regarding the space $\mathscr{L}(\mathscr{H})$ as an inner product space with respect to the Hilbert-Schmidt inner product $\langle B, C\rangle := \mathrm{Tr}(B^\dagger C)$, we note that a finite collection of operators $\mathscr{Q} := \{Q_x\}_{x=0}^{K-1}$ is linearly independent if and only if the Gram matrix

$$\boldsymbol{M}(\mathscr{Q}) = [M_{x,x'}]_{x,x'=0}^{K-1}, \quad M_{x,x'} := \langle Q_x, Q_{x'}\rangle = \mathrm{Tr}(Q_x^\dagger Q_{x'}), \tag{62}$$

is positive definite. Since $\boldsymbol{M}(\mathscr{Q})$ in general is positive semi-definite, we thus know that $\mathscr{Q}$ is linearly independent if and only if all the eigenvalues of $\boldsymbol{M}(\mathscr{Q})$ are non-zero, and consequently, if and only if $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$.

In the following we shall prove that $\{A_x\}_{x=0}^{d-1}$, as constructed via (60), satisfies the purity condition for all U, except for a subset of Haar measure zero. The general idea of the proof is as follows. For a sufficiently large N, we consider a specific subset $\mathscr{Q} \subset \{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N=0}^{d-1}$, and we note that $\det \boldsymbol{M}(\mathscr{Q})$ is a polynomial in the matrix elements of U. By Lemma 7, we can thus conclude that either $\det \boldsymbol{M}(\mathscr{Q}) = 0$ on the whole of $\mathbb{U}(D)$, or $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$ for all U except for a subset of measure zero. If we moreover let $\mathscr{Q}$ contain precisely D^2 elements, $|\mathscr{Q}| = D^2$, then this would mean that either $\mathscr{Q}$ does not span $\mathscr{L}(\mathscr{H})$ for any U, or $\mathscr{Q}$ spans $\mathscr{L}(\mathscr{H})$ for almost all U. To show the latter, it thus suffices to find one single unitary U such that $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$. The bulk of the proof below is focused on determining such a unitary, and subset $\mathscr{Q}$, yielding $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$.

An important building block for the construction of the particular unitary operator is the set of generalized Pauli-operators, or shift and clock operators [34–37]. On a Hilbert space $\mathscr{H}$ with dimension D, and an orthonormal basis $\{|n\rangle\}_{n=0}^{D-1}$, we define the operators

$$\Lambda_1 := \sum_{n=0}^{D-1} |(n+1) \mathrm{mod} D\rangle\langle n|, \quad \Lambda_3 := \sum_{n=0}^{D-1} \omega^n |n\rangle\langle n|, \quad \omega := e^{i2\pi/D}, \quad U_{jk} := \Lambda_1^j \Lambda_3^k. \tag{63}$$

(The reason for the, at first sight maybe odd-looking, numbering in the subscripts is that Λ_1 can be regarded as the counterpart to the Pauli-operator σ_1, and Λ_3 the counterpart to σ_3.) We recall that the set of unitary operators $\{U_{jk}\}_{j,k=0}^{d-1}$ forms a basis for $\mathscr{L}(\mathscr{H})$, and that $\langle U_{jk}, U_{j'k'}\rangle = \mathrm{Tr}(U_{jk}^\dagger U_{j'k'}) = D\delta_{jj'}\delta_{kk'}$, and moreover that $\Lambda_3\Lambda_1 = \omega\Lambda_1\Lambda_3$, which in turn leads to

$$U_{j'k'} U_{jk} = \omega^{k'j - j'k} U_{jk} U_{j'k'}. \tag{64}$$

We moreover note that

$$U_{jk}^\dagger = \omega^{jk} U_{(D-j) \mathrm{mod} D, (D-k) \mathrm{mod} D}. \tag{65}$$

Let us here recall that we want a subset $\mathscr{Q}$ made of positive semi-definite operators, Q_x, such that $\det M(\mathscr{Q}) \neq 0$, and thus $\mathscr{Q}$ is linearly independent. For this purpose, we will in the following consider a decomposition (Lemma 8) of Hermitian operators, R, in terms of the operators $\{U_{jk}\}_{j,k=0}^{D-1}$. Next, we pick an operator R (Lemma 10) with particular properties, which we will use for the construction of $\mathscr{Q}$. More precisely, in the following we wish to find a positive semi-definite operator, R, that is bounded by the identity, and has non-zero overlaps with all the operators U_{jk}, i.e., all the expansion coefficients in the $\{U_{jk}\}_{j,k=0}^{D-1}$ basis should be non-zero. By the virtue of being positive semi-definite, the hypothetical operator R has to be Hermitian, i.e., $R^\dagger = R$. For this reason, it is useful to take a closer look on the effect, as described by (65), of the Hermitian conjugation on the basis elements U_{jk}. Suppose now that $D \geq 3$ is an odd number. This means that we can partition the index set $\{0,\ldots,D-1\}$ into the three subsets $\{0\}$, $\{1,\ldots,(D-1)/2\}$, $\{(D+1)/2,\ldots,D-1\}$. Under the mapping $n \mapsto (D-n) \bmod D$ the set $\{0\}$ is mapped to itself, while the two sets $\{1,\ldots,(D-1)/2\}$ and $\{(D+1)/2,\ldots,D-1\}$ are mapped into each other. In view of (65) one can thus conclude that every Hermitian operator is uniquely determined by the expansion coefficients corresponding to, e.g., the basis elements

$$U_{00}, \quad \{U_{n,0}\}_{n=1}^{(D-1)/2}, \quad \{U_{0,m}\}_{m=1}^{(D-1)/2}, \quad \{U_{n,m}\}_{n=1,m=1}^{n=(D-1)/2,m=(D-1)/2}, \quad \{U_{n,m}\}_{n=1,m=(D+1)/2}^{n=(D-1)/2,m=D-1}, \tag{66}$$

while the expansion coefficients of the remaining basis-elements are fixed by the Hermiticity and the resulting map (65). By the mapping (65) it also follows that for a Hermitian operator, the expansion coefficient corresponding to U_{00} has to be real. By the above consideration, we can conclude the following lemma.

Lemma 8. *For a finite-dimensional complex Hilbert space with odd dimension $D \geq 3$, every Hermitian operator R can be uniquely expanded as*

$$\begin{aligned} R =& rU_{00} + \sum_{n=1}^{(D-1)/2} [a_n U_{n,0} + a_n^* U_{D-n,0}] + \sum_{m=1}^{(D-1)/2} [b_m U_{0,m} + b_m^* U_{0,D-m}] \\ &+ \sum_{n=1}^{(D-1)/2} \sum_{m=1}^{(D-1)/2} [A_{n,m} U_{n,m} + A_{n,m}^* \omega^{nm} U_{D-n,D-m}] \\ &+ \sum_{n=1}^{(D-1)/2} \sum_{m=(D+1)/2}^{D-1} [B_{n,m} U_{n,m} + B_{n,m}^* \omega^{nm} U_{D-n,D-m}], \end{aligned} \tag{67}$$

where $r \in \mathbb{R}$ and $(a_n)_{n=1}^{(D-1)/2}, (b_m)_{m=1}^{(D-1)/2} \in \mathbb{C}^{(D-2)/2}$ and $(A_{n,m})_{n=1,m=1}^{n=(D-1)/2,m=(D-1)/2}, (B_{n,m})_{n=1,m=(D+1)/2}^{n=(D+1)/2,m=D-1} \in \mathbb{C}^{(D-2)/2 \times (D-1)/2}$. Moreover, by choosing the above coefficients to be non-zero, it follows that all the expansion coefficients of R in the basis $\{U_{jk}\}_{j,k=0}^{D-1}$ are non-zero.

Remark: The even-dimensional case is more involved since in this case, not only 0 is invariant under the map $n \mapsto (D-n) \mathrm{mod} D$, but also $D/2$. For this reason we here focus on the more straightforward odd-dimensional case.

The next step, in Lemma 10 below, consists of choosing coefficients r, a_n, b_m, $A_{n,m}$, and $B_{n,m}$ to obtain an operator R with the desired properties to construct $\mathscr{Q}$. However, in order to prove this lemma we first need the following observation, which we state without proof.

Lemma 9. *Let U be a unitary operator on a finite-dimensional Hilbert space, and θ a real number, then*

$$0 \leq \frac{1}{2}\mathbb{1} + \frac{1}{4}e^{i\theta}U + \frac{1}{4}e^{-i\theta}U^\dagger \leq \mathbb{1}. \tag{68}$$

Lemma 10. *On every finite-dimensional Hilbert space of odd dimension $D \geq 3$, there exists an operator R, such that $\mathbb{1} \geq R \geq 0$, and $\mathrm{Tr}(U_{j,k}^\dagger R) \neq 0$ for all $j,k \in \{0,\ldots,D-1\}$, where U_{jk} are as defined in (63).*

Proof. By combining Lemma 8 with Lemma 9, and the observation that $U_{D-n,0} = U_{n,0}^\dagger$, $U_{0,D-m} = U_{0,m}^\dagger$ and $\omega^{nm}U_{D-n,D-m} = U_{n,m}$, we find that we can obtain an R with the desired properties, if we choose $(a_n)_{n=1}^{(D-1)/2}, (b_m)_{m=1}^{(D-1)/2} \in \mathbb{C}^{(D-2)/2}$ and $(A_{n,m})_{n=1,m=1}^{n=(D-1)/2,m=(D-1)/2}, (B_{n,m})_{n=1,m=(D+1)/2}^{n=(D+1)/2,m=D-1} \in \mathbb{C}^{(D-2)/2\times(D-1)/2}$ and $r \in \mathbb{R}$ in Lemma 8 as non-zero numbers, such that

$$r := 2\sum_{n=1}^{(D-1)/2}|a_n| + 2\sum_{m=1}^{(D-1)/2}|b_m| + 2\sum_{n=1}^{(D-1)/2}\sum_{m=1}^{(D-1)/2}|A_{nm}| + 2\sum_{n=1}^{(D-1)/2}\sum_{m=(D+1)/2}^{D-1}|B_{n,m}| \leq \frac{1}{2}. \tag{69}$$

□

Having established the existence of an operator R with non-zero overlaps with all the operators U_{jk}, which at the same time satisfies that $\mathbb{1} \geq R \geq 0$, we are in the position to construct the subset $\mathscr{Q} \subset \{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1}\}_{x_1,\ldots,x_N=0}^{d-1}$.

Lemma 11. *Let $\mathscr{H}$ be a finite-dimensional complex Hilbert space with odd dimension $D \geq 3$. Let $\mathscr{H}_A$ be a finite-dimensional complex Hilbert space with dimension $d \geq 5$. Let $\{|a_x\rangle\}_{x=0}^{d-1}$ be an orthonormal basis of $\mathscr{H}_A$, and $|a\rangle \in \mathscr{H}_A$ normalized. Then, there exists a unitary operator U on $\mathscr{H} \otimes \mathscr{H}_A$, and a subset $\mathscr{Q} \subset \{A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1}\}_{x_1,\ldots,x_{2D-1}=0}^{d-1}$, with $|\mathscr{Q}| = D^2$, such that $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$, where $\boldsymbol{M}(\mathscr{Q})$ is as defined in (62), and where $A_x := \langle a_x|U|a\rangle$ for $x = 0,\ldots,d-1$.*

Proof. Since $D \geq 3$ is odd, we begin this proof by using Lemma 10 in order to construct the subset $\mathscr{Q}$. We know that there exists an operator R on $\mathscr{H}$ such that $\mathbb{1} \geq R \geq 0$ and $\mathrm{Tr}(U_{j,k}^\dagger R) \neq 0$ for all $j,k = 0,\ldots,D-1$, where the latter implies that

$$R = \sum_{j,k=0}^{D-1} \xi_{j,k}U_{j,k}, \quad \xi_{j,k} = \frac{1}{D}\mathrm{Tr}(U_{j,k}^\dagger R) \neq 0. \tag{70}$$

Because of $\mathbb{1} \geq R \geq 0$ both, $\sqrt{R}$ and $\sqrt{1-R}$, are well defined, and since $d \geq 5$, we can define

$$V := \frac{1}{2}|a_0\rangle\langle a| \otimes \sqrt{R} + \frac{1}{2}|a_1\rangle\langle a| \otimes \Lambda_1 + \frac{1}{2}|a_2\rangle\langle a| \otimes \sqrt{1-R} + \frac{1}{2}|a_3\rangle\langle a| \otimes \Lambda_3 + \frac{1}{2}|a_4\rangle\langle a| \otimes \mathbb{1}. \tag{71}$$

We note that $V^\dagger V = |a\rangle\langle a| \otimes \mathbb{1}$, and thus V is a partial isometry.

By virtue of being a partial isometry, V can be extended to a unitary operator U on $\mathscr{H} \otimes \mathscr{H}_A$, that satisfies

$$\begin{aligned} A_0 &:= \langle a_0|U|a\rangle = \langle a_0|V|a\rangle = \sqrt{R}, \quad A_1 := \langle a_1|U|a\rangle = \langle a_1|V|a\rangle = \Lambda_1, \\ A_3 &:= \langle a_3|U|a\rangle = \langle a_3|V|a\rangle = \Lambda_3, \quad A_4 := \langle a_4|U|a\rangle = \langle a_4|V|a\rangle = \mathbb{1}. \end{aligned} \tag{72}$$

(The above puts no particular restrictions on $A_x := \langle a_x|U|a\rangle$ for $x \geq 5$.) In the following, we consider a particular index subset $I \subset \{0,\ldots,d-1\}^{\times(2D-1)}$ of the form

$$(x_1,\ldots,x_{2D-1}) = (\underbrace{4,\ldots,4}_{2D-2-j-k},\underbrace{3,\ldots,3}_{k},\underbrace{1,\ldots,1}_{j},0), \quad j,k = 0,\ldots,D-1, \tag{73}$$

thus leaving us with operators on the form

$$A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1} = A_4^{\dagger 2D-2-k-j} A_3^{\dagger k} A_1^{\dagger j} A_0^\dagger A_0 A_1^j A_3^k A_4^{2D-2-k-j} = U_{jk}^\dagger R U_{jk}, \tag{74}$$

for $j,k = 0,\ldots,D-1$. We let

$$\begin{aligned}\mathscr{Q} :=& \{A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1}\}_{(x_1,\ldots x_{2D-1})\in I},\\ =& \{U_{jk}^\dagger R U_{jk}\}_{j,k=0}^{D-1}.\end{aligned} \tag{75}$$

By these constructions, it is clear that $|\mathscr{Q}| = D^2$. Next, we wish to show that $\mathscr{Q}$ is a linearly independent set. We have

$$\begin{aligned}Q_{j,k} :=& U_{jk}^\dagger R U_{jk},\\ & \text{[By (70)]}\\ =& \sum_{j',k'=0}^{D-1} \xi_{j',k'} U_{jk}^\dagger U_{j'k'} U_{jk},\\ & \text{[By (64)]}\\ =& \sum_{j',k'=0}^{D-1} \xi_{j',k'} \omega^{k'j-j'k} U_{j'k'}.\end{aligned} \tag{76}$$

Hence, if we let $(c_{j,k})_{j,k=0}^{D-1} \in \mathbb{C}^{D\times D}$, then

$$\sum_{j,k=0}^{D-1} c_{j,k} Q_{j,k} = \sum_{j',k'=0}^{D-1} \xi_{j',k'} \Big(\sum_{j,k=0}^{D-1} c_{j,k} \omega^{k'j-j'k} \Big) U_{j'k'}. \tag{77}$$

In order to show that $\mathscr{Q} = \{Q_{j,k}\}_{j,k=0}^{D-1}$ is a linearly independent set, we want to check that Eq. (77) vanishes only when all $c_{j,k}$ are zero. Since $\{U_{j'k'}\}_{j',k'=0}^{D-1}$ is a basis, we can conclude from Eq. (77) that $\sum_{j,k=0}^{D-1} c_{j,k} Q_{j,k} = 0$ implies that $\xi_{j',k'} \sum_{j,k=0}^{D-1} c_{j,k} \omega^{k'j-j'k} = 0$, for all $j',k' = 0,\ldots,D-1$. From (70), we know that $\xi_{j',k'} \neq 0$, and thus it follows that

$$\boldsymbol{\omega}_{k'}^\dagger C \boldsymbol{\omega}_{j'} = \sum_{j,k=0}^{D-1} c_{j,k} \omega^{k'j-j'k} = 0, \quad j',k' = 0,\ldots,D-1, \tag{78}$$

where $C := [c_{j,k}]_{j,k=0}^{D-1}$ and $\boldsymbol{\omega}_{j'} := (1, \omega^{j'}, \omega^{2j'}, \ldots, \omega^{(D-1)j'})^t$. We note that $\{\boldsymbol{\omega}_{j'}\}_{j'=0}^{D-1}$ forms a basis of $\mathbb{C}^D$. Hence, $[\boldsymbol{\omega}_{k'}^\dagger C \boldsymbol{\omega}_{j'}]_{k',j'=0}^{D-1}$ is nothing but the matrix-representation of C in the basis $\{\boldsymbol{\omega}_{j'}\}_{j'=0}^{D-1}$, and thus (78) implies that $C = 0$, and consequently $c_{j,k} = 0$. Hence, we can conclude that $\{Q_{j,k}\}_{j,k=0}^{D-1}$ is a linearly independent set. As in (62), we can define $\boldsymbol{M}(\mathscr{Q}) = [M_{jk,j'k'}]_{j,k,j',k'=0}^{D-1}$ with $M_{jk,j'k'} := \mathrm{Tr}(Q_{j,k}^\dagger Q_{j',k'})$. From the linear independence of $\mathscr{Q}$, it follows that $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$. □

Finally, we assemble the above results in order to prove that purity is a typical property of a set of operators $\{A_x\}_{x=0}^{d-1}$ of a finite-dimensional Hilbert space with odd dimension.

Proposition 12. *Let $\mathscr{H}$ be a finite-dimensional complex Hilbert space with odd dimension $D \geq 3$. Let $\mathscr{H}_A$ be a finite-dimensional complex Hilbert space with dimension $d \geq 5$. Let $\{|a_x\rangle\}_{x=0}^{d-1}$ be an orthonormal basis of $\mathscr{H}_A$, and $|a\rangle \in \mathscr{H}_A$ normalized. For each unitary operator U on $\mathscr{H} \otimes \mathscr{H}_A$, let $\{A_x\}_{x=0}^{d-1}$ be defined by $A_x := \langle a_x|U|a\rangle$ for $x = 0,\ldots,d-1$. Then, $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3 for all unitary operators U on $\mathscr{H} \otimes \mathscr{H}_A$, except for a subset of Haar-measure zero.*

Proof. For any fixed subset $\mathscr{Q} \subset \{A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1}\}_{x_1,\ldots,x_{2D-1}=0}^{d-1}$, with corresponding matrix $\boldsymbol{M}(\mathscr{Q})$ as defined by (62), it is the case that $\det \boldsymbol{M}(\mathscr{Q})$ is a polynomial in the real and imaginary parts of the matrix elements in a matrix representation of U (where one may note that since $\boldsymbol{M}(\mathscr{Q})$ is positive semi-definite, it follows that $\det \boldsymbol{M}(\mathscr{Q})$ is real-valued). Hence, by Lemma 7, we know that either $\det \boldsymbol{M}(\mathscr{Q}) = 0$ for all U, or $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$ for all U except for a subset of Haar measure 0. By Lemma 11, there exists a unitary U on $\mathscr{H} \otimes \mathscr{H}_A$, and a subset $\mathscr{Q} \subset \{A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1}\}_{x_1,\ldots,x_{2D-1}=0}^{d-1}$, with $|\mathscr{Q}| = D^2$, such that $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$. Hence, we can conclude that $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$ for all U, except for a subset of measure zero. If $\det \boldsymbol{M}(\mathscr{Q}) \neq 0$, then this implies that the elements of $\mathscr{Q}$ are linearly independent. Since $|\mathscr{Q}| = D^2 = \dim \mathscr{L}(\mathscr{H})$, it moreover follows that $\mathrm{Sp}(\mathscr{Q}) = \mathscr{L}(\mathscr{H})$. Consequently, $\mathrm{Sp}\Big(\{A_{x_1}^\dagger \cdots A_{x_{2D-1}}^\dagger A_{x_{2D-1}} \cdots A_{x_1}\}_{x_1,\ldots,x_{2D-1}=0}^{d-1}\Big) = \mathscr{L}(\mathscr{H})$. By Proposition 6 it follows that $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition. We can conclude that for all unitary operators on $\mathscr{H} \otimes \mathscr{H}_A$, except for a subset of Haar measure zero, it is the case that $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition.

□

E. Rate of decay

As we have seen in Section V, the CMI of an MPS decays exponentially to zero when the matrices of the MPS satisfy the purity condition (see Def. 3). Moreover, the transfer operator of an injective MPS, $\mathbb{E}$, decays to its unique fixed point exponentially fast at a rate lower bounded by the gap of the channel. The decay rate is often referred to as the correlation length. One could expect that there exists a relation between the correlation length and the decay rate of the classical CMI of Theorem 4. In this section, we consider two examples which give clear evidence of the absence of such relation.

Consider an MPS of the form of Eq. (1) such that all matrices A_{x_i} have rank one. After a measurement of system B, the reduced state $\Psi_C(x_B)$ becomes pure for any measurement outcome (see Eq. (14)). Therefore, the average von Neumann entropy of $\Psi_C(x_B)$, and thus the post-measurement CMI according to Eq. (12), is zero instantaneously even if region B is a single site. On the other hand, the correlation length need not be zero. For example, given a collection of transition probabilities, $\{P(x_j|x_i)\}_{x_i,x_j=0}^{d-1}$, and an orthonormal basis, $\{|x_i\rangle\}_{x_i=0}^{d-1}$, the repeated application of the transfer operator of an MPS with $A_{x_{ij}} = \sqrt{P(x_j|x_i)}|x_j\rangle\langle x_i|$ effectively implements a classical Markov process with transition probability $P(x_j|x_i)$ such that

$$\mathbb{E}^{\circ N}(\chi) = \sum_{x_1,\ldots,x_N=0}^{d-1} P(x_2|x_1)\cdots P(x_N|x_{N-1})\langle x_1|\chi|x_1\rangle|x_N\rangle\langle x_N|.$$

Nothing prevents this Markov chain to have a slow convergence to its equilibrium distribution.

Conversely, consider an MPS with matrices A_{x_i} proportional to unitary operators. As we have discussed in Section VI A, this implies that there is no decay of the von Neumann entropy, and thus the CMI remains constant. However, an injective MPS always has a finite correlation length. As a concrete example, consider

$$A_{x_{ij}} := \frac{1}{D}\sum_{k=0}^{D-1} e^{2\pi i \frac{kx_j}{D}}|k\rangle\langle (k+x_i) \bmod D|,$$

where $\{|k\rangle\}_{k=0}^{D-1}$ is an orthonormal basis. One can easily check that $A_{x_{ij}}$ are proportional to unitary operators, and hence the average von Neumann entropy, $\langle S[\Psi_C(x_B)]\rangle$, does not decay. The transfer operator of this MPS is the replacement map that replaces any input state, χ, with the maximally mixed state, i.e.,

$$\mathbb{E}(\chi) = \sum_{x_i,x_j=0}^{D-1} A_{x_{ij}}\chi A_{x_{ij}}^\dagger = \frac{\mathrm{Tr}\,\chi}{D}\mathbb{1}.$$

In Refs. 23 and 24, the decay of classical and quantum correlations is also studied. There, the authors introduce an entanglement measure called Localizable Entanglement (LE). The LE is defined as the maximal amount of entanglement that can be created on average between two spins at positions i and j of a chain by performing local measurements on the other spins. It is easy to note that the LE is similar to the scenario that we are considering in this paper (see Eq. (13)). Indeed, the difference is simply that the LE optimises over the basis of the measurement, while we pick a concrete basis. For the case when the measured spins are spin-1/2, it is shown in Refs. 23 and 24 that the connected correlation function provides a lower bound on the LE.

F. Examples

In this section we consider examples that illustrate some features of the process under study. As a prototypical example, we look at the AKLT model and obtain the exact convergence rate in a specific basis. Then, we consider MPSs with strictly contractive transfer operator and pure fixed point. This second example shows that primitivity of the transfer operator is not a necessary condition for the exponential convergence of the post-measurement CMI. In the last example that we construct, the purity condition is violated up to a fixed length $|B| = N$, but satisfied thereafter.

1. AKLT state

The first state we want to consider is the 1D AKLT model. The AKLT state defined on a chain has a well-known MPS description with bond dimension $D = 2$ and physical dimension $d = 3$. The matrices in the MPS picture are given by

$$A_0 = -\frac{1}{\sqrt{6}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad A_+ = \sqrt{\frac{1}{3}}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \quad A_- = -\sqrt{\frac{1}{3}}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}. \tag{79}$$

We take the $\{x_i\} = \{0,+,-\}$ as the basis for our physical space. It can be seen by inspection that the transfer operator, $\mathbb{E}(\chi) = \sum_{x_i} A_{x_i} \chi A_{x_i}^\dagger$, has a unique stationary state $\rho = \mathbb{1}/2$. In the infinite chain setting, the probability of a measurement outcome $x_B = x_1, \ldots, x_N$ is given by

$$p_\Psi(x_B) = \frac{1}{2} \mathrm{Tr}\left[A_{x_N} \cdots A_{x_1} A_{x_1}^\dagger \cdots A_{x_N}^\dagger\right]. \tag{80}$$

We want to calculate the average entropy on C after measurement of system B, i.e., we need to estimate

$$\langle S[\Psi_C(x_B)]\rangle = \sum_{x_B=0,+,-} p_\Psi(x_B) S\left[\frac{A_{x_N} \cdots A_{x_1} A_{x_1}^\dagger \cdots A_{x_N}^\dagger}{\mathrm{Tr}\left(A_{x_N} \cdots A_{x_1} A_{x_1}^\dagger \cdots A_{x_N}^\dagger\right)}\right]. \tag{81}$$

We note two scenarios: $p_\Psi(x_B) = 0$ and $p_\Psi(x_B) \neq 0$. Since $A_+ A_+ = A_- A_- = 0$, we get that whenever the string x_B contains two (or more) successive $+$ (or $-$), then $p_\Psi(x_B) = 0$. In other words, the only strings that give a $p_\Psi(x_B) \neq 0$ are those with an alternating sequence (Ex: $+-+-$), possibly interspersed with 0's. However, the only string with non-zero entropy is the one with all 0's because any alternating sequence has rank one. This string will occur with probability $1/3^N$. We get that

$$\langle S[\Psi_C(x_B)]\rangle = \frac{1}{3^N} S[\Psi_C(x_0)], \tag{82}$$

where $x_0 := 0, \ldots, 0$. Hence, the AKLT model in the standard basis has a post-measurement CMI that is exponentially decaying in the size of B for large A and C. The correlation length is coincidentally the same as the classical CMI decay in this basis.

Let us now consider a change of basis. The 1D AKLT state is also given by the MPS representation with matrices

$$\tilde{A}_0 := \sqrt{\frac{1}{3}}\hat{\sigma}_x = \sqrt{\frac{1}{3}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tilde{A}_1 := \sqrt{\frac{1}{3}}\hat{\sigma}_y = \sqrt{\frac{1}{3}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tilde{A}_2 := \sqrt{\frac{1}{3}}\hat{\sigma}_z = \sqrt{\frac{1}{3}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where $\hat{\sigma}_i$ are the Pauli matrices. The Pauli matrices are unitary, and thus the average von-Neumann entropy of $\Psi_C(x_B)$ (Eq. (81)) is constant, namely $\log 2$. In other words, the post-measurement CMI of the AKLT chain when measured in the basis corresponding to the Pauli matrices is not decaying. This is consistent with the discussion in Section VI A because the AKLT chain is in the Haldane phase, which is a SPT phase protected by the $Z_2 \times Z_2$ symmetry generated by the π rotations around three orthogonal axes.

2. *Strictly contractive map with a pure fixed point*

As mentioned in Section II, translationally invariant MPSs are injective if and only if the transfer operator $\mathbb{E}$, defined in (2), is primitive [30], where primitivity means that the channel possesses a unique full-rank fixed point. Primitivity in turn guarantees the existence of a gapped parent Hamiltonian and exponential decay of correlations [13, 38].

In view of the essential role played by primitivity for the decay of correlations, one may ask how it relates to the purity condition for the exponential decay of the CMI, in the sense of Theorem 4. In this section, we present an example which shows that primitivity is *not* a necessary condition for the exponential decay of the CMI. In other words, we consider an MPS with a non-primitive transfer operator, which nevertheless yields and exponentially decaying CMI due to purity.

Consider an MPS of the form of Eq. (1) which has a strictly contractive transfer operator, $\mathbb{E}$, with a pure fixed point, denoted by $|\phi\rangle$. Let us recall that a channel, Φ, is strictly contractive if there exists a number $0 \leq \alpha < 1$ such that $\|\Phi(\chi_1) - \Phi(\chi_2)\|_1 \leq \alpha \|\chi_1 - \chi_2\|_1$ for all density operators χ_1 and χ_2. Note that if a channel is strictly contractive, then the fixed point is unique. Note further that, since the fixed point of $\mathbb{E}$ is pure, the transfer operator is, by definition, not primitive. Let us also assume that $F = \sqrt{\mathbb{E}^{*|C|}(R)}$ is full rank. Under these assumptions, our aim is to find an exponentially decaying bound of the average von Neumann entropy of the reduced post-measurement state (see Eq. (15)). We start by using the concavity of the entropy and obtain

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq S\left(\frac{1}{K^2}\sum_{x_B} F A_{x_N} \cdots A_{x_1} \sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger\right) = S\left(\frac{F\mathbb{E}^N(\sigma)F^\dagger}{\mathrm{Tr}\left[F\mathbb{E}^N(\sigma)F^\dagger\right]}\right),$$

where $K^2 = \mathrm{Tr}\left[F\mathbb{E}^N(\sigma)F^\dagger\right]$. The purity of the fixed point of $\mathbb{E}$ allows us to transform the above inequality to

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq S\left(\frac{F\mathbb{E}^N(\sigma)F^\dagger}{\mathrm{Tr}\left[F\mathbb{E}^N(\sigma)F^\dagger\right]}\right) = \left|S\left(\frac{F\mathbb{E}^N(\sigma)F^\dagger}{\mathrm{Tr}\left[F\mathbb{E}^N(\sigma)F^\dagger\right]}\right) - S\left(\frac{F\mathbb{E}^N(|\phi\rangle\langle\phi|)F^\dagger}{\mathrm{Tr}\left[F\mathbb{E}^N(|\phi\rangle\langle\phi|)F^\dagger\right]}\right)\right|,$$

We can further bound the average von Neumann entropy using the Fannes-Audenaert inequality [39, 40]. This yields

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq t\log(D-1) + H_B(t), \tag{83}$$

where H_B is the binary entropy, i.e., $H_B(t) = -t\log(t) - (1-t)\log(1-t)$, with $H_B(0) := 0$ and $H_B(1) := 0$, and where t is defined as

$$t := \frac{1}{2}\left\| \frac{F\mathbb{E}^N(\sigma)F^\dagger}{\mathrm{Tr}[F\mathbb{E}^N(\sigma)F^\dagger]} - \frac{F\mathbb{E}^N(|\phi\rangle\langle\phi|)F^\dagger}{\mathrm{Tr}[F\mathbb{E}^N(|\phi\rangle\langle\phi|)F^\dagger]} \right\|_1, \tag{84}$$

with $\|\cdot\|_1$ denoting the trace norm.

For any full-rank F, and any pair of density operators χ_1, χ_2 on a finite-dimensional Hilbert space, one can show that

$$\left\| \frac{F\chi_2F^\dagger}{\mathrm{Tr}(F\chi_2F^\dagger)} - \frac{F\chi_1F^\dagger}{\mathrm{Tr}(F\chi_1F^\dagger)} \right\|_1 \leq 2\left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\chi_1 - \chi_2\|_1, \tag{85}$$

where $\nu_1(F)$ and $\nu_D(F)$ denote the largest and the smallest singular values of F, and where we note that $\nu_D(F) > 0$ since F is full-rank on a finite-dimensional space.

By combining (84) and (85) with an iterative use of strict contractivity of $\mathbb{E}$, we find an upper-bound on t such that

$$\begin{aligned} t &\leq \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\mathbb{E}^N(\sigma) - \mathbb{E}^N(|\phi\rangle\langle\phi|)\|_1, \\ &\leq \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \alpha^N \|\sigma - |\phi\rangle\langle\phi|\|_1. \end{aligned} \tag{86}$$

If it would be possible to choose $\alpha = 0$, then $\|\mathbb{E}(\chi_1) - \mathbb{E}(\chi_2)\|_1 = 0$, and thus $\mathbb{E}(\chi_1) = \mathbb{E}(\chi_2)$, which implies that the convergence is not only exponential, but immediate. Hence, without loss of generality, we may in the following assume that $0 < \alpha < 1$.

The next step consists of bounding the binary entropy, $H_B(t)$. For that, we define the function $g(t) := t - t\log t$, with $g(0) := 0$. One can show that g is monotonically increasing on $t \in [0,1]$ and satisfies $H_B(t) \leq g(t)$ for $0 \leq t \leq 1$. These two properties of g together with inequality (86) lead to

$$\begin{aligned} H_B(t) \leq{}& \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \alpha^N \|\sigma - |\phi\rangle\langle\phi|\|_1 \\ &- \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \alpha^N \|\sigma - |\phi\rangle\langle\phi|\|_1 \log\left[\left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\sigma - |\phi\rangle\langle\phi|\|_1\right] \\ &- \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\sigma - |\phi\rangle\langle\phi|\|_1 N\alpha^N \log\alpha. \end{aligned} \tag{87}$$

By combining (83) with (87), and again using inequality (86), we find that

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq c_1\alpha^N + c_2 N\alpha^N, \tag{88}$$

where c_1 and c_2 are defined as

$$c_1 := \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\sigma - |\phi\rangle\langle\phi|\|_1 \left[\log(D-1) + 1 - \log\left(\left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\sigma - |\phi\rangle\langle\phi|\|_1\right)\right],$$

$$c_2 := \left(\frac{\nu_1(F)}{\nu_D(F)}\right)^4 \|\sigma - |\phi\rangle\langle\phi|\|_1 (-\log\alpha).$$

The right-hand side of Eq. (88) decays exponentially to zero when N grows to infinity because $0 < \alpha < 1$. Hence, we have found an exponentially-decaying bound on the post-measurement CMI for an MPS with a transfer operator that is strictly contractive and has a pure fixed point. This shows that primitivity of the transfer operator is not a necessary condition for the CMI to converge.

3. Jordan blocks

Here, we construct a simple example of a two-element set $\{A_0, A_1\}$ that has a nontrivial correctable subspace (in the sense of Section VI B) for small enough N, but where the purity condition nevertheless holds. For the Hilbert space $\mathscr{H}$, we let $\dim \mathscr{H} = D+1$ and let $|0\rangle, \ldots, |D-1\rangle, |D\rangle$ be an orthonormal basis of $\mathscr{H}$. We define the projector $P := \sum_{k=0}^{D-1} |k\rangle\langle k|$ and the operators

$$A_0 := \sum_{k=0}^{D-1} |k+1\rangle\langle k|, \qquad \text{and} \qquad A_1 := |D\rangle\langle D|. \tag{89}$$

We note that A_0 is a Jordan block with zeros on the diagonal, and that

$$A_0^\dagger A_0 = \sum_{k=0}^{D-1} |k\rangle\langle k| = P, \tag{90}$$

while $A_1^\dagger A_1 = A_1 = |D\rangle\langle D|$. Thus, $A_0^\dagger A_0 + A_1^\dagger A_1 = \mathbb{1}$. Moreover, observe that

$$\begin{aligned} P A_0^\dagger A_0 P &= \lambda_0 P, \quad \lambda_0 = 1, \\ P A_1^\dagger A_1 P &= \lambda_1 P, \quad \lambda_1 = 0. \end{aligned} \tag{91}$$

Hence, for a single site, the correctable subspace is $\mathscr{C}_1 := \text{span}\{|0\rangle, \ldots, |D-1\rangle\}$.

Now consider the case of several sites, where we construct the sequence $A_{x_N} \cdots A_{x_1}$. It is not difficult to see that

$$A_0^N = \sum_{k=0}^{D-N} |k+N\rangle\langle k|. \tag{92}$$

Hence, the correctable subspace decreases the dimension with one step along the sequence, until it is exhausted. As a consequence, the CMI in this example will be exponentially decaying with a pre-factor that grows exponentially in D. Examples that do not have a block diagonal structure can also be constructed. We note that the example above is similar in spirit to a bosonic annihilation operator. Indeed, in the infinite system case where the operators A_i are bosonic creation and annihilation operators, the purity condition no longer makes sense, and the theory breaks down.

VII. OUTLOOK

We have shown that the amplitudes of an injective MPS in a specific local basis follow a quasi-local Gibbs distribution with exponentially decaying tails if the matrices associated to the MPS satisfy a particular 'purity-condition'. The purity condition reflects the fact that no information can be preserved in the virtual subspace on average, upon measurements. Our proof makes extensive use of the theory of random matrix products.

A number of open questions remains. Perhaps the most obvious is whether the methods used in this paper can be applied in higher dimensions or in the context of matrix product operators, and whether this leads to new insights or algorithmic improvements. In the setting of matrix product operators, the purity condition would no longer be sufficient to prevent information transmission along the chain. There one would likely have to bound the stochastic process upon measurements from above and below. Some recent progress in this direction has been communicated to us [41].

Another place where the present tools might be applied is in the rigorous analysis of the Wave Function Monte Carlo algorithm. A first attempt to achieve this has been made in Ref. 22, yet some work remains to be done in connecting these mathematical results to more realistic physical settings and particular examples. Yet another extension would be to continuous MPSs [42].

On a more technical level, it would be valuable to get a better handle on the decay rate of the stochastic process. In particular, whether there exists a closed form expression as is the case for the correlation length (as the spectral gap of the transfer operator).

There are also open questions related to the purity condition, such as whether it can be easily checked or not for a given MPS. In this investigation we have shown that the purity condition in essence is necessary and sufficient for the exponential decay of the quantum conditional mutual information, $I_{\Phi_B(\Psi)}(A:C|B)$, while we only have proved that purity is a sufficient condition for the exponential decay of the classical conditional mutual information, $I_{p_\Psi}(A:C|B)$. An open question is thus whether purity in essence also is a necessary condition for the latter, or whether there exists another weaker condition that would be both necessary and sufficient.

ACKNOWLEDGMENTS

We thank T. Benoist for clarifying some details in Ref. 20. We thank David Gross for helpful discussions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1-390534769. This work was completed while MJK was at the University of Cologne. We also thank an anonymous referee for pointing out improvements of Proposition 1 and Lemma 2, which strengthened the results and simplified the proofs.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Appendix A: Elements of the proof of Theorem 4

Here we present a detailed account of the proof of Theorem 4. For convenience of presentation, we have divided the statement of Theorem 4 into two parts. The first part, that purity implies (31), is proved in A 1, and stated in a more detailed version in Theorem 18. The second part of Theorem 4, that (32) implies purity, is the focus of section A 2, and is formalized in Theorem 24.

1. Proof of the first part of Theorem 4

In order to show Theorem 4 in Section V, we use two key bounds, one on the average von Neumann entropy and a second on the average purity. Here, we state these bounds in the form of two lemmas. In the following we let $\|Q\| := \sup_{\|\psi\|=1} \|Q|\psi\rangle\|$ denote the standard operator norm.

Lemma 13. *Let $\{\rho_x\}_{x=0}^{M-1}$ be a collection of density operators on a Hilbert space $\mathscr{H}$, with $D = \dim \mathscr{H}$, and $\{p(x)\}_{x=0}^{M-1}$ be real numbers such that $p(x) \geq 0$ and $\sum_{x=0}^{M-1} p(x) = 1$. Then,*

$$\sum_{x=0}^{M-1} p(x) S(\rho_x) \leq -Q \log Q + Q\left[\log\left(D-1\right)+1\right], \tag{A1}$$

where we refer to Q as the the average purity and define it as

$$Q := 1 - \sum_{x=0}^{M-1} p(x) \|\rho_x\|. \tag{A2}$$

Proof. To begin with, let us first consider a single density operator, ρ_x, and define the channel

$$\Gamma(\rho_x) := |\phi\rangle\langle\phi|\rho_x|\phi\rangle\langle\phi| + \Phi^{\perp}\rho_x\Phi^{\perp},$$

where $|\phi\rangle$ is a pure state in $\mathscr{H}$, and $\Phi^{\perp} := \mathbb{1} - |\phi\rangle\langle\phi|$. The channel Γ is mixing-enhancing, i.e., $S(\rho_x) \leq S[\Gamma(\rho_x)]$. Moreover, Γ transforms any input state into a block-diagonal state, which implies that for any function, f, and any input state, ρ, it holds that $f[\Gamma(\rho)] = f\left(|\phi\rangle\langle\phi|\rho|\phi\rangle\langle\phi|\right) + f\left(\Phi^{\perp}\rho\Phi^{\perp}\right)$. Using these two properties of Γ, we obtain

$$\begin{aligned} S(\rho_x) &\leq H_B[q(x)] + q(x) S\left[\frac{\Phi^{\perp}\rho_x\Phi^{\perp}}{\mathrm{Tr}\left(\Phi^{\perp}\rho_x\right)}\right], \\ &\leq H_B[q(x)] + q(x)\log\left(\dim\mathscr{H}-1\right), \end{aligned}$$

where we have defined $q(x) := \mathrm{Tr}(\Phi^{\perp}\rho_x) = 1 - \langle\phi|\rho_x|\phi\rangle$ for $x = 0,\ldots,M$, and recall that $H_B(t) = -t\log t - (1-t)\log(1-t)$ is the binary entropy. Note that we can choose $|\phi\rangle$ to be the normalized eigenvector corresponding to the largest eigenvalue of ρ_x, which we denote as $\lambda_1^{\downarrow}(\rho_x)$. Then, we have $q(x) = 1 - \lambda_1^{\downarrow}(\rho_x) = 1 - \|\rho_x\|$.

Considering now the whole set of density operators, $\{\rho_x\}_{x=0}^{M}$, we have by the concavity of the entropy that

$$\sum_{x=0}^{M-1} p(x) S(\rho_x) \leq H_B(Q) + Q\log\left(\dim\mathscr{H}-1\right), \tag{A3}$$

where Q is defined in Eq. (A2).

One can next bound the binary entropy, as $H_B(t) \leq t - t\log t$ on $0 \leq t \leq 1$. By combining this observation with (A3), we obtain (A1).

□

The average purity, Q, defined in Eq. (A2) can be bounded if one considers some structure on the density operators and the probabilities. In particular, taking $\rho_x = \Psi_C(x_B)$ and $p(x) = p_\Psi(x_B)$ (see Eq. (1) and Eq. (7)), the average purity is $Q = 1 - K^{-2}\sum_{x_N,\ldots,x_1=0}^{d-1} \|FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\|$, where recall that $\sigma = \mathbb{E}^{|A|}(L)$ and $F^\dagger F = \mathbb{E}^{*|C|}(R)$. An upper and a lower bound on Q are stated and shown in the following lemma.

Lemma 14. *Let $\{A_x\}_{x=0}^{d-1}$ be a collection of operators on a Hilbert space, $\mathscr{H}$, with $D := \dim\mathscr{H} \leq +\infty$, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. Let σ be a density operator on $\mathscr{H}$, and F an operator on $\mathscr{H}$, such that $F^\dagger F \leq \mathbb{1}$. Then,*

$$\tfrac{1}{K^2}\sum_{x_B}\lambda_2^\downarrow\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right) \leq Q \leq \tfrac{D-1}{K^2}\sum_{x_B}\lambda_2^\downarrow\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right), \tag{A4}$$

where $x_B := x_1,\ldots,x_N$, and where the sum over x_B spans all of $\{0,\ldots,d-1\}^N$. Moreover, K is a normalization constant such that

$$K^2 := \sum_{x_B}\mathrm{Tr}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right); \tag{A5}$$

and Q is

$$Q = 1 - \frac{1}{K^2}\sum_{x_B}\left\|FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right\|. \tag{A6}$$

Proof. Consider a positive semi-definite operator, $\rho \geq 0$, and define a function, L, on ρ such that

$$L(\rho) := \sum_{j=2}^{D}\lambda_j^\downarrow(\rho).$$

This function $L(\rho)$ can be upper and lower bounded as

$$\lambda_2^\downarrow(\rho) \leq L(\rho) \leq (D-1)\lambda_2^\downarrow(\rho). \tag{A7}$$

Moreover, it holds that

$$\lambda_1^\downarrow(\rho) + L(\rho) = \mathrm{Tr}(\rho). \tag{A8}$$

If we introduce in Eq. (A8) the positive operator $\rho := K^{-2}FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger$ and we sum over all possible values of $x_B := x_1,\ldots,x_N$, we obtain

$$\begin{aligned}\frac{1}{K^2}\sum_{x_B}\lambda_1^\downarrow\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right) + \frac{1}{K^2}\sum_{x_B}L\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right)& \\ = \frac{1}{K^2}\sum_{x_B}\mathrm{Tr}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right),& \\ = 1,&\end{aligned}$$

where the last equality holds due to the definition of K in Eq. (A5). This implies that

$$\begin{aligned}\frac{1}{K^2}\sum_{x_B}L\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right) &= 1 - \frac{1}{K^2}\sum_{x_B}\lambda_1^\downarrow\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right), \\ &= 1 - \frac{1}{K^2}\sum_{x_B}\left\|FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger\cdots A_{x_N}^\dagger F^\dagger\right\|, \\ &= Q.\end{aligned}$$

Using the bounds of Eq. (A7), we finish the proof, since we obtain the bounds on Q in Eq. (A4).

□

Recall that we throughout this investigation assume that log denotes the natural logarithm. We state the following two lemmas without proof.

Lemma 15. *Let*

$$H_B(t) := -t\log t - (1-t)\log(1-t), \quad 0 < t < 1, \tag{A9}$$

and $H_B(0) := 0$ *and* $H_B(1) := 0$*. Let*

$$g(t) := t - t\log t, \quad 0 < t \leq 1, \tag{A10}$$

and $g(0) := 0$*. Then,* g *is monotonically increasing on* $[0,1]$*, and*

$$H_B(t) \leq g(t), \quad 0 \leq t \leq 1. \tag{A11}$$

Lemma 16.

$$-t\log t \leq \frac{1}{\varepsilon}t^{1-\varepsilon}, \quad 0 \leq t \leq 1, \quad 0 < \varepsilon < 1. \tag{A12}$$

We recall that if the channel $\mathbb{E}$ is primitive, then it follows that $\mathbb{E}$ has a unique full-rank fixed point ρ [30]. With the replacement-map $\mathscr{R}(\sigma) := \rho\,\mathrm{Tr}(\sigma)$, the fact that every initial state σ converges to ρ can be expressed as $\lim_{N\to\infty}\mathbb{E}^N = \mathscr{R}$. Since the underlying Hilbert space is finite-dimensional, we can express the convergence in terms of any norm. It is convenient to express the convergence in terms of the norm

$$\|\mathscr{F}\|_{1:1} := \sup_{\|Q\|_1=1} \|\mathscr{F}(Q)\|_1, \tag{A13}$$

and thus $\lim_{N\to\infty}\|\mathbb{E}^N - \mathscr{R}\|_{1:1} = 0$, where $\|Q\|_1 := \mathrm{Tr}\sqrt{Q^\dagger Q}$ is the trace norm.

Lemma 17. *Let* $\{A_x\}_{x=0}^{d-1}$ *be operators on a Hilbert space* $\mathscr{H}$*, with* $D := \dim\mathscr{H} < +\infty$*, such that* $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$*, and* $\mathbb{E}(\cdot) := \sum_{x=0}^{d-1} A_x \cdot A_x^\dagger$ *is primitive. Then, there exists a real number* $r > 0$ *and a natural number* N_0 *such that*

$$\langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle \geq r, \quad \forall |A|,|C|,\ \forall \|R\| = 1, \|L\| = 1,\ \forall |B| \geq N_0. \tag{A14}$$

Proof. For the map $\mathscr{R}(\sigma) := \rho\,\mathrm{Tr}(\sigma)$ with ρ the unique fixed point ρ of $\mathbb{E}$, we first observe that

$$\begin{aligned}
\Big|\langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle - \langle R|\rho|R\rangle\Big| &= \Big|\,\mathrm{Tr}\Big(|R\rangle\langle R|\big(\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|) - \rho\big)\Big)\Big|, \\
&\leq \big\||R\rangle\langle R|\big\|\,\big\|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|) - \rho\big\|_1, \\
&= \Big\|\mathbb{E}^{|B|}\big(\mathbb{E}^{|A|+|C|}(|L\rangle\langle L|)\big) - \mathscr{R}\big(\mathbb{E}^{|A|+|C|}(|L\rangle\langle L|)\big)\Big\|_1, \\
&\leq \big\|\mathbb{E}^{|B|} - \mathscr{R}\big\|_{1:1}\,\|\mathbb{E}^{|A|+|C|}(|L\rangle\langle L|)\|_1, \\
&= \big\|\mathbb{E}^{|B|} - \mathscr{R}\big\|_{1:1}.
\end{aligned} \tag{A15}$$

Since $\mathbb{E}$ is assumed to be primitive, it follows that $\mathbb{E}$ has a unique full rank fixed point ρ. Since $\mathscr{H}$ is assumed to be finite-dimensional, it follows that the minimal eigenvalue of ρ is such that $\lambda_{\min}(\rho) > 0$. By (A15), it follows that

$$\begin{aligned}
\lambda_{\min}(\rho) - \big\|\mathbb{E}^{|B|} - \mathscr{R}\big\|_{1:1} &\leq \langle R|\rho|R\rangle - \big\|\mathbb{E}^{|B|} - \mathscr{R}\big\|_{1:1}, \\
&\leq \langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle.
\end{aligned} \tag{A16}$$

Since $\lim_{|B|\to\infty}\big\|\mathbb{E}^{|B|} - \mathscr{R}\big\|_{1:1} = 0$ and $\lambda_{\min}(\rho) > 0$, it follows that there exists an r such that $\lambda_{\min}(\rho) > r > 0$ and a N_0, such that

$$\langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle \geq r, \quad \forall |B| \geq N_0. \tag{A17}$$

One should note that r and N_0 are independent of $|A|$, $|C|$, and all normalized $|R\rangle$ and $|L\rangle$. □

Theorem 4 in the main text follows as a direct corollary of Theorem 18 below with $\kappa := \gamma^{1-\varepsilon}$ and $c := c_\varepsilon$ for any fixed $0 < \varepsilon < 1$. In essence, we use the bound $f(N) \leq \bar{c}\gamma^N$ in Proposition 5 in order to prove the bound in Theorem 18, and thus it is the same γ that appears in both bounds. The reason for the transition from γ to $\gamma^{1-\varepsilon}$ is loosely speaking due to a leading order term proportional to $|B|\gamma^{|B|}$. This term appears in a bound on the CMI and can be accommodated by an arbitrarily small sacrifice of the rate in the exponential decay. However, since we here are not only interested in the asymptotics, but rather wish to achieve a general bound valid for all values of $|B|$, the construction in the proof becomes more elaborate.

Theorem 18. *For a set of operators $\{A_x\}_{x=0}^{d-1}$ on a Hilbert space $\mathscr{H}$ with $D := \dim \mathscr{H} \geq 2$, and normalized $|R\rangle, |L\rangle \in \mathscr{H}$, let Ψ be the MPS as defined in (1) on a region $\Lambda = ABC$. The set $\{A_x\}_{x=0}^{d-1}$ is such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$ satisfies the purity condition in Definition 3, and is such that $\mathbb{E}(\cdot) := \sum_{x=0}^{d-1} A_x \cdot A_x^\dagger$ is primitive. For the constant γ as guaranteed by Proposition 5, and for every $0 < \varepsilon < 1$, there exists a constant $c_\varepsilon \geq 0$ such that*

$$I_{p_\Psi}(A:C|B) \leq I_{\Phi_B(\Psi)}(A:C|B) \leq c_\varepsilon \gamma^{|B|(1-\varepsilon)}, \quad |B| = 1,2,\dots . \tag{A18}$$

The constant γ is independent of $|A|$, $|B|$, $|C|$, $|L\rangle$, $|R\rangle$ and ε. The constant c_ε is independent of $|A|$, $|B|$, $|C|$, $|L\rangle$ and $|R\rangle$, but may depend on ε.

Proof. We first note that

$$\begin{aligned} I_{\Phi_B(\Psi)}(A:C|B) &= \langle S[\Psi_A(x_B)]\rangle_{p_\Psi(x_B)} + \langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}, \\ &= 2\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)}, \end{aligned} \tag{A19}$$

where we recall that $p_\Psi(x_\Lambda) = \langle x_\Lambda|\Psi|x_\Lambda\rangle$, and where the state $\Psi_X(x_B)$ is the reduced state in region X of the post-measurement state, $\Psi(x_B)$, and $\langle S[\Psi(x)]\rangle_{p_{\Psi(x)}}$ is the average von Neumann entropy

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} = \sum_{x_B} p_\Psi(x_B) S(\Psi_C(x_B)), \tag{A20}$$

with

$$\begin{aligned} p_\Psi(x_B) :=& \frac{1}{K^2} \operatorname{Tr}\left[A_{x_N}\cdots A_{x_1} \mathbb{E}^{|A|}(L) A_{x_1}^\dagger \cdots A_{x_N}^\dagger \mathbb{E}^{*|C|}(R)\right], \\ =& \frac{1}{K^2} \operatorname{Tr}\left[F A_{x_N}\cdots A_{x_1} \sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger\right], \end{aligned} \tag{A21}$$

and

$$\Psi_C(x_B) := \frac{1}{p_\Psi(x_B) K^2} F A_{x_N}\cdots A_{x_1} \sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger, \tag{A22}$$

where

$$\begin{aligned} F :=& \sum_{x_{|BC|},\dots,x_{N+1}} |x_{x_{|BC|},\dots,x_{N+1}}\rangle\langle R| A_{x_{|BC|},\dots,x_{N+1}}, \\ \sigma :=& \mathbb{E}^{|A|}(L). \end{aligned} \tag{A23}$$

The equality (A19) follows from the fact that the post-measurement state is pure, and thus the reduced states on regions A and C are isospectral (up to zero eigenvalues). With $\rho_{x_B} := \Psi_C(x_B)$ in Lemma 13, we know that

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq H_B(Q) + Q\log(D-1), \tag{A24}$$

with

$$Q := 1 - \sum_{x_B} p_\Psi(x_B) \|\Psi_C(x_B)\|. \tag{A25}$$

By Lemma 15 we know that the function $g(t) = t - t\log t$ is monotonically increasing on the interval $[0,1]$ and satisfies $H_B(t) \leq g(t)$. By combining this observation with (A24), we get

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq g(Q) + Q\log(D-1) = -Q\log Q + Q + Q\log(D-1). \tag{A26}$$

By Lemma 14 we furthermore know that

$$Q \leq \frac{D-1}{K^2} \sum_{x_B} \lambda_2^\downarrow\left(F A_{x_N}\cdots A_{x_1} \sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger\right), \tag{A27}$$

where $\lambda_j^{\downarrow}(O)$ are the eigenvalues of an operator O in non-increasing order, i.e., $\lambda_1^{\downarrow}(O) \geq \cdots \geq \lambda_D^{\downarrow}(O)$. Similarly, we let in the following $\nu_j^{\downarrow}(O)$ denote the singular values of O in non-increasing order $\nu_1^{\downarrow}(O) \geq \cdots \geq \nu_D^{\downarrow}(O)$. Then, recalling that for any operator O, we have $\lambda_j(OO^\dagger) = \nu_j(O)^2$, we get

$$\begin{aligned}
Q \leq & \frac{D-1}{K^2} \sum_{x_B} \lambda_2^{\downarrow}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger\right), \\
\leq & \frac{D-1}{K^2} \sum_{x_B} \sqrt{\lambda_1^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger)\lambda_2^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^\dagger \cdots A_{x_N}^\dagger F^\dagger)}, \\
= & \frac{D-1}{K^2} \sum_{x_B} \nu_1^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})\nu_2^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma}), \\
= & \frac{D-1}{K^2} f(N), \\
& \text{[By Proposition 5]} \\
\leq & \frac{D-1}{K^2}\bar{c}\gamma^N, \\
= & \frac{D-1}{\langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle}\bar{c}\gamma^{|B|},
\end{aligned} \tag{A28}$$

where recall that $N = |B|$ and the constant $\bar{c}$ and $0 < \gamma < 1$ are independent of $\sigma := \mathbb{E}^{|A|}(L)$ and F as in (16), and consequently are independent of $|A|$ and $|C|$ (as well as of $|B|$).

By Lemma 17, there exist constants $r > 0$ and N_0 such that $\langle R|\mathbb{E}^{|A|+|B|+|C|}(|L\rangle\langle L|)|R\rangle \geq r$ for all $|B| \geq N_0$. By Lemma 17 we know that r and N_0 do not depend on $|A|, |B|, |C|, |R\rangle, |L\rangle$. By combining this observation with (A28), we can conclude that

$$Q \leq \tilde{c}\gamma^{|B|}, \quad \text{with} \quad \tilde{c} := \frac{D-1}{r}\bar{c}, \quad \forall |B| \geq N_0, \tag{A29}$$

where we note that $\tilde{c}$ and N_0 do not depend on $|A|, |B|, |C|, |R\rangle, |L\rangle$. By inspection of the definition of Q in (A25), one can see that

$$Q \leq 1 \tag{A30}$$

is trivially true. By combining (A29) and (A30), we thus get

$$Q \leq t, \quad \forall |B| \geq N_0, \quad \text{with} \quad t := \min\left[1, \tilde{c}\gamma^{|B|}\right], \tag{A31}$$

where t by necessity is contained in the interval $[0,1]$.

We next combine (A26) and (A31) with the monotonicity of g to obtain

$$\begin{aligned}
\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq & g(Q) + Q\log(D-1), \\
& \text{[Monotonicity of } g\text{, Lemma 15, together with (A31)]} \\
\leq & g(t) + t\log(D-1), \\
= & -t\log t + t\left[1 + \log(D-1)\right], \\
& \text{[By Lemma 16]} \\
\leq & \frac{1}{\varepsilon}t^{1-\varepsilon} + t\left[1 + \log(D-1)\right], \\
& \text{[By } t \leq t^{1-\varepsilon}, \quad 0 \leq t \leq 1, \quad 0 < \varepsilon < 1\text{]} \\
\leq & \left[\frac{1}{\varepsilon} + 1 + \log(D-1)\right] t^{1-\varepsilon}.
\end{aligned} \tag{A32}$$

Since $t = \min\left[1, \tilde{c}\gamma^{|B|}\right] \leq \tilde{c}\gamma^{|B|}$ we get $t^{1-\varepsilon} \leq \tilde{c}^{1-\varepsilon}\gamma^{|B|(1-\varepsilon)}$, and thus

$$\langle S[\Psi_C(x_B)]\rangle_{p_\Psi(x_B)} \leq \left[\frac{1}{\varepsilon} + 1 + \log(D-1)\right] \tilde{c}^{1-\varepsilon}\gamma^{|B|(1-\varepsilon)}. \tag{A33}$$

By combining this with (A19) we get

$$\begin{aligned} I_{\Phi_B(\Psi)}(A:C|B) \leq & \tilde{c}_\varepsilon \gamma^{|B|(1-\varepsilon)}, \quad \forall |B| \geq N_0, \\ \text{with} \quad \tilde{c}_\varepsilon = & 2\left[\frac{1}{\varepsilon}+1+\log(D-1)\right]\tilde{c}^{1-\varepsilon}. \end{aligned} \tag{A34}$$

Finally we should remove the restriction that $|B| \geq N_0$. By (A19) and (A20) we can conclude that $I_{\Phi_B(\Psi)}(A:C|B) = 2\sum_{x_B} p_\Psi(x_B) S\big(\Psi_C(x_B)\big) \leq 2\log D$, where the last inequality follows since $\Psi_C(x_B)$ is (up to zero eigenvalues) isospectral to the density operator in (14), and thus the entropy of these two states are equal. The state in (14) is a density operator on $\mathscr{H}$, which has dimension D, and thus the entropy is bounded by $\log D$. Let

$$c_\varepsilon := \max\left(\tilde{c}_\varepsilon, 2\log(D)\gamma^{-(N_0-1)(1-\varepsilon)}\right). \tag{A35}$$

One can confirm that this guarantees that

$$I_{\Phi_B(\Psi)}(A:C|B) \leq c_\varepsilon \gamma^{|B|(1-\varepsilon)} \tag{A36}$$

for all $|B| = 1,2,\ldots$. The resulting constant c_ε is independent of $|A|$, $|B|$, $|C|$, $|L\rangle$ and $|R\rangle$. By combining (A36) with

$$I_{p_\Psi}(A:C|B) = I_{\Phi_\Lambda(\Psi)}(A:C|B) \leq I_{\Phi_B(\Psi)}(A:C|B), \tag{A37}$$

we obtain

$$I_{p_\Psi}(A:C|B) \leq c_\varepsilon \gamma^{|B|(1-\varepsilon)}. \tag{A38}$$

□

By combining Lemma 2 with Theorem 18 (for the inequality $I_{p_\Psi}(A:C|B) \leq c_\varepsilon \gamma^{|B|(1-\varepsilon)}$), and defining $\kappa := \gamma^{1-\varepsilon}$ and $c := c_\varepsilon$ for some arbitrary but fixed $0 < \varepsilon < 1$, we get the following.

Corollary 19. *For a set of operators $\{A_x\}_{x=0}^{d-1}$ on a Hilbert space $\mathscr{H}$ with $D := \dim \mathscr{H} \geq 2$, and normalized $|R\rangle, |L\rangle \in \mathscr{H}$, let Ψ be the MPS as defined in (1) on a region $\Lambda = ABC$. The set $\{A_x\}_{x=0}^{d-1}$ is such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$, satisfies the purity condition in Definition 3, and is such that $\mathbb{E}(\cdot) := \sum_{x=0}^{d-1} A_x \cdot A_x^\dagger$ is primitive. Let $p_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) = \langle x_\Lambda|\Psi|x_\Lambda\rangle$ be the classical restriction of Ψ, and assume that this restriction is such that $p_{1,\ldots,|\Lambda|}(x_1,\ldots,x_{|\Lambda|}) > 0$ for all $x_1,\ldots,x_{|\Lambda|}$. Let $p^\ell_{1,\ldots,|\Lambda|}$ be as defined in (19-21). Then, there exist constants, $0 \leq c$ and $0 < \kappa < 1$, such that*

$$S(p_{1,\ldots,|\Lambda|} \| p^\ell_{1,\ldots,|\Lambda|}) \leq c|\Lambda|\kappa^\ell, \quad 1 \leq \ell \leq |\Lambda| - 2. \tag{A39}$$

2. Proof of the second part of Theorem 4

In this section we prove the second part of Theorem 4, i.e., that the exponential decay of the quantum CMI $I_{\Phi_B(\Psi)}(A:C|B)$ implies purity of the set of matrices $\{A_x\}_{x=0}^{d-1}$ associated to $|\Psi\rangle$. In order to do this, we show several bounds that all combined will allow us to bound $f(N)$, and thus the quantum CMI.

The first step is to find a lower bound to the binary entropy.

Lemma 20. *Define $H_B : [0,1] \to \mathbb{R}$ by*

$$\begin{aligned} H_B(\lambda) :=& -\lambda\log\lambda - (1-\lambda)\log(1-\lambda), \quad 0 < \lambda < 1, \\ H_B(0) :=& 0, \quad H_B(1) := 0. \end{aligned} \tag{A40}$$

Then,

$$H_B(\lambda) \geq 4\log(2)\lambda(1-\lambda), \quad 0 \leq \lambda \leq 1. \tag{A41}$$

Proof. Define

$$f(\lambda) := \frac{H_B(\lambda)}{\lambda(1-\lambda)} = g(\lambda) + g(1-\lambda), \qquad g(\lambda) := -\frac{\log(1-\lambda)}{\lambda}. \tag{A42}$$

It is clear that f is symmetric around $\lambda = 1/2$, i.e., it is symmetric under the map $\lambda \mapsto 1-\lambda$. In the following, we will prove that g is a convex function on the interval $[0,1]$, with the consequence that f also is a convex function.

Consider the function $r(\lambda) := e^{-\lambda^2/2-\lambda}$. One can confirm that $r''(\lambda) \geq 0$ for all $\lambda \geq 0$, and thus r is convex on $[0,+\infty)$. Consequently, the tangent of r at $\lambda = 0$ is a lower bound to r on $[0,+\infty)$. Since the tangent at $\lambda = 0$ is $1-\lambda$, we can conclude that

$$e^{-\lambda^2/2-\lambda} \geq 1-\lambda, \quad \lambda \geq 0. \tag{A43}$$

Assuming that $1 > \lambda > 0$, we can rewrite (A43) as

$$g(\lambda) \geq \frac{\lambda}{2} + 1, \tag{A44}$$

with the result that

$$g''(\lambda) = -2\lambda^{-2}(1-\lambda)^{-1} + \lambda^{-1}(1-\lambda)^{-2} + 2\lambda^{-2}g(\lambda) \geq \frac{\lambda}{(1-\lambda)^2} \geq 0. \tag{A45}$$

Hence g is convex on $0 < \lambda < 1$. As mentioned above, it thus follows that f is a convex function. Since f moreover is symmetric under the map $\lambda \mapsto 1-\lambda$, it follows that the minimum is attained at $\lambda = 1/2$. Hence, $f(\lambda) \geq f(1/2) = 4\log 2$, which in turn yields (A41). □

While the above lemma bounds the binary entropy, the following lemma relates it to the von Neuman entropy of a density operator.

Lemma 21. *Let ρ be a density operator on a finite-dimensional Hilbert space. Then*

$$H_B\big(\lambda_1^{\downarrow}(\rho)\big) \leq S(\rho), \tag{A46}$$

where H_B is as defined in (A40), and $\lambda_1^{\downarrow}(\rho)$ denotes the largest eigenvalue of ρ.

Proof. Recall that a vector $\boldsymbol{a} \in \mathbb{R}^N$ majorizes a vector $\boldsymbol{b} \in \mathbb{R}^N$ (denoted $\boldsymbol{b} \prec \boldsymbol{a}$) if $\sum_{j=1}^k \boldsymbol{b}_j^{\downarrow} \leq \sum_{j=1}^k \boldsymbol{a}_j^{\downarrow}$ for $k = 1,\ldots,N-1$, and $\sum_{j=1}^N \boldsymbol{b}_j^{\downarrow} = \sum_{j=1}^N \boldsymbol{a}_j^{\downarrow}$, where $\boldsymbol{a}^{\downarrow}$ denotes the vector that we obtain by permuting the components of $\boldsymbol{a}$, such that they occur in a non-increasing order $\boldsymbol{a}_1^{\downarrow} \geq \boldsymbol{a}_2^{\downarrow} \geq \cdots \geq \boldsymbol{a}_N^{\downarrow}$, and analogous for $\boldsymbol{b}^{\downarrow}$. Also recall that if $\boldsymbol{a}$ and $\boldsymbol{b}$ can be regarded as probability distributions, then $\boldsymbol{b} \prec \boldsymbol{a}$ implies that the Shannon entropy of $\boldsymbol{a}$ is lower than the Shannon entropy of $\boldsymbol{r}$, i.e., $H(\boldsymbol{a}) \leq H(\boldsymbol{b})$ [43].

Let N be the dimension of the Hilbert space, and consider the vector of ordered eigenvalues $\boldsymbol{b} := \big(\lambda_1^{\downarrow}(\rho),\ldots,\lambda_N^{\downarrow}(\rho)\big)$. We also consider the vector $\boldsymbol{a} := \big(\lambda_1^{\downarrow}(\rho), 1-\lambda_1^{\downarrow}(\rho), 0, \ldots, 0\big)$. We notice that both $\boldsymbol{a}$ and $\boldsymbol{b}$ can be regarded as probability distributions. In both cases, $\lambda_1^{\downarrow}(\rho) \geq 1-\lambda_1^{\downarrow}(\rho)$ and $\lambda_1^{\downarrow}(\rho) < 1-\lambda_1^{\downarrow}(\rho)$, one can confirm that $\boldsymbol{b} \prec \boldsymbol{a}$, and consequently $H(\boldsymbol{a}) \leq H(\boldsymbol{b})$. The claim of the lemma follows by the observations that $H(\boldsymbol{a}) = H_B\big(\lambda_1^{\downarrow}(\rho)\big)$ and $H(\boldsymbol{b}) = S(\rho)$. □

Next, we consider a collection of density operators and use Lemmas 20 and 21 to bound the average von Neumann entropy of these density operators.

Lemma 22. *Let $\{\rho_x\}_{x=0}^{M-1}$ be density operators on a finite-dimensional Hilbert space, and let $\{p_x\}_{x=0}^{M-1}$ be such that $p_x \geq 0$ and $\sum_{x=0}^{M-1} p_x = 1$. Then,*

$$\sum_{x=0}^{M-1} p_x \sqrt{\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x)} \leq \frac{1}{2\sqrt{\log(2)}} \sqrt{\sum_{x=0}^{M-1} p_x S(\rho_x)}. \tag{A47}$$

Proof. We first note that due to the convexity of $x \mapsto x^2$, we have

$$\begin{aligned} \left(\sum_{x=0}^{M-1} p_x \sqrt{\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x)}\right)^2 &\leq \sum_{x=0}^{M-1} p_x \left(\sqrt{\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x)}\right)^2, \\ &= \sum_{x=0}^{M-1} p_x \lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x). \end{aligned} \tag{A48}$$

Next we note that $1-\lambda_1^{\downarrow}(\rho_x)=\sum_{k=2}^{D}\lambda_k^{\downarrow}(\rho_x)\geq\lambda_2^{\downarrow}(\rho_x)$, which implies $\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x)\leq\lambda_1^{\downarrow}(\rho_x)\big(1-\lambda_1^{\downarrow}(\rho_x)\big)$. This, combined with (A48), yields

$$\begin{aligned}
4\log(2)\left(\sum_{x=0}^{M-1}p_x\sqrt{\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x)}\right)^2 \leq & 4\log(2)\sum_{x=0}^{M-1}p_x\lambda_1^{\downarrow}(\rho_x)\lambda_2^{\downarrow}(\rho_x),\\
&\leq\sum_{x=0}^{M-1}p_x 4\log(2)\lambda_1^{\downarrow}(\rho_x)\Big(1-\lambda_1^{\downarrow}(\rho_x)\Big),\\
&\quad\text{[By Lemma 20]}\\
&\leq\sum_{x=0}^{M-1}p_x H_B\big(\lambda_1^{\downarrow}(\rho_x)\big),\\
&\quad\text{[By Lemma 21]}\\
&\leq\sum_{x=0}^{M-1}p_x S(\rho_x),
\end{aligned}\tag{A49}$$

which implies (A47). □

The last result that we need before stating the second part of Theorem 4 is a bound on the function $f(N)$ in terms of the average entropy, which is obtained in the following lemma.

Proposition 23. *For a set of operators $\{A_x\}_{x=0}^{d-1}$ on a Hilbert space $\mathscr{H}$ with $D:=\dim\mathscr{H}\geq 2$, and normalized $|R\rangle,|L\rangle\in\mathscr{H}$, let Ψ be the MPS as defined in (1) on a region $\Lambda=ABC$. The set $\{A_x\}_{x=0}^{d-1}$ is such that $\sum_{x=0}^{d-1}A_x^{\dagger}A_x=\mathbb{1}$. Then,*

$$f(N)\leq\frac{1}{2\sqrt{\log(2)}}\sqrt{\sum_{x_B}p_\Psi(x_B)S[\Psi_C(x_b)]},\tag{A50}$$

where $x_B=x_1,\ldots,x_N$ and

$$f(N):=\sum_{x_B}v_1^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})v_2^{\downarrow}(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma}),\tag{A51}$$

and where $p_\Psi(x_B)$ is as defined in (A21), $\Psi_C(x_b)$ as in (A22), as well as F and σ as in (A23).

Proof. First recall that if $v_k^{\downarrow}$ denotes the k:th singular value in non-increasing order, and $\lambda_k^{\downarrow}$ denotes the k:th eigenvalue in non-decreasing order, then $v_k^{\downarrow}(Q)=\sqrt{\lambda_k^{\downarrow}(Q^{\dagger}Q)}$. With this observation in mind, we find that

$$\begin{aligned}
f(N)=&\sum_{x_B}\sqrt{\lambda_1^{\downarrow}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^{\dagger}\cdots A_{x_N}^{\dagger}F^{\dagger}\right)\lambda_2^{\downarrow}\left(FA_{x_N}\cdots A_{x_1}\sigma A_{x_1}^{\dagger}\cdots A_{x_N}^{\dagger}F^{\dagger}\right)},\\
=&K^2\sum_{x_B}p_\Psi(x_B)\sqrt{\lambda_1^{\downarrow}\big(\Psi_C(x_b)\big)\lambda_2^{\downarrow}\big(\Psi_C(x_b)\big)},\\
&\text{[By Lemma 22]}\\
\leq&K^2\frac{1}{2\sqrt{\log(2)}}\sqrt{\sum_{x_B}p_\Psi(x_B)S[\Psi_C(x_b)]},\\
=&\operatorname{Tr}\left(F^{\dagger}F\mathbb{E}^N(\sigma)\right)\frac{1}{2\sqrt{\log(2)}}\sqrt{\sum_{x_B}p_\Psi(x_B)S[\Psi_C(x_b)]},\\
&\left[\quad F^{\dagger}F\leq 1,\quad \mathbb{E}^N(\sigma)\quad\text{is a density operator since }\mathbb{E}\text{ is a channel}\quad\right]\\
\leq&\frac{1}{2\sqrt{\log(2)}}\sqrt{\sum_{x_B}p_\Psi(x_B)S[\Psi_C(x_b)]}.
\end{aligned}\tag{A52}$$

□

Finally, in the theorem below we use all bounds derived previously to prove that an exponential decay of $I_{\Phi_B(\Psi)}(A:C|B)$ implies purity. One may note that Theorem 4 assumes that the MPS is injective, while this is not strictly speaking needed for Theorem 24, where the relevant assumption rather is that $\sigma:=\mathbb{E}^{|A|}(L)$, and $F^{\dagger}F=\mathbb{E}^{*|C|}(R)$ are full rank operators.

Theorem 24. *For a set of operators $\{A_x\}_{x=0}^{d-1}$, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$, on a Hilbert space $\mathscr{H}$ with $D := \dim \mathscr{H} \geq 2$, and normalized $|R\rangle, |L\rangle \in \mathscr{H}$, let Ψ be the MPS as defined in (1) on a region $\Lambda = ABC$. Suppose that $|R\rangle$, $|L\rangle$, $|A|$, and $|C|$ are such that $\sigma := \mathbb{E}^{|A|}(L)$, and $F^\dagger F = \mathbb{E}^{*|C|}(R)$ are full rank operators, where $\mathbb{E}(\cdot) := \sum_{x=0}^{d-1} A_x \cdot A_x^\dagger$. Moreover suppose that there exist constants $\tilde{c}$ and $0 \leq \tilde{\kappa} < 1$, such that*

$$I_{\Phi_B(\Psi)}(A:C|B) \leq \tilde{c}\tilde{\kappa}^{|B|}, \quad |B| = 1, 2, \ldots. \tag{A53}$$

Then, $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3.

Proof. We first note that

$$I_{\Phi_B(\Psi)}(A:C|B) = 2\sum_{x_B} p_\Psi(x_B) S[\Psi_C(x_b)], \tag{A54}$$

with $p_\Psi(x_B)$ as in (A21) and $\Psi_C(x_b)$ as in (A22). By comparing with Proposition 23 we can thus conclude that

$$f(N) \leq \frac{1}{2\sqrt{2\log(2)}} \sqrt{I_{\Phi_B(\Psi)}(A:C|B)}, \tag{A55}$$

with $f(N)$ as in (A51). By assumption, there exist constants $\tilde{c}$ and $0 \leq \tilde{\kappa} < 1$, such that (A53) holds. Combining (A53) and (A55) yields $f(N) \leq c\gamma^{|B|}$, with $c := \frac{1}{2\sqrt{2\log(2)}}\tilde{c}$, and $\gamma := \tilde{\kappa}^{1/2}$. Since we moreover assume that $\sigma := \mathbb{E}^{|A|}(L)$, and $F^\dagger F = \mathbb{E}^{*|C|}(R)$ are full rank operators, we can conclude that the conditions of Proposition 5 are satisfied, and thus $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3. □

We can conclude from Theorems 18 and 24 that purity in all essence is a necessary and sufficient condition for the exponential decay of $I_{\Phi_B(\Psi)}(A:C|B)$, and thus the typical post-measurement state $\Psi(x_B)$ loosely speaking approaches a pure product state. Concerning the classical CMI $I_{p_\Psi}(A:C|B)$, purity is only stated as a sufficient condition for exponential decay. In relation to the question whether purity also is necessary, one can note that for a normalized $|\psi\rangle \in \mathscr{H}_A \otimes \mathscr{H}_B \otimes \mathscr{H}_C$, it is the case that

$$I_{p_\psi}(A:C|B) = I_{\Phi_\Lambda(\psi)}(A:C|B) = 0 \tag{A56}$$

if and only if $\Phi_B(|\psi\rangle\langle\psi|)$ can be written

$$\begin{aligned} \Phi_B(|\psi\rangle\langle\psi|) &= \sum_{x_B} p^B(x_B) |\chi_{x_B}\rangle\langle\chi_{x_B}|, \\ |\chi_{x_B}\rangle &= \sum_{x_A, x_C} e^{i\theta(x_A, x_B, x_C)} \sqrt{p^A_{x_B}(x_A)} \sqrt{p^C_{x_B}(x_C)} |x_A\rangle|x_C\rangle, \end{aligned} \tag{A57}$$

where $\{p^B(x_B)\}$ is a probability distribution, and for each x_B it is the case that $\{p^A_{x_B}(x_A)\}_{x_A}$ and $\{p^C_{x_B}(x_C)\}_{x_C}$ are probability distributions, and $\theta(x_A, x_B, x_C) \in \mathbb{R}$. One can also realize that the states $|\chi_{x_B}\rangle$ in some sense are typically not product states, because of the arbitrary phase factors $e^{i\theta(x_A, x_B, x_C)}$, and thus $I_{\Phi_B(\psi)}(A:C|B) \neq 0$. This appears to leave room for the possibility that there may exist cases of exponential decay of $I_{p_\psi}(A:C|B)$, even though $I_{\Phi_B(\psi)}(A:C|B)$ does not decay, and that there thus may exist a weaker condition than purity for the exponential decay of $I_{p_\psi}(A:C|B)$. However, if this indeed can happen for MPSs with a fixed finite-dimensional virtual space, is an open question which we leave for future investigations.

Appendix B: Notions from probability theory

As mentioned in the main text, and in the proof overview, the proof of Proposition 5 relies on various probabilistic concepts. Here, we briefly review the pertinent notions, and also collect the technical results that we will need at various points along the proof. Throughout these derivations we will use bold letters, such as $\boldsymbol{x}$, $\boldsymbol{Y}$, etc, to denote random variables and random operators (where 'random variables' by default are real-valued measurable functions on the underlying probability space, while 'random operators' are operator-valued measurable functions). In the following $E(\boldsymbol{x})$, $E(\boldsymbol{y})$, etc, denote the expectation value, and $E(\boldsymbol{y}|\boldsymbol{x})$ denotes the expectation value of $\boldsymbol{y}$ conditioned on $\boldsymbol{x}$. One should keep in mind that $E(\boldsymbol{y}|\boldsymbol{x})$ is a random variable (due to $\boldsymbol{x}$). One should also keep in mind the general relation $E\big(E(\boldsymbol{y}|\boldsymbol{x})\big) = E(\boldsymbol{y})$.

1. Almost surely

When we say that a relation for one, or several, random variables holds almost surely (*a.s.*), it means that the relation is true apart from a set of probability zero. Put differently, the relation is true with probability one. For example, $\boldsymbol{x} = \boldsymbol{y}$ *a.s.* means that $P(\{\omega \in \Omega : \boldsymbol{x}(\omega) = \boldsymbol{y}(\omega)\}) = 1$, where Ω denotes the underlying sample space, and ω an element of the sample space.

As examples, one can consider various notions that intuitively remain true even if 'a few' points are excluded. For example, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are such that $\boldsymbol{x} \leq \boldsymbol{y}$ then (if the expectations exist) $E(\boldsymbol{x}) \leq E(\boldsymbol{y})$. This conclusion still holds, even if the inequality only holds almost everywhere (see e.g. Theorem 4.4 in chapter 2 of Ref. 44).

Lemma 25. *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are non-negative random variables, then $\boldsymbol{x} \leq \boldsymbol{y}$ a.s implies $E(\boldsymbol{y}) \leq E(\boldsymbol{x})$.*

Another statement in a similar spirit is the following. (The claim of the lemma is contained in Theorem 4.4 in chapter 2 of Ref. 44.)

Lemma 26. *If $\boldsymbol{x}$ is a non-negative random variable, then $\boldsymbol{x} = 0$ a.s. if and only if $E(\boldsymbol{x}) = 0$.*

For a random variable $\boldsymbol{x}$, we define the positive and negative components by $\boldsymbol{x}^+ := \max(\boldsymbol{x}, 0)$, $\boldsymbol{x}^- := \max(-\boldsymbol{x}, 0)$. By this construction, it is the case that $\boldsymbol{x}^+ \geq 0$, $\boldsymbol{x}^- \geq 0$, $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-$. The expectation value $E(\boldsymbol{x})$ of a random variable, $\boldsymbol{x}$, is defined as $E(\boldsymbol{x}) = E(\boldsymbol{x}^+) - E(\boldsymbol{x}^-)$ if at least one of $E(\boldsymbol{x}^+)$ and $E(\boldsymbol{x}^-)$ is finite.

Lemma 27. *If $\boldsymbol{x}$ is a random variable such that $\boldsymbol{x} \geq 0$ a.s. and $E(\boldsymbol{x}) = 0$, then $\boldsymbol{x} = 0$ a.s.*

Proof. Since $\boldsymbol{x} \geq 0$ almost surely, we can conclude that $\boldsymbol{x}^- = 0$ almost surely. Since $\boldsymbol{x}^-$ by construction is a non-negative random variable, it follows by Lemma 26 that $E(\boldsymbol{x}^-) = 0$. We can conclude that $E(\boldsymbol{x})$ is well defined, and $E(\boldsymbol{x}) = E(\boldsymbol{x}^+) - E(\boldsymbol{x}^-)$ and thus $E(\boldsymbol{x}) = 0$ implies $E(\boldsymbol{x}^+) = E(\boldsymbol{x}^-) = 0$. Since $\boldsymbol{x}^+$ by construction is non-negative, Lemma 26 implies $\boldsymbol{x}^+ = 0$ *a.s.* We can thus conclude that $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^- = 0$ *a.s.* □

2. Stochastic convergence of real-valued sequences

A sequence of random variables $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ is said to converge almost surely to a random variable $\boldsymbol{x}_\infty$ (denoted $\lim_{N\to\infty} \boldsymbol{x}_N = \boldsymbol{x}_\infty$ *a.s.*) if

$$P(\{\omega \in \Omega : \lim_{N\to\infty} \boldsymbol{x}_N(\omega) = \boldsymbol{x}_\infty(\omega)\}) = 1. \tag{B1}$$

As mentioned above, ω is an element of the underlying sample space, Ω, and $\boldsymbol{x}_N(\omega)$ is a specific realization of the stochastic process. (If one thinks of an infinite sequence of coin-tosses, then $\boldsymbol{x}_N$ vaguely stands for all possible sequences of coin-tosses, while $\boldsymbol{x}_N(\omega)$ means a specific sequence of heads and tails.) What (B1) essentially says is that if we look at the set of all sequences $\boldsymbol{x}_N(\omega)$ and $\boldsymbol{x}_\infty(\omega)$, such that $\boldsymbol{x}_N(\omega)$ actually do converge to $\boldsymbol{x}_\infty(\omega)$, then this set has probability 1.

In these derivations, we will often start with a process that converges $\lim_{N\to\infty} \boldsymbol{x}_N = \boldsymbol{x}_\infty$ almost surely, but we want to show that $\lim_{N\to\infty} E(\boldsymbol{x}_N) = E(\boldsymbol{x}_\infty)$. This is not generally true, but as a consequence of Lebesgues dominated convergence theorem (see e.g., Theorem 5.3 in chapter 2 of Ref. 44) we have the following.

Proposition 28. *Suppose that $\boldsymbol{x}_N$, $\boldsymbol{x}_\infty$ and $\boldsymbol{y}$ are random variables such that $|\boldsymbol{x}_N| \leq \boldsymbol{y}$ for all N, where $E(\boldsymbol{y}) < +\infty$, and that $\boldsymbol{x}_N \to \boldsymbol{x}_\infty$ a.s., then*

$$\lim_{N\to\infty} E(\boldsymbol{x}_N) = E(\boldsymbol{x}_\infty). \tag{B2}$$

In the special case that $\boldsymbol{y}$ is equal to a constant C, one obtains the following special case (sometimes referred to as the bounded convergence theorem).

Proposition 29. *Suppose that $\boldsymbol{x}_N$, $\boldsymbol{x}_\infty$ are random variables, and there exists a constant $C < +\infty$, such that $|\boldsymbol{x}_N| \leq C$ for all N. If $\boldsymbol{x}_N \to \boldsymbol{x}_\infty$ a.s., then*

$$\lim_{N\to\infty} E(\boldsymbol{x}_N) = E(\boldsymbol{x}_\infty). \tag{B3}$$

The following lemma is a consequence of the Borel-Cantelli Lemma (and is included in Theorem 3.1 in chapter 5 of Ref. 44).

Lemma 30. *Let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ and $\boldsymbol{x}$ be random variables such that $\sum_{n=1}^{\infty} P(|\boldsymbol{x}_n - \boldsymbol{x}| > \varepsilon) < +\infty$ for all $\varepsilon > 0$ (sometimes referred to as complete convergence of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ to $\boldsymbol{x}$). Then, $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges almost surely to $\boldsymbol{x}$.*

The above can be used to obtain the following.

Lemma 31. *Let $(\boldsymbol{r}_N)_{N\in\mathbb{N}}$ be a sequence of random variables, such that $\boldsymbol{r}_N \geq 0$, and such that the expectations values $E(\boldsymbol{r}_N)$ exist and are finite. Suppose that there exists a number R such that $\lim_{k\to\infty} E\Big(\sum_{N=1}^{k} \boldsymbol{r}_N\Big) = R < +\infty$, then $\lim_{N\to\infty} \boldsymbol{r}_N = 0$ a.s.*

Proof. We first want to note that if $(a_N)_{N\in\mathbb{N}}$ is a sequence of real numbers such that $a_N \geq 0$, and if $A_k := \sum_{N=1}^{k} a_N$ is such that $\lim_{k\to\infty} A_k = R < +\infty$, then we have $\lim_{N\to\infty} a_N = 0$. With $a_N := E(\boldsymbol{r}_N)$, and $A_k := \sum_{N=1}^{k} a_N = E\Big(\sum_{N=1}^{k} \boldsymbol{r}_N\Big)$, it thus follows, by the assumptions of the lemma, that $\lim_{N\to\infty} E(\boldsymbol{r}_N) = \lim_{N\to\infty} a_N = 0$. By assumption, $\boldsymbol{r}_N \geq 0$ and $E(\boldsymbol{r}_N)$ are well defined and finite. Hence, by Markov's inequality, it follows that $P(\boldsymbol{r}_N > \varepsilon) \leq E(\boldsymbol{r}_N)/\varepsilon$ for all $\varepsilon > 0$. Consequently, $\sum_{N=1}^{k} P(\boldsymbol{r}_N > \varepsilon) \leq 1/\varepsilon E\Big(\sum_{N=1}^{k} \boldsymbol{r}_N\Big)$, and thus

$$\sum_{N=1}^{\infty} P(|\boldsymbol{r}_N| > \varepsilon) \leq \frac{1}{\varepsilon} \lim_{N\to\infty} E\left(\sum_{N=1}^{k} \boldsymbol{r}_N\right) = \frac{R}{\varepsilon} < +\infty, \quad \forall \varepsilon > 0. \tag{B4}$$

Hence, $(\boldsymbol{r}_N)_{N\in\mathbb{N}}$ converges completely to 0. By Lemma 30, we can conclude that $(\boldsymbol{r}_N)_{N\in\mathbb{N}}$ converges almost surely to 0. □

3. Stochastic convergence of operator-valued sequences

Convergence of various sequences of operators play an important role in this investigation. Since we here exclusively will deal with finite-dimensional spaces, one may argue that the distinction between 'random variables' and 'random operators' is not very dramatic. For the sake of clarity, we will nevertheless throughout these derivations make a distinction of the these two types and, to further this, we will use small bold letters, $\boldsymbol{x}$, $\boldsymbol{y}$, etc to denote random variables, while capital bold letters $\boldsymbol{X}$, $\boldsymbol{Y}$, etc denote random operators.

Here, we briefly recall that, on finite-dimensional Hilbert spaces, all norms are metrically equivalent. Due to the metrical equivalence in finite dimensions (see e.g., Corollary 5.4.5 in Ref. 45), we do not need to make a distinction between different norms when we discuss convergences of sequences of operators, and we can equivalently consider the element-wise convergence of the elements of the matrix-representation in some arbitrary basis. In what follows we will switch between these equivalent manifestations of convergence without any further comments. For the operator norms, we will mainly be using the supremum norm, $\|O\| := \sup_{\||\psi\rangle\|=1} \|O|\psi\rangle\|$, and the trace-norm, $\|O\|_1 := \mathrm{Tr}\sqrt{O^\dagger O}$, but also the Hilbert-Schmidt norm, $\|O\|_2 := \sqrt{\mathrm{Tr}(O^\dagger O)}$.

Let us now consider a sequence of random operators $(\boldsymbol{X}_N)_{N\in\mathbb{N}}$ and $\boldsymbol{X}_\infty$ on a finite-dimensional Hilbert space. We interpret the convergence $\boldsymbol{X}_N \to \boldsymbol{X}_\infty$ *a.s.* as

$$\lim_{N\to\infty} \|\boldsymbol{X}_N - \boldsymbol{X}_\infty\| = 0 \quad a.s., \tag{B5}$$

or equivalently for any other operator norm (since the underlying Hilbert space is finite-dimensional), or as

$$\lim_{N\to\infty} \langle k|\boldsymbol{X}_N|k'\rangle = \langle k|\boldsymbol{X}_\infty|k'\rangle \quad a.s. \quad \forall k,k' = 1,\ldots,D. \tag{B6}$$

The following is a counterpart of Proposition 29, which can be obtained by applying Proposition 29 to the real and imaginary matrix components with respect to a basis, i.e., $\mathrm{Re}\langle k|\boldsymbol{X}_N|k'\rangle$ and $\mathrm{Im}\langle k|\boldsymbol{X}_N|k'\rangle$.

Proposition 32. *Suppose that $\boldsymbol{X}_N$ and $\boldsymbol{X}_\infty$ are random operators on a complex Hilbert space with finite dimension, and that there exists a constant $C < +\infty$, such that $\|\boldsymbol{X}_N\| \leq C$ for all N. If $\boldsymbol{X}_N \to \boldsymbol{X}_\infty$ a.s., then*

$$\lim_{N\to\infty} E(\boldsymbol{X}_N) = E(\boldsymbol{X}_\infty). \tag{B7}$$

4. Martingales

Our primary interest in martingales is that they allow for statements concerning the stochastic convergence of sequences of random variables. However, in order to connect to the manner that these convergence-theorems typically are phrased in the literature, we need to briefly discuss some technical concepts. (For a more thorough introduction, see, e.g., chapter 10 in Ref. 44.)

Consider a sequence of random variables $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ on a probability space $(\Omega, \mathscr{F}, P)$, where Ω is the sample space, $\mathscr{F}$ is a σ-algebra (the event space), and P a probability measure. We also consider a filtration, i.e., a non-decreasing sequence of σ-subalgebras $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$. A sequence $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ of random variables is said to be adapted to $(\mathscr{F}_N)_{N\in\mathbb{N}}$ if each $\boldsymbol{y}_N$ is measurable with respect to $\mathscr{F}_N$. A sequence $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ is a martingale with respect to $(\mathscr{F}_N)_{N\in\mathbb{N}}$ if $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ is adapted to

$(\mathscr{F}_N)_{N\in\mathbb{N}}$, satisfies $E(\boldsymbol{y}_{N+1}|\mathscr{F}_N)=\boldsymbol{y}_N$ *a.s.*, as well as $E(|\boldsymbol{y}_N|)<+\infty$. Intuitively, $\mathscr{F}_N$ stands for the information available to us at step N. In our setting, this information corresponds to variables $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ (which are assumed to also be random variables on the same underlying probability space $(\Omega,\mathscr{F},P)$). More precisely, $\mathscr{F}_N:=\sigma(\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1)$, which denotes the σ-algebra generated by $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ and often is referred to as the natural filtration of $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$. Since in the following we exclusively will use the natural filtrations, we will employ the more succinct notation $E(\boldsymbol{y}_{N+1}|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1):=E(\boldsymbol{y}_{N+1}|\mathscr{F}_N)$, with $\mathscr{F}_N:=\sigma(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$. We moreover say that $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ is a martingale with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ if

$$\begin{aligned} E(\boldsymbol{y}_{N+1}|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1)=&\boldsymbol{y}_N \quad a.s., \quad \text{and} \quad E(|\boldsymbol{y}_N|)<+\infty, \\ &\text{with} \quad \boldsymbol{y}_N=f_N(\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1), \end{aligned} \tag{B8}$$

for (Borel measurable) functions f_N. The construction with the functions f_N guarantees that $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ is adapted to the natural filtration of $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$. The following proposition is obtained as a special case of Theorem 12.1 in chapter 10 of Ref. 44.

Proposition 33. *Let $(\boldsymbol{y}_N)_{N\in\mathbb{N}}$ be a martingale with respect to another process $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$, and suppose that there exists a real number C such that $|\boldsymbol{y}_N|\leq C$ for all $N\in\mathbb{N}$, then there exists a random variable $\boldsymbol{y}_\infty$ such that*

$$\lim_{N\to\infty}\boldsymbol{y}_N=\boldsymbol{y}_\infty \quad a.s. \quad \textit{and} \quad E(|\boldsymbol{y}_\infty|)<+\infty. \tag{B9}$$

As a technical remark concerning the relation to Theorem 12.1 in chapter 10 of Ref. 44, one may note that the condition $|\boldsymbol{y}_N|\leq C$ implies that $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ is uniformly integrable.

Our main interest is not these 'standard' real-valued martingales, but rather operator-valued martingales. It is again worth recalling that we here only consider finite-dimensional spaces, and hence we can represent each operator as a finite matrix with respect to some choice of basis. With this in mind, we say that an operator-valued process $(\boldsymbol{Y}_N)_{N\in\mathbb{N}}$ on a finite-dimensional Hilbert space is an operator-valued martingale with respect to a stochastic process $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ if each of $\mathrm{Re}\langle k|\boldsymbol{Y}_N|k'\rangle$ and $\mathrm{Im}\langle k|\boldsymbol{Y}_N|k'\rangle$ are a martingale with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ for some fixed orthonormal basis $\{|k\rangle\}_{k=1}^D$. One may note that the condition that $E(|\mathrm{Re}\langle k|\boldsymbol{Y}_N|k'\rangle|)<+\infty$ and $E(|\mathrm{Im}\langle k|\boldsymbol{Y}_N|k'\rangle|)<+\infty$ in the finite-dimensional case is equivalent to

$$E(\|\boldsymbol{Y}_N\|)<+\infty. \tag{B10}$$

Similarly, the conditions

$$\begin{aligned} E\big(\mathrm{Re}\langle k|\boldsymbol{Y}_{N+1}|k'\rangle\big|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\big)&=\mathrm{Re}\langle k|\boldsymbol{Y}_{N+1}|k'\rangle \ \ a.s., \\ E\big(\mathrm{Im}\langle k|\boldsymbol{Y}_{N+1}|k'\rangle\big|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\big)&=\mathrm{Im}\langle k|\boldsymbol{Y}_{n+1}|k'\rangle \ \ a.s., \end{aligned} \tag{B11}$$

can equivalently be stated as

$$E(\boldsymbol{Y}_{N+1}|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1)=\boldsymbol{Y}_N \ \ a.s. \tag{B12}$$

In a similar manner, Proposition 33 can be applied to the real and imaginary components of an operator-valued martingale, which yields the following 'operator counterpart' to Proposition 33.

Proposition 34. *Let $(\boldsymbol{Y}_N)_{N\in\mathbb{N}}$ be an operator-valued martingale on a finite-dimensional complex Hilbert space with respect to a real-valued process $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$, and suppose that there exists real number C such that $\|\boldsymbol{Y}_N\|\leq C$ for all $N\in\mathbb{N}$, then there exists a random operator, $\boldsymbol{Y}_\infty$, such that*

$$\lim_{N\to\infty}\boldsymbol{Y}_N=\boldsymbol{Y}_\infty \quad a.s. \quad \textit{and} \quad E(\|\boldsymbol{Y}_\infty\|)<+\infty. \tag{B13}$$

Appendix C: Stochastic process of measurements

Consider a set of operators $\{A_x\}_{x=0}^{d-1}$ on a finite-dimensional Hilbert space, $\mathscr{H}$, with dimension $D:=\dim\mathscr{H}$, such that $\sum_{x=0}^{d-1}A_x^\dagger A_x=\mathbb{1}$. We introduce the stochastic process $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ with a joint distribution such that, for each N, the marginal distribution of $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ is given by

$$P(\boldsymbol{x}_N=x_N,\ldots,\boldsymbol{x}_1=x_1)=\frac{1}{D}\,\mathrm{Tr}\left(A_{x_N}\cdots A_{x_1}A_{x_1}^\dagger\cdots A_{x_N}^\dagger\right). \tag{C1}$$

This means that $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N$ can be interpreted as the outcomes of a sequence of measurements, where the initial state is maximally mixed, i.e., $\sigma=\mathbb{1}/D$. Note that when we in the following refer to an expectation value, the underlying probability distribution is assumed to be (C1) unless otherwise stated. It will be useful to note that

$$P(\boldsymbol{x}_{N+1}=x_{N+1}|\boldsymbol{x}_N=x_N,\ldots,\boldsymbol{x}_1=x_1)=\frac{\mathrm{Tr}(A_{x_1}^\dagger\cdots A_{x_N}^\dagger A_{x_{N+1}}^\dagger A_{x_{N+1}}A_{x_N}\cdots A_{x_1})}{\mathrm{Tr}(A_{x_1}^\dagger\cdots A_{x_N}^\dagger A_{x_N}\cdots A_{x_1})}, \tag{C2}$$

where $P(y|x) := P(y,x)/P(x)$ denotes the conditional probability.

Based on the process $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$, we define sequences of random operators

$$\begin{aligned}
&\boldsymbol{A}_N := A_{\boldsymbol{x}_N},\\
&\boldsymbol{W}_N := \boldsymbol{A}_N \cdots \boldsymbol{A}_1 = A_{\boldsymbol{x}_N} \cdots A_{\boldsymbol{x}_1},\\
&\boldsymbol{M}_N := \begin{cases} \frac{\boldsymbol{W}_N^\dagger \boldsymbol{W}_N}{\mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N)} & \text{if } \mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N) \neq 0,\\ 0 & \text{if } \mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N) = 0. \end{cases}
\end{aligned} \tag{C3}$$

One should note that $\boldsymbol{M}_N$ is Hermitian, i.e., $\boldsymbol{M}_N^\dagger = \boldsymbol{M}_N$. Moreover, $\boldsymbol{M}_N$ is positive semi-definite, and either has trace 1 or trace 0. Hence, $\boldsymbol{M}_N$ is either a density operator, or the zero operator. One may further note that

$$\mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N) = \mathrm{Tr}(A_{\boldsymbol{x}_1}^\dagger \cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N} \cdots A_{\boldsymbol{x}_1}) = P(\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) D, \tag{C4}$$

from which we can conclude that $\boldsymbol{M}_N = 0$ with probability zero, i.e., $\boldsymbol{M}_N$ is almost surely a density operator.

As a side-remark, one might note that $\boldsymbol{M}_N$ is *not* the post-measurement state of the measurement process. The post-measurement state would rather be $\boldsymbol{\rho}_N := \boldsymbol{W}_N \boldsymbol{W}_N^\dagger / \mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N)$. However, $\boldsymbol{M}_N$ and $\boldsymbol{\rho}_N$ have the same non-zero eigenvalues (as can be seen by a singular value decomposition of $\boldsymbol{W}_N$). The main reason for why it is convenient to use $\boldsymbol{M}_N$, rather than $\boldsymbol{\rho}_N$, is that on $\boldsymbol{M}_N$ we can directly utilize $\sum_x A_x^\dagger A_x = \mathbb{1}$, which for example is used in the proof of the martingale property in Lemma 35.

In the following it will be useful to observe that since $\boldsymbol{A}_N$ is a (deterministic) function of $\boldsymbol{x}_N$ [as seen by (C3)] it is the case that

$$E(\boldsymbol{A}_N | \boldsymbol{x}_N) = A_{\boldsymbol{x}_N} = \boldsymbol{A}_N. \tag{C5}$$

Analogously, $E(\boldsymbol{W}_N | \boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = \boldsymbol{W}_N$, and similarly $E(\boldsymbol{M}_N | \boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = \boldsymbol{M}_N$.

Appendix D: Elements of the proof of Proposition 5

1. $\lim_{N\to\infty} \boldsymbol{M}_N = \boldsymbol{M}_\infty$ *a.s.*

The purpose of this section is to show that $\boldsymbol{M}_N$ has limit operator $\boldsymbol{M}_\infty$ in a sufficiently strong sense, and that this limit operator has 'nice' properties. We do this by first showing that $\boldsymbol{M}_N$ is a martingale relative to the sequence of measurement outcomes $\boldsymbol{x}_N$. This in turn yields almost sure convergence to limiting operator $\boldsymbol{M}_\infty$. Recall that the underlying probability distribution is assumed to be (C1), and that all expectations are taken with respect to this distribution.

Lemma 35. *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a finite-dimensional complex Hilbert space, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. Then, $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$, defined by (C3), is an operator-valued martingale with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ with distribution (C1).*

Proof. From the fact that each $\boldsymbol{M}_N$ is a density operator, or the zero operator, it follows that $\|\boldsymbol{M}_N\| \leq 1$, and thus in particular that $E(\|\boldsymbol{M}_N\|) \leq 1 < +\infty$. Moreover, by the construction in (C3), it is the case that $\boldsymbol{M}_N$ (and thus the matrix-elements with respect to a given basis) are functions of $\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1$. By (C2) and $\sum_{x_{N+1}=0}^{d-1} A_{x_{N+1}}^\dagger A_{x_{N+1}} = \mathbb{1}$ we find that

$$\begin{aligned}
&E(\boldsymbol{M}_{N+1} | \boldsymbol{x}_N = x_N, \ldots, \boldsymbol{x}_1 = x_1)\\
&\quad = \sum_{x_{N+1}} \frac{A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_{N+1}}^\dagger A_{x_{N+1}} A_{x_N} \cdots A_{x_1}}{\mathrm{Tr}(A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_{N+1}}^\dagger A_{x_{N+1}} A_{x_N} \cdots A_{x_1})} P(\boldsymbol{x}_{N+1} = x_{N+1} | \boldsymbol{x}_N = x_N, \ldots, \boldsymbol{x}_1 = x_1),\\
&\quad = E(\boldsymbol{M}_N | \boldsymbol{x}_N = x_N, \ldots, \boldsymbol{x}_1 = x_1).
\end{aligned}$$

We can conclude that $E(\boldsymbol{M}_{N+1} | \boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = E(\boldsymbol{M}_N | \boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = \boldsymbol{M}_N$. Hence, $(\boldsymbol{M}_N)_N$ is a martingale sequence with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$. □

Lemma 36. *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a finite-dimensional complex Hilbert space, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. Let $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ be as defined in (C3) with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ and distributed as in (C1). Then, there exists a random operator, $\boldsymbol{M}_\infty$,*

such that

$$\lim_{N\to\infty} \boldsymbol{M}_N = \boldsymbol{M}_\infty \quad a.s., \tag{D1}$$

$$\boldsymbol{M}_\infty \text{ is almost surely a density operator,} \tag{D2}$$

$$\lim_{N\to\infty} E(\boldsymbol{M}_N) = E(\boldsymbol{M}_\infty), \tag{D3}$$

$$\lim_{N\to\infty} E(\|\boldsymbol{M}_N\|) = E(\|\boldsymbol{M}_\infty\|), \tag{D4}$$

$$E(\|\boldsymbol{M}_\infty\|) < +\infty. \tag{D5}$$

Proof. By Lemma 35 we know that $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ is a martingale with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$. From the fact that each $\boldsymbol{M}_N$ is a density operator, or the zero operator, it follows that

$$\|\boldsymbol{M}_N\| \leq 1, \quad \forall N \in \mathbb{N}. \tag{D6}$$

By Proposition 34 it follows that there exists a random operator, $\boldsymbol{M}_\infty$, such that

$$\lim_{N\to\infty} \boldsymbol{M}_N = \boldsymbol{M}_\infty \quad a.s., \tag{D7}$$

with $E(\|\boldsymbol{M}_\infty\|) < +\infty$. By combining (D7) with (D6), Proposition 32 yields $E(\boldsymbol{M}_N) \to E(\boldsymbol{M}_\infty)$. Moreover, (D7) yields $\lim_{N\to\infty} \|\boldsymbol{M}_N\| = \|\boldsymbol{M}_\infty\|$ *a.s.* By this observation together with (D6), Proposition 29 with $\boldsymbol{x}_N := \|\boldsymbol{M}_N\|$ and $\boldsymbol{x}_\infty := \|\boldsymbol{M}_\infty\|$ yields $E(\|\boldsymbol{M}_N\|) \to E(\|\boldsymbol{M}_\infty\|)$. Finally, we should show that $\boldsymbol{M}_\infty$ almost surely is a density operator, i.e., that $\boldsymbol{M}_\infty \geq 0$ almost surely, and that $\operatorname{Tr}\boldsymbol{M}_\infty = 1$ almost surely. From (D7) it follows that $\lim_{N\to\infty}\langle\psi|\boldsymbol{M}_N|\psi\rangle = \langle\psi|\boldsymbol{M}_\infty|\psi\rangle$ *a.s.* Since $\langle\psi|\boldsymbol{M}_N|\psi\rangle \geq 0$, it follows that $\langle\psi|\boldsymbol{M}_\infty|\psi\rangle \geq 0$ *a.s.* Analogously, since $\operatorname{Tr}\boldsymbol{M}_N = 1$ almost surely, it follows that $\lim_{N\to\infty}\operatorname{Tr}\boldsymbol{M}_N = \operatorname{Tr}\boldsymbol{M}_\infty = 1$ *a.s.* Hence, $\boldsymbol{M}_\infty$ is almost surely a density operator. □

2. If $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition, then $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$ *a.s.*

The purpose of this section is to show that the limit operator $\boldsymbol{M}_\infty$, more or less always, is a rank-one operator whenever $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition. The first step (Lemma 37) is to show that the difference between the operators $\boldsymbol{M}_{N+p}$ and $\boldsymbol{M}_N$ tends to vanish as N increases, even when conditioned on all the measurement outcomes $\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1$.

Lemma 37. *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a finite-dimensional complex Hilbert space, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. Let $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ be as defined in (C3) with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ and distributed as in (C1). Then,*

$$\lim_{N\to\infty} E\Big(\|\boldsymbol{M}_{N+p} - \boldsymbol{M}_N\| \Big| \boldsymbol{x}_N, \ldots, \boldsymbol{x}_1\Big) = 0 \quad a.s. \tag{D8}$$

Proof. Recall that $\boldsymbol{M}_N$ is a deterministic function of $\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1$, and thus

$$E(\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = \boldsymbol{M}_N. \tag{D9}$$

A direct consequence is that $\boldsymbol{M}_N$ and $\boldsymbol{M}_{N+p}$ are independent when conditioned on $\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1$, and thus

$$E(\boldsymbol{M}_{N+p}\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1) = E(\boldsymbol{M}_{N+p}|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)E(\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1). \tag{D10}$$

By expanding $E\big((\boldsymbol{M}_{N+p} - \boldsymbol{M}_N)^2\big)$ one obtains cross-terms such as

$$\begin{aligned} E(\boldsymbol{M}_{N+p}\boldsymbol{M}_N) &= E\Big(E(\boldsymbol{M}_{N+p}\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)\Big), \\ &= E\Big(E(\boldsymbol{M}_{N+p}|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)E(\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)\Big), \end{aligned} \tag{D11}$$

where the last equality follows by the conditional independence in (D10). By combining these observations with the martingale property, as shown in Lemma 35, with (D9), $E\big((\boldsymbol{M}_{N+p} - \boldsymbol{M}_N)^2\big)$ results in

$$\begin{aligned} &E\big((\boldsymbol{M}_{N+p} - \boldsymbol{M}_N)^2\big) \\ &\quad = E(\boldsymbol{M}_{N+p}^2) + E(\boldsymbol{M}_N^2) - 2E\Big(E(\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)E(\boldsymbol{M}_N|\boldsymbol{x}_N, \ldots, \boldsymbol{x}_1)\Big), \\ &\quad = E(\boldsymbol{M}_{N+p}^2) - E(\boldsymbol{M}_N^2), \end{aligned} \tag{D12}$$

where we in the last step have used (D9).

By the observation that $\boldsymbol{M}_N$ is Hermitian, it follows that $\mathrm{Tr}E(\boldsymbol{M}_{N+p}^2) = E(\|\boldsymbol{M}_{N+p}\|_2^2)$, $\mathrm{Tr}E(\boldsymbol{M}_N^2) = E(\|\boldsymbol{M}_N\|_2^2)$ and $\mathrm{Tr}E\big((\boldsymbol{M}_{N+p}-\boldsymbol{M}_N)^2\big) = E\big(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2\big)$, which with (D12) yields

$$E(\|\boldsymbol{M}_{N+p}\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2) \geq 0. \tag{D13}$$

By using (D12), we next observe that

$$\begin{aligned}\sum_{N=0}^{p-1} E(\boldsymbol{M}_{N+k+1}^2) - \sum_{N=0}^{p-1} E(\boldsymbol{M}_N^2) &= \sum_{N=0}^{k} E(\boldsymbol{M}_{N+p}^2) - \sum_{N=0}^{k} E(\boldsymbol{M}_N^2), \\ &= \sum_{N=0}^{k} E\big((\boldsymbol{M}_{N+p}-\boldsymbol{M}_N)^2\big), \\ &= E\left(\sum_{N=0}^{k} E\Big((\boldsymbol{M}_{N+p}-\boldsymbol{M}_N)^2 \Big| \boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\Big)\right),\end{aligned} \tag{D14}$$

where we in the last step use the general relation $E\big(E(\boldsymbol{y}|\boldsymbol{x})\big) = E(\boldsymbol{y})$. Recall that $\|O\|_2^2 := \mathrm{Tr}(O^2)$. By applying the trace to (D14) we obtain

$$E\left(\sum_{N=0}^{k} E\Big(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2 \Big| \boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\Big)\right) = \sum_{N=0}^{p-1}\Big[E(\|\boldsymbol{M}_{N+k+1}\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2)\Big].$$

By Lemma 36 we know that $\lim_{N\to\infty}\boldsymbol{M}_N = \boldsymbol{M}_\infty$ almost surely. From this observation it follows that $\lim_{N\to\infty}\|\boldsymbol{M}_{N+k+1}\|_2^2 = \|\boldsymbol{M}_\infty\|_2^2$ *a.s.* Next, we note that $\boldsymbol{M}_N$ is a density operator, or the zero operator, and thus it follows that $\|\boldsymbol{M}_N\|_2^2 \leq 1$. With $\boldsymbol{x}_N := \|\boldsymbol{M}_N\|_2^2$ and $\boldsymbol{x}_\infty := \|\boldsymbol{M}_\infty\|_2^2$, it follows by Proposition 29 that

$$\lim_{N\to\infty} E(\|\boldsymbol{M}_{N+k+1}\|_2^2) = E(\|\boldsymbol{M}_\infty\|_2^2). \tag{D15}$$

By Lemma 36 we know that $\boldsymbol{M}_\infty$ is almost surely a density operator, from which it follows that $\|\boldsymbol{M}_\infty\|_2^2 \leq 1$ *a.s.* With $\boldsymbol{x} := \|\boldsymbol{M}_\infty\|_2^2$ and $\boldsymbol{y} := 1$ in Lemma 25, we get

$$E(\|\boldsymbol{M}_\infty\|_2^2) \leq 1. \tag{D16}$$

With $p := k+1$ in the inequality (D13), it follows that $E(\|\boldsymbol{M}_{N+k+1}\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2) \geq 0$, which implies $E(\|\boldsymbol{M}_\infty\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2) \geq 0$. Hence,

$$\begin{aligned}\lim_{k\to\infty} E\left(\sum_{N=0}^{k} E\Big(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2 \Big| \boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\Big)\right) &= \sum_{N=0}^{p-1}\Big[E(\|\boldsymbol{M}_\infty\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2)\Big], \\ &=: R(p).\end{aligned} \tag{D17}$$

By $E(\|\boldsymbol{M}_\infty\|_2^2) - E(\|\boldsymbol{M}_N\|_2^2) \geq 0$, it follows that $R(p) \geq 0$ and, by the inequality (D16), it follows that $R(p) \leq p < +\infty$. Define $\boldsymbol{r}_N := E(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2|\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1)$. Note that $\boldsymbol{r}_N \geq 0$. Moreover, $E(\boldsymbol{r}_N) = E(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2)$. Since $\boldsymbol{M}_N$ is either a density operator, or the zero operator, it follows that $\|\boldsymbol{M}_N\|_2 \leq 1$, and thus $\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2 \leq (\|\boldsymbol{M}_{N+p}\|_2 + \|\boldsymbol{M}_N\|_2)^2 \leq 4$. We conclude that $E(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2) \leq 4$, which together with $E(\boldsymbol{r}_N) = E(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2^2)$ yields $E(\boldsymbol{r}_N) \leq 4$. By Eq. (D17), there exists a number $R(p)$ such that $\lim_{k\to\infty} E(\sum_{N=0}^{k}\boldsymbol{r}_N) = R(p) < +\infty$. All the conditions of Lemma 31 are thus satisfied and it yields

$$\lim_{N\to\infty} E\Big(\|\boldsymbol{M}_{N+p}(\omega)-\boldsymbol{M}_N(\omega)\|_2^2 \Big| \boldsymbol{x}_N(\omega),\ldots,\boldsymbol{x}_1(\omega)\Big) = 0. \tag{D18}$$

Next, we note that $x \mapsto x^2$ is a convex function, and thus by Jensen's inequality $(E(X))^2 \leq E(X^2)$. By combining this observation with (D18) we obtain

$$\lim_{N\to\infty} E\Big(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|_2 \Big| \boldsymbol{x}_N,\ldots,\boldsymbol{x}_1\Big) = 0 \quad a.s. \tag{D19}$$

By the general relation between the supremum norm and the Hilbert-Schmidt norm, $\|R\| \leq \|R\|_2$, we get $E(\|\boldsymbol{R}\|) \leq E(\|\boldsymbol{R}\|_2)$, and thus (D19) yields (D8). □

The following lemma provides a reformulation of the purity condition that is better suited for the proof-technique that we employ.

Lemma 38. *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a complex finite-dimensional Hilbert space, $\mathscr{H}$. Then, $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition if and only if the following condition holds:*

$$\begin{aligned}&\textit{If } O \textit{ is an operator on } \mathscr{H} \textit{ such that}\\ &O^\dagger A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} O \propto O^\dagger O, \quad \forall N \in \mathbb{N}, \quad \forall (x_1,\ldots,x_N) \in \{0,\ldots,d-1\}^{\times N},\\ &\textit{then } \operatorname{rank}(O) = 1.\end{aligned} \tag{D20}$$

Proof. We start proving the direction that, if $\{A_x\}_{x=0}^{d-1}$ satisfies condition (D20), then $\{A_x\}_{x=0}^{d-1}$ also satisfies the purity condition. Suppose that condition (D20) holds. For the subset of operators $O = P$ for projectors P, we thus find that condition (30) holds, and hence $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition.

Conversely, we wish to show that, if $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition, then $\{A_x\}_{x=0}^{d-1}$ also satisfies condition (D20). Hence, assume that $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition. Let O be any operator on $\mathscr{H}$ such that

$$O^\dagger A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} O \propto O^\dagger O \tag{D21}$$

for all N and all $x_1,\ldots,x_N$. We next note that $OO^\dagger$ is positive semi-definite, and let $(OO^\dagger)^\ominus$ denote the inverse on the support of $OO^\dagger$, such that $(OO^\dagger)^\ominus OO^\dagger = OO^\dagger (OO^\dagger)^\ominus = P$, where P is the projector onto the support of $OO^\dagger$. Multiplying (D21) from the left with $(OO^\dagger)^\ominus O$ and from the right with $O^\dagger (OO^\dagger)^\ominus$ results in $P A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} P \propto P$. Since the purity condition is assumed to hold, it follows that $\operatorname{rank}(P) = 1$. However, $\operatorname{rank}(P) = \operatorname{rank}(OO^\dagger) = \operatorname{rank}(O)$. We can thus conclude that if $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition, then $\{A_x\}_{x=0}^{d-1}$ also satisfies condition (D20). □

In the following lemma we use the convergence in (D19) to show that $\boldsymbol{M}_\infty$ almost surely is a rank-one operator. A key-step in the proof is the equality (D23) below, which with (D19) and the observation that $\|\sqrt{\boldsymbol{M}_N}\| \leq 1$ yields the limit in (D24). Given that we know that $\lim_{N\to\infty} \boldsymbol{M}_N = \boldsymbol{M}_\infty$ *a.s.*, it seems reasonable that we in the limit $N \to \infty$ obtain the proportionality in (D28). The latter does via Lemma 38 imply the desired result that $\boldsymbol{M}_\infty$ almost surely is a rank-one operator. However, there is a complication to this reasoning, namely the sequence of unitary operators, $\boldsymbol{U}_N$. These unitary operators are the result of a polar decomposition of the operators $A_{x_N} \cdots A_{x_1} / \sqrt{\operatorname{Tr}(A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1})}$, and we have very little control of the sequence $(\boldsymbol{U}_N)_{N\in\mathbb{N}}$, and in particular whether it possesses a limit $\boldsymbol{U}_\infty$. However, we can mend this issue by using the fact that the set of unitary operators on a finite-dimensional Hilbert space is sequentially compact. Recall that a topological space, C, is sequentially compact if, for every sequence $(x_j)_{j\in\mathbb{N}} \subset C$, there exists a subsequence $(x_{j_k})_{k\in\mathbb{N}}$ such that x_{j_k} converges to an element in C. On a finite-dimensional complex Hilbert space with dimension D, the set of unitary operators, $U(D)$, forms a sequentially compact (as well as compact) space. Hence, whenever we have a sequence $(U_j)_{j\in\mathbb{N}}$ in $U(D)$, then there exists a subsequence $(U_{j_k})_{k\in\mathbb{N}}$ such that U_{j_k} converges to an element in $U(D)$.

Lemma 39. *With the assumptions in Lemma 36, let $\boldsymbol{M}_\infty$ be the random operator guaranteed by Lemma 36. If $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition, then*

$$\operatorname{rank}(\boldsymbol{M}_\infty) = 1 \quad a.s. \tag{D22}$$

Proof. In order to prove this lemma, let us start defining $M_{x_N,\ldots,x_1} := A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1} / \operatorname{Tr}(A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N} \cdots A_{x_1})$, and thus we have $\boldsymbol{M}_N = M_{\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1}$. With a unitary operator $U_{x_N,\ldots,x_1}$, we make a polar decomposition such that $U_{x_N,\ldots,x_1}\sqrt{M_{x_N,\ldots,x_1}} = A_{x_N}\cdots A_{x_1}/\sqrt{\operatorname{Tr}(A_{x_1}^\dagger \cdots A_{x_N}^\dagger A_{x_N}\cdots A_{x_1})}$, and define $\boldsymbol{U}_N := U_{\boldsymbol{x}_N,\ldots,\boldsymbol{x}_1}$. Then, we have

$$\begin{aligned}
&E\Big(\|\boldsymbol{M}_{N+p} - \boldsymbol{M}_N\| \Big| \boldsymbol{x}_N = x_N, \ldots, \boldsymbol{x}_1 = x_1\Big)\\
&\quad= \sum_{x_{N+p},\ldots,x_{N+1}} \Big\| \sqrt{M_{x_N,\ldots,x_1}} U_{x_N,\ldots,x_1}^\dagger A_{x_{N+1}}^\dagger \cdots A_{x_{N+p}}^\dagger A_{x_{N+p}} \cdots A_{x_{N+1}} U_{x_N,\ldots,x_1} \sqrt{M_{x_N,\ldots,x_1}}\\
&\qquad - M_{x_N,\ldots,x_1} \operatorname{Tr}(A_{x_{N+1}}^\dagger \cdots A_{x_{N+p}}^\dagger A_{x_{N+p}} \cdots A_{x_{N+1}} U_{x_N,\ldots,x_1} M_{x_N,\ldots,x_1} U_{x_N,\ldots,x_1}^\dagger)\Big\|,\\
&\quad= E\left(\sum_{x'_p,\ldots,x'_1} \Big\| \sqrt{\boldsymbol{M}_N} \boldsymbol{U}_N^\dagger A_{x'_1}^\dagger \cdots A_{x'_p}^\dagger A_{x'_p} \cdots A_{x'_1} \boldsymbol{U}_N \sqrt{\boldsymbol{M}_N}\right.\\
&\qquad \left. - \boldsymbol{M}_N \operatorname{Tr}(A_{x'_1}^\dagger \cdots A_{x'_p}^\dagger A_{x'_p} \cdots A_{x'_1} \boldsymbol{U}_N \boldsymbol{M}_N \boldsymbol{U}_N^\dagger)\Big\| \;\Big|\; \boldsymbol{x}_N = x_N, \ldots, \boldsymbol{x}_1 = x_1\right),
\end{aligned}$$

where we in the second equality have renamed the indices $x_{N+1},\dots,x_{N+p}$ to $x'_1,\dots,x'_p$. Consequently,

$$E\Big(\|\boldsymbol{M}_{N+p}-\boldsymbol{M}_N\|\Big|\boldsymbol{x}_N,\dots,\boldsymbol{x}_1\Big)=\sum_{x'_p,\dots,x'_1}\Big\|\sqrt{\boldsymbol{M}_N}\boldsymbol{U}_N^\dagger A_{x'_1}^\dagger\cdots A_{x'_p}^\dagger A_{x'_p}\cdots A_{x'_1}\boldsymbol{U}_N\sqrt{\boldsymbol{M}_N} - \boldsymbol{M}_N\operatorname{Tr}(A_{x'_1}^\dagger\cdots A_{x'_p}^\dagger A_{x'_p}\cdots A_{x'_1}\boldsymbol{U}_N\boldsymbol{M}_N\boldsymbol{U}_N^\dagger)\Big\|, \tag{D23}$$

where we have used that $\boldsymbol{M}_N$ and $\boldsymbol{U}_N$ are deterministic functions of $\boldsymbol{x}'_N,\dots,\boldsymbol{x}'_1$. Since $\boldsymbol{M}_N$ is a density operator, or the zero operator, it follows that $\|\sqrt{\boldsymbol{M}_N}\|\leq 1$. By combining this observation with (D23), and with Lemma 37, it follows that

$$\lim_{N\to\infty}\sum_{x_p,\dots,x_1}\Big\|\boldsymbol{M}_N\boldsymbol{U}_N^\dagger A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_N\boldsymbol{M}_N - \boldsymbol{M}_N^2\operatorname{Tr}(A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_N\boldsymbol{M}_N\boldsymbol{U}_N^\dagger)\Big\|=0\quad a.s. \tag{D24}$$

Next, we recall that Lemma 36 guarantees that $\lim_{N\to\infty}\boldsymbol{M}_N=\boldsymbol{M}_\infty$ *a.s.*, where $\boldsymbol{M}_\infty$ almost surely is a density operator. Let $\omega\in\Omega$ be such that $\lim_{N\to\infty}\boldsymbol{M}_N(\omega)=\boldsymbol{M}_\infty(\omega)$, where $\boldsymbol{M}_\infty(\omega)$ is a density operator, and the limit in (D24) holds. The latter implies a sequence of unitary operators $(\boldsymbol{U}_N(\omega))_{N\in\mathbb{N}}\subset U(D)$. By the sequential compactness of $U(D)$, it follows that there exists a subsequence $\big(\boldsymbol{U}_{N_k}(\omega)\big)_{k\in\mathbb{N}}$ and an element $\boldsymbol{U}_\infty(\omega)\in U(D)$, such that $\lim_{k\to\infty}\boldsymbol{U}_{N_k}(\omega)=\boldsymbol{U}_\infty(\omega)$. It still remains true that $\lim_{k\to\infty}\boldsymbol{M}_{N_k}(\omega)=\boldsymbol{M}_\infty(\omega)$, and similarly the limit in (D24) remains true with N replaced with N_k. With the definition

$$B_k(\omega):=\sum_{x_p,\dots,x_1}\Big\|\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}^\dagger(\omega)A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_{N_k}(\omega)\boldsymbol{M}_{N_k}(\omega) - \boldsymbol{M}_{N_k}^2(\omega)\operatorname{Tr}\big[A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_{N_k}(\omega)\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}(\omega)^\dagger\big]\Big\|,$$

it thus follows by (D24) that $B_k(\omega)\to 0$. Define

$$B_\infty(\omega):=\sum_{x_p,\dots,x_1}\Big\|\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega) - \boldsymbol{M}_\infty^2(\omega)\operatorname{Tr}[A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger]\Big\|. \tag{D25}$$

Next we wish to show that $B_k(\omega)\to B_\infty(\omega)$. By the inverted triangle inequality, a rearrangement, and the triangle inequality, one obtains

$$\begin{aligned}|B_\infty(\omega)-B_k(\omega)|\leq\sum_{x_p,\dots,x_1}&\Big\|\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\\&-\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}^\dagger(\omega)A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_{N_k}(\omega)\boldsymbol{M}_{N_k}(\omega)\Big\|\\+\sum_{x_p,\dots,x_1}&\Big\|\boldsymbol{M}_\infty^2(\omega)\operatorname{Tr}\big(A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger\big)\\&-\boldsymbol{M}_{N_k}^2(\omega)\operatorname{Tr}\big(A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_{N_k}(\omega)\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}(\omega)^\dagger\big)\Big\|\end{aligned} \tag{D26}$$

The goal is to utilize the fact that $\boldsymbol{M}_{N_k}(\omega)\to\boldsymbol{M}_\infty(\omega)$, and thus that $\boldsymbol{M}_{N_k}^2(\omega)\to\boldsymbol{M}_\infty^2(\omega)$, and similarly that $\boldsymbol{U}_{N_k}(\omega)\to\boldsymbol{U}_\infty(\omega)$. To this end, in the first sum in (D26), inside the norm, one can subtract and add $\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}^\dagger(\omega)A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)$. Similarly in the second sum, we subtract and add $\boldsymbol{M}_{N_k}^2(\omega)\operatorname{Tr}(A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger)$ inside of the norm. One can repeatedly use the triangle inequality, subtractions and additions in the similar spirit as above, and general relations such as $\|AB\|\leq\|A\|\|B\|$, $|\operatorname{Tr}(AB)|\leq\|A\|\|B\|_1$, as well as observations such as $\|\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\|\leq 1$, $\big\|\boldsymbol{M}_{N_k}(\omega)\boldsymbol{U}_{N_k}^\dagger(\omega)\big\|\leq 1$,

$\|\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger\|_1 = \|\boldsymbol{M}_\infty(\omega)\|_1 = 1$, $\sum_{x_p,\dots,x_1}\|A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\| \leq d^p$, and $\|\boldsymbol{M}_{N_k}^2(\omega)\| \leq 1$ to show that

$$\begin{aligned}|B_\infty(\omega) - B_n(\omega)| \leq\ & d^p\|\boldsymbol{M}_\infty(\omega) - \boldsymbol{M}_{N_k}(\omega)\| + d^p\|\boldsymbol{U}_\infty(\omega)^\dagger - \boldsymbol{U}_{N_k}^\dagger(\omega)\| \\ & + d^p\|\boldsymbol{M}_\infty(\omega) - \boldsymbol{M}_{N_k}(\omega)\| + d^p\|\boldsymbol{U}_\infty(\omega) - \boldsymbol{U}_{N_k}(\omega)\| \\ & + d^p\|\boldsymbol{M}_\infty^2(\omega) - \boldsymbol{M}_{N_k}^2(\omega)\| + d^p\|\boldsymbol{U}_\infty(\omega) - \boldsymbol{U}_{N_k}(\omega)\|_1 \\ & + d^p\|\boldsymbol{M}_\infty(\omega) - \boldsymbol{M}_{N_k}(\omega)\|_1 + d^p\|\boldsymbol{U}_\infty(\omega)^\dagger - \boldsymbol{U}_{N_k}(\omega)^\dagger\|_1, \\ =:& C_k(\omega),\end{aligned}$$

and thus $C_k(\omega) \to 0$. We can conclude that $B_\infty(\omega) - B_k(\omega) \leq |B_\infty(\omega) - B_k(\omega)| \leq C_k(\omega)$, which implies $0 \leq B_\infty(\omega) \leq B_k(\omega) + C_k(\omega)$. By combining this observation with $B_k(\omega) \to 0$ and $C_k(\omega) \to 0$, as well as with the definition of $B_\infty(\omega)$ in (D25), we can conclude that

$$\begin{aligned}B_\infty(\omega) = & \sum_{x_p,\dots,x_1}\Big\|\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega) \\ & - \boldsymbol{M}_\infty^2(\omega)\operatorname{Tr}\left(A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty(\omega)^\dagger\right)\Big\|, \\ = & \ 0.\end{aligned} \tag{D27}$$

This in turn implies

$$\boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty^\dagger(\omega)A_{x_1}^\dagger\cdots A_{x_p}^\dagger A_{x_p}\cdots A_{x_1}\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega) \propto \boldsymbol{M}_\infty(\omega)\boldsymbol{U}_\infty^\dagger(\omega)\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega). \tag{D28}$$

Since we have assumed the purity condition, it follows by Lemma 38, with $O := \boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)$, that $\operatorname{rank}(\boldsymbol{M}_\infty(\omega)) = \operatorname{rank}(\boldsymbol{U}_\infty(\omega)\boldsymbol{M}_\infty(\omega)) = 1$. Since this holds for almost all elements ω in the sample space, we can conclude that $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$ *a.s.* □

3. $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$ *a.s.* implies the purity condition

While we in Appendix D 2 demonstrated that the purity condition is sufficient for $\boldsymbol{M}_\infty$ being a rank-one operator, we here show that it also is a necessary condition. The idea is to assume that $\boldsymbol{M}_\infty$ has rank one, but that the purity condition does not hold. The latter means that there exists a projector P with $\operatorname{rank}(P) > 1$, while still $PA_{x_1}^\dagger\cdots A_{x_N}^\dagger A_{x_N}\cdots A_{x_1}P \propto P$. The latter is then showed to imply $P\boldsymbol{M}_\infty(\omega)P \propto P$. However, since $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$, the only possibility is that $P\boldsymbol{M}_\infty(\omega)P = 0$. This turns out to be in contradiction with $\sum_{x=0}^{d-1}A_x^\dagger A_x = \mathbb{1}$.

Lemma 40. *Let $\{A_x\}_{x=0}^{d-1}$ be operators on a finite-dimensional complex Hilbert space, $\mathscr{H}$, such that $\sum_{x=0}^{d-1}A_x^\dagger A_x = \mathbb{1}$. Let $(\boldsymbol{M}_N)_{N\in\mathbb{N}}$ be as defined in (C3) with respect to $(\boldsymbol{x}_N)_{N\in\mathbb{N}}$ and distributed as in (C1). Let $\boldsymbol{M}_\infty := \lim_{N\to\infty}\boldsymbol{M}_N$ a.s., as guaranteed by Lemma 36. If $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$ a.s., then $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3.*

Proof. We proceed via a proof by contradiction, and thus assume that $\{A_x\}_{x=0}^{d-1}$ is such that $\operatorname{rank}(\boldsymbol{M}_\infty) = 1$ *a.s*, but that the purity condition does *not* hold. The latter means that there exists a projector P such that

$$PA_{x_1}^\dagger\cdots A_{x_N}^\dagger A_{x_N}\cdots A_{x_1}P \propto P, \qquad \forall N\in\mathbb{N}, \quad \forall(x_1,\dots,x_N)\in\{0,\dots,d-1\}^{\times N}, \tag{D29}$$

but $\operatorname{rank}(P) > 1$. Recall that

$$\boldsymbol{M}_N = \begin{cases} \dfrac{A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1}}{\operatorname{Tr}(A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1})} & \text{if } \operatorname{Tr}(A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1}) \neq 0, \\ 0 & \text{if } \operatorname{Tr}(A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1}) = 0, \end{cases} \tag{D30}$$

where we note that $\operatorname{Tr}(A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1}) = 0$ if and only if $A_{\boldsymbol{x}_1}^\dagger\cdots A_{\boldsymbol{x}_N}^\dagger A_{\boldsymbol{x}_N}\cdots A_{\boldsymbol{x}_1} = 0$. By (D29), it thus follows that $P\boldsymbol{M}_N P \propto P$. Let $\omega\in\Omega$ be such that $\lim_{N\to\infty}\boldsymbol{M}_N(\omega) = \boldsymbol{M}_\infty(\omega)$. Consequently,

$$\lim_{N\to\infty}\|P\boldsymbol{M}_N(\omega)P - P\boldsymbol{M}_\infty(\omega)P\| = 0. \tag{D31}$$

By $P\boldsymbol{M}_N P \propto P$, we know that there exists a proportionality constant, $a_N(\omega)$, for each N and ω, such that

$$P\boldsymbol{M}_N(\omega)P = a_N(\omega)P. \tag{D32}$$

Next we use the general relation $|\operatorname{Tr}(AB)| \leq \|A\|_1 \|B\|$ to show

$$\begin{aligned} |a_N(\omega)\operatorname{Tr}(P) - \operatorname{Tr}(PM_\infty(\omega)P)| =& \Big| \operatorname{Tr}\Big(\mathbb{1}\big(a_N(\omega)P - PM_\infty(\omega)P\big)\Big)\Big|, \\ \leq& D\|PM_N(\omega)P - PM_\infty(\omega)P\| \to 0, \end{aligned} \tag{D33}$$

where we have used $\|\mathbb{1}\|_1 = D$, (D32) and (D31). With $a_\infty(\omega) := \operatorname{Tr}(PM_\infty(\omega))/\operatorname{Tr}(P)$, we can thus conclude that $\lim_{N\to\infty}|a_N(\omega) - a_\infty(\omega)| = 0$. Hence,

$$\begin{aligned} \|a_\infty(\omega)P - PM_\infty(\omega)P\| =& \|a_\infty(\omega)P - a_N(\omega)P + a_N(\omega)P - PM_\infty(\omega)P\|, \\ \leq& |a_\infty(\omega) - a_N(\omega)| + \|PM_N(\omega)P - PM_\infty(\omega)P\| \to 0. \end{aligned}$$

We can thus conclude that $PM_\infty(\omega)P \propto P$, and hence $PM_\infty P \propto P$ *a.s.* However, since M_∞ by assumption is rank-one *a.s.*, and $\operatorname{rank}(P) > 1$, the only possibility is that the proportionality constant is zero, i.e., that $PM_\infty P = 0$ *a.s.* Next, we note that $E(M_N) = \mathbb{1}/D$. By (D3) in Lemma 36, we know that $E(M_N) \to E(M_\infty)$, and thus $E(M_\infty) = \mathbb{1}/D$. However, this is in contradiction with $PM_\infty P = 0 \quad a.s.$ □

4. $w(N)$ goes to zero exponentially if and only if the purity condition holds

In Appendices D 2 and D 3, we have shown that $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition if and only if $\operatorname{rank}(M_\infty) = 1$. Here we show that the latter in turn is equivalent to $\lim_{N\to\infty} w(N) = 0$, and that this in turn is equivalent to $w(N)$ converging exponentially fast to zero.

If M_∞ has rank one, i.e., $\operatorname{rank}(M_\infty) = 1$, then it follows that $\|M_\infty\| = 1$ and we can relate $\|M_\infty\| - \|M_N\| = 1 - \|M_N\|$ to the eigenvalues of M_N and the singular values of W_N. The latter directly connects to the definition of $w(N)$ in (D36). To this end, we introduce the following notation. For a general operator, O, on a space of finite dimension, D, let $\nu_1^\downarrow(O) \geq \cdots \geq \nu_D^\downarrow(O)$ be the ordered singular values of O. Similarly, for a Hermitian operator, J, let $\lambda_1^\downarrow(J) \geq \cdots \geq \lambda_D^\downarrow(J)$ be the ordered eigenvalues of J.

The fact that $w(N)$ converges to zero does, of course, not guarantee that $w(N)$ converges exponentially fast to zero. The latter we obtain by first showing that $w(N)$ is submultiplicative, i.e., $w(N+M) \leq w(N)w(M)$, which implies that $\log w(N)$ is subadditive.

We obtain the submultiplicativity by rewriting $w(N)$ in terms of the norm of the second order exterior power of $A_{x_N}\cdots A_{x_1}$. In order to introduce the exterior power of an operator, consider a Hilbert space, $\mathscr{H}$, with an orthonormal basis, $\{|j\rangle\}_{j=1}^D$, and $D = \dim\mathscr{H}$. On the product space $\mathscr{H}\otimes\mathscr{H}$, we construct the swap-operator, $S := \sum_{j,k=1}^D |j\rangle\langle k|\otimes|k\rangle\langle j|$, where one may note that $S^2 = \mathbb{1}\otimes\mathbb{1}$ and $S^\dagger = S$. We also define the projector $P_A := (\mathbb{1}\otimes\mathbb{1} - S)/2$ onto the the anti-symmetric subspace of $\mathscr{H}\otimes\mathscr{H}$. For an operator O on $\mathscr{H}$, we define the exterior power (of degree two) of O as $\wedge^2(O) := P_A[O\otimes O]P_A$. A consequence of this definition is that $\|\wedge^2(O)\| = \nu_1^\downarrow(O)\nu_2^\downarrow(O)$. By comparing these definitions with (D36) below, we can conclude that $w(N) = \sum_{x_1,\ldots,x_N=0}^{d-1}\|\wedge^2(A_{x_N}\cdots A_{x_1})\|$.

One may note that the above construction presumes that O is an operator from one space to itself. However, due to the operator F [see e.g. (16)], we need a generalization to mappings from one space to another, $O : \mathscr{H}_1 \to \mathscr{H}_2$. However, analogous to P_A, we can for these two spaces let $P_A^{(1)}$ and $P_A^{(2)}$ be the projectors onto the anti-symmetric subspaces of $\mathscr{H}_1\otimes\mathscr{H}_1$ and $\mathscr{H}_2\otimes\mathscr{H}_2$, respectively. Based on these, we define the generalization $\wedge^2(O) := P_A^{(2)}[O\otimes O]P_A^{(1)}$. For this generalization it remains true that $\|\wedge^2(O)\| = \nu_1^\downarrow(O)\nu_2^\downarrow(O)$. By comparing with (43) we see that $f(N) = \sum_{x_N,\ldots,x_1=1}^{d-1}\|\wedge^2(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})\|$. A further observation that also holds for the generalization is

$$\|\wedge^2(O)\| \leq \|O\|^2. \tag{D34}$$

Moreover, if $O_A : \mathscr{H}_1 \to \mathscr{H}_2$ and $O_B : \mathscr{H}_2 \to \mathscr{H}_3$, then $\wedge^2(O_BO_A) = \wedge^2(O_B)\wedge^2(O_A)$, and consequently $\|\wedge^2(O_BO_A)\| \leq \|\wedge^2(O_B)\|\|\wedge^2(O_A)\|$.

The exponential decay of $w(N)$ is obtained by combining $\lim_{N\to\infty} w(N) = 0$ with the submultiplicativity of $\log w(N)$ and Fekete's subadditivity lemma. Fekete's Lemma is commonly attributed to Ref. 46. For a proof, see Lemma 1.2.1 in Ref. 47, and for a historical overview, see Section 1.10 in Ref. 47.

Lemma 41 (Fekete's subadditive lemma)**.** *Let $(a_N)_{N\in\mathbb{N}}$ be a subadditive sequence of real numbers, i.e., $a_{N+M} \leq a_N + a_M$. Then the limit $\lim_{N\to\infty} a_N/N$ is well defined (but may be $-\infty$) and*

$$\lim_{N\to\infty}\frac{a_N}{N} = \inf_{N\in\mathbb{N}}\frac{a_N}{N}. \tag{D35}$$

Proposition 42. *Let $\{A_x\}_{x=0}^{d-1}$ be linear operators on a finite-dimensional Hilbert space, such that $\sum_{x=0}^{d-1} A_x^\dagger A_x = \mathbb{1}$. Define*

$$w(N) := \sum_{x_1,\ldots,x_N=0}^{d-1} \nu_1^\downarrow(A_{x_N}\cdots A_{x_1})\nu_2^\downarrow(A_{x_N}\cdots A_{x_1}). \tag{D36}$$

The following statements are equivalent:

1. *$\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3.*
2. *$\lim_{N\to\infty} w(N) = 0$.*
3. *There exist real constants $C' \geq 0$ and $0 < \gamma < 1$ such that*

$$w(N) \leq C'\gamma^N, \quad \forall N \in \mathbb{N}. \tag{D37}$$

Proof. **1 $\Rightarrow$ 2:** Let $\boldsymbol{M}_N$ be as defined in (C3). We first distinguish the two cases that $\boldsymbol{M}_N$ is a density operator, or that it is the zero operator. In the case that $\boldsymbol{M}_N$ is a density operator, it follows that $1 = \mathrm{Tr}(\boldsymbol{M}_N) \geq \lambda_1^\downarrow(\boldsymbol{M}_N) + \lambda_2^\downarrow(\boldsymbol{M}_N)$, and thus $1 \geq 1 - \lambda_1^\downarrow(\boldsymbol{M}_N) \geq \lambda_2^\downarrow(\boldsymbol{M}_N) \geq 0$. By noting that $\|\boldsymbol{M}_N\| = \lambda_1^\downarrow(\boldsymbol{M}_N)$, we thus get $\sqrt{\|\boldsymbol{M}_N\|(1-\|\boldsymbol{M}_N\|)} \geq \sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)}$. Since $\boldsymbol{M}_N$ is assumed to be a density operator, it moreover follows that $\|\boldsymbol{M}_N\| \leq 1$, and thus

$$\sqrt{|1-\|\boldsymbol{M}_N\||} \geq \sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)}. \tag{D38}$$

In the case that $\boldsymbol{M}_N$ is the zero operator, then (D38) is trivially true.

By Lemma 36, we know that $\boldsymbol{M}_\infty$ almost surely is a density operator. By Lemma 39, we also know that $\boldsymbol{M}_\infty$ almost surely is a rank-one operator. Hence, $\boldsymbol{M}_\infty$ almost surely corresponds to a pure state. Consequently, $\|\boldsymbol{M}_\infty\| = 1$ *a.s.* Combining this observation with the inverted triangle inequality yields

$$\sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|} \geq \sqrt{|\|\boldsymbol{M}_\infty\| - \|\boldsymbol{M}_N\||} = \sqrt{|1-\|\boldsymbol{M}_N\||} \quad a.s. \tag{D39}$$

Combining (D38) with (D39) yields

$$\sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|} \geq \sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)} \quad a.s. \tag{D40}$$

We next observe that

$$\nu_k^\downarrow\left(\frac{\boldsymbol{W}_N}{\sqrt{\mathrm{Tr}(\boldsymbol{W}_N^\dagger\boldsymbol{W}_N)}}\right) = \sqrt{\lambda_k^\downarrow\left(\frac{\boldsymbol{W}_N^\dagger\boldsymbol{W}_N}{\mathrm{Tr}(\boldsymbol{W}_N^\dagger\boldsymbol{W}_N)}\right)} = \sqrt{\lambda_k^\downarrow(\boldsymbol{M}_N)}. \tag{D41}$$

Thus, (D40) and (D41) yields

$$\sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|} \geq \frac{\nu_1^\downarrow(\boldsymbol{W}_N)\,\nu_2^\downarrow(\boldsymbol{W}_N)}{\mathrm{Tr}(\boldsymbol{W}_N^\dagger\boldsymbol{W}_N)} \quad a.s., \tag{D42}$$

which, by Lemma 25, results in

$$E\big(\sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|}\big)D \geq E\left(\frac{\nu_1^\downarrow(\boldsymbol{W}_N)\,\nu_2^\downarrow(\boldsymbol{W}_N)}{\mathrm{Tr}(\boldsymbol{W}_N^\dagger\boldsymbol{W}_N)}\right)D = w(N). \tag{D43}$$

By Lemma 36, we know that $\boldsymbol{M}_N \to \boldsymbol{M}_\infty$ almost surely. Since the underlying Hilbert space is finite-dimensional, we then have $\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\| \to 0$ *a.s.*, and consequently

$$\boldsymbol{x}_N := \sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|} \to 0 \quad a.s. \tag{D44}$$

We next observe that $\boldsymbol{M}_N$ is a density operator, or the zero operator, and thus $\|\boldsymbol{M}_N\| \leq 1$. Hence, $\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\| \leq \|\boldsymbol{M}_\infty\| + \|\boldsymbol{M}_N\| \leq 1 + \|\boldsymbol{M}_\infty\|$, which yields

$$\boldsymbol{x}_N = \sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|} \leq \sqrt{1+\|\boldsymbol{M}_\infty\|} \leq 1 + \|\boldsymbol{M}_\infty\| =: \boldsymbol{y}. \tag{D45}$$

By Lemma 36 we know that $E(\|\boldsymbol{M}_\infty\|) < +\infty$, and thus $E(\boldsymbol{y}) = 1 + E(\|\boldsymbol{M}_\infty\|) < +\infty$. By using this observation and Eqns. (D44) and (D45) into Proposition 28, we can conclude that $\lim_{N\to\infty} E(\sqrt{\|\boldsymbol{M}_\infty - \boldsymbol{M}_N\|}) = 0$. By combining this with (D43), it follows that $\lim_{N\to\infty} w(N) = 0$. Hence, we can conclude that statement 1 implies statement 2.

2 ⇒ 3: We first make the observation that

$$\|\wedge^2(A_{x_{N+M}} \cdots A_{x_1})\| \leq \|\wedge^2(A_{x_{N+M}} \cdots A_{x_{N+1}})\| \|\wedge^2(A_{x_N} \cdots A_{x_1})\|, \tag{D46}$$

which in turn yields $w(N+M) \leq w(M)w(N)$. Hence, w is submultiplicative, and thus $\log w(N)$ is subadditive. By statement 2 we know that $\lim_{N\to\infty} w(N) = 0$. It follows that there exists an $N_0 \in \mathbb{N}$ such that $\log w(N_0) < 0$. Hence, since $\log w(N)$ is subadditive, it follows by Lemma 41 that

$$0 > \frac{\log w(N_0)}{N_0} \geq \inf_N \frac{\log w(N)}{N} = \lim_{N\to\infty} \frac{\log w(N)}{N}. \tag{D47}$$

In the case that the limit is finite, let $l := \lim_{N\to\infty} 1/N \log w(N)$. By definition of the limit, we know that for any $\varepsilon > 0$, there exists an N_ε such that $[\log w(N)]/N - l \leq \varepsilon$ for all $N \geq N_\varepsilon$. We choose an arbitrary but fixed $\varepsilon > 0$, and thus $w(N) \leq \gamma^N$ for all $N \geq N_\varepsilon$, where $\gamma := e^{l+\varepsilon}$. Define $C' := \max\{1, \max_{N=1,\ldots,N_\varepsilon} w(N)/N\}$, and thus (D37) holds.

Finally consider the case that $\lim_{N\to\infty} 1/N \log w(N) = -\infty$. This means that for every $a > 0$ there exists an N_a such that $[\log w(N)]/N \leq -a$ for all $N \geq N_a$, which we can easily rewrite as $w(N) \leq e^{-aN}$. Hence, with $\gamma := e^{-a}$ and $C' := \max\{1, \max_{N=1,\ldots,N_a} w(N)/N\}$ we again obtain (D37). We can conclude that statement 2 implies statement 3.

3 ⇒ 2: This implication is trivial.

2 ⇒ 1: In our first step, we show that $\lim_{N\to\infty} w(N) = 0$ implies that $\|\boldsymbol{M}_\infty\| = 1$ *a.s.* We first observe that if η is a density operator on a complex Hilbert space with finite dimension D, then $1 - \|\eta\| \leq \sqrt{(D-1)D}\sqrt{\lambda_1^\downarrow(\eta)\lambda_2^\downarrow(\eta)}$. We know that $\boldsymbol{M}_N$ is either a density operator, or the zero operator, and thus $1 - \|\boldsymbol{M}_N\| \geq 0$. We moreover know that $\boldsymbol{M}_N$ almost surely is a density operator. With $\boldsymbol{x} := 1 - \|\boldsymbol{M}_N\|$ and $\boldsymbol{y} := \sqrt{D(D-1)}\sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)}$, we can use the above observations to conclude that $0 \leq \boldsymbol{x} \leq \boldsymbol{y}$ *a.s.* Moreover, by Lemma 25, we obtain

$$1 - E(\|\boldsymbol{M}_N\|) = E(1 - \|\boldsymbol{M}_N\|) \leq D\sqrt{\frac{D-1}{D}} E\left(\sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)}\right). \tag{D48}$$

Next we note that the observation in (D41) yields

$$E\left(\sqrt{\lambda_1^\downarrow(\boldsymbol{M}_N)\lambda_2^\downarrow(\boldsymbol{M}_N)}\right) D = E\left(\frac{\nu_1^\downarrow(\boldsymbol{W}_N)\,\nu_2^\downarrow(\boldsymbol{W}_N)}{\mathrm{Tr}(\boldsymbol{W}_N^\dagger \boldsymbol{W}_N)}\right) D = w(N). \tag{D49}$$

By combining (D48) and (D49), one obtains $1 - E(\|\boldsymbol{M}_N\|) \leq w(N)\sqrt{(D-1)/D}$. By the assumption that $\lim_{N\to\infty} w(N) = 0$, it follows that $\lim_{N\to\infty} E(\|\boldsymbol{M}_N\|) = 1$. By (D4) in Lemma 36, we know that $E(\|\boldsymbol{M}_N\|) \to E(\|\boldsymbol{M}_\infty\|)$. We can thus conclude that $E(\|\boldsymbol{M}_\infty\|) = 1$. With $\boldsymbol{x} := 1 - \|\boldsymbol{M}_\infty\|$, it follows that $E(\boldsymbol{x}) = 0$. Since $\boldsymbol{M}_\infty$ is almost surely a density operator, it follows that $1 \geq \|\boldsymbol{M}_\infty\|$ almost surely. Hence, $\boldsymbol{x} = 1 - \|\boldsymbol{M}_\infty\| \geq 0$ *a.s.* By combining this observation and $E(\boldsymbol{x}) = 0$ with Lemma 27, we obtain $\boldsymbol{x} = 0$ *a.s.*, and thus $\|\boldsymbol{M}_\infty\| = 1$ *a.s.* By Lemma 36, we know that $\boldsymbol{M}_\infty$ almost surely is a density operator. If $\boldsymbol{M}_\infty$ is a density operator, then $\boldsymbol{M}_\infty$ is a rank one operator if and only if $\|\boldsymbol{M}_\infty\| = 1$. We can thus conclude that $\mathrm{rank}(\boldsymbol{M}_\infty) = 1$ *a.s.* By Lemma 40, this implies that $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition in Definition 3. Hence, statement 2 implies statement 1.

□

5. Generalization to F and σ

The entire proof has up to this point concerned the exponential decay of $w(N)$, while we actually wish to find conditions for the exponential decay of $f(N)$. Here, we find necessary as well as sufficient conditions for the exponential decay of $f(N)$. We state a slightly more elaborate version of Proposition 5.

Proposition 43. *Let $\{A_x\}_{x=1}^{d-1}$ be linear operators on the finite-dimensional complex Hilbert space $\mathscr{H}$, such that $\sum_{x=1}^{d-1} A_x^\dagger A_x = \mathbb{1}$. For an operator σ on $\mathscr{H}$, and an operator $F : \mathscr{H} \to \mathscr{H}'$ for a finite-dimensional complex Hilbert space $\mathscr{H}'$, define*

$$f(N) := \sum_{x_N,\ldots,x_1=1}^{d-1} \nu_1^\downarrow(FA_{x_N} \cdots A_{x_1}\sqrt{\sigma}) \nu_2^\downarrow(FA_{x_N} \cdots A_{x_1}\sqrt{\sigma}). \tag{D50}$$

If $\{A_x\}_{x=1}^{d-1}$ satisfies the purity condition in Definition 3, then there exist real constants $0 \leq \overline{c}$ and $0 < \gamma < 1$, which satisfy

$$f(N) \leq \overline{c}\gamma^N, \quad \forall N \in \mathbb{N}, \tag{D51}$$

for all density operators σ, and all F such that $F^\dagger F \leq \mathbb{1}$. Conversely, if f is defined with respect to some full-rank operators σ and $F^\dagger F$, such that there exist constants $0 \leq \overline{c}_{\sigma,F}$ and $0 < \gamma < 1$ which fulfill

$$f(N) \leq \overline{c}_{\sigma,F}\gamma^N, \quad \forall N \in \mathbb{N}, \tag{D52}$$

then $\{A_x\}_{x=1}^{d-1}$ satisfies the purity condition.

Proof. We begin by proving the first claim of the proposition. We first note that

$$\begin{aligned} f(N) &= \sum_{x_1,\ldots,x_N} \|\wedge^2(FA_{x_N}\cdots A_{x_1}\sqrt{\sigma})\|, \\ &\leq \|\wedge^2(F)\|\|\wedge^2(\sqrt{\sigma})\| w(N), \\ &\leq \|F\|^2\|\sqrt{\sigma}\|^2 w(N), \\ &\leq w(N), \end{aligned} \tag{D53}$$

where w is as defined in (D36), and where the next to last inequality follows from (D34). The last inequality follows since σ is assumed to be a density operator, and thus $\|\sqrt{\sigma}\| \leq 1$, and similarly $F^\dagger F \leq \mathbb{1}$ implies $\|F\| \leq 1$. If $\{A_x\}_{x=0}^{d-1}$ satisfies the purity condition, then it follows by Proposition 42 that $w(N) \leq C'\gamma^N$. By combining this observation with (D53), we obtain (D51) with $\overline{c} := C'$. Note that Proposition 42 makes no reference to F or σ, and thus C', and consequently $\overline{c}$, is independent of these.

Next, we turn to the second claim of the proposition. For this purpose, we first note that since $F^\dagger F$ and σ (and thus $\sqrt{\sigma}$) are full-rank operators on a finite-dimensional space, it follows that $(F^\dagger F)^{-1}$ and $\sqrt{\sigma}^{-1}$ exist. With w as defined in (D36), we thus find

$$w(N) = \sum_{x_1,\ldots,x_N} \|\wedge^2(A_{x_N}\cdots A_{x_1})\| \leq \|\wedge^2\big((F^\dagger F)^{-1}F^\dagger\big)\|\|\wedge^2(\sqrt{\sigma}^{-1})\| f(N). \tag{D54}$$

Hence, with $\overline{c}'_{\sigma,F} := \|\wedge^2\big((F^\dagger F)^{-1}F^\dagger\big)\|\|\wedge^2(\sqrt{\sigma}^{-1})\|$, we get $w(N) \leq \overline{c}'_{\sigma,F} f(N)$. Combined with the assumption (D52), it follows

$$w(N) \leq \overline{c}_{\sigma,F}\overline{c}'_{\sigma,F}\gamma^N, \quad \forall N \in \mathbb{N}, \tag{D55}$$

where w is as defined in Proposition 42, and $0 < \gamma < 1$. With $C' := \overline{c}_{\sigma,F}\overline{c}'_{\sigma,F}$ in Proposition 42, it follows that $\{A_x\}_{x=1}^{d-1}$ satisfies the purity condition in Definition 3. □

[1] M. B. Hastings, Solving gapped Hamiltonians locally, Phys. Rev. B **73**, 085115 (2006).
[2] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).
[3] U. Schollwöck, The density-matrix renormalization group, Rev. Mod. Phys. **77**, 259 (2005).
[4] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Annals of Physics **326**, 96–192 (2011).
[5] J. C. Xavier, J. A. Hoyos, and E. Miranda, Adaptive density matrix renormalization group for disordered systems, Phys. Rev. B **98**, 195115 (2018).
[6] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Matrix product density operators: Simulation of finite-temperature and dissipative systems, Phys. Rev. Lett. **93**, 207204 (2004).
[7] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, Time-evolution methods for matrix-product states, Annals of Physics **411**, 167998 (2019).
[8] J. Almeida, M. A. Martin-Delgado, and G. Sierra, DMRG applied to critical systems: spin chains, AIP Conference Proceedings **918**, 261 (2007).
[9] S. Bravyi and D. Gosset, Polynomial-time classical simulation of quantum ferromagnets, Phys. Rev. Lett. **119**, 100503 (2017).
[10] M. Jerrum and A. Sinclair, Polynomial-time approximation algorithm for the ising model, SIAM Journal on Computing **22**, 1087 (1993).
[11] F. Martinelli and E. Olivieri, Approach to equilibrium of glauber dynamics in the one phase region. I. The attractive case, Comm. Math. Phys. **161**, 447 (1994).
[12] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[13] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Info. Comput. **7**, 401–430 (2007).
[14] O. K. Kozlov, Gibbs description of a system of random variables, Probl. Peredachi Inf. **10**, 94 (1974), English translation: Problems of Information Transmission, 1974, **10**:3, 258–265.
[15] W. Brown and D. Poulin, Quantum markov networks and commuting hamiltonians (2012), arXiv:1206.0755 [quant-ph].
[16] K. Kato and F. G. S. L. Brandão, Quantum approximate Markov chains are thermal, Comm. Math. Phys. **370**, 117–149 (2019).
[17] K. Kato and F. G. S. L. Brandão, Locality of edge states and entanglement spectrum from strong subadditivity, Phys. Rev. B **99**, 195124 (2019).
[18] M. J. Kastoryano, A. Lucia, and D. Perez-Garcia, Locality at the boundary implies gap in the bulk for 2d peps, Comm. Math. Phys. **366**, 895–926 (2019).
[19] W. De Roeck, C. Maes, K. Netočnỳ, and M. Schütz, Locality and nonlocality of classical restrictions of quantum spin systems with applications to quantum large deviations and entanglement, Journal of Mathematical Physics **56**, 023301 (2015).
[20] T. Benoist, M. Fraas, Y. Pautrat, and C. Pellegrini, Invariant measure for quantum trajectories, Prob. Theory and Related Fields **174**, 307–334 (2019).
[21] H. Maassen and B. Kümmerer, Purification of quantum trajectories, in *Dynamics & Stochastics*, Lecture Notes–Monograph Series, Vol. Volume 48, edited by D. Denteneer, F. d. Hollander, and E. Verbitskiy (Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2006) pp. 252–261.
[22] T. Benoist, M. Fraas, Y. Pautrat, and C. Pellegrini, Invariant measure for stochastic Schrödinger equations, Annales Henri Poincaré **22**, 347–374 (2021), arXiv:1907.08485 [math-ph].
[23] F. Verstraete, M. Popp, and J. I. Cirac, Entanglement versus correlations in spin systems, Phys. Rev. Lett. **92**, 027901 (2004).
[24] M. Popp, F. Verstraete, M. A. Martín-Delgado, and J. I. Cirac, Localizable entanglement, Phys. Rev. A **71**, 042306 (2005).
[25] T. B. Wahl, D. Pérez-García, and J. I. Cirac, Matrix product states with long-range localizable entanglement, Phys. Rev. A **86**, 062314 (2012).
[26] E. Knill, R. Laflamme, and L. Viola, Theory of quantum error correction for general noise, Phys. Rev. Lett. **84**, 2525 (2000).
[27] D. V. Else, I. Schwarz, S. D. Bartlett, and A. C. Doherty, Symmetry-protected phases for measurement-based quantum computation, Phys. Rev. Lett. **108**, 240505 (2012).
[28] R. Movassagh and J. Schenker, An ergodic theorem for homogeneously distributed quantum channels with applications to matrix product states (2019), arXiv:1909.11769 [quant-ph].
[29] R. Movassagh and J. Schenker, Theory of ergodic quantum processes (2020), arXiv:2004.14397 [quant-ph].
[30] M. Sanz, D. Perez-Garcia, M. M. Wolf, and J. I. Cirac, A quantum version of Wielandt's inequality, IEEE Transactions on Information Theory **56**, 4668–4673 (2010).
[31] M. B. Hastings, How quantum are non-negative wavefunctions?, Journal of Math. Phys. **57**, 015210 (2016).
[32] P. Bougerol, *Products of Random Matrices with Applications to Schrödinger Operators*, Progress in Probability (Birkhäuser Boston, 2012).
[33] I. Nechita and C. Pellegrini, Random repeated quantum interactions and random invariant states, Probability Theory and Related Fields **152**, 299–320 (2010).
[34] J. J. Sylvester, *A word in Nonions* (Cambridge University Press, 1909).
[35] H. Weyl, Quantenmechanik und gruppentheorie, Quantenmechanik und Gruppentheorie, Zeitschrift für Physik **46**, 1 (1927).
[36] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover Publications, 1950).
[37] J. Schwinger, Unitary operator bases, Proceedings of the National Academy of Sciences of the United States of America **46**, 570 (1960).
[38] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, Comm. Math. Phys. **144**, 443 (1992).
[39] M. Fannes, A continuity property of the entropy density for spin lattice systems, Comm. Math. Phys. **31**, 291 (1973).
[40] K. M. R. Audenaert, A sharp continuity estimate for the von neumann entropy, Journal of Physics A: Mathematical and Theoretical **40**, 8127–8136 (2007).
[41] C.-F. Chen, K. Kato, and F. G. S. L. Brandão, Matrix product density operators: when do they have a local parent hamiltonian? (2020), arXiv:2010.14682 [quant-ph].
[42] F. Verstraete and J. I. Cirac, Continuous matrix product states for quantum fields, Phys. Rev. Lett. **104**, 190405 (2010).
[43] A. Wehrl, General properties of entropy, Rev. Mod. Phys. **50**, 221 (1978).
[44] A. Gut, *Probability: A Graduate Course*, Springer Texts in Statistics (Springer New York, 2010).
[45] R. Horn and C. Johnson, *Matrix Analysis*, Matrix Analysis (Cambridge University Press, 2013).
[46] M. Fekete, Uber die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten, Mathematische Zeitschrift **17**, 228 (1923).
[47] J. M. Steele, *Probability Theory and Combinatorial Optimization* (Society for Industrial and Applied Mathematics, 1997).

Part II

Minimising statistical errors in gate calibration

Probabilities are intrinsic in quantum mechanics. Indeed, the theory of quantum mechanics states with which probability a concrete outcome of a measurement is obtained, but cannot predict the value of the outcome with certainty. As a direct consequence, a quantum experimentalist needs to carry out several runs of the same experiment. Then, a statistical treatment of the collected data should be conducted to estimate the measured quantities. This fact creates a dialogue between the disciplines of quantum mechanics and statistical theory, a dialogue that we will explore in this part of the thesis.

The second part of this thesis has two chapters. Within Chapter 3, we introduce the notions of statistics and gate calibration that are fundamental for our findings in Chapter 4.

Chapter 4 builds on the work presented in Ref. [39] about calibration of quantum-gate sets. In their work, a protocol called Gate Set Calibration protocol (GSC) is introduced and used to extract systematic gate errors. Although in quantum mechanics physical measurements have an intrinsic uncertainty, statistical errors are not considered in Ref. [39]. In this thesis, we conduct a statistical analysis of GSC in order to optimise the protocol. Not only is our goal to minimise statistical errors, but also to require physical realisability of GSC. For the latter, we restrict the implementation of GSC to involve a small number of different quantum gates as well as a small sequence length.

Chapter 3

Foundations

The first chapter of this part contains the theoretical foundations behind the research presented in Chapter 4. In Section 3.1, we review concepts about statistical theory. We start in Section 3.1.1 brushing up notions of multivariate probability distributions, such as random vector and the central limit theorem. We focus on the probability distributions that will be used in Chapter 4, namely the multivariate Gaussian distribution and the multivariate binomial distribution.

Section 3.1.2 is an introduction to the theory of design of experiments. After beginning with a general definition of the concept 'process', we review the goals of this theory. We then put the focus on nuisance factors, which are the factors with which we deal in Chapter 4.

To finish this chapter, we give a brief overview of the field of gate calibration in Sec. 3.2. Tomography protocols, as well as randomized benchmarking protocols, are explained. We also identify the scenarios where each calibration approach is more suitable.

3.1 Statistics

Within this section, we introduce basic notions of statistics that are essential for a clear understanding of the research presented in Chapter 4. Section 3.1.1 is about the branch of statistics that considers scenarios with more than one random variable. Afterwards, we explain the foundations of the branch of statistics called design of experiments.

3.1.1 Multivariate probability distributions

Many interesting problems in physics take place in a multidimensional space and involve vectors whose coordinates are correlated. A statistical analysis of these problems requires to deal with random variables which form a vector. Here, we review the basic notions of the branch of statistics that considers more than one random variable and constructs multivariate models. Along this section we will assume that random variables are continuous unless stated differently.

Consider k scalar-valued random variables, $X_1, \ldots, X_k$, on the same probability space with probability density functions $p(x_i)$ for all $i = 1, \ldots, k$. A *multivariate random vector* (or simply *random vector*), $\vec{X}$, is a k-dimensional vector with random variables as components, i.e., $\vec{X} = (X_1, \ldots, X_k)$.

A random vector gives rise to a *multivariate probability density function*, $p(x_1, \ldots, x_k)$, defined as the joint probability density function of all components of the random vector. Then, we write the joint probability of finding each random variable, X_i, in a concrete interval, $[a_i, b_i]$, as

$$\Pr[a_1 \leq X_1 \leq b_1 \text{ and } \ldots \text{ and } a_k \leq X_k \leq b_k] = \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} p(x_1, \ldots, x_k) dx_1 \cdots dx_k,$$

which, by definition, satisfies

$$\int_{x_1} \cdots \int_{x_k} p(x_1, \ldots, x_k) dx_1 \cdots dx_k = 1.$$

The probability density functions of each component of the random vector, $p(x_i)$, are known as *marginal probability density functions*.

Analogously to univariate probability distributions, multivariate probability distributions are often parametrized by a couple of quantities with a practical interpretation. The *mean vector* of a random vector, $\vec{X}$, is $\mathrm{E}[\vec{X}] = (\mathrm{E}[X_1], \ldots, \mathrm{E}[X_k])$, which we will also sometimes denote as $\vec{\mu}$. The correlation between the random variables that form a random vector is specified with a matrix called *covariance matrix*. The covariance matrix, Σ, of a random vector, $\vec{X} = (X_1, \ldots, X_k)$, is a $k \times k$ matrix with entries

$$\Sigma_{ij} = \mathrm{cov}(X_i, X_j) = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_j]\mathrm{E}[X_j].$$

Note that the diagonal entries of Σ are the variances of the random variables X_i.

When the random variables X_i are independent, the multivariate probability density function factorises and can be written as $p(x_1, \ldots, x_k) = p(x_1) \cdots p(x_k)$. Furthermore, in this case the covariance matrix is diagonal.

An important probability distribution is the Gaussian distribution, which can be extended to the multivariate case straightforwardly. Given a k-dimensional random vector, $\vec{X}$, it is said that $\vec{X}$ is distributed according to a *multivariate Gaussian distribution* if any linear combination of its k components is distributed according to a (univariate) Gaussian distribution. We write $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, where $\vec{\mu}$ and Σ are the mean vector and the covariance matrix of $\vec{X}$, respectively.

The relevance of the (multivariate) Gaussian distribution comes from the well-known (multivariate) *Central Limit Theorem* (CLT). This theorem states that the mean value of many independent samples of a random vector, with finite mean and covariance matrix, is also a random vector. Furthermore, its probability distribution converges to a (multivariate) Gaussian distribution in the limit of a large number of samples.

Because of the CLT, the Gaussian distribution appears in many processes in nature that depend on multiple factors, such as neural activity, gene expression or measurement errors.

In quantum physics, the so-called binomial distribution, a discrete probability distribution, is also crucial since it appears in the context of quantum measurements. Consider an observable, $\hat{A}$, with two outcomes, a_1 and a_2, and an experimental setting that requires N measurements of the observable A. The data obtained in the experiment is distributed according to a binomial distribution. Specifically, the amount of times the measurement device returns the outcome a_i gives rise to a binomial distribution such that $B(N, p_i)$, where p_i is the probability of obtaining the outcome a_i. In the limit of large N, the binomial distribution can be approximated by a Gaussian distribution with parameters $\mathcal{N}(Np_i, Np_i(1-p_i))$ due to CLT.

When each random variable that constitutes a random vector gives rise to a (univariate) binomial distribution, we say that the random vector is distributed according to a *multivariate binomial distribution*.

Random vectors can be part of algebraic transformations during statistical analysis and create new vectors. A vector $\vec{Y} = M\vec{X} + \vec{v}$ is also a $k-$dimensional random vector if it is an affine transformation of the random vector $\vec{X}$, where M is a $k \times k$ matrix and $\vec{v}$ is a k-dimensional vector [40].

In the particular case that the random vector $\vec{X}$ is distributed according to a multivariate Gaussian distribution such that $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, then the random vector $\vec{Y}$ also gives rise to a multivariate Gaussian distribution with mean vector $M\vec{\mu} + \vec{v}$, and covariance matrix $M\Sigma M^T$, i.e., $\vec{Y} \sim \mathcal{N}(M\vec{\mu} + \vec{v}, M\Sigma M^T)$.

Affine transformations of multivariate Gaussian distributions will be a crucial tool to improve the design of a calibration protocol in Ch. 4, but, first, we need to learn about design of experiments.

3.1.2 Design of experiments

Experimentation is required not only in science but also in industry and engineering. The efficiency of such experiments can benefit from statistical theory to obtain more robust and precise conclusions. Applying statistical theory to experimentation consists in planning an experiment to collect the appropriated data, and conducting statistical analysis to extract valid

and objective conclusions from the data. Therefore, there are two stages where statistics play a crucial role in any experiment: the design of the experiment and the statistical analysis of the collected data. This section is devoted to the former: the design of experiments (DOE). The material presented here is based on Ref. [41].

Consider a system in some initial conditions that undergoes a *process*. Figure 3.1 is a schematic description of a general process, where the initial conditions are referred to as *input variables*. The word process here refers to any transformation, including operations from industrial processing, such as changes caused by machines, or from quantum computing, such as applications of quantum gates and measurements. Two different kind of factors affect the process: the *controllable* and the *uncontrollable* factors. As their name suggests, the latter are caused by variables that one cannot modify, while the former can be adjusted. The process transforms the input into an *output* characterised by one or more response variables.

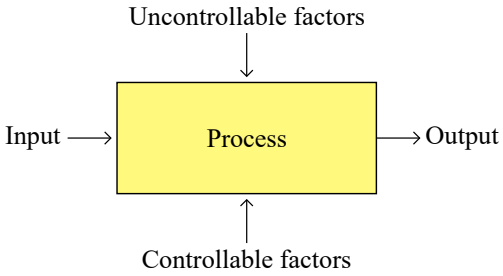

Figure 3.1: Schematic description of a process.

An *experiment* can be defined as a series of runs of a process. Thus, the output of an experiment is usually a data set.

In order to improve a process, DOE suggests to carry out experiments with the following goals:

1. Determining the most influential factors on the output,
2. Determining how to adjust the controllable factors such that the output's response variables are always near the desired value,
3. Determining how to adjust the controllable factors such that the variability of the response variables is small,
4. Determining how to adjust the controllable factors such that the effect of the uncontrollable factors is minimal.

We focus in this thesis on the last two goals of DOE. In particular, we put attention on the branch of DOE that takes care of *nuisance factors*, i.e., factors whose effect on the

output is significant enough to be considered even if we are not interested in their influence. Nuisance factors are classified as *controllable*, *uncontrollable* and *noise factors*. Controllable and uncontrollable nuisance factors can be measured. Moreover, the former can be adjusted and thus blocked, while the latter cannot be modified, so they need to be addressed with statistical analysis. A nuisance factor is called a noise factor when it is uncontrollable and varies naturally in each run of the process. This variability makes it impossible to control a noise factor only with one run of the process, but it might be addressed at the experiment level (i.e., doing several runs of a process). Noise factors are often managed by setting the controllable (design) factors of the process in a range that minimises the variability of the output responses.

Consider as an example the scenario that we study in Chapter 4: calibrating a quantum gate called CNOT. As devices are not perfect, the application of a CNOT is affected by uncontrollable nuisance factors and, instead of a CNOT gate, a perturbed version of the gate is applied. To account for these uncontrollable nuisance factors, we build a model for the perturbed gate that considers error parameters. Noise factors are also present in the calibration process due to quantum measurements. The output of a quantum measurement is a random variable, and thus it induces statistical errors. This statistical error is a noise factor because it cannot be controlled in a single process, and it varies in each run. However, we can reduce the statistical error by realising many runs of the process and, as we show in Chapter 4, by adjusting controllable factors of the calibration process. Before delving into the application of DOE on a gate calibration protocol, we review basic concepts of gate calibration in the next section.

3.2 Gate calibration

Theoretically, a quantum computation is a unitary matrix performed on a vector. In a laboratory, however, everything becomes messy. The vector corresponds to a physical system, which can spontaneously change its state due to undesired interactions with the environment. The unitary matrix corresponds to a series of pulses that must be precisely selected and performed to obtain the expected result. In order to consider all these factors, a model that characterises the quantum computation is designed. Then, experiments are carried out, and the data obtained is studied to choose the optimal values for the model parameters. This process is known as gate calibration and is indispensable in quantum computing. Within this section, we give an overview of gate calibration.

To design suitable gate calibration protocols, one must consider that errors in a quantum computer can have different origins. We can distinguish between coherent errors and incoherent errors. The former occur in the devices, while the latter arise from the interaction between

quantum systems and the environment. In order to have a quantum computer that applies operations close to the ideal unitary gates, both types of errors, incoherent and coherent, must be small enough [42, 43].

Maybe the most intuitive gate calibration strategy consists of fully characterising the quantum process, including its errors. Quantum process tomography (QPT) [44] and Gate Set Tomography (GST) [45] are the two most well-known methods to realise this task. QPT characterises a collection of states on which the process of interest has been applied, and uses this information to learn about the process. The characterisation of the resulting states is done via quantum state tomography [46, 47], which reconstructs a quantum state by performing measurements on a collection of identical states. QPT runs, however, into a self-consistency problem since it needs fiducial measurements and initial states, which, in turn, require already calibrated quantum gates.

The self-consistency problem of QPT can be solved using GST. GST has a similar spirit to QPT but includes State-Preparation and Measurement (SPAM) errors in its error model. Tomography methods provide a complete characterisation of the quantum process and can distinguish between individual gate errors and errors on the global gate set. Nevertheless, they are infeasible for large system sizes, since the number of measurements scales exponentially with the number of qubits.

To overcome the lack of scalability of tomography methods, calibration techniques that work with a partial characterisation of the quantum process, such as Randomized Benchmarking (RB) protocols, are used. Nowadays, RB refers to a family of calibration protocols [48, 49, 50, 51, 52, 53, 54, 55] based on the original proposal by Dankert et al. [56]. The idea behind these protocols consists of three steps. First, preparing an input state, then applying a squences of random gates, and finally measuring the resulting state. These three steps are repeated using sequences of different lengths. While errors in the gates are amplified with increasing sequence length, SPAM errors are independent of the sequence length. Moreover, the measured quantities only depend linearly on SPAM errors. These differences on the dependencies allow extracting an average error rate of the quantum process disentangled from SPAM errors by exponentially fitting the measured quantities versus the sequence length. RB protocols have been used widely because of their apparent advantages: scalability and easy implementation. Nevertheless, RB protocols are unsuitable for experiments requiring a detailed characterisation of individual quantum operations or distinguishing between incoherent and coherent errors.

Further calibration methods make use of ideas from both GST and RB [57, 58]. RB tomography [59], randomised linear GST [60] and compressive GST [61] are, for example,

calibration protocols that work on a trade-off between the information extracted from the quantum process and the resource requirements.

Another example of a calibration protocol that is a middle way between GST and RB is GSC, a protocol introduced in Ref. [39]. Given a set of available quantum gates, initial states and measurements, GSC constructs short gate sequences that allow all coherent gate errors to be extracted by solving an inverse problem. This protocol offers the flexibility to either be self-consistent or simplified for scenarios where single-qubit gates are already calibrated. Within Chapter 4, we review GSC in detail, and we apply the statistical theory presented in Sec. 3.1 to improve the design of this calibration protocol.

Chapter 4

Results

A fundamental property of quantum mechanics is its intrinsic uncertainty: responses of measurements can only be predicted probabilistically. Therefore, experiments aimed to extract information from a quantum system have to be designed so as to reduce the expected statistical uncertainty. The branch of statistics concerned with such questions is the Design of Experiments (DOE). Among the main objectives of DOE is the optimisation of statistical error, as well as validity, reliability and replicability. DOE has been applied in different quantum information problems such as parameter estimation [62, 63], and quantum state and process tomography [64, 65, 66]. In this chapter, we address the task of implementing DOE in a calibration protocol for quantum gate sets.

Specifically, we will work with the Gate Set Calibration protocol (GSC) introduced in Ref. [39] for the purpose of identifying coherent errors in unitary quantum gates. Given a set of available quantum gates, initial states and measurements, GSC constructs short gate sequences that allow all coherent gate errors to be extracted. This protocol can be considered as a middle way between randomized benchmarking and gate set tomography, as it provides more detailed information than the former using fewer measurements than the latter [50, 67].

In order to optimise GSC, we introduce parameters in the sequences that GSC builds and perform an optimisation over these parameters to minimise the statistical error of GSC.

This chapter is structured as follows. In Sec. 4.1, a simplified version of GSC is introduced and the notation is established. In Sec. 4.2 we present the theoretical framework under which the optimisation of GSC will be conducted. We analyse the statistics of GSC and define a measure of uncertainty that will allow us to compare the statistical error of different versions of GSC. Section 4.3 is dedicated to the optimisation of GSC by minimising the statistical error with respect to the measure of uncertainty. The version of GSC resulting from the optimisation is further modified to take into account the feasibility of the protocol. We verify in Sec. 4.3.1 that the reduction of the statistical error also holds when imperfect measurements are considered. An outlook is presented in Sec. 4.4. Note that this chapter has its own appendix.

4.1 The gate set calibration protocol

In Ref. [39], a new protocol for characterising and calibrating quantum gates, the GSC, has been introduced. This section will introduce the precise notation and present a slightly simplified version of the protocol. In particular, we restrict attention to the specific case of calibrating a two-qubit CNOT gate subject to coherent errors.

The GSC consists of the following steps. First, the qubits are initialised in a known state. Then, a gate sequence from a set of available gates is applied to the initial state, and finally, a measurement is performed. All coherent gate errors can be extracted by repeating this process for different sequences, initial states and measurements, as we will see below.

Assume that we aim to implement a set of unitary quantum gates. In reality, the hardware will realise perturbed gates, which we still assume to be unitary.

Consider concretely a situation where only gates in the set $\{\mathrm{CNOT}, \mathrm{X}_\theta^{(1)}, \mathrm{Y}_\theta^{(1)}, \mathrm{X}_\theta^{(2)}, \mathrm{Y}_\theta^{(2)}\}$ can be applied, namely a global CNOT and arbitrary single-qubit rotations around X- and Y-axis. Here, the superscripts indicate on which qubit the gate is applied, and the gates are defined as

$$\mathrm{CNOT} := |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes \sigma_1, \;\; X_\theta := e^{-i\frac{\theta}{2}\sigma_1}, \;\; Y_\theta := e^{-i\frac{\theta}{2}\sigma_2},$$

where σ_1, σ_2, σ_3 are the Pauli matrices. We assume that all single-qubit gates are already calibrated, and thus here, we want to use GSC to tune a perturbed CNOT gate, which we denote as $\tilde{\mathrm{CNOT}}$. The deviation from the optimal gate is measured by the *coherent error operator* associated with the gate. In the particular case of the CNOT gate we define

$$E := \mathrm{CNOT}^{-1}\tilde{\mathrm{CNOT}}. \tag{4.1}$$

In order to analyse the perturbed gate, we expand the error operator in the Pauli basis

$$E(\vec{p}) = \mathbb{I} - i\sum_{k=1}^{15} p_k \tau_k,$$

for suitable coefficients $\vec{p}$, which we call *error parameters*, and for $\tau_k = \tau_{ij} = \sigma_i \otimes \sigma_j$ with $\sigma_0 = \mathbb{I}$. The choice of the prefactor of the sum will ensure later that the unitarity of E to first order implies that p_k are real.

The ultimate goal of GSC is to obtain the error parameters, $\vec{p}$. To estimate the error parameters, GSC considers a number of experimental settings, which we label with the letter s. These settings can vary in the initial state, the gate sequence, and the measurement. We will always take $\rho = |00\rangle\langle 00|$ as initial state and perform a measurement $M_s \in \{\sigma_0 \otimes \sigma_3, \sigma_3 \otimes \sigma_0\}$.

Thus, for each setting one can experimentally estimate the quantity

$$R_s(\vec{p}) = \mathrm{Tr}\left[(G_{l_s,s}\cdots\tilde{\mathrm{CNOT}}(\vec{p})\cdots G_{1,s})|00\rangle\langle 00|(G_{1,s}^{\dagger}\cdots\tilde{\mathrm{CNOT}}(\vec{p})\cdots G_{l_s,s}^{\dagger})M_s\right], \tag{4.2}$$

which we will refer to as the *measurement response*. Here, l_s is the length of each sequence and $G_{j,s} \in \{\mathrm{X}_\theta^{(1)}, \mathrm{Y}_\theta^{(1)}, \mathrm{X}_\theta^{(2)}, \mathrm{Y}_\theta^{(2)}\}$ are the single-qubit gates of sequence s. Note that Eq. (4.2) does not restrict the number of CNOT gates of each sequence.

Following the framework of Ref. [39], we now make the assumption that the error parameters are small, and thus assume that the vector of responses depends only linearly on $\vec{p}$. We write

$$\vec{R}(\vec{p}) = \vec{R}(\vec{0}) + L\vec{p}, \tag{4.3}$$

where we define the matrix elements of L as

$$L_{su} := \left.\frac{\partial R_s(\vec{p})}{\partial p_u}\right|_{\vec{p}=\vec{0}}. \tag{4.4}$$

These matrix elements, L_{su}, as well as the vector $\vec{R}(\vec{0})$ can be computed analytically. Therefore, given the perturbed responses, $\vec{R}(\vec{p})$, we obtain the gate errors parameters by solving the inverse problem

$$\vec{p} = L^{-1}\left[\vec{R}(\vec{p}) - \vec{R}(\vec{0})\right]. \tag{4.5}$$

To ensure that the matrix L is non-singular, the number of responses must be equal to or greater than the number of error parameters. When this is not the case, additional responses can be added by considering more sequences, initial states or measurements. Moreover, the condition number of L must not be too large so that computing L^{-1} is numerically stable.

Having introduced the basic setup, we are now able to explain more clearly the motivation of this research. Table 4.1 summarises the combination of sequences and measurements used in Ref. [39] to calibrate a perturbed CNOT employing GSC. In the third column of Tab. 4.1, we see that the settings chosen consider only responses distributed around zero. Put differently, $R_s(\vec{0}) = 0$ for all $s = 1, \ldots, 15$. It turns out that, if one looks at statistics, this is also the point of maximal uncertainty since the variance of the responses is maximal for $R_s(\vec{0}) = 0$ (see Fig. 4.1). In fact, it is not obvious what the optimal point is, i.e., which settings of GSC minimise the statistical error and, at the same time, ensure stability of the numerical inverse of L and implementability of the protocol. Our research looks into this trade-off between reducing the statistical error, but still having a well-conditioned matrix L. Although we only consider the calibration of a two-qubit gate, the statistical analysis performed in the following sections can be generalised to higher dimensions straightforward.

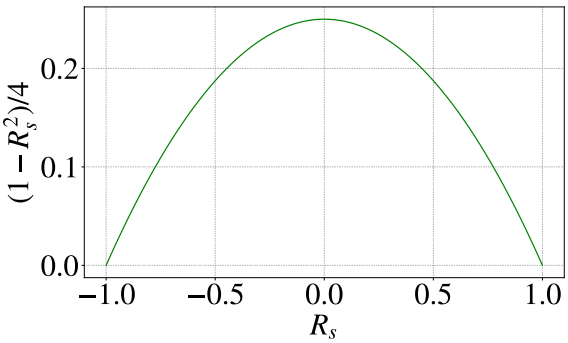

Figure 4.1: Variance of a estimated response, R_s^*, times the number of repetitions, N, as a function of the expectation value, R_s (see Eq. (4.6)).

4.2 The statistics of GSC

Quantum mechanics is unable to predict with certainty future events, such as measurement outcomes. Therefore, as we do in this section, it is essential to consider the statistics of the problem one is investigating to minimise statistical errors. Here, we study the statistics of GSC by determining the probability distribution according to which the measurement responses are distributed, as well as the error parameters. Subsequently, we define a quality measure that will allow us to compare the statistical error between different versions of GSC.

Before starting the discussion about GSC statistics, let us establish that all estimated quantities will be denoted by a superscript $*$. Assuming independent measurements, the outcome of a measurement with only two possible outcomes is described by a binomial distribution, $B(N, q)$, where N is the number of repetitions and q the probability of obtaining the outcome. In the scenario presented in Section 4.1, fifteen measurements are performed, one per setting, and each one is repeated N times. Therefore, the overall outcome is described by a multivariate binomial distribution. The CLT, reviewed in Sec. 3.1.1, states that binomial distribution can be approximated by a Gaussian distribution such that $\mathcal{N}(Nq, Nq(1-q))$ in the limit of sufficiently large N. Our interest is, in particular, focused on spin qubits, which are considered in Ref. [39] and where a high number of repetitions is easy to obtain. Therefore, we can assume that the overall GSC outcome can be approximated by a multivariate Gaussian distribution. We assume specifically that the estimated vector of responses is distributed as $\vec{R}^*(\vec{p}) \sim \mathcal{N}(\vec{R}(\vec{p}), \Sigma)$, where

$$\Sigma_{su} = \frac{1 - R_s^2(\vec{p})}{4N}\delta_{su}, \tag{4.6}$$

with $R_s(\vec{p})$ defined in Eq. (4.2). Note that the covariance matrix Σ is diagonal, i.e., settings are independent of each other.

s	1st gate	2nd gate	3rd gate	4th gate	M_s	$R_s(\vec{p})/2$
1	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$			τ_{12}	$0 - p_5 + p_{10}$
2	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT			τ_{12}	$0 - p_4 - p_7$
3	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$			τ_{12}	$0 - p_6 - p_9$
4	$\mathrm{Y}^{(1)}_{\pi/2}$	CNOT			τ_{12}	$0 - p_8 - p_{11}$
5	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$			τ_3	$0 - p_1 - p_{13}$
6	CNOT	$\mathrm{Y}^{(2)}_{\pi/2}$			τ_3	$0 - p_2 - p_{14}$
7	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$		τ_{12}	$0 - p_2 + p_{10}$
8	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$		τ_{12}	$0 - p_6 - p_{13}$
9	$\mathrm{Y}^{(1)}_{\pi/2}$	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$		τ_{12}	$0 + p_{10} + p_{13}$
10	$\mathrm{X}^{(1)}_{\pi/2}$	$\mathrm{X}^{(2)}_{\pi/2}$	CNOT		τ_3	$0 - p_7 - p_{13}$
11	$\mathrm{Y}^{(1)}_{\pi/2}$	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$		τ_{12}	$0 - p_2 - p_6$
12	$\mathrm{Y}^{(1)}_{\pi/2}$	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT		τ_3	$0 - p_{11} - p_{14}$
13	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$		τ_3	$0 + p_3 + p_{15}$
14	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$	τ_{12}	$0 - p_3 - 2p_{12} - p_{15}$
15	$\mathrm{X}^{(1)}_{\pi/2}$	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$	τ_3	$0 - p_6 + p_{15}$

Table 4.1: Settings used in Ref. [39] to calibrate a $\tilde{\mathrm{CNOT}}$ gate (Eq. (4.1)) via GSC. Each row corresponds to one setting, where gates are applied from left to right and followed by a measurement (either $\tau_3 = \sigma_0 \otimes \sigma_3$ or $\tau_{12} = \sigma_3 \otimes \sigma_0$). The last column contains the measurement responses, which depend on the error parameters, $\vec{p}$ (see Eqns. (4.2) and (4.3)).

The estimated error parameters, $\vec{p}^*$, are computed with Eq. (4.5) using the estimated responses, $\vec{R}^*$, obtained from the lab. Therefore, $\vec{p}^*$ is a random vector distributed according to a multivariate Gaussian distribution with $\vec{p}^* \sim \mathcal{N}(\vec{p}, \Sigma_{\vec{p}^*})$, where

$$\Sigma_{\vec{p}^*} := L^{-1}\Sigma L^{-T}, \tag{4.7}$$

with $L^{-T} := (L^{-1})^T$. In App. 4.A, we give numerical evidences that support this theory.

In order to quantify the statistical error, we need to construct a measure of uncertainty. We take our quality measure to be the mean squared distance between $\vec{p}^*$ and $\vec{p}$, i.e.,

$$\langle D^2 \rangle := \langle ||\vec{p}^* - \vec{p}||^2 \rangle = \sum_{s=1}^{15} \langle (p_s^* - p_s)^2 \rangle = \sum_{s=1}^{15} \Sigma^{(ss)}_{\vec{p}^* - \vec{p}}\,, \tag{4.8}$$

$$= \mathrm{Tr}\, \Sigma_{\vec{p}^* - \vec{p}} = \mathrm{Tr}(L^{-1}\Sigma L^{-T}), \tag{4.9}$$

where $\Sigma_{\vec{p}^* - \vec{p}}$ is the covariance matrix of the random vector $\vec{p}^* - \vec{p}$ with matrix elements $\Sigma^{(su)}_{\vec{p}^* - \vec{p}}$. Note that $\langle \vec{p}^* - \vec{p} \rangle = \vec{0}$ since $\langle \vec{p}^* \rangle = \vec{p}$, which we have used in the second equality. Note further that $\Sigma_{\vec{p}^* - \vec{p}} = \Sigma_{\vec{p}^*}$ is given in Eq. (4.7). We extend the definition of $\langle D^2 \rangle$ to singular L by defining it to be ∞ in this case.

Now that we have characterised the statistics of GSC and defined a quality measure, we are able to minimise the statistical error of GSC. In the next section, we will look for a set of sequences, initial states and measurements which achieves a small mean squared distance under the restrictions that, first, L is a well-conditioned matrix and, second, the settings of GSC are physically implementable.

4.3 Optimising GSC

In this section, we employ the statistical characterisation we have developed in Sec. 4.2 to optimise the calibration protocol GSC. To do so, we introduce some parameters in the settings of GSC and minimise the statistical error over these parameters. Aiming for a physical implementation of GSC, we further modify the optimised settings to have a small number of different gates. Before all this, however, we start analysing the version of GSC proposed in Ref. [39] with the quality measure defined in Sec. 4.2.

In Ref. [39], GSC is introduced and applied to calibrate a perturbed two-qubit CNOT gate, as we have summarised in Sec. 4.1. Evaluating the statistical error of this version of the protocol in terms of the mean squared distance defined in Eq. (4.8), we obtain that $\langle D^2 \rangle \approx 1.8/N$.

As we have already mentioned in Sec. 4.2, we can assume that each estimated response, R_s^*, is distributed according to a Gaussian distribution such that $\mathcal{N}(R_s(\vec{p}), \Sigma_{ss})$ with $R_s(\vec{p})$ and Σ_{ss} defined in Eqns. (4.2) and (4.6), respectively. In Sec. 4.1, we have pointed out that, for the settings chosen in Ref. [39], each estimated response is distributed around zero when $\vec{p} = \vec{0}$ (see the last column of Tab. 4.1). The calibration of the CNOT is, therefore, performed at the point of maximal uncertainty, i.e., where the variance of each response reaches its maximum, $\Sigma_{ss} = 1/(4N)$, as we can see in Fig. 4.1. Here we have restricted the attention to the diagonal elements of Σ because all settings are independent of each other.

We propose to slightly modify GSC in order to benefit from working at a point with smaller statistical error. To this end, we allow a different angle for each rotation of each GSC sequence so that the calibration of a $\tilde{\text{CNOT}}$ is performed using the sequences in Tab. 4.2. Our idea is to numerically minimise the mean squared distance of these settings over the angles $\theta_1, \ldots, \theta_{25}$. Motivation for this optimisation strategy can be found in App. 4.C.

We use a local derivative-free optimiser specified as NLOPT_LN_BOBYQA in the NLopt library [68], which, according to the documentation, is derived from the BOBYQA subroutine introduced in Ref. [69]. This algorithm optimises the objective function by iteratively constructing its quadratic approximation. We set boundary constraints such that $\theta_i \in [0, 2\pi]$ for all $i = 1, \ldots, 25$. A stopping tolerance for the optimiser is set to 10^{-10} at the angles. Note that a local optimiser such as BOBYQA does not always converge to a global minimum. In order to gain confidence that our result is indeed a global minimum, we have repeated the optimisation 10^4 times with random starting points. In roughly one in 30 runs, the optimiser stopped at values within a factor of 10^{-5} of the smallest mean squared error observed. The code developed for this project is available in [70] and, in Appendix 4.D, we present essential parts of the code.

We obtain a minimum mean squared distance of $\langle D^2 \rangle \approx 0.84/N$ with the angles shown in Fig. 4.2. The numerical values of the angles are in Appendix 4.E. This result means that one can achieve a reduction of the statistical error of GSC by a factor of 0.47 with respect to the version of GSC proposed in Ref. [39].

The new version of GSC requires sixteen single-qubit rotations for a maximal sequence depth of four gates. Specifically, we need one rotation for each different angle and, for some angles, we need two rotations: one rotation around the X-axis and one rotation around the Y-axis. Note that, although all rotations $\mathrm{X}_{n\phi}$ and $\mathrm{Y}_{n\phi}$, where n is an integer, can be implemented by applying n times the rotation X_ϕ or Y_ϕ, respectively, this implies a cost in the sequence depth. Indeed, we can reduce the number of single-qubit rotations to twelve but with a cost of increasing the sequence depth to seven gates.

s	1st gate	2nd gate	3rd gate	4th gate	M_s
1	CNOT	$\mathrm{X}^{(1)}_{\theta_{15}}$			τ_{12}
2	$\mathrm{X}^{(1)}_{\theta_1}$	CNOT			τ_{12}
3	CNOT	$\mathrm{Y}^{(1)}_{\theta_{16}}$			τ_{12}
4	$\mathrm{Y}^{(1)}_{\theta_2}$	CNOT			τ_{12}
5	CNOT	$\mathrm{X}^{(2)}_{\theta_{17}}$			τ_3
6	CNOT	$\mathrm{Y}^{(2)}_{\theta_{18}}$			τ_3
7	$\mathrm{X}^{(1)}_{\theta_3}$	CNOT	$\mathrm{X}^{(1)}_{\theta_{19}}$		τ_{12}
8	$\mathrm{X}^{(1)}_{\theta_4}$	CNOT	$\mathrm{Y}^{(1)}_{\theta_{20}}$		τ_{12}
9	$\mathrm{Y}^{(1)}_{\theta_5}$	CNOT	$\mathrm{X}^{(1)}_{\theta_{21}}$		τ_{12}
10	$\mathrm{X}^{(1)}_{\theta_6}$	$\mathrm{X}^{(2)}_{\theta_7}$	CNOT		τ_3
11	$\mathrm{Y}^{(1)}_{\theta_8}$	CNOT	$\mathrm{Y}^{(1)}_{\theta_{22}}$		τ_{12}
12	$\mathrm{Y}^{(1)}_{\theta_9}$	$\mathrm{Y}^{(2)}_{\theta_{10}}$	CNOT		τ_3
13	$\mathrm{Y}^{(2)}_{\theta_{11}}$	CNOT	$\mathrm{X}^{(2)}_{\theta_{23}}$		τ_3
14	$\mathrm{X}^{(1)}_{\theta_{12}}$	CNOT	CNOT	$\mathrm{Y}^{(1)}_{\theta_{24}}$	τ_{12}
15	$\mathrm{X}^{(1)}_{\theta_{13}}$	$\mathrm{Y}^{(2)}_{\theta_{14}}$	CNOT	$\mathrm{X}^{(2)}_{\theta_{25}}$	τ_3

Table 4.2: Settings to optimise the calibration of a $\mathrm{C\tilde{N}OT}$ gate in Eq. (4.1) via GSC. A numerical optimisation over $\theta_1, \ldots, \theta_{25}$ allows to reduce the statistical error with respect to the GSC with settings in Tab. 4.1. For clarity, the responses of the measurements are in Appendix 4.B.

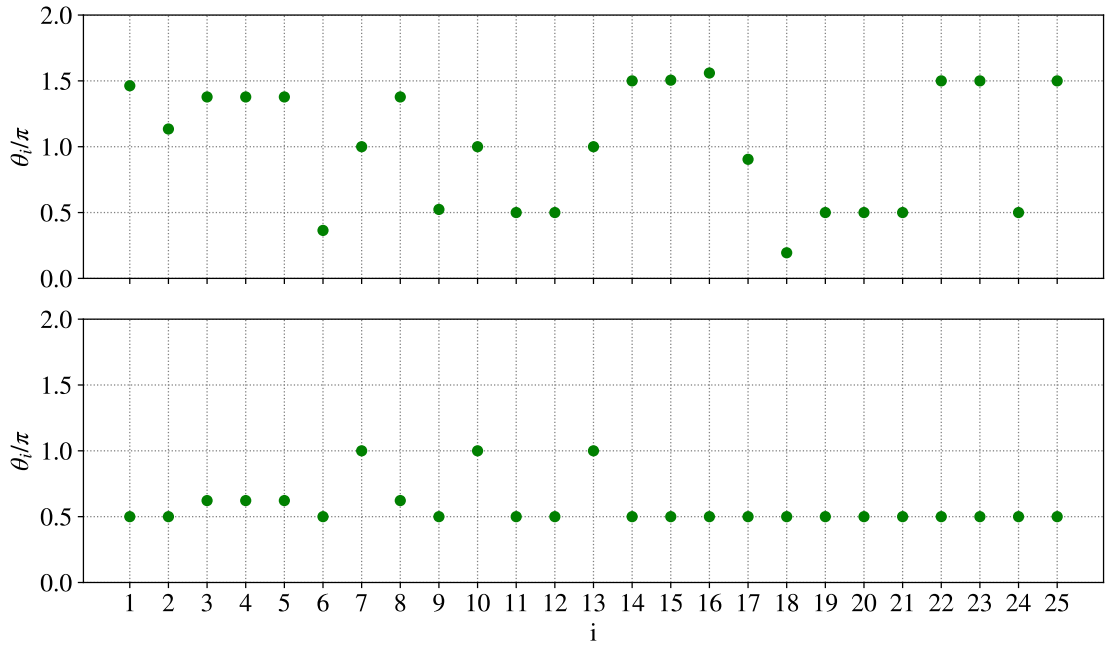

Figure 4.2: Above, angles that minimise the mean squared distance of the calibration of a $\widetilde{\text{CNOT}}$ using settings in Tab. 4.2 (Eq. (4.8)). Below, angles proposed here to optimise GSC. The angles below are the same angles as above after a folding transformation (i.e., mapping $\theta_i \longrightarrow 2\pi - \theta_i$ for all $\theta_i > \pi$) and setting $\theta_1 = \theta_2 = \theta_6 = \theta_9 = \theta_{15} = \theta_{16} = \theta_{17} = \theta_{18} = \pi/2$. Since the mean squared distance is invariant under folding transformation and independent of θ_1, θ_2, θ_6, θ_9, θ_{15}, θ_{16}, θ_{17}, θ_{18}, both angle combinations achieve the minimum mean squared distance, $\langle D^2 \rangle \approx 0.84/N$. The numerical values of the angles are in Appendix 4.E.

From an experimental point of view, tuning and calibrating sixteen different single-qubit rotations is undesirable. Hence, we would like to reduce the number of angles without increasing significantly the mean squared distance. We have observed numerically that the mean squared distance is invariant under the following transformations of the set of *optimal* angles (the transformations are not symmetries of the objective function at non-optimal points):

1. *degeneracy*: any change of the angles θ_1, θ_2, θ_6, θ_9, θ_{15}, θ_{16}, θ_{17}, and θ_{18} as long as L remains non-singular;

2. *global reflection*: $\theta_i \longrightarrow 2\pi - \theta_i$ for all $i = 1, \ldots, 25$;

3. *local reflection*: $\theta_i \longrightarrow 2\pi - \theta_i$ for one $i \in \{1, \ldots, 25\}$ as long as L remains non-singular;

4. *folding transformation*: $\theta_i \longrightarrow 2\pi - \theta_i$ for all $\theta_i > \pi$.

It would be tempting to set all the angles listed under 1. to zero, thereby eliminating the corresponding gates. However, it turns out that this choice does lead to a singular L and will therefore not be considered here. Still, a significant reduction of complexity can be achieved by making use of the above symmetries. The folding transformation allows us to reduce the

number of different rotations to thirteen for a maximal sequence depth of four gates without cost in the mean squared distance. At the price of increasing the sequence depth to seven gates, one can further reduce to eleven distinct rotations.

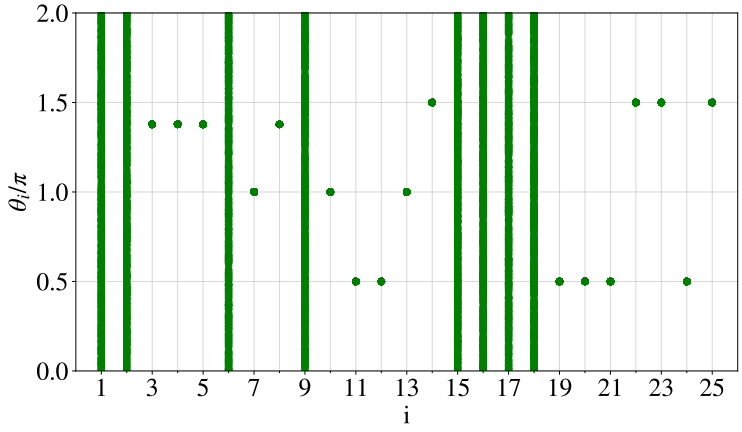

Figure 4.3: Combinations of angles that give a mean squared distance $\langle D^2 \rangle \approx 0.84/N$. The mean squared distance is independent of θ_1, θ_2, θ_6, θ_9, θ_{15}, θ_{16}, θ_{17}, θ_{18}, and thus we can set these angles at any value in $[0, 2\pi]$ without any cost in the statistical error.

To reduce even more the number of necessary gates, we use the fact that the minimum mean squared distance is degenerate. In Fig. 4.3, we can see that the minimum mean squared distance is achieved regardless of the value of the angles θ_1, θ_2, θ_6, θ_9, θ_{15}, θ_{16}, θ_{17} and θ_{18}. As we can unfortunately not set these angles simply to zero, we set them to $\pi/2$, which is a single-qubit gate already required to implement GSC (see Fig. 4.2). The mean squared distance remains at $\langle D^2 \rangle \approx 0.84/N$, and now GSC needs only four rotations and a maximal sequence depth of five gates.

Ultimately, we propose a GSC protocol that calibrates a $\widetilde{\text{CNOT}}$ gate with a statistical error 0.47 times smaller than in Ref. [39]. The maximal sequence depth is five and the set of allowed gates is $\{\text{CNOT}, \text{X}^{(1)}_{\pi/2}, \text{Y}^{(1)}_{\pi/2}, \text{X}^{(2)}_{\pi/2}, \text{Y}^{(2)}_{\pi/2}, \text{X}^{(1)}_{\theta}, \text{Y}^{(1)}_{\theta}\}$, where $\theta = 0.62208\pi$, as we see in Tab. 4.3.

4.3.1 GSC with imperfect measurements

In all previous sections, the measurements performed in GSC are assumed to be perfect. In other words, we have assumed that the measurement outcome obtained by the measurement device always corresponds to the state after the measurement. However, in the laboratory, devices are imperfect and, therefore, there is a non-zero error probability, i.e., a probability of obtaining a measurement outcome that differs from the true outcome corresponding to the post-measurement state. With the intention of building a model as realistic as possible, we introduce imperfect measurements in GSC along this section.

s	1st gate	2nd gate	3rd gate	4th gate	5th gate	M_s	$R_s(\vec{p})/2$
1	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$				τ_{12}	$0 - p_5 + p_{10}$
2	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT				τ_{12}	$0 - p_4 - p_7$
3	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$				τ_{12}	$0 - p_6 - p_9$
4	$\mathrm{Y}^{(1)}_{\pi/2}$	CNOT				τ_{12}	$0 - p_8 - p_{11}$
5	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$				τ_3	$0 - p_1 - p_{13}$
6	CNOT	$\mathrm{Y}^{(2)}_{\pi/2}$				τ_3	$0 - p_2 - p_{14}$
7	$\mathrm{X}^{(1)}_{\theta}$	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$			τ_{12}	$0 - \sin\theta p_2 - \cos\theta p_5 + p_{10}$
8	$\mathrm{X}^{(1)}_{\theta}$	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$			τ_{12}	$0 - p_6 - \cos\theta p_9 - \sin\theta p_{13}$
9	$\mathrm{Y}^{(1)}_{\theta}$	CNOT	$\mathrm{X}^{(1)}_{\pi/2}$			τ_{12}	$0 - \cos\theta p_5 + p_{10} + \sin\theta p_{13}$
10	$\mathrm{X}^{(1)}_{\pi/2}$	$\mathrm{X}^{(2)}_{\pi/2}$	$\mathrm{X}^{(2)}_{\pi/2}$	CNOT		τ_3	$0 + p_4 - p_7$
11	$\mathrm{Y}^{(1)}_{\theta}$	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$			τ_{12}	$0 - \sin\theta p_2 - p_6 - \cos\theta p_9$
12	$\mathrm{Y}^{(1)}_{\pi/2}$	$\mathrm{Y}^{(2)}_{\pi/2}$	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT		τ_3	$0 + p_8 - p_{11}$
13	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$			τ_3	$0 + p_3 + p_{15}$
14	$\mathrm{X}^{(1)}_{\pi/2}$	CNOT	CNOT	$\mathrm{Y}^{(1)}_{\pi/2}$		τ_{12}	$0 - p_3 - 2p_{12} - p_{15}$
15	$\mathrm{X}^{(1)}_{\pi/2}$	$\mathrm{X}^{(1)}_{\pi/2}$	$\mathrm{Y}^{(2)}_{\pi/2}$	CNOT	$\mathrm{X}^{(2)}_{\pi/2}$	τ_3	$0 - p_3 + p_{15}$

Table 4.3: Settings proposed to reduce the statistical error in the calibration of a $\tilde{\mathrm{CNOT}}$ using GSC. The set of allowed gates is $\{\mathrm{CNOT}, \mathrm{X}^{(1)}_{\pi/2}, \mathrm{Y}^{(1)}_{\pi/2}, \mathrm{X}^{(2)}_{\pi/2}, \mathrm{Y}^{(2)}_{\pi/2}, \mathrm{X}^{(1)}_{\theta}, \mathrm{Y}^{(1)}_{\theta}\}$ with $\theta = 0.62208\pi$.

Consider the scenario presented in Sec. 4.1: calibrating a perturbed gate, $\tilde{\text{CNOT}}$, (see Eq. (4.1)) using GSC. All measurements performed in GSC are Pauli observables, and thus they have only two outcomes, the positive and the negative outcome. We now consider an imperfect measurement device and allow a different error probability for each outcome. In other words, the measurement device gives the correct outcome $i \in \{\pm\}$ with a probability F_i, and an incorrect outcome with probability $1 - F_i$.

The probability to read a positive outcome on the measurement device after measuring M_s is then

$$\begin{aligned} Q_s(+) :=& F_+ q_s + (1 - F_-)(1 - q_s), \\ =& \frac{1}{2}\left[1 + F_+ - F_- + R_s\left(F_+ + F_- - 1\right)\right], \end{aligned}$$

where q_s is the probability of the positive outcome, i.e., $q_s = (1 + R_s)/2$, and R_s is defined in Eq. (4.2). Recall that R_s depends on the error parameters, and so do q_s and $Q_s(+)$.

The probability that the measurement device gives a negative outcome after performing the measurement M_s is

$$\begin{aligned} Q_s(-) =& F_-(1 - q_s) + (1 - F_+)q_s, \\ =& \frac{1}{2}\left[1 + F_- - F_+ - R_s\left(F_+ + F_- - 1\right)\right]. \end{aligned}$$

It is easy to check that $Q_s(+) + Q_s(-) = 1$.

The *imperfect measurement responses* are hence given by

$$\begin{aligned} \tilde{R}_s(\vec{p}) &= Q(+) - Q(-), \\ &= F_+ - F_- + R_s(F_+ + F_- - 1). \end{aligned}$$

Then, the matrix elements of L can be written as

$$\tilde{L}_{ru} = \left.\frac{\partial \tilde{R}_r(\vec{p})}{\partial p_u}\right|_{\vec{p}=\vec{0}}, \tag{4.10}$$

$$= (F_+ + F_- - 1)L_{ru}. \tag{4.11}$$

For imperfect measurements we also use the mean squared distance as measure of uncertainty to compare different versions of GSC. The mean squared distance considering imperfect measurements becomes

$$\langle \tilde{D}^2 \rangle = \text{Tr}(\tilde{L}^{-1}\tilde{\Sigma}\tilde{L}^{-T}), \tag{4.12}$$

where the matrix elements of the covariance matrix of the perturbed measurement responses, $\tilde{\Sigma}$, are $\tilde{\Sigma}_{su} = \frac{1}{4N}(1 - \tilde{R}_s^2)\delta_{su}$.

Let $F_+ = 0.99$ and $F_- = 0.98$ and consider the GSC presented in Ref. [39], which uses the sequences in Tab. 4.1. The statistical error measured by the mean squared distance in Eq. (4.12) is $\langle \tilde{D}^2 \rangle \approx 2.0/N$.

Allow now a different angle for each GSC rotation, i.e., consider the sequences in Tab. 4.2. With the optimisation procedure described for the case of perfect measurement, we minimise the mean squared distance in Eq. (4.12) over the angles $\theta_1, \ldots, \theta_{25}$. The angles in Fig. 4.4 achieve a minimal mean squared distance at $\langle \tilde{D}^2 \rangle \approx 0.89/N$. This means a statistical error 0.45 times smaller than in Ref. [39], but requires seventeen different rotations with a maximal sequence depth of four gates. Assuming a cost of a sequence depth of seven gates, we can decrease the number of rotations to thirteen.

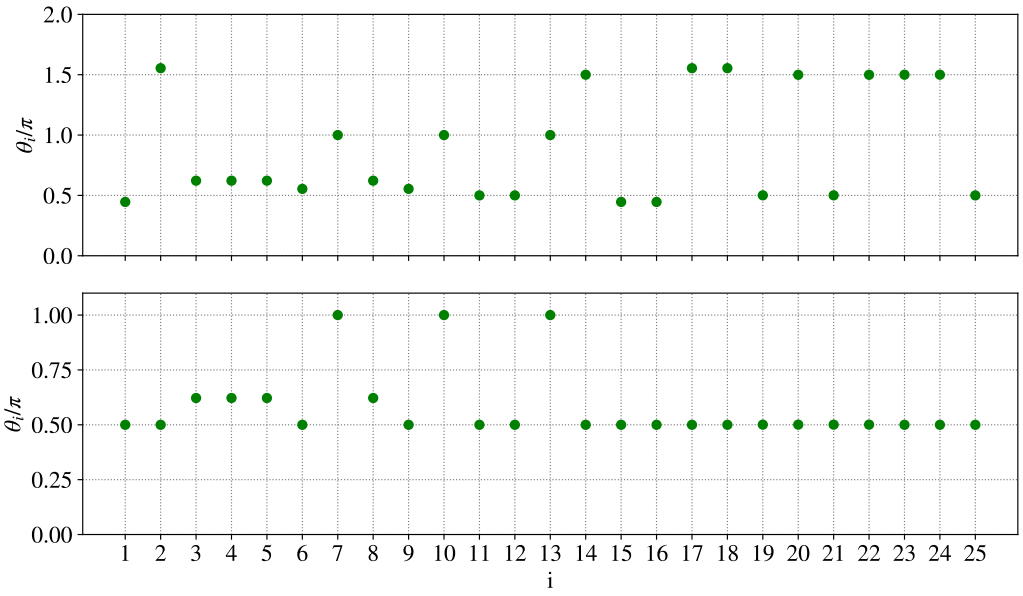

Figure 4.4: Above, angles that minimise the mean squared distance in Eq. (4.12), which considers imperfect measurements in the calibration of a $\tilde{\text{CNOT}}$ using sequences in Tab. 4.2. Below, angles proposed here to optimise GSC considering imperfect measurement. The angles below are the same angles as above after a folding transformation (i.e., mapping $\theta_i \longrightarrow 2\pi - \theta_i$ for all $\theta_i > \pi$) and setting $\theta_1 = \theta_2 = \theta_6 = \theta_9 = \theta_{15} = \theta_{16} = \theta_{17} = \theta_{18} = \pi/2$.

Analogous to the case with perfect measurements, the number of rotations can be reduced by considering symmetries of the optimal mean squared distance. We have numerically shown that the minimal $\langle \tilde{D}^2 \rangle$ is invariant under folding transformation as well as global and local reflection of the *optimal* angles. After considering the folding transformation, the number of rotations reduces to twelve without any change in the mean squared distance and a maximal sequence depth of four gates.

If twelve different single-qubit rotations are undesirable to perform in the lab, we can set the angles $\theta_1 = \theta_2 = \theta_6 = \theta_9 = \theta_{15} = \theta_{16} = \theta_{17} = \theta_{18} = \pi/2$. In contrast to the case of per-

fect measurements, the mean squared distance for imperfect measurement is not degenerated. Therefore, setting these angles implies an increase of the statistical error to $\langle \tilde{D}^2 \rangle \approx 0.90/N$. This means, however, still a reduction of the statistical error by a factor of 0.45 with respect to the statistical error in Ref. [39]. Note that the final settings suggested for GSC are the same in both cases, considering perfect and imperfect measurements.

In summary, we propose to calibrate a perturbed CNOT using GSC with the sequences in Tab. 4.3 and $\theta = 0.62208\pi$. This version of GSC requires a maximal sequence depth of five gates and a set of allowed gates consisting of $\{\mathrm{CNOT}, \mathrm{X}^{(1)}_{\pi/2}, \mathrm{Y}^{(1)}_{\pi/2}, \mathrm{X}^{(2)}_{\pi/2}, \mathrm{Y}^{(2)}_{\pi/2}, \mathrm{X}^{(1)}_{\theta}, \mathrm{Y}^{(1)}_{\theta}\}$, which contains two more single-qubit gates than the set proposed in Ref. [39]. Nevertheless, it reduces the statistical error by a factor of 0.47 for perfect measurements and a factor of 0.45 for imperfect measurements, with $F_+ = 0.99$ and $F_- = 0.98$. All results are compiled in Tab. 4.4.

	$[F_+, F_-]$	$\langle D^2 \rangle \cdot N$	# rotations	Maximal sequence depth
Angles of Ref. [39]	$[1, 1]$	1.8	2	4
Optimal angles	$[1, 1]$	0.84	17	4
Folded angles	$[1, 1]$	0.84	12	4
Modified angles	$[1, 1]$	0.84	4	5
Angles of Ref. [39]	$[0.99, 0.98]$	2.0	2	4
Optimal angles	$[0.99, 0.98]$	0.89	17	4
Folded angles	$[0.99, 0.98]$	0.89	12	4
Modified angles	$[0.99, 0.98]$	0.90	4	5

Table 4.4: Summary of numerical results, where the statistical error is measured by the mean squared distance, $\langle D^2 \rangle$. The last two columns show the number of single-qubit rotations that must be available and the maximal sequence depth for each version of GSC.

4.4 Conclusions and outlook

We have seen that optimising the design of an experiment is crucial for minimising statistical errors. In particular, we have used DOE to propose a new version of GSC that reduces the statistical error by a factor of 0.47 for perfect measurements and by a factor of 0.45 for imperfect measurements. To achieve this, one initial state, two Pauli measurements and six single-qubit rotations must be available.

Several questions remain open. Although we have reduced the statistical error, a physical explanation of the optimal sequences is still missing. We also do not know why the mean squared distance is degenerate for perfect measurements and why the degeneracy breaks down for imperfect measurements (even when $F_+ = F_- \neq 1$).

Our optimisation has been based on parmeterising the GSC sequences, but different initial states and measurements could also be considered. Furthermore, we have used fifteen settings to solve fifteen error parameters. Considering additional settings, however, and making a statistical treatment of the redundant information could further reduce the statistical error.

Finally, it would also be interesting to put our results into practice, and thus experimentally calibrate a $\tilde{\text{CNOT}}$ with the version of GSC proposed in this chapter.

Appendix

4.A Histograms of the error parameters

In Sec. 4.2, we have statistically analysed the calibration protocol GSC. Particularly assuming that the measurement responses follow a Gaussian distribution, we have determined the probability distribution of the error parameters. Here, we present some histograms of simulated data that support our statistical characterisation of GSC.

Making use of the CLT, we have determined in Sec. 4.2 that the vector of estimated responses, $\vec{R}^*(\vec{p})$, follows a multivariate Gaussian distribution such that $\vec{R}^*(\vec{p}) \sim \mathcal{N}(\vec{R}(\vec{p}), \Sigma)$, where the matrix elements of the Σ are given in Eq. (4.6). Consequently, the vector of estimated error parameters, $\vec{p}^*$, is also distributed according to a multivariate Gaussian distribution with $\vec{p}^* \sim \mathcal{N}(\vec{p}, \Sigma_{\vec{p}^*})$, where $\Sigma_{\vec{p}^*}$ is defined in Eq. (4.7). Recall that $\vec{p}^*$ can be computed with the affine transformation in Eq. (4.5) using the vector of estimated responses, $\vec{R}(\vec{p})^*$.

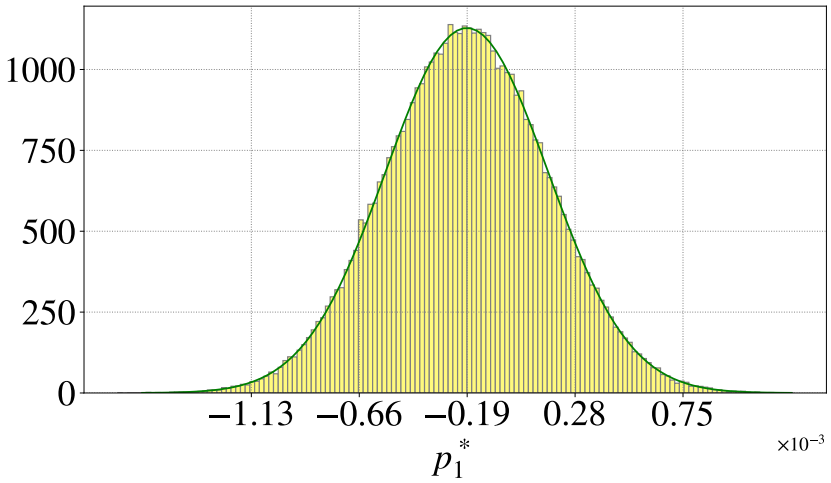

Figure 4.A.1: Histogram of the estimated error parameter p_1^* and, in green, Gaussian distribution with parameters $\mathcal{N}(p_1, \Sigma_{\vec{p}^*}^{(11)})$, where $\Sigma_{\vec{p}^*}^{(11)}$ is the first diagonal element of $\Sigma_{\vec{p}^*}$ defined in Eq. (4.7).

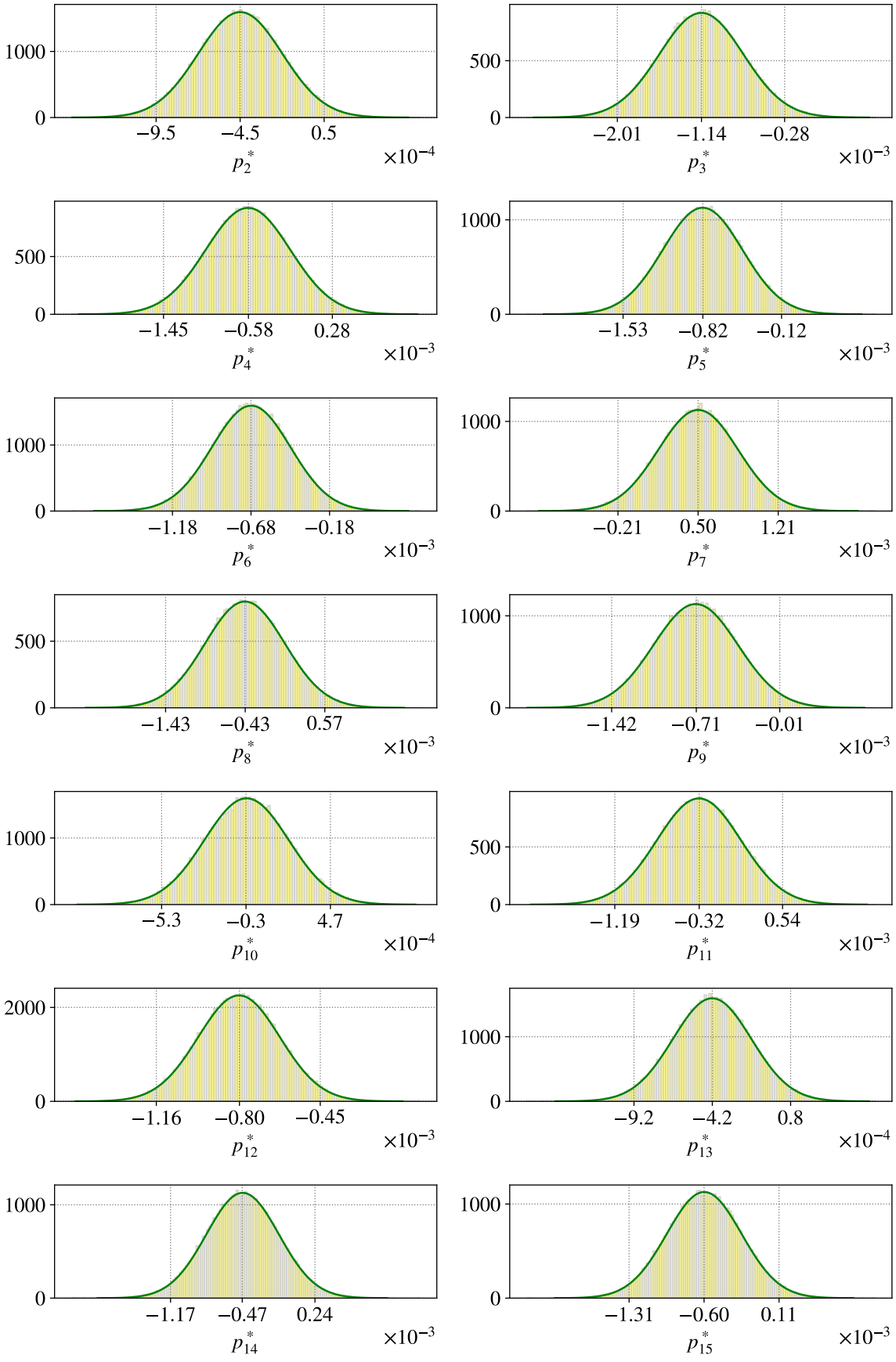

Figure 4.A.2: Histograms of the error parameters $p_2^*, \ldots, p_{15}^*$. In green, Gaussian distributions with parameters $\mathcal{N}(p_s, \Sigma_{\vec{p}^*}^{(ss)})$, where $\Sigma_{\vec{p}^*}^{(ss)}$ are the diagonal elements of $\Sigma_{\vec{p}^*}$ defined in Eq. (4.7).

To confirm the mean vector and the covariance matrix of the multivariate Gaussian distribution of $\vec{p}^*$ numerically, we simulate the calibration of a perturbed CNOT gate using the GSC introduced in Ref. [39]. Assuming the calibration with GSC is repeated $N = 10^6$ times in the lab, we perform the simulation 10^5 times. We compute the vector of estimated error parameters in each simulation and save it. With this data, we build a histogram for each estimated error parameter, p_s^* for $s = 1, \ldots, 15$. In Fig. 4.A.1, we show that the histogram of the error parameter p_1^* is compatible with the theoretically predicted Gaussian distribution $\mathcal{N}(p_1, \Sigma_{\vec{p}^*}^{(11)})$, where $\Sigma_{\vec{p}^*}^{(su)}$ are the matrix elements of $\Sigma_{\vec{p}^*}$. Analogous histograms of the estimated error parameters $p_2^*, \ldots, p_{15}^*$ are plotted apart for clarity in Fig. 4.A.2.

4.B Responses of Table 4.2

In Sec. 4.3, we have considered GSC with the settings summarised in Tab. 4.2. We present here the measurement responses corresponding to this protocol.

Using Eqns. (4.2) and (4.3), we obtain that the responses of the settings in Tab. 4.2 are given by

$$
\begin{aligned}
R_1(\vec{p}) &= \cos\theta_{15} - 2\cos\theta_{15}p_5 + 2\cos\theta_{15}p_{10},\\
R_2(\vec{p}) &= \cos\theta_1 - 2\cos\theta_1 p_4 - 2\cos\theta_1 p_7,\\
R_3(\vec{p}) &= \cos\theta_{16} - 2\cos\theta_{16}p_6 - 2\cos\theta_{16}p_9,\\
R_4(\vec{p}) &= \cos\theta_2 - 2\cos\theta_2 p_8 - 2\cos\theta_2 p_{11},\\
R_5(\vec{p}) &= \cos\theta_{17} - 2\cos\theta_{17}p_1 - 2\cos\theta_{17}p_{13},\\
R_6(\vec{p}) &= \cos\theta_{18} - 2\cos\theta_{18}p_2 - 2\cos\theta_{18}p_{14},\\
R_7(\vec{p}) &= \cos\theta_{19}\cos\theta_3 - 2\cos\theta_{19}\cos\theta_3 p_2 - 2\cos\theta_{19}\cos\theta_3 p_4\\
&\quad - 2\cos\theta_3\cos\theta_{19}p_5 - 2\cos\theta_{19}\cos\theta_3 p_7 + 2\cos\theta_{19}p_{10},\\
R_8(\vec{p}) &= \cos\theta_{20}\cos\theta_4 - 2\cos\theta_{20}\cos\theta_4 p_4 - 2\cos\theta_{20}p_6\\
&\quad - 2\cos\theta_{20}\cos\theta_4 p_7 - 2\cos\theta_4\cos\theta_{20}p_9 - 2\cos\theta_{20}\cos\theta_4 p_{13},\\
R_9(\vec{p}) &= \cos\theta_{21}\cos\theta_5 - 2\cos\theta_5\cos\theta_{21}p_5 - 2\cos\theta_{21}\cos\theta_5 p_8\\
&\quad + 2\cos\theta_{21}p_{10} - 2\cos\theta_{21}\cos\theta_5 p_{11} + 2\cos\theta_{21}\cos\theta_5 p_{13},\\
R_{10}(\vec{p}) &= \cos\theta_6\cos\theta_7 - 2\cos\theta_6 p_7 - 2\cos\theta_7\cos\theta_6 p_4\\
&\quad - 2\sin\theta_7 p_{13} - 2p_1\cos\theta_6\sin\theta_7,\\
R_{11}(\vec{p}) &= \cos\theta_{22}\cos\theta_8 - 2\cos\theta_{22}p_6 - 2\cos\theta_8\cos\theta_{22}p_9\\
&\quad - 2\cos\theta_{22}\cos\theta_8 p_{11} - 2\cos\theta_{22}\cos\theta_8 p_8 - 2\cos\theta_{22}\cos\theta_8 p_2,\\
R_{12}(\vec{p}) &= \cos\theta_{10}\cos\theta_9 - 2\cos\theta_9\cos\theta_{10}p_2 - 2\cos\theta_{10}\cos\theta_9 p_8\\
&\quad - 2\cos\theta_9 p_{11} - 2\cos\theta_{10}p_{14},
\end{aligned}
$$

$$
\begin{aligned}
R_{13}(\vec{p}) = &\cos\theta_{11}\cos\theta_{23} - 2\cos\theta_{11}\cos\theta_{23}p_1 - 2\cos\theta_{23}\cos\theta_{11}p_2 \\
&+ 2\cos\theta_{11}\cos\theta_{23}p_3 - 2\cos\theta_{11}\cos\theta_{23}p_{13} \\
&- 2\cos\theta_{23}\cos\theta_{11}p_{14} + 2\cos\theta_{11}\cos\theta_{23}p_{15}, \\
R_{14}(\vec{p}) = &\cos\theta_{12}\cos\theta_{24} - 2\cos\theta_{12}\cos\theta_{24}p_3 - 2\cos\theta_{24}\cos\theta_{12}p_4 \\
&- 2\cos\theta_{24}\cos\theta_{12}p_5 - 2\cos\theta_{12}\cos\theta_{24}p_6 \\
&- 2\cos\theta_{24}\cos\theta_{12}p_7 - 2\cos\theta_{12}\cos\theta_{24}p_8 \\
&- 2\cos\theta_{12}\cos\theta_{24}p_9 - 2\cos\theta_{12}\cos\theta_{24}p_{11} \\
&+ 2\cos\theta_{24}\cos\theta_{12}p_{10} - 4\cos\theta_{12}\cos\theta_{24}p_{12} \\
&- 2\cos\theta_{12}\cos\theta_{24}p_{15}, \\
R_{15}(\vec{p}) = &\cos\theta_{13}\cos\theta_{14}\cos\theta_{25} - 2\cos\theta_{25}\cos\theta_{13}p_7 - 2\cos\theta_{14}\cos\theta_{25}\cos\theta_{13}p_4 \\
&- 2\cos\theta_{25}\cos\theta_{14}p_{14} - 2\cos\theta_{13}\cos\theta_{25}\cos\theta_{14}p_2 \\
&- 2\cos\theta_{14}\cos\theta_{25}p_{13} - 2\cos\theta_{13}\cos\theta_{14}\cos\theta_{25}p_1 \\
&- 2\cos\theta_{13}\cos\theta_{25}p_6 + 2\cos\theta_{14}\cos\theta_{25}p_{15} \\
&+ 2\cos\theta_{13}\cos\theta_{14}\cos\theta_{25}p_3.
\end{aligned}
$$

4.C Statistical error as a function of the rotation angle

The optimisation of GSC performed in Sec. 4.3 has been carried out by allowing a different angle for each single-qubit rotation. Nevertheless, GSC permits not only changing the sequences of the settings but also the measurements and the initial states. Here, we present some motivation for our choice of optimisation strategy.

As we have mentioned in the main text, the goal of this project is to optimise GSC by minimising statistical errors. We intend, at the same time, to propose a new version of GSC that is physically feasible, and thus we aim to introduce small modifications on the original version of GSC. For this reason, we start considering the same set-up as in Ref. [39], i.e., the settings in Tab. 4.1. In this version of GSC all single-qubit gates are $\pi/2$-rotations, so a simple change on the protocol is considering a different rotation angle, which we denote as θ. Computing the mean squared distance of GSC, $\langle D^2 \rangle$ defined in Eq. (4.8), for different values of θ, we can plot the statistical error of GSC as a function of the rotation angle. In Fig. 4.C.1, we find that the choice made in Ref. [39], i.e., $\theta = \pi/2$, does not give the minimum mean squared distance, but the minimum $\langle D^2 \rangle$ is achieved for $\theta = 1.428\pi$ does.

In order to further reduce the statistical error, we have considered in Sec. 4.3 not only one rotation angle, but a different angle for each single-qubit rotation.

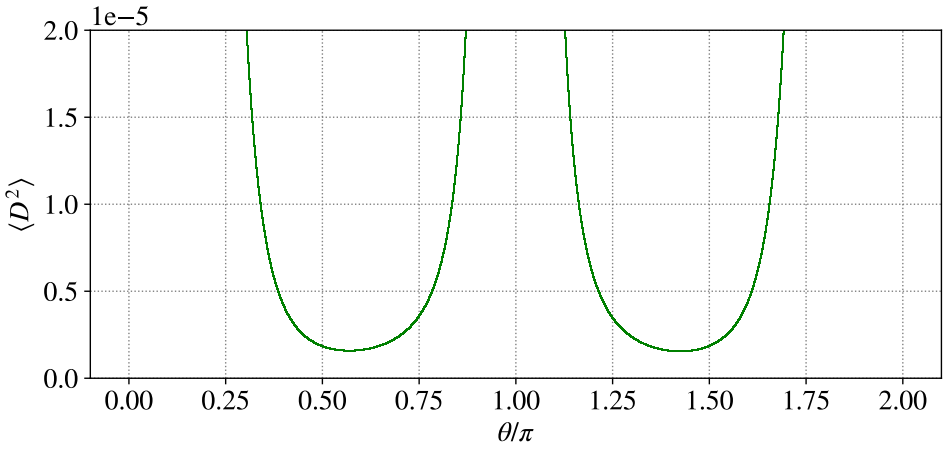

Figure 4.C.1: Mean squared distance, $\langle D^2 \rangle$ defined in Eq. (4.8), as a function of the single-qubit-rotation angle, θ.

4.D Optimisation code

Section 4.3 has been devoted to optimise GSC. Below we present the code that we developed and used for this optimisation.

```
import numpy as np
import nlopt as nlopt

import functions.basic_functions as bf
import functions.fake_data_functions as fd
import functions.optimisation_functions as opt

# OPTIMISATION OVER 25 ANGLES
M = 1*10**4 # Number of optimisations
F = [1, 1] # Measurement fidelities. If F=[1,1], perfect measurements.

list_result_code = [] # Initialise list for stopping criteria
list_opt_mean_distance2 = [] # Initialise list for mean squared distance
list_opt_vec_theta = [] # Initialise list for angles

for m in range(M):
    # When the optimiser runs into a RoundoffLimited error, the value of theta is skipped.
    try:
        # The initial point of the optimiser is chosen randomly.
        vec_theta0 = np.random.random(25)*2*np.pi
```

```
# Check if the initial value gives a non-singular L.
inv = bf.is_invertible(np.array(opt.L(vec_theta0,F)))
while inv == False:
    vec_theta0 = np.random.random(25) * 2 * np.pi
    inv = bf.is_invertible(opt.L(vec_theta0))

# (LOCAL) OPTIMISATION

# Choice of the optimiser and number of parameters to optimise
optL = nlopt.opt(nlopt.LN_BOBYQA, 25)

# Set max time for each optimisation
optL.set_maxtime(0.25*60*60)

# Set 0 as lower bound for the angles
optL.set_lower_bounds(np.zeros(25))

# Set 2*pi as upper bound for the angles
optL.set_upper_bounds([2*np.pi for i in range(25)])

# Choose the function called opt.mysuperfunc1 to minimise
optL.set_min_objective(opt.mysuperfunc1)

# Set tolerance for the angles
optL.set_xtol_rel(1e-10)

# Start optimisation with vec_theta0 as initial point and keep the optimal angles
opt_vec_theta = optL.optimize(vec_theta0)

# Keep optimised mean squared distance
opt_mean_distance2 = optL.last_optimum_value()

# Keep the stopping criteria
result_code_L = optL.last_optimize_result()

list_opt_vec_theta.append(opt_vec_theta)
list_opt_mean_distance2.append(opt_mean_distance2)
list_result_code.append(result_code_L)
```

```
    # Skip values of theta that produce a RoundoffLimited error
    except nlopt.RoundoffLimited:
        print("RoundoffLimited error at m=", m, ".")
        print("vec_theta0 =", vec_theta0)
        print("Value skipped.")
```

The code above optimises a function called `opt.mysuperfunc1`, which we have defined as

```
def mysuperfunc1(vec_theta, grad):
    N = 10**6 # Number of runs of the "experiment in the lab"
    F = [1, 1] # Measurement fidelities

    return mean_distance2(vec_theta, N, F)
```

where

```
def mean_distance2(vec_theta, N, F):
    # Constructing the matrix L
    matrix_L = L(vec_theta, F)

    # Constructing covariance matrix of responses
    cov_matrix = cov_matrixR(vec_theta, N, F)

    # If L is non-singular, the function computes the mean squared distance
    if bf.is_invertible(np.array(matrix_L))==True:
        D2 = np.matrix.trace(np.linalg.inv(matrix_L) @ cov_matrix @ np.matrix.transpose(np.
            ↪ linalg.inv(matrix_L)))
        return D2
    else: # If L is singular, the function returns infinity.
        return mth.inf
```

with the function `L(vec_theta,F)` computing the L matrix defined in Eq. (4.10).

4.E Optimal angles

In Sec. 4.3 we have optimised the calibration of a $\widetilde{\text{CNOT}}$ via GSC with sequences in Tab. 4.2 considering perfect and imperfect measurements. Here, the concrete values of $\theta_1, \ldots, \theta_{25}$ that achieve the minimal mean squared distance are presented for both cases.

With the optimisation process presented in Sec. 4.3, we have obtained a minimal squared distance in the case of perfect measurements (Eq. (4.8)) of $\langle D^2 \rangle \approx 0.84/N$ with the angles

$$
\begin{aligned}
\theta_1/\pi &= 1.3864, & \theta_2/\pi &= 0.3743,\\
\theta_3/\pi &= 1.3779, & \theta_4/\pi &= 1.3779,\\
\theta_5/\pi &= 1.3779, & \theta_6/\pi &= 1.5722,\\
\theta_7/\pi &= 1.0000, & \theta_8/\pi &= 1.3779,\\
\theta_9/\pi &= 0.4003, & \theta_{10}/\pi &= 1.0000,\\
\theta_{11}/\pi &= 0.5000, & \theta_{12}/\pi &= 1.5000,\\
\theta_{13}/\pi &= 1.0000, & \theta_{14}/\pi &= 0.5000,\\
\theta_{15}/\pi &= 1.0310, & \theta_{16}/\pi &= 0.2618,\\
\theta_{17}/\pi &= 0.3087, & \theta_{18}/\pi &= 0.8718,\\
\theta_{19}/\pi &= 0.5000, & \theta_{20}/\pi &= 0.5000,\\
\theta_{21}/\pi &= 0.5000, & \theta_{22}/\pi &= 1.5000,\\
\theta_{23}/\pi &= 0.5000, & \theta_{24}/\pi &= 0.5000,\\
\theta_{25}/\pi &= 1.5000.
\end{aligned}
$$

These angles are plotted in Fig. 4.2.

Using the same optimisation process, but the mean squared distance in Eq. (4.12), which considers imperfect measurements, we have obtained a minimum mean squared distance at $\langle \tilde{D}^2 \rangle \approx 0.89/N$ with the angles

$$
\begin{aligned}
\theta_1/\pi &= 0.4459, & \theta_2/\pi &= 0.4459,\\
\theta_3/\pi &= 1.3779, & \theta_4/\pi &= 1.3779,\\
\theta_5/\pi &= 0.6221, & \theta_6/\pi &= 0.5541,\\
\theta_7/\pi &= 1.0000, & \theta_8/\pi &= 0.6221,\\
\theta_9/\pi &= 0.5541, & \theta_{10}/\pi &= 1.0000,\\
\theta_{11}/\pi &= 1.5000, & \theta_{12}/\pi &= 0.5000,\\
\theta_{13}/\pi &= 1.0000, & \theta_{14}/\pi &= 1.5000,\\
\theta_{15}/\pi &= 1.5541, & \theta_{16}/\pi &= 0.4459,\\
\theta_{17}/\pi &= 1.5541, & \theta_{18}/\pi &= 0.4459,\\
\theta_{19}/\pi &= 1.4993, & \theta_{20}/\pi &= 0.5007,\\
\theta_{21}/\pi &= 0.5007, & \theta_{22}/\pi &= 1.4993,\\
\theta_{23}/\pi &= 1.5000, & \theta_{24}/\pi &= 0.5000,\\
\theta_{25}/\pi &= 1.5000.
\end{aligned}
$$

Figure 4.4 contains a plot of these angles.

Conclusions

This chapter summarises the results presented along the thesis and discusses future research as well as open questions. In accordance with the structure of the thesis, we divide the chapter into two sections, where each presents the conclusion of one part.

Classical restrictions of matrix product states

Summary

Physical interactions are local, and this locality is one of the most relevant factors that enforces which states are likely for low-energy physical systems. Matrix product states are an efficient representation of ground states of any gapped Hamiltonian. Not only do MPSs provide a complete framework for the most physically relevant Hamiltonians, but simulations based on MPSs are also highly precise and efficient. The efficiency and high precision of MPSs rely on the accurate description of correlations within the state, which are restricted to be local.

Other approaches to simulate many-body physics, such as quantum Monte-Carlo simulations, use heuristic assumptions on the state characterising the system. Although convergence is not guaranteed, quantum Monte-Carlo simulations usually succeed when the system is characterised by a locally restricted Gibbs Ansatz. The notion of locality used in quantum Monte-Carlo algorithm consists in assuming that the Hamiltonian describing the system is local, and hence the corresponding Gibbs state is also local.

With the research presented in Ch. 2, we have contributed in connecting these two notions of locality. Our main result consists in proving that classical restrictions of generic MPSs are quasi-locally Gibbsian. The proof of this result has been divided in the following steps. First, we have shown that any probability distribution defined on a one-dimensional lattice with open boundary conditions is exponentially close to a local Gibbs distribution, if the CMI between any tripartition of the lattice decays rapidly in the width of the middle region. Then, the CMI of classical restrictions of injective MPSs is demonstrated to decay if the purity condition is fulfilled. Finally, we have shown that the purity condition is a generic property in a concrete probabilistic model.

Outlook

The first question that naturally arises concerning our research presented in Ch. 2 is whether the results obtained are also valid for tensor networks of higher dimensions. A generalisation of our results in the context of matrix product density operators [71] or continuous MPSs [72] would also be interesting.

Additionally, the connection that we have established between classical restrictions of MPSs and Gibbs states could provide new understanding and algorithm improvements in quantum simulation. For example, could this connection be used for a rigorous convergence proof of the Monte Carlo wave function method [73]?

Within Ch. 2 we have proved that the CMI decays exponentially if the purity condition is satisfied, but no concrete decay rate has been given. It would be desirable to have a better insight into the dependencies of the decay rate and, even more, to obtain a closed expression.

There are also open questions regarding the purity condition. We have shown that the purity condition is necessary and sufficient for the quantum CMI, but only a sufficient statement has been proved for the classical CMI. One could wonder, then, if the purity condition is also a necessary condition in the classical case or if a weaker condition which is sufficient and necessary exists. On another note, is it easy to check whether a given MPS satisfies the purity condition?

Finally, our proofs make extensive use of the theory of random matrix products, and thus we wonder if these tools could also be advantageous in other fields.

Minimising statistical errors in gate calibration

Summary

Statistics is an essential tool in experimental science. Either before an experiment via DOE or afterwards via statistical analysis, statistics can be used to improve the results of experiments.

DOE is of significant importance in quantum mechanics because responses of measurements are predicted only probabilistically. In other words, every measurement outcome is a random variable distributed according to a concrete probability distribution with an associated variance. Moreover, when an experiment extracts more than one measurement response, we need to consider random vectors with possibly correlated components in the statistical analysis. Carefully planning DOE allows for the global statistical error to be minimised.

Chapter 4 encloses the application of DOE on the calibration protocol called GSC [39], and, in particular, in the calibration of a perturbed CNOT gate. We have proposed a modification on the calibration protocol that implies having two more single-qubit rotations available and allowing for a sequence length of five gates instead of four. In return, these changes on the protocol reduce the statistical error of the calibration by a factor of 0.47 when perfect measurements are considered and by a factor of 0.45 for imperfect measurements.

Outlook

The research conducted in Ch. 2 has opened some questions. We have reduced the statistical error of GSC by finding a trade-off between numerical stability and uncertainty. Nevertheless, we do not theoretically understand why the settings proposed here give a smaller statistical error than the sequences proposed in Ref. [39].

During the statistical analysis, we have numerically observed that, in the case of perfect measurements, the minimal statistical error does not depend on some of the rotations forming the sequences. It would be useful to have more insights into this property and understand why the dependency is different for perfect and imperfect measurements.

Our optimisation protocol has only considered modifications on the single-qubit rotations. The settings of GSC, however, can also vary in its initial state and measurement. Including variation on them could be a research line for further investigations.

We have assumed single-qubit rotations to be calibrated. A follow-up project could consider self-consistent calibration, i.e., calibrating the CNOT gate and the single-qubit rotations required for this calibration. How precise should single-qubit rotations be so that our GSC version still reduces the statistical error despite the cost of the number of rotations and the increased sequence length?

In order to obtain the error parameters, we have solved an inverse problem where the number of equations was the same as the number of parameters. One could consider more settings in the GSC such that the problem is overdetermined. Statistical analysis could be then applied to the redundant information, which may eventually lead to an extra reduction in the statistical error.

Finally, an experimental implementation of the version of GSC proposed in Ch. 4 would be interesting follow-up project.

Acknowledgments

In one session with my psychologist, I described doing a PhD like being in a tsunami. Ok, maybe I was having a hard time and not seeing the whole picture, but although it has not always been like this, there have been moments where I felt sinking. If now I am writing the acknowledgements of my PhD thesis, so if I arrived at a safe haven, it has been with the help of many.

Johan, thanks. You have been an essential part of this journey. Either before the pandemic, turning around, taking a pen and writing in the middle of a blank page, or clicking the zoom button during pandemic times and exchanging the pen and the blank page for the mouse and the screen of your laptop. Both actions are so natural for you. Your dedication to students, group-mates, family and simply to any person around you is admirable. Thanks for helping me *surviving* this tsunami.

(Big) David, thanks for adopting me. Having the opportunity to work with you has been inspiring and motivating. I have appreciated your clear and practical guidance as well as your attentive and open-minded leadership. Moreover, I am grateful that you pushed me and convinced me to start a project that involved coding, as this made me discover its fun.

Thank you, Michael, for supervising my challenging first project. Your expertise in formulating research questions and new ideas has been fundamental.

Working in the quantum-information research group in Cologne has been a great experience. Not only have I learned from all of you about physics, but also about fellowship. Especially, many thanks to Mariela for taking care of all of us with patience, joy and, above all, warmth. You have made me feel that nothing can go wrong when you are around.

Many thanks as well to experimentalists in the Bluhm group in Aachen. Your feedback on the second part of the thesis has definitely enhanced the project.

An meine Freunde, a les meves amigues. Alícia i Sandra sou violetes a la meva barca, alçant les veles quan fa falta o celebrant el cel obert quan la mar és en calma. Miriam, you continuously teach me about life. Kilian and Raphael, thanks for sharing experiences (and for reading parts of my thesis!).

Querida señorita Birke, gracias de corazón por mostrarme que existen otras formas. Sin usted, la última frase de estos agradecimientos nunca hubiera sido posible.

Gràcies Papa per transmetre'm la passió per la ciència i el coneixement. Mama, gràcies per estar sempre a l'altra banda del telèfon o agafant un avió si fos necessari. Sou un exemple de coratge, valentia i dedicació. Xarma, Nano, Josef, Sara, imprescindibles. Minis, avui què voleu un conte o un poema? Gràcies per sempre mostrar entusiasme per desvirtualitzar-me. Us estimo.

Dr. Niklitas, vielen Dank. Gràcies for not giving up when I had already dropped out. Wir sind ein Team (de tres).

La última frase és per mi. Agraeixo la meva resiliència.

Bibliography

[1] P. W. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". In: *SIAM Journal on Computing* 26.5 (Oct. 1997), pp. 1484–1509.

[2] S. K. Liao. et al. "Satellite-to-ground quantum key distribution". In: *Nature* 549 (2017), pp. 43–47.

[3] D. P. DiVincenzo. "The Physical Implementation of Quantum Computation". In: *Fortschritte der Physik* 48.9-11 (Sept. 2000), pp. 771–783.

[4] N. Friis et al. "Observation of Entangled States of a Fully Controlled 20-Qubit System". In: *Phys. Rev. X* 8 (2 Apr. 2018), p. 021012.

[5] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage. "Trapped-ion quantum computing: Progress and challenges". In: *Applied Physics Reviews* 6.2 (2019), p. 021314.

[6] N. K. Bernardes et al. "High Resolution non-Markovianity in NMR". In: *Scientific Reports* 6 (Mar. 2016).

[7] Y. Wu et al. "A programmable two-qubit solid-state quantum processor under ambient conditions". In: *npj Quantum Information* 5 (Dec. 2019), p. 9.

[8] F. Arute et al. "Quantum Supremacy using a Programmable Superconducting Processor". In: *Nature* 574 (2019), pp. 505–510.

[9] H.-S. Zhong et al. "Quantum computational advantage using photons". In: *Science* 370.65-23 (2020), pp. 1460–1463.

[10] "Wikipedia. Observable Universe". `https://en.wikipedia.org/wiki/Observable_universe`. Accessed: 29.01.2022.

[11] J. Eisert, M. Cramer, and M. B. Plenio. "Colloquium: Area laws for the entanglement entropy". In: *Reviews of Modern Physics* 82.1 (Feb. 2010), pp. 277–306.

[12] M. B. Hastings. "Solving gapped Hamiltonians locally". In: *Physical Review B* 73.8 (Feb. 2006).

[13] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac. "Area Laws in Quantum Systems: Mutual Information and Correlations". In: *Physical Review Letters* 100.7 (Feb. 2008).

[14] M. B. Hastings. "Entropy and entanglement in quantum ground states". In: *Phys. Rev. B* 76 (3 July 2007), p. 035114.

[15] I. Arad, A. Kitaev, Z. Landau, and U. Vazirani. "An area law and sub-exponential algorithm for 1D systems". arXiv: quant-ph/1301.1162. 2013.

[16] V. Vedral. "Mean-field approximations and multipartite thermal correlations". In: *New Journal of Physics* 6 (Feb. 2004), pp. 22–22.

[17] K. G. Wilson. "The renormalization group: Critical phenomena and the Kondo problem". In: *Rev. Mod. Phys.* 47 (4 Nov. 1975), pp. 773–840.

[18] F. Verstraete, V. Murg, and J. Cirac. "Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems". In: *Advances in Physics* 57.2 (Mar. 2008), pp. 143–224.

[19] R. Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states". In: *Annals of Physics* 349 (Nov. 2014), pp. 117–158.

[20] D. Perez-Garcia, F. Verstraete, M. Wolf, and J. Cirac. "Matrix Product State Representations". In: *Quantum Information & Computation* 7 (July 2007), pp. 401–430.

[21] M. Fannes, B. Nachtergaele, and R. F. Werner. "Finitely correlated states on quantum spin chains". In: *Communications in Mathematical Physics* 144.3 (1992), pp. 443–490.

[22] U. Schollwöck. "The density-matrix renormalization group in the age of matrix product states". In: *Annals of Physics* 326.1 (Jan. 2011), pp. 96–192.

[23] M. B. Hastings. "Solving gapped Hamiltonians locally". In: *Physical Review B* 73.8 (Feb. 2006).

[24] M. Sanz Ruiz. "Tensor Networks in Condensed Matter". Dissertation. München: Technische Universität München, 2011.

[25] F. Verstraete and J. I. Cirac. "Matrix product states represent ground states faithfully". In: *Phys. Rev. B* 73 (9 Mar. 2006), p. 094423.

[26] N. Schuch and J. I. Cirac. "Matrix product state and mean-field solutions for one-dimensional systems can be found efficiently". In: *Phys. Rev. A* 82 (1 July 2010), p. 012314.

[27] Z. Landau, U. Vazirani, and T. Vidick. "A polynomial-time algorithm for the ground state of 1D gapped local Hamiltonians". arXiv: quant-ph/1307.5143. 2013.

[28] I. Affleck, T. G. Kennedy, E. H. Lieb, and H. Tasaki. "Valence bond ground states in isotropic quantum antiferromagnets". In: *Communications in Mathematical Physics* 115 (1988), pp. 477–528.

[29] F. G. S. L. Brandão and M. Horodecki. "Exponential Decay of Correlations Implies Area Law". In: *Communications in Mathematical Physics* 333.2 (Nov. 2014), pp. 761–798.

[30] M. Sanz, D. Perez-Garcia, M. M. Wolf, and J. I. Cirac. "A Quantum Version of Wielandt's Inequality". In: *IEEE Transactions on Information Theory* 56.9 (Oct. 2010), pp. 4668–4673.

[31] D. M. Greenberger, M. A. Horne, and A. Zeilinger. "Going Beyond Bell's Theorem". In: *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. Ed. by M. Kafatos. Dordrecht: Springer Netherlands, 1989, pp. 69–72.

[32] W. Dür, G. Vidal, and J. I. Cirac. "Three qubits can be entangled in two inequivalent ways". In: *Physical Review A* 62.6 (Nov. 2000).

[33] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki. "Rigorous results on valence-bond ground states in antiferromagnets". In: *Phys. Rev. Lett.* 59 (7 Aug. 1987), pp. 799–802.

[34] Y. Aragonés-Soria, J. Åberg, C.-Y. Park, and M. J. Kastoryano. "Classical restrictions of generic matrix product states are quasi-locally Gibbsian". In: *Journal of Mathematical Physics* 62.9 (Oct. 2021), p. 093511.

[35] O. K. Kozlov. "Gibbs Description of a System of Random Variables". In: *Probl. Peredachi Inf.* 10.3 (1974). English translation: Problems of Information Transmission, 1974, **10**:3, 258–265, pp. 94–103.

[36] P. Bougerol. "Products of Random Matrices with Applications to Schrödinger Operators". Progress in Probability. Birkhäuser Boston, 2012.

[37] T. Benoist, M. Fraas, Y. Pautrat, and C. Pellegrini. "Invariant measure for quantum trajectories". In: *Prob. Theory and Related Fields* 174 (2019), pp. 307–334.

[38] H. Maassen and B. Kümmerer. "Purification of quantum trajectories". In: *Dynamics & Stochastics*. Ed. by D. Denteneer, F. d. Hollander, and E. Verbitskiy. Vol. Volume 48. Lecture Notes–Monograph Series. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2006, pp. 252–261.

[39] P. Cerfontaine, R. Otten, and H. Bluhm. "Self-Consistent Calibration of Quantum-Gate Sets". In: *Physical Review Applied* 13.4 (Apr. 2020).

[40] Y. Tong. "The Multivariate Normal Distribution". Springer, New York, NY, 1990.

[41] D. C. Montgomery. "Design and Analysis of Experiments". Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

[42] D. Aharonov and M. Ben-Or. "Fault-Tolerant Quantum Computation with Constant Error". In: *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*. STOC '97. El Paso, Texas, USA: Association for Computing Machinery, 1997, pp. 176–188.

[43] J. Preskill. "Reliable quantum computers". In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454.1969 (Jan. 1998), pp. 385–410.

[44] J. F. Poyatos, J. I. Cirac, and P. Zoller. "Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate". In: *Phys. Rev. Lett.* 78 (2 Jan. 1997), pp. 390–393.

[45] S. T. Merkel et al. "Self-consistent quantum process tomography". In: *Physical Review A* 87.6 (June 2013).

[46] J. B. Altepeter et al. "Ancilla-Assisted Quantum Process Tomography". In: *Phys. Rev. Lett.* 90 (19 May 2003), p. 193601.

[47] G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi. "Quantum Tomography". arXiv: quant-ph/0302028. 2003.

[48] B. Lévi, C. C. López, J. Emerson, and D. G. Cory. "Efficient error characterization in quantum information processing". In: *Phys. Rev. A* 75 (2 Feb. 2007), p. 022314.

[49] E. Knill et al. "Randomized benchmarking of quantum gates". In: *Phys. Rev. A* 77 (1 Jan. 2008), p. 012307.

[50] E. Magesan, J. M. Gambetta, and J. Emerson. "Scalable and Robust Randomized Benchmarking of Quantum Processes". In: *Phys. Rev. Lett.* 106 (18 May 2011), p. 180504.

[51] E. Magesan, J. M. Gambetta, and J. Emerson. "Characterizing quantum gates via randomized benchmarking". In: *Phys. Rev. A* 85 (4 Apr. 2012), p. 042311.

[52] A. Carignan-Dugas, J. J. Wallman, and J. Emerson. "Characterizing universal gate sets via dihedral benchmarking". In: *Phys. Rev. A* 92 (6 Dec. 2015), p. 060302.

[53] J. J. Wallman, M. Barnhill, and J. Emerson. "Robust characterization of leakage errors". In: *New Journal of Physics* 18.4 (Apr. 2016), p. 043021.

[54] A. Carignan-Dugas, K. Boone, J. J. Wallman, and J. Emerson. "From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters". In: *New Journal of Physics* 20.9 (Oct. 2018), p. 092001.

[55] K. Boone, A. Carignan-Dugas, J. J. Wallman, and J. Emerson. "Randomized benchmarking under different gate sets". In: *Phys. Rev. A* 99 (3 Mar. 2019), p. 032329.

[56] C. Dankert, R. Cleve, J. Emerson, and E. Livine. "Exact and approximate unitary 2-designs and their application to fidelity estimation". In: *Phys. Rev. A* 80 (1 July 2009), p. 012304.

[57] J. Eisert et al. "Quantum certification and benchmarking". In: *Nature Reviews Physics* 2.7 (June 2020), pp. 382–390.

[58] M. Kliesch and I. Roth. "Theory of Quantum System Certification". In: *PRX Quantum* 2.1 (Jan. 2021).

[59] S. Kimmel et al. "Robust Extraction of Tomographic Information via Randomized Benchmarking". In: *Phys. Rev. X* 4 (1 Mar. 2014), p. 011050.

[60] Y. Gu, R. Mishra, B.-G. Englert, and H. K. Ng. "Randomized Linear Gate-Set Tomography". In: *PRX Quantum* 2 (3 Aug. 2021), p. 030328.

[61] R. Brieger, I. Roth, and M. Kliesch. "Compressive gate set tomography". arXiv: quant-ph/2112.05176. 2021.

[62] L. Ruppert, D. Virosztek, and K. Hangos. "Optimal parameter estimation of Pauli channels". In: *Journal of Physics A: Mathematical and Theoretical* 45.26 (June 2012), p. 265305.

[63] G. Ballo, K. M. Hangos, and D. Petz. "Convex Optimization-Based Parameter Estimation and Experiment Design for Pauli Channels". In: *IEEE Transactions on Automatic Control* 57.8 (2012), pp. 2056–2061.

[64] R. Kosut, I. A. Walmsley, and H. Rabitz. "Optimal experiment design for quantum state and process tomography and hamiltonian parameter estimation". arXiv: quant-ph/0411093. 2004.

[65] J. Suzuki. "Quantum-state estimation problem via optimal design of experiments". arXiv: quant-ph/2012.11949. 2020.

[66] J. Nunn et al. "Optimal experiment design for quantum state tomography: Fair, precise, and minimal tomography". English. In: *Physical Review A* 81.4 (Apr. 2010), p. 042109.

[67] R. Blume-Kohout et al. "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography". In: *Nature Communications* 8.1 (Feb. 2017).

[68] S. G. Johnson. "NLopt nonlinear-optimization package". `https://nlopt.readthedocs.io/en/latest/`. [Online; accessed 28-December-2021].

[69] M. J. D. Powell. "The BOBYQA algorithm for bound constrained optimization without derivatives". Cambridge, England: Department of Applied Mathematics and Theoretical Physics, technical report, NA2009/06, 2009.

[70] Y. Aragonés-Soria. "Minimising the statistical error in the calibration of quantum-gate sets". `https://github.com/yaiza-aragonessoria/min_stat_error_GSC.git`. 2022.

[71] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac. "Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems". In: *Physical Review Letters* 93.20 (Nov. 2004).

[72] F. Verstraete and J. I. Cirac. "Continuous Matrix Product States for Quantum Fields". In: *Physical Review Letters* 104.19 (May 2010).

[73] J. Dalibard, Y. Castin, and K. Mølmer. "Wave-function approach to dissipative processes in quantum optics". In: *Phys. Rev. Lett.* 68 (5 Feb. 1992), pp. 580–583.

[74] M. A. Nielsen and I. L. Chuang. "Quantum Computation and Quantum Information: 10th Anniversary Edition". 10th. USA: Cambridge University Press, 2011.

[75] B. Zeng, X. Chen, D. Zhou, and X.-G. Wen. "Quantum Information Meets Quantum Matter". New York, United States: Springer-Verlag New York Inc., 2019.

[76] R. Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states". In: *Annals of Physics* 349 (Nov. 2014), pp. 117–158.

[77] R. D. Somma, S. Boixo, H. Barnum, and E. Knill. "Quantum Simulations of Classical Annealing Processes". In: *Phys. Rev. Lett.* 101 (13 Oct. 2008), p. 130504.

[78] M. Kieferová and N. Wiebe. "Tomography and generative training with quantum Boltzmann machines". In: *Phys. Rev. A* 96 (6 Dec. 2017), p. 062327.

[79] J. Biamonte et al. "Quantum machine learning". In: *Nature* 549.7671 (Oct. 2017), pp. 195–202.

[80] F. G. Brandao and K. M. Svore. "Quantum Speed-Ups for Solving Semidefinite Programs". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 415–426.

[81] J. Hammersley and D. Handscomb. "Monte Carlo Methods". Vol. 40. Methuen & Co, 1964.

[82] M. Suzuk. "Quantum Monte Carlo Methods in Equilibrium and Nonequilibrium Systems". Springer, Berlin, Heidelberg, 1986.

[83] H. De Raedt and W. von der Linden. "Quantum lattice problems". In: *Binder K. (eds) The Monte Carlo Method in Condensed Matter Physics. Topics in Applied Physics*. Vol. 71. Springer, Berlin, Heidelberg, 1992.

[84] A. N. Chowdhury, G. H. Low, and N. Wiebe. "A Variational Quantum Algorithm for Preparing Quantum Gibbs States". arXiv: quant-ph/2002.00055. 2020.

[85] N. Wiebe, A. Kapoor, and K. M. Svore. "Quantum Deep Learning". In: *Quantum Info. Comput.* 16.7–8 (May 2016), pp. 541–587.

[86] M.-H. Yung and A. Aspuru-Guzik. "A quantum–quantum Metropolis algorithm". In: *Proceedings of the National Academy of Sciences* 109.3 (2012), pp. 754–759.

[87] D. B. Kaplan, N. Klco, and A. Roggero. "Ground States via Spectral Combing on a Quantum Computer". arXiv: quant-ph/1709.08250. 2017.

[88] A. Riera, C. Gogolin, and J. Eisert. "Thermalization in Nature and on a Quantum Computer". In: *Phys. Rev. Lett.* 108 (8 Feb. 2012), p. 080402.

[89] M. Motta et al. "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution". In: *Nature Physics* 16.2 (Nov. 2019), pp. 205–210.

Formalia

Kurzzusammenfassung

Diese Arbeit besteht aus zwei unabhängigen Teilen. Im ersten Teil untersuchen wir den Begriff der Lokalität im Zusammenhang mit Matrizproduktzuständen (MPS, engl. *Matrix Product State*), während wir im zweiten Teil den statistischen Fehler in einem Kalibrierungsprotokoll für Quantengattersets minimieren.

Im ersten Teil dieser Arbeit betrachten wir einen MPS auf einem eindimensionalen Gitter und untersuchen die Norm der quadrierten Amplituden in Bezug auf eine lokale orthonormale Basis. Wir nennen diese Norm die klassische Restriktion des MPS. Konkret stellen wir die Frage, wann die klassische Restriktion eines MPS quasi-lokal Gibbssch ist, d.h. exponentiell gut durch eine Gibbs-Verteilung mit einem lokalen Hamiltonian approximiert ist. Wir beweisen, dass die klassische Restriktion eines injektiven MPS quasi-lokal Gibbssch ist, wenn die mit dem MPS assoziierten Matrizen eine 'Reinheits'-Bedingung erfüllen, eine Bedingung, die zuvor in der Theorie der zufälligen Matrixprodukte etabliert wurde. Unser Ergebnis verbindet zwei Begriffe von Lokalität: Lokalität von Korrelationen in einem MPS und Lokalität von Wechselwirkungen in dem Hamiltonian, die die entsprechende Gibbs-Verteilung erzeugt.

Der Beweis für unser Ergebnis besteht aus zwei Schritten. Zunächst zeigen wir für einen auf einem Gitter definierten MPS, für welchen die Reinheits-Bedingung gilt, dass die bedingte Transinformation (CMI, engl. *Conditional Mutual Information*) einer beliebigen zusammenhängenden Dreiteilung des Gitters mit der Breite der mittleren Partition schnell abnimmt. Anschließend wird gezeigt, dass dieses Abklingen der klassischen CMI impliziert, dass die Wahrscheinlichkeitsverteilung, die mit der klassischen Einschränkung des MPS einhergeht, quasi-lokal Gibbssch ist. Im Rahmen dieser Untersuchung stellen wir Beobachtungen rund um die Reinheits-Bedingung vor. Wir untersuchen, wie 'typisch' die Reinheits-Bedingung ist, indem wir ein probabilistisches Modell konstruieren und zeigen, dass in diesem Modell die Reinheits-Bedingung im Allgemeinen erfüllt ist. Darüber hinaus zeigen wir, dass die Verletzung der Reinheits-Bedingung einen verallgemeinerten Begriff der Fehlerkorrektur im virtuellen Raum ermöglicht, was den generischen Charakter der Reinheits-Bedingung unterstreicht.

Im zweiten Teil dieser Arbeit betrachten wir ein Protokoll zur Kalibrierung von Quantengattersets und führen eine statistische Analyse durch, um statistische Fehler zu minimieren.

Die Kalibrierung von Quantengattern ist eine notwendige Hürde auf dem Weg zu einem zuverlässigen Quantencomputer. In einer kürzlich erschienenen Arbeit wurde das Quantengatterset-Kalibrierungsprotokoll (GSC, engl. *Gate Set Calibration protocol*) vorgestellt und verwendet, um kohärente Fehler aus multi-qubit Quantengattern zu extrahieren. Wir bauen auf dieser Studie in zweierlei Hinsicht auf. Erstens berücksichtigen wir die Unsicherheit jeder Messung im Protokoll, indem wir eine statistische Analyse durchführen. Zweitens optimieren wir die statistische Unsicherheit, während wir gleichzeitig verlangen, dass das Protokoll nur eine geringe Anzahl von unterschiedlichen Gattern umfasst, was die physikalische Realisierbarkeit fördert. Wir zeigen numerisch, dass der statistische Fehler, der bei der Kalibrierung eines CNOT-Gatters entsteht, um mehr als den Faktor zwei verringert wird, wenn man nur zwei weitere 1-Qubit-Gatter zum GSC hinzufügt.

Erklärung zur Dissertation

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten – noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen

Y. Aragonés-Soria, J. Åberg, C.-Y. Park, and M. J. Kastoryano. "Classical restrictions of generic matrix product states are quasi-locally Gibbsian". In: *Journal of Mathematical Physics* 62.9 (Oct. 2021), p. 093511

Köln, 14. Februar 2022
Ort, Datum

Yaiza Aragonés Soria