UNIVERSITÀ DEGLI STUDI DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN FISICA

Error mitigation and performace trade-offs in scaling variational quantum algorithms

Settore scientifico-disciplinare: **FIS/02 Fisica teorica e modelli matematici**

DOTTORANDO
Giulio Crognaletti

COORDINATORE
PROF. Francesco Longo

SUPERVISORE DI TESI
PROF. Angelo Bassi

CO-SUPERVISORE DI TESI
DR. Michele Grossi

ANNO ACCADEMICO 2024/2025

UNIVERSITY OF TRIESTE

Abstract

Department of Phyisics

Doctor of Philosophy

Error mitigation and performance trade-offs in scaling variational quantum algorithms

by Giulio CROGNALETTI

Quantum computers promise significant computational advantages across domains such as optimization, cryptography, and quantum simulation. However, the potential of current devices is constrained by noise arising from imperfect hardware implementations. By combining the potential for noise resilience of parameterized circuits with the power of classical optimizers, Variational Quantum Algorithms are among the most promising in this context.

In this Thesis, two practical applications of this framework are explored. In the first, we study quantum phase transitions in the one-dimensional J_1–J_2 Heisenberg spin chain. This is achieved by mapping the presence of a phase transition into the crossing of low-lying excitations, and subsequently using level spectroscopy to identify the crossings. More specifically, by leveraging the system's symmetries, we constrain the optimization to selected symmetry sectors, enabling efficient low-energy computations and accurate estimation of the critical point. Moreover, the adaptation of Quantum Error Mitigation techniques, such as Zero-Noise Extrapolation, to this setting allows reliable results even in the presence of realistic noise. The second application introduces a variational approach to correcting gate calibration errors. By exploiting structured states, such as stabilizer states, we isolate errors arising from miscalibrations and compensate for them robustly under additional sources of noise.

Despite its potential, the cost concentration or barren plateau phenomenon remains one of the main limitations of this framework, severely limiting its scalability, especially in the presence of noise. The primary contribution of this Thesis is to provide a general mathematical formulation describing cost concentration under arbitrary noise processes. In particular, it unifies and extends existing models, elucidating how circuit structure influences trainability and noise resilience, an essential step for the development of scalable and reliable variational algorithms. Towards this goal, it identifies a class of architectures termed Quantum Residual Networks, that are able to mitigate cost concentration, thus preserving meaningful gradients at scale.

Contents

Chapter 1

Introduction

Quantum computers hold the promise of delivering substantial speedups across a broad range of computational tasks. By harnessing the intrinsic principles of quantum mechanics, they are expected to address problems that remain intractable for classical architectures, spanning areas such as optimization, cryptography, and the simulation of complex quantum systems. Nevertheless, the realization of their full potential is currently hindered by the pervasive presence of noise, arising from the imperfect implementation of existing quantum hardware.

While it is widely acknowledged that fault-tolerant quantum computation will ultimately depend on Quantum Error Correction (QEC), wherein errors are detected and corrected dynamically during computation, the technological requirements for large-scale QEC remain far beyond the capabilities of current quantum devices. For this reason, Quantum Error Mitigation (QEM) has been developed as an intermediate strategy to alleviate the impact of noise on near-term applications. Typically relying on classical post-processing of noisy measurement outcomes, QEM techniques provide a practical balance between accuracy and resource efficiency. Although approximate in nature, they have become indispensable tools in the present era. Within this context, hybrid quantum-classical schemes such as Variational Quantum Algorithms (VQAs), which combine the expressive power of quantum circuits with the robustness of classical optimization, have emerged as especially promising. Their conceptual simplicity, flexibility across diverse problem classes, and compatibility with QEM approaches make them a compelling framework for exploiting the capabilities of current noisy devices.

The first part of this Thesis presents two applications of this framework in the presence of realistic noise. As a first application, VQAs are employed to investigate quantum phase transitions, a cornerstone of condensed matter theory. This is achieved through *level spectroscopy*, a technique that maps the identification of critical points to the analysis of the low-lying excited-state structure of the system. Focusing on the one-dimensional J_1-J_2 Heisenberg spin chain, we exploit the model's symmetries to *constrain* the optimization to selected symmetry sectors, thereby enabling the efficient computation of low-energy excitations. Furthermore, by augmenting standard optimization routines with symmetry information, we show how standard QEM methods like Zero-Noise Extrapolation can be easily adapted to make them suitable for this kind of application.

Secondly, the VQA framework is used to correct gate calibration errors in imperfect devices. In particular, we employ highly structured quantum states, such as stabilizer states, to *isolate* errors arising from gate miscalibration from other sources of noise. This approach enables efficient compensation that remains robust even in the presence of unwanted interactions between the system and its environment.

Despite their successes on small-scale systems, general VQAs face inherent optimization challenges that limit their scalability. Among the most critical is the *cost concentration* (or *barren plateau*) phenomenon, characterized by an exponential flattening of the optimization landscape as the system size increases, rendering optimization practically unfeasible. The

presence of noise can further exacerbate this issue by inducing a stronger form concentration and amplifying trainability problems. Understanding the mechanisms underlying cost concentration and their interplay with noise is therefore essential for the development of scalable and reliable variational algorithms.

As a major milestone toward this goal, the main contribution of this Thesis is the development of a mathematical framework for modeling cost concentration under arbitrary noise processes. This unified formulation captures the phenomenon in its most general form, encompassing existing models in the literature as special limiting cases, and offering new theoretical insights into the mechanisms that govern trainability in noisy quantum systems. An important outcome of this framework is the establishment of formal connections between circuit structure and noise resilience, which in turn provide practical guidelines for designing trainable quantum models. As an example, the formalism allows for the identification of a novel class of architectures, termed Quantum Residual Networks (QResNets), which exhibit intrinsic robustness against cost concentration, thereby preserving meaningful gradients as the system size increases.

In summary, we aim to advance both the theoretical and practical foundations of variational quantum computing by addressing three interrelated challenges: the characterization of cost concentration, the mitigation of noise effects, and the adaptation of variational frameworks to physically meaningful quantum simulations. Through a combination of analytical modeling, algorithmic design, and numerical experimentation, it provides a unified perspective on the trade-offs between expressivity, trainability, and robustness that define the path toward scalable quantum computation.

The Thesis is organized as follows. In Chapter 2 we briefly introduce noisy quantum computation, QEM and VQAs. In Chapter 3 and Chapter 4 we present two applications of the VQA framework. In particular in Chapter 3 we introduce a variational algorithm to detect phase transitions on a quantum computer and in Chapter 4 we present a variational approach to the correction of calibration errors. Then in Chapter 5 the problem of cost concentration is introduced and finally a general treatment of cost concentration under arbitrary noise is given in Chapter 6.

The research presented in this thesis has resulted in the following publications:

1. **Publication**. G. Crognaletti, G. Di Bartolomeo, M. Vischi and L.L. Viteritti. *Equivariant Variational Quantum Eigensolver to detect Phase Transitions through Energy Level Crossings Quantum Sci. Technol. 10 015048, 2024.*

2. **Preprint I**. G. Di Bartolomeo, G. Crognaletti, A. Bassi and M. Vischi. *Mitigating Coherent Errors through a Decoherence-Resistant Variational Framework employing Stabilizer States. arXiv.2510.20445 [quant-ph].*

3. **Preprint II**. G. Crognaletti, M. Grossi and A. Bassi. *Estimates of loss function concentration in noisy parametrized quantum circuits. arXiv.2410.01893 [quant-ph].*

Chapter 2

Quantum computation in the presence of noise

The term *quantum computing* refers to a computational model based on quantum mechanics, which uses properties of quantum systems to perform calculations. There are many equivalent ways to formalize this idea, from Adiabatic Quantum Computing (AQC) [3] to Measurement Based Quantum Computing (MBQC) [11], but the most widely used is the circuital model of quantum computation. Here, we introduce this model in the ideal setting. Subsequently, an introduction to open quantum systems is provided and applied in Section 2.3 to the modelling of imperfect, or *noisy* quantum computation. Finally, the Variational Quantum Algorithm (VQA) framework is presented.

2.1 Quantum Computing in a nutshell

In the following, a brief overview of the main aspects of the circuital model is given, from the definition of a quantum bit (or *qubit*) to the properties of complex circuits necessary for the subsequent Chapters.

Qubits and registers

Similarly to the classical case, where the fundamental unit of information is the bit, which can take only the binary values of 0 and 1, in quantum computing the fundamental unit of information is the *qubit*, which can be formalized as the quantum state $|\psi\rangle$ of a 2-level system. The state of the qubit can then be the expressed as a linear combination of two orthonormal vectors, often called the *computational basis*:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{2.1}$$

where $\alpha, \beta \in \mathbb{C}$ satisfy $|\alpha|^2 + |\beta|^2 = 1$. It is common practice to label the computational basis set as $\{|0\rangle, |1\rangle\}$ in accordance with classical notation. Moreover, where not specified otherwise, the computational basis will coincide with the eigen-basis of the Pauli-Z operator for the qubit. As classical bits are arranged and manipulated in arrays called *registers*, also qubits are organized in *quantum registers*, made up of collections of qubits. Let $\mathcal{H}$ be the Hilbert space of those qubits, i.e. $|\psi\rangle \in \mathcal{H}$. Then, the quantum state representing the newly formed register consisting of n qubits will be described by a vector $|\Psi\rangle \in \mathcal{H}^{\otimes n}$. One of the key differences with the classical case is that the dimension of $\mathcal{H}^{\otimes n}$ increases exponentially in n, more precisely $\dim(\mathcal{H}^{\otimes n}) = 2^n$, which means that classically representing the state of a quantum register of n qubits *exactly*, quickly becomes classically intractable.

Quantum circuits

In quantum mechanics, quantum states $|\Psi\rangle$ evolve by unitary operators, usually denoted by U, determined by the Schrödinger equation[1]

$$\partial_t |\Psi_t\rangle = -iH_g |\Psi_t\rangle , \tag{2.2}$$

by appropriately choosing the generating Hamiltonian H_g and a time t through the relation

$$U = \exp\big(-itH_g\big). \tag{2.3}$$

For this reason, quantum computers perform calculations by transforming the registers' state using suitable unitary transformations.

Similarly to the classical case, where complex functions of the registers can be decomposed in simple *elementary* gates involving few bits (e.g. NAND, XOR, etc.), also general transformations of the quantum registers U can be realized by combining single qubit and 2-qubit unitary *quantum* gates from an elementary gate set into circuits. Indeed, it can be proven that the infinite elementary set composed of all single qubit gates and the CNOT gate suffice to compose any unitary transformation U acting on $\mathcal{H}^{\otimes n}$, i.e. they form a *universal set* for quantum computation [77]. Instead, for finite elementary quantum gates sets the result holds just as an approximation [77]. The set of fundamental gates is not unique, and different gates can be considered to be elementary depending on the specific hardware under consideration.

Interestingly, quantum gates can also be *parameterized*, i.e. the transformation applied is a function of an external parameter θ. Such gates, generally denoted by U_θ, are defined as in Eq. (2.3) by the relation $\theta \propto t$ and the choice of a suitable generating Hamiltonian, commonly a Pauli string.

Putting everything together, we can describe the action of a general circuit as a sequence of L unitary, possibly parametrized transformations, acting on an initial state $|\Psi_0\rangle$ as

$$|\Psi_{\boldsymbol{\theta}}\rangle = U_{\theta_L} U_{\theta_{L-1}} \cdots U_{\theta_1} |\Psi_0\rangle , \tag{2.4}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_L)$. The order of the operations is determined up to permutation of contiguous and commuting operations. Since, as discussed, many of such operations are local, we can give additional structure to Eq. (2.4) by considering the notion of a *circuit moment*.

Definition 1 (Circuit moment)**.** Given a quantum circuit as in Eq. (2.4), we define a circuit moment as a subset of operations that can be executed simultaneously, i.e. they are contiguous and do not share any qubit. More technically, we can define a moment by the product $\prod_{i=i_s}^{i_e} U_{\theta_i}$, where the indices i_s and i_e satisfy

$$\text{supp } U_{\theta_i} \cap \text{supp } U_{\theta_j} = \varnothing \ \forall i, j \in [i_s, i_e],$$

and supp U denotes the subsystem where U acts non trivially on.

This concept allows us to introduce another important concept for the following chapters, that of *depth* of a circuit.

Definition 2 (Circuit Depth)**.** Given a quantum circuit as in Eq. (2.4), its depth is defined as the minimum number of moments needed to include all U_{θ_i}.

The study of the behaviour of different classes of quantum circuits in terms of the system size n and their depth will be central to across this Thesis.

[1] From here on, we set $\hbar = 1$.

Measurement

Finally, after evolving the initial state using a suitable unitary transformation, the registers (or suitable subsets of qubits) are measured in the computational basis[2] for a certain number of times or *shots*, in order to approximate the expectation value $\langle\Psi_\theta|H\Psi_\theta\rangle$ for a given observable H, which constitutes the result of the calculation. If we let N_s denote the number of shots for a given estimation, and $\{|s_i\rangle\}_{i=1}^{N_s}$ on the computational basis state obtained after each measurement, such expectation value is approximated by

$$\langle\Psi_{\boldsymbol{\theta}}|H|\Psi_{\boldsymbol{\theta}}\rangle \approx \frac{1}{N_s}\sum_{i=1}^{N_s}\langle s_i|H|s_i\rangle, \tag{2.5}$$

which is an *unbiased* estimator with a statistical uncertainty scaling as $1/\sqrt{N_s}$.

2.2 Open quantum systems

While the state-vector description of a quantum state $|\Psi\rangle$ is convenient for ideal, closed quantum systems, when the interaction with surrounding environment is not negligible, it is necessary to use a density matrix formulation, where quantum states are defined by operators $\rho : \mathcal{H} \to \mathcal{H}$ which satisfy the axioms of positive semidefiniteness, i.e. $\rho \geq 0$ and normalization, i.e. $\mathrm{Tr}\{\rho\} = 1$. Such an operator describes a *statistical mixture* of pure states, which can be obtained using the spectral decomposition of ρ.

In this formalism, quantum evolution is described by linear *super-operators*, that map density operators into density operators. As an example, the unitary evolution defined by Eq. (2.3) can be equivalently described in this formalism by the super-operator $\mathcal{U}$ acting on the density matrix ρ as $\mathcal{U}(\rho) = U\rho U^\dagger$. We can get the form of $\mathcal{U}$ in terms of the generating Hamiltonian H_g and t using the Liouville-Von Neumann equation, which extends Schrödinger equation to this setting, namely

$$\partial_t\rho = -i[H_g, \rho], \tag{2.6}$$

which allows to express $\mathcal{U}$ as the exponential

$$\mathcal{U} = e^{-it\mathcal{H}}, \text{ where } \mathcal{H}(\rho) = [H, \rho], \tag{2.7}$$

which more closely resembles Eq. (2.3). However, when a system interacts with its surrounding environment and the environmental degrees of freedom are ignored, the system's evolution becomes non-unitary. In such a scenario the evolution is better described by a more general equation for the system's density matrix called Gorini-Kossakowski-Sudarshan-Lindblad (GKSL) master equation [38, 59]. A direct consequence of such a non-unitary evolution is decoherence, namely the decay over time of the off-diagonal entries of the system's density matrix in a certain basis. In other words, the system transitions from a pure state to a mixed state.

2.2.1 The GKSL master equation

Consider a quantm system of interest S coupled to an environment E. The total Hilbert space is given by the tensor product $\mathcal{H}_{tot} = \mathcal{H}_s \otimes \mathcal{H}_e$, and the total Hamiltonian can be written

[2] In the context of quantum computing, projective measurement in the computational basis are in fact equivalent to any other type of measurement. Even POVM measurement of the register can be carried out by adding some ancillary qubits. For further details, see Ref. [77]

as

$$H_{tot} = H_s + H_e + \lambda H_{int}, \tag{2.8}$$

where H_s and H_e denote the Hamiltonians of the system and the environment, respectively, and H_{int} represents the interaction Hamiltonian. The dimensionless coupling constant λ controls the strength of the interaction. The joint evolution of system and environment is governed by the unitary super-operator $\mathcal{U}_{tot}(t)$, acting on the composite state $\rho_{tot} = \rho_s \otimes \rho_e$. After time t, the total state evolves as

$$\rho_{tot}(t) = \mathcal{U}_{tot}(t)\rho_{tot},$$

and the system S and the envirnoment E become entangled. Since in practice we are rarely able (or interested) to track all environmental degrees of freedom, we generally focus on the reduced state of the system, obtained by tracing out the environment, namely

$$\rho_s \mapsto \mathrm{Tr}_e \left\{ \mathcal{U}_{tot}(\rho_s \otimes \rho_e) \right\}. \tag{2.9}$$

The reduced dynamics resulting from this partial trace is, in general, extremely complicated and does not preserve the simple unitary form of evolution. The loss of information into the environment manifests itself as effective non-unitarity in the dynamics of ρ_s.

Under suitable physical approximations, most notably the weak-coupling $\lambda \ll 1$ and Markovian assumptions, the dynamics of the system can nonetheless be described in a compact and universal form. In this regime, one assumes that the environment acts as a memoryless reservoir, such that the future evolution of the system depends only on its present state and not on its history. Mathematically, this corresponds to a one-parameter semigroup of completely positive, trace-preserving (CPTP) maps $\{\Gamma_t\}_{t\geq 0}$ satisfying the semigroup composition law

$$\Gamma_{s+t} = \Gamma_s \circ \Gamma_t = \Gamma_t \circ \Gamma_s, \quad t, s \geq 0. \tag{2.10}$$

The generator of such a semigroup defines the so-called *Lindbladian* or *GKSL* master equation, named after the authors Gorini, Kossakowski, Sudarshan and Lindblad who independently derived its most general form [59, 38]. In particular, it takes the form

$$\partial_t \rho = -i[H_s, \rho] + \sum_k \gamma_k \left(L_k \rho L_k^\dagger - \frac{1}{2} \left\{ L_k^\dagger L_k, \rho \right\} \right) = -i\mathcal{H}_s + \mathcal{D}, \tag{2.11}$$

where the first term describes the coherent evolution generated by H_s, while the second term, often referred to as the *dissipator* $\mathcal{D}$, encodes the irreversible processes induced by the coupling to the environment. The GKSL equation generalizes Eq. (2.6), which is recovered in the limit $\gamma_k \to 0$ for all k.

Importantly, the structure of the Lindbladian is not arbitrary: it is precisely the one required to ensure that the dynamical maps Γ_t are *completely positive*, meaning that $\Gamma_t \otimes \mathbb{1}_d$ acts positively on all extended systems of arbitrary dimension d. This property guarantees physical consistency even when the system is part of a larger entangled state.

A direct calculation also confirms that the trace of the density matrix is preserved under this evolution, namely

$$\mathrm{Tr}\{\partial_t \rho\} = -i\,\mathrm{Tr}\{[H_s, \rho]\} + \sum_k \gamma_k \,\mathrm{Tr}\left\{L_k \rho L_k^\dagger\right\} - \frac{\gamma_k}{2}\,\mathrm{Tr}\left\{\left\{L_k^\dagger L_k, \rho\right\}\right\} = 0. \tag{2.12}$$

This ensures that the evolution remains trace-preserving and that the density matrix can be consistently interpreted as a statistical mixture of quantum states at all times. The maps Γ_t

generated in this way therefore define quantum operations that are both completely positive and trace-preserving, a class of maps collectively referred to as *quantum channels*. These channels constitute the fundamental mathematical objects describing the allowed transformations of quantum states under noise and decoherence. In subsequent Sections, we shall analyze their structural properties, including Kraus representations and contractivity, which play a crucial role in later Chapters.

2.2.2 Useful properties of quantum channels

In this section, we will briefly introduce the Kraus decomposition of CPTP maps, as well as the concept of *duality*, which allows to define the Heisenberg picture evolution for open systems, encoded in completely positve and *unital* (CPU) maps.

Kraus decomposition and unitality

Any quantum channel Φ acting on the space of density operators can be represented in the so-called *Kraus decomposition* form [54], which provides an explicit operator-sum realization of the channel. In this representation, the action of Φ is expressed as

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{2.13}$$

where the operators $\{K_i\}_i$, known as Kraus operators, satisfy the *completeness* relation

$$\sum_i K_i^\dagger K_i = \mathbb{1}$$

to ensure trace preservation. This decomposition highlights that any physical quantum operation can be interpreted as a probabilistic mixture of elementary transformations $\rho \mapsto K_i \rho K_i^\dagger$. This form provides an intuitive link between the abstract notion of completely positive maps and concrete physical processes such as measurements, decoherence, or dissipation. A central feature distinguishing different classes of quantum dynamical maps is the notion of *unitality*.

Definition 2.2.1 (Unitality)**.** A completely positive (CP) map Φ is said to be *unital* if it preserves the identity operator, i.e.

$$\Phi(\mathbb{1}) = \mathbb{1}. \tag{2.14}$$

This property plays a crucial role in characterizing the physical processes described by Φ. Unital maps correspond to dynamics that do not create or destroy population in the maximally mixed state: if the system is initially in the state $\rho = \mathbb{1}/d$, it remains invariant under a unital evolution. Notably, unitary channels describing closed-system quantum mechanics are also unital. Examples of non-unitary, unital maps used in noisy quantum computation will be provided in Section 2.3.1.

The property of *unitality* of a channel can be directly expressed in terms of its Kraus operators. In particular, a unital channel must satisfy the complementary relation

$$\sum_i K_i K_i^\dagger = \mathbb{1},$$

imposing that the maximally mixed state remains invariant under its action. Whenever this condition is violated, the channel is said to be *non-unital*. Physically, such non-unital dynamics typically corresponds to dissipative processes that relax the system towards a non-maximally mixed stationary state. For instance, in the amplitude-damping channel,

the steady state is the pure ground state, indicating that the environment acts as a sink for excitations.

Duality and the Hilbert–Schmidt scalar product.

Another important conceptual and mathematical structure in the study of quantum channels is the notion of *duality* under the Hilbert–Schmidt scalar product. Formally, consider the space $\mathcal{B}$ of bounded operators on a finite-dimensional Hilbert space $\mathcal{H}$. Then $\mathcal{B}$ can itself be endowed with a scalar product, namely the Hilbert–Schmidt scalar product

$$\langle A, B\rangle_{HS} = \mathrm{Tr}\{A^\dagger B\}. \tag{2.15}$$

With respect to this scalar product, any linear super-operator Λ acting on $\mathcal{B}$ admits a unique dual (or adjoint) map $\Lambda^\dagger$ defined implicitly by the relation

$$\mathrm{Tr}\{A\,\Lambda(B)\} = \mathrm{Tr}\{\Lambda^\dagger(A)\,B\}, \quad \forall A, B \in \mathcal{B}. \tag{2.16}$$

This duality allows one to characterize dynamical maps and their generators in terms of their adjoint actions on observables, providing a direct connection between the Schrödinger and Heisenberg pictures of quantum mechanics. Indeed, if a quantum channel Φ acts on density operators in the Schrödinger picture, its dual $\Phi^\dagger$ acts on observables, describing their time evolution in the Heisenberg picture.

Observation 1. A completetly positive (CP) map Φ is trace preserving if and only if its dual map $\Phi^\dagger$ is CP and unital, i.e.

$$\mathrm{Tr}\{\Phi(\rho)\} = \mathrm{Tr}\{\rho\}\ \forall\rho \quad \Leftrightarrow \quad \Phi^\dagger(\mathbb{1}) = \mathbb{1}. \tag{2.17}$$

Hence, the dual of a completely positive trace-preserving (CPTP) channel is always completely positive and unital (CPU), and vice versa.

2.2.3 Contractivity of CP maps

In this Section, we briefly introduce some relevant properties of completely positive (CP) maps. These will be especially useful in the trace preserving case (CPTP), i.e. quantum channels, and the unital case (CPU), i.e. the corresponding adjoint action with respect to the Hilbert-Schmidt scalar product. In particular, the following results apply to density matrices in the computational basis, namely $(\rho)_{ij} = \langle i|\,\rho\,|j\rangle$. In general, we denote by $\mathbb{M}_n$ the space of $n \times n$ matrices, which we can endow with a norm as follows.

Definition 2.2.2. (Schatten norm) Given $A \in \mathbb{M}_n$ and $p \in [1, \infty]$, we define the Schatten p-norm $\|\cdot\|_p : \mathbb{M}_n \to \mathbb{R}$ as

$$\|A\|_p = \mathrm{Tr}\left\{\left(\sqrt{A^\dagger A}\right)^p\right\}^{1/p}. \tag{2.18}$$

A crucial property of Schatten norms is Hölder inequality.

Theorem 2.2.1 (Hölder inequality)**.** Let $A, B \in \mathbb{M}_n$, and let $p, q \in [1, \infty]$ such that $1/p + 1/q = 1$. Then the Hilbert-Schmidt scalar product modulus is upper bounded by

$$|\mathrm{Tr}\{A^\dagger B\}| \leq \|A\|_p \|B\|_q\,.$$

As a special case, for $p = 2$ we get back the Hilbert-Schmidt norm, and Hölder inequality reduces to the familiar Cauchy-Schwarz inequality.

As a direct consequence of Definition 2.2.2, an induced norm on linear operators acting on $\mathbb{M}_n$ can be defined.

Definition 2.2.3. (Induced norm) Given a linear operator $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ we define the induced $p \to q$ norm as

$$\|\Lambda\|_{p\to q} := \sup_{A:\|A\|_p=1} \|\Lambda(A)\|_q. \tag{2.19}$$

Depending on the properties of such induced norms, we might refer to the map Λ as contractive or strictly contractive. In particular, we will use the following definitions.

Definition 2.2.4. (Contractivity of linear maps) A linear operator $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ is said to be *contractive* with respect to the p norm if $\|\Lambda\|_{p\to p} \leq 1$, and similarly to be *strictly contractive* if $\|\Lambda\|_{p\to p} < 1$.

Linear maps that are also CPTP are known to always be contractive with respect to the 1-norm [87, 45], but are in general not contractive for other p-norms.

Specific classes of quantum channels can be shown to be contractive with respect to a wider variety of norms. In particular, we have that, for $p \geq 2$, *unitality* of the map is a necessary and sufficient condition for contractivity. (see Theorem II.4 in [86]). Within unital channels, unitary transformation $\mathcal{U}$ always saturate the bound, as they have the additional property of being norm-preserving, i.e. $\|\mathcal{U}(A)\|_p = \|U^\dagger A U\|_p = \|A\|_p$. Finally, if we reduce the action of the channel to the subset $\mathbb{H}_0 \subset \mathbb{M}_n$ of Hermitian, traceless matrices, then the unitality property is no longer a necessary condition for contractivity. Indeed, any single qubit channel $\mathcal{N}$ is contractive in this setting, i.e. $\|\mathcal{N}|_{\mathbb{H}_0}\|_{p\to p} \leq 1 \forall p$ if $n = 2$. For single qubit channels we can even be more explicit, as showed in the following Lemma.

Lemma 2.2.1 (Single qubit channel normal form)**.** Let $\{P_i\}$ be the normalized (with respect to the Schatten 2-norm) Pauli basis of $\mathbb{M}_2$, and $\mathcal{N}$ be a single qubit channel. Then there exist unitary matrices U, V such that $\mathcal{N}'(\cdot) = U^\dagger \mathcal{N}(V^\dagger \cdot V) U$ satisfies

$$\mathcal{N}'(P_0) = P_0 + \sum_{i>0} t_i P_i, \quad \mathcal{N}'(P_i) = \lambda_i P_i \; \forall i > 0 \tag{2.20}$$

where $\sum_{i>0}(t_i + \lambda_i \alpha_i)^2 \leq 1$, $\forall \alpha_i \in \mathbb{R}$ s.t. $\sum_{i>0} \alpha_i^2 \leq 1$.

Proof. It has been shown in Refs. [12, 66] that any single qubit quantum channel can be cast in the canonical form of Eq. (2.20) by means of a change of basis. Furthermore, the constraint on the parameters follow, analogously to Ref. [5], by considering that any single qubit state must have bounded purity, namely $\mathrm{Tr}\{\rho^2\} \leq 1$, and since $\mathcal{N}'$ is a channel, the same must hold for $\mathcal{N}'(\rho)$. Since any qubit state can be decomposed in terms of $\{P_i\}_i$ as $\rho = 1/\sqrt{2} P_0 + 1/\sqrt{2} \sum_{i>0} \alpha_i P_i$, $\alpha_i \in \mathbb{R}$, we get

$$\mathrm{Tr}\{\rho^2\} = \frac{1 + \sum_{i>0} \alpha_i^2}{2} \leq 1, \quad \mathrm{Tr}\{\mathcal{N}'(\rho)^2\} = \frac{1 + \sum_{i>0}(t_i + \lambda_i \alpha_i)^2}{2} \leq 1 \tag{2.21}$$

respectively, which concludes the proof. □

Since in the rest of the Thesis, especially Chapter 5 and Chapter 6, only the reduction $\Lambda|_{\mathbb{H}_0}$ of linear maps Λ to the traceless hyperplane $\mathbb{H}_0 \subset \mathbb{M}$ will be relevant for the analysis, we will use, with a slight abuse of terminology, the term *contractive* and *strictly contractive* to refer to $\Lambda|_{\mathbb{H}_0}$ without explicit mention of such restriction.

Finally, considering instead CPU maps, the most relevant result is Kadison-Schwarz inequality, which for our purposes, can be stated as follows.

Theorem 2.2.2 (Kadison-Schwarz inequality [91])**.** Let $A, B \in \mathbb{H}_0$, and $\Lambda : \mathbb{M}_n \to \mathbb{M}_n$ be a CPU map. Then

$$\Lambda(A)\Lambda(B) \leq \Lambda(AB). \tag{2.22}$$

2.3 Quantum Errors and Their Mitigation

In what follows, we apply the above tools to describe quantum computation in the presence of noise. In particular, we outline the principal distinctions between coherent and incoherent noise, introduce representative noise models in explicit Kraus form, and discuss strategies designed to mitigate their effects.

2.3.1 Error modelling in Quantum Computing

During the execution of an algorithm there are two main sources of noise, namely, *coherent* and *incoherent* errors. Coherent errors can be thought of reflecting systematic errors in the instrumentation used to drive the quantum system, and thus give constant contribution, generally modeled by a unitary transformation. On the other hand, incoherent errors are due to interaction between the quantum processor and the enviroment, and therefore are naturally described by quantum channels.

Coherent noise

Coherent noise arises from miscalibrations of control parameters, and can be unitarily modelled as errors that consistently rotate the quantum state in an unintended but deterministic manner. For example, consider a single-qubit rotation gate

$$U_\theta = e^{-i\theta H_g/2}, \tag{2.23}$$

where H_g denotes the generating Hamiltonian of U. If the control electronics introduce a small fractional over-rotation ϵ, the implemented operation becomes

$$\tilde{U}_\theta = e^{-i(\theta+\epsilon)H_g/2} = U_\theta\, e^{-i\epsilon H_g/2}. \tag{2.24}$$

Such coherent deviations can interfere constructively or destructively across circuit layers, potentially amplifying their impact. Such amplification can nonetheless be exploited to detect and compensate them. We show a possible technique in Chapter 4.

Incoherent noise

In contrast, incoherent noise corresponds to random, non-unitary processes that lead to irreversible loss of quantum coherence. These are typically modeled using *factorized* quantum channels, namely as

$$\mathcal{E} = \bigotimes_{i=1}^{n} \mathcal{E}_i\,,$$

$\mathcal{E}_i$ denote single qubit channels. Some canonical examples include:

- **Depolarizing channel:**

$$\mathcal{E}_{\text{dep}}(\rho) = (1-p)\rho + \frac{p}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right), \tag{2.25}$$

  with Kraus operators $E_0 = \sqrt{1-p}\,\mathbb{1}$, $E_1 = \sqrt{\frac{p}{3}}\,X$, $E_2 = \sqrt{\frac{p}{3}}\,Y$, $E_3 = \sqrt{\frac{p}{3}}\,Z$. This model captures isotropic loss of information, effectively replacing the qubit state with the maximally mixed state with probability p. This is an example of a unital and strictly contractive channel.

- **Dephasing channel:**
$$\mathcal{E}_\phi(\rho) = (1-p)\rho + p\, Z\rho Z, \tag{2.26}$$
corresponding to Kraus operators $E_0 = \sqrt{1-p}\,\mathbb{1}$, $E_1 = \sqrt{p}\,Z$. This channel randomizes the relative phase between computational basis states, destroying off-diagonal elements of ρ in the Z basis. This is an example of a unital, yet non strictly contractive channel.

- **Amplitude damping channel:**
$$\mathcal{E}_{\text{amp}}(\rho) = E_0\rho E_0^\dagger + E_1\rho E_1^\dagger, \tag{2.27}$$
with
$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad E_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix},$$
describing relaxation from $|1\rangle$ to $|0\rangle$ with probability γ. This is an example of a non-unital channel.

These elementary noise maps provide convenient abstractions for theoretical and numerical studies, though real devices often exhibit correlated noise, which cannot be captured by this simplified model. To address this problem, in Chapter 6 noise will be modelled by a generic quantum channel.

Additionally, state preparation and measurement (SPAM) errors are also present, which happen at the very beginning and the very end of a quantum algorithm, and are generally modelled by suitable choice of a faulty intitial state $\tilde{\rho}$, and confusion matrix-like approaches respectively.

2.3.2 Error Mitigation Strategies

In the past years, Quantum Error Correction (QEC) has emerged as the ultimate tool against noise and decoherence. While QEC is a wide and diversified field, in general terms, it leverages smart encoding of quantum information into highly entangled states of multiple physical qubits. By actively detecting and correcting errors during computation without directly measuring the encoded information, QEC enables the preservation of quantum coherence over extended periods of time. Thus, the development of efficient error-correcting codes, such as the surface code and stabilizer codes, has been pivotal in moving quantum technologies toward fault-tolerant operation, making QEC a cornerstone of scalable quantum information processing [104, 53].

However, it is worth emphasizing that due to their significant hardware requirements, QEC remains infeasible on near-term quantum hardware [27]. In contrast, Quantum Error Mitigation (QEM) offers a practical alternative. In this framework, the impact of noise is reduced *without* actively intervening on the hardware, but by classically post-processing the results of several circuit runs to obtain better estimates of the measured quantities, making it far more suitable for current noisy quantum devices.

As an example, Zero-Noise Extrapolation (ZNE) [106] is among the most widely adopted mitigation strategies for near term devices. The essential idea is to estimate the noiseless observable value $\langle H\rangle^\star$ by evaluating the same quantum circuit at multiple effective noise strengths λ, and subsequently extrapolating to the zero-noise limit. This idea is pictorially represented in Fig. 2.1

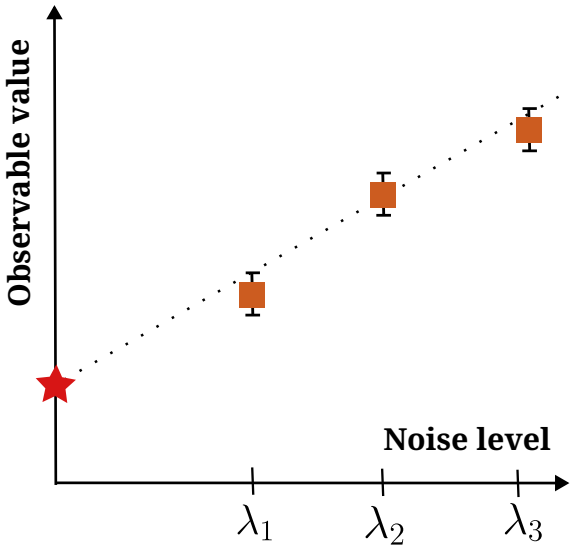

FIGURE 2.1: Pictorial representation of Zero Noise Extrapolation. Noisy estimates at different noise strengths represented by squares, are used to extrapolate to the zero-noise limit represented by the star.

The notion of noise scaling, can be realized in many ways. The most intuitive example relies on *gate-time stretching*. If we let t_g be the time normally required to implement a physical a gate U, we can describe this procedure by artificially prolonging this time to $\tau > t_g$. While this still produces the same unitary gate U, the increase of the gate-time prolongs the run time of the circuit, and thus the interaction time with the environment, increasing the effective noise strength. In this setting, the scaling factor λ can be taken as the ratio τ/t_g. Indeed, such a manipulation can be realized experimentally through various techniques [106, 58, 36].

Now, if $\langle H\rangle_\lambda$ denotes the measured expectation value at noise scaling factor λ, one can assume that

$$\langle H\rangle_\lambda - \langle H\rangle^* \approx a_1\lambda + a_2\lambda^2 + O(\lambda^3). \tag{2.28}$$

where only the first term meaningfully contribute. Then, performing the circuit measurements at several scaled noise levels λ_i and fitting the resulting values to a polynomial model, an extrapolated estimate $\langle H\rangle_{\lambda=0}$ of $\langle H\rangle^*$ can be obtained.

Common variants include *linear* or *Richardson extrapolation*, where the coefficients are chosen to cancel leading-order noise terms. ZNE requires the ability to reliably control the noise scaling and assumes approximate smoothness of $\langle H\rangle_\lambda$, conditions that can typically be satisfied on well-calibrated devices.

Another prominent method is Probabilistic Error Cancellation (PEC) [106], which seeks to reconstruct an unbiased estimator of the ideal quantum circuit output by probabilistically sampling from an ensemble of noisy circuits. Indeed, this techniques hinges on the decomposition of noise channel $\mathcal{E}$ as a linear combination of implementable operations, in order to stochastically reconstruct the action of its (nonphysical) inverse. In particular, under these assumptions, we have $\mathcal{E}^{-1} = \sum_i \eta_i \mathcal{O}_i$, with real coefficients η_i and noisy operations $\mathcal{O}_i$. Exploiting this structure, expectation values of an ideal observable can be recovered as a weighted average over noisy runs. However, the price for this exact correction is an exponential growth in sampling overhead, which makes it far more expensive than easier procedures such as ZNE.

Additional mitigation strategies include *symmetry verification*, where measurement outcomes violating known conserved quantities are discarded; and *measurement error mitigation*, where classical readout noise is corrected by applying a pre-characterized calibration matrix. In practice, hybrid combinations of multiple techniques are often employed to achieve a balance between computational overhead and accuracy.

2.4 Variational quantum algorithms

The main idea behind Variational Quantum Algorithms (VQAs) is to use a parameterized quantum channel $\Phi_{\boldsymbol{\theta}}$, often called the *ansatz*, which depends on some parameter vector $\boldsymbol{\theta} \in \Theta$, and then find the configuration $\boldsymbol{\theta}^*$ that minimize some cost function $C(\boldsymbol{\theta})$. Cost functions are in general defined as the expectation value of some observable, i.e.

$$C(\boldsymbol{\theta}) = \mathrm{Tr}\{\Phi_{\boldsymbol{\theta}}(\rho_0)H\} \tag{2.29}$$

where ρ_0 is the initial state, $\rho_{\boldsymbol{\theta}} = \Phi_{\boldsymbol{\theta}}(\rho_0)$ is the output state and H is an observable called the *cost* operator, which characterizes the problem to be solved. In recent years, a vast variety of anstäze have been introduced in the literature to solve a wide variety of problems, from chemistry, to optimization. A common feature that links many of them is the *layered* structure, where the total channel $\Phi^L_{\boldsymbol{\theta}}$ is built by repeating the same, simpler channel. While some examples of non-unitary anstäze have been discussed in the literature [48] and will be touched upon in Chapter 6, most applications leverage unitary circuits to parameterize the ansatz. Indeed it is useful to formalize this idea in this setting.

Definition 2.4.1 (Layered ansatz)**.** An ansatz $\mathcal{U}^L_{\boldsymbol{\theta}}$ is said to be *layered* with L layers if it can be decomposed as

$$\mathcal{U}^L_{\boldsymbol{\theta}} = \mathcal{U}_{\theta_L}\mathcal{U}_{\theta_{L-1}}\cdots\mathcal{U}_{\theta_1} \tag{2.30}$$

where each *layer* $\mathcal{U}_{\theta_l}$ share a common structure, i.e. they only differ by the value of θ_l, and the parameter vector $\boldsymbol{\theta}$ of $\mathcal{U}^L$ is given by $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_L)$.

Here, non-unitary channels are generally only included to model the presence of noise on near term devices. This is achieved using non-parameterized channels $\mathcal{E}_l$.

Definition 2.4.2 (Noisy layered ansatz)**.** An ansatz $\Phi^L_{\boldsymbol{\theta}}$ is said to be *noisy* and *layered* with L layers if it can be decomposed as

$$\Phi^L_{\boldsymbol{\theta}} = \mathcal{E}_L\mathcal{U}_{\theta_L}\mathcal{E}_{L-1}\mathcal{U}_{\theta_{L-1}}\cdots\mathcal{E}_1\mathcal{U}_{\theta_1} \tag{2.31}$$

where $\mathcal{U}_{\theta_l}$ are unitary transformations. Furthermore, if the noise channel does not depend on the layer, i.e. $\mathcal{E}_l = \mathcal{E}\ \forall l$, then $\Phi^L_{\boldsymbol{\theta}}$ is said to be *uniform*.

Even though computing $C(\boldsymbol{\theta})$ requires the evaluation of the quantum circuit, the optimization can still be performed by classical means using optimizers, that use it as a black box. Indeed, the Variational Quantum Algorithm defining feature is a *hybrid* quantum-classical loop aiming at iteratively minimizing $C(\boldsymbol{\theta})$. To achieve this goal, two important classes of optimizers can be identified, namely gradient free and gradient-based algorithms. Gradient-free optimizers only require calculation of the cost function, and as such can be used out of the box in VQAs. Gradient based algorithms can also be used (e.g. Algorithm 1), provided a method to compute gradients of the quantum circuit. In fact most parameterized quantum circuits are *differentiable*, and their gradients can be approximated on a quantum computer by several methods (e.g. parameter shift rule [123], finite difference approach). This procedure is pictorially represented in Fig. 2.2.

Algorithm 1 VQA hybrid quantum-classical loop (using gradient descent)

$\boldsymbol{\theta}^{(0)} \leftarrow \boldsymbol{\theta}_0$
$k \leftarrow 0$
while $k < K$ **do**
  $\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} - \eta \nabla C|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}$
  $k \leftarrow k + 1$
end while

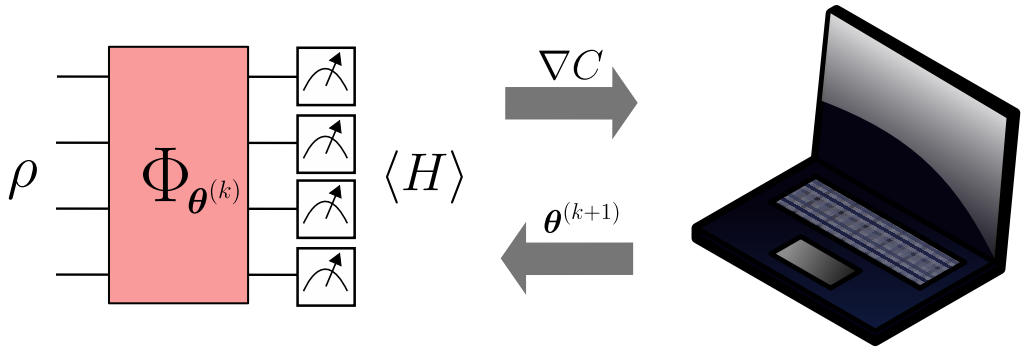

FIGURE 2.2: Structure of the VQA framework, using a gradient based optimizer. The initial state ρ is evolved using Φ in order to compute the mean value of the observable H, which constitutes the cost function $C(\boldsymbol{\theta})$. Information about the gradient ∇C is passed to the classical optimizer, which updates the parameters.

Depending on the nature of the problem at hand, one might be interested either in the optimal value of the parameters $\boldsymbol{\theta}^*$ or the optimal value of the cost function $C(\boldsymbol{\theta}^*)$. To give an example, the problem of finding the ground state of a quantum system is fully encoded in the energy function $C(\boldsymbol{\theta}) = \mathrm{Tr}\{\rho_{\boldsymbol{\theta}} H\}$ of the system, which is of type Eq. 2.29. In this case, the Hamiltonian is the cost operator of this problem, and we are interested in approximating the ground state energy $E_0 \approx C(\boldsymbol{\theta}^*)$. For such applications, this framework is generally termed Variational Quantum Eigensolver (VQE). This is the case of the analysis of Chapter 3. Instead, in Chapter 4 we present an example in which the VQA framework is used to correct coherent errors by leveraging knowledge of the optimal parameters $\boldsymbol{\theta}^*$, without requiring an accurate estimation of $C(\boldsymbol{\theta}^*)$.

Even if many optimization techniques can be used, finding a satisfactory solution is by no means guaranteed with any of them. A first issue to consider is that of the *existence* of such solution to the optimization problem. Indeed, given a cost operator, there exists a subset $\mathbb{U}_s$ of the unitary group $\mathbb{U}(2^n)$ of unitaries that is said to *solve* it, i.e.

$$\mathrm{Tr}\{\mathcal{U}(\rho)H\} \leq \mathrm{Tr}\{\mathcal{V}(\rho)H\} \quad \mathcal{U} \in \mathbb{U}_s, \mathcal{V} \in \mathbb{U}(2^n).$$

However, generic ansätze $\mathcal{U}_{\boldsymbol{\theta}}$ used to seek such solution will generally span only a subset $\mathbb{U}_\Theta \subset \mathbb{U}(2^n)$, i.e. $\mathbb{U}_\Theta = \{\mathcal{U}_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}$, where Θ is the domain of the parameterization. Then, the optimal solution is attainable if and only if $\mathbb{U}_\Theta \cap \mathbb{U}_s \neq \varnothing$. In this case, the ansatz is said to be *complete* with respect to the problem.

In order to maximize the probability to have a complete model, one might think to maximize the exploratory power of the ansatz, i.e. its *expressivity*. In other words, the ansatz should span almost uniformly almost all of the unitary group. While on the one hand, this

in principle improves the quality of the optimal solution attainable, on the other hand it generally leads to a harder optimization step. Indeed, increasing expressivity of a model is linked to the *vanishing gradients* problem, connected to the cost concentration or *barren plateau* phenomenon, introduced in Chapter 5 and studied in depth in Chapter 6.

Despite these difficulties, VQAs are often a very fruitful option, especially in the noisy setting. In fact, thanks to the ability of adapting the value of the parameters, variational circuits show greater noise resilience compared to other type of algorithms [28], which is a property particularly desirable in the near term.

Chapter 3

Variational estimation of phase transitions

The conceptual aim of this Chapter is to demonstrate that condensed matter techniques, such as level spectroscopy, can be combined with variational quantum algorithms to detect quantum phase transitions. In particular, we show that, using the level spectroscopy technique [26, 98, 78] and symmetry-preserving ansätze, knowledge of specific excited states can be obtained and leveraged to detect phase transitions that are challenging to characterize by directly measuring order parameters due to significant finite-size effects. Furthermore, the sensitivity of our approach to noise is evaluated through numerical simulations of noisy quantum circuits. Despite the explicit breaking of variational state symmetries in the presence of noise [108], we demonstrate that by integrating standard error mitigation techniques, such as Zero Noise Extrapolation (ZNE) [106], we can successfully restore the desired physical properties.

3.1 From phase transition to ground state estimation

Exploring phase transitions is crucial for understanding the fundamental behavior of matter [94]. These transitions are usually identified by analyzing order parameters that describe specific phase changes. However, accurate estimates of transition points require large systems approaching the thermodynamic limit, which typically is highly computationally expensive. An alternative approach involves utilizing level spectroscopy [26, 98, 78]. In this framework, the ground state phase transition is detected by the energy level crossings of low-lying excited states. Indeed, if the quantum numbers of the ground state do not change across the phase transition, but those of the low-energy excited states do change, then a correspondence between excited energy level crossings and the phase transitions is expected [75, 76, 97, 96]. The benefit of this method lies in its insensitivity to system size, enabling highly accurate predictions of the phase transition point even with small clusters. For several one-dimensional models, this mapping is well established, and higly accurate estimations of phase transition points can be gained by analyzing the behavior of the first few energy levels. Moreover, it seems to be a promising approach also for two-dimensional spin systems [118, 31, 80].

While the VQA approach is designed to only obtain ground state estimations, it can be extended to describe low-lying excited states as well. For instance, this can be achieved by constructing suitable variational states where specific symmetry sectors can be specified [52, 72, 114, 80, 79]. Indeed, Hamiltonians of interest often have certain symmetries (e.g. conservation of total spin, momentum), which can be used to decompose the system Hilbert space into symmetry sectors labeled by these conserved quantities. The lowest-energy state in each sector corresponds to the minimal excitation compatible with that symmetry constraint, i.e. the lowest-energy configuration that cannot relax to the true ground

state without breaking the conserved quantity. Hence, these sector ground states represent a low-energy excitations of the full model.

In the next Sections, we show how, by studing the symmetries of the J_1-J_2 Heisenberg model on a chain, we are able to efficiently implement an instance of the VQA framework adequate for the implementation on near-term quantum devices to estimate low-lying excited states, and use that information to detect its phase transition exploiting level spectroscopy.

3.2 The J_1-J_2 Heisenberg model

In the following Section, we focus on the J_1-J_2 Heisenberg model on a chain, where a mapping from phase transitions and level crossing is well established [31]. The Hamiltonian of the model reads

$$H_{J_1\text{-}J_2} = J_1 \sum_{r=1}^{n} \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+1} + J_2 \sum_{r=1}^{n} \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+2} \,, \tag{3.1}$$

where $\boldsymbol{S}_r = (S_r^x, S_r^y, S_r^z)$ is the spin 1/2 operator at site r, composed of the usual Pauli matrices here denoted by S_r^α, $\alpha \in \{x, y, z\}$ and $J_1, J_2 \geq 0$ are the antiferromagnetic couplings between nearest and next-nearest neighbors sites, respectively. For $J_2 = 0$, Eq. (3.1) reduces to the one-dimensional Heisenberg model and the ground-state properties can be computed using the Bethe ansatz [34]. However, when $J_2 > 0$ there are no exact solutions. Still, its phase diagram is well established by numerical and analytical calculations [121, 55]. In particular, a Berezinskii–Kosterlitz–Thouless (BKT) transition [49] at $(J_2/J_1)_c = 0.24116(7)$ separates a gapless region at small values of J_2/J_1 by a gapped state at large values of the frustration ratio. The detection of a BKT transition is especially complicated by the direct computation of an order parameter. Indeed, large size effects are present and a huge number of sites is necessary to give a meaningful estimation of the phase transition point [121, 114, 55, 115]. Contrarily, by employing level spectroscopy, a small number of sites ($n \sim 30$) is sufficient to achieve a very accurate estimation [26].

The eigenstates of the Hamiltonian in Eq. (3.1) can be classified according to the total spin S and the momentum k, respectively the quantum numbers of $\boldsymbol{S}^2$, where $\boldsymbol{S} = \sum_{r=1}^{n} \boldsymbol{S}_r$, and of the one-site translation operator T (assuming periodic boundary conditions). The transition point can be detected as the energy level crossing between the first singlet ($S = 0$) and triplet ($S = 1$) excited states with momentum π with respect to the ground state.

It is worth noting that defining variational states which are eigenstates of the total spin operator $\boldsymbol{S}^2$ requires considerable computational effort when using classical methods like Tensor Networks and Neural-Network Quantum States (NQS). Indeed, for NQS the wave function amplitude is typically represented in the eigenbasis of Pauli Z operators, which is not $SU(2)$-symmetric and therefore inherently breaks total spin symmetry. To address this, it is necessary to define the amplitude in a spin-symmetric basis, as proposed in Refs. [112, 111]. Concerning Tensor Networks, they can be constructed with $SU(2)$-symmetric tensors, allowing for a larger bond dimensions than what is possible when only $U(1)$ symmetry is imposed [105, 100, 127, 37]. However, due to the complexity of incorporating non-abelian symmetries, both Tensor Network and NQS approaches typically restrict computations to the S^z symmetry sectors, where highly accurate results can still be achieved [13, 80, 79, 92, 93, 115, 116, 43].

Conversely, as we show in Section 3.3.1, preparing translational- and spin-equivariant quantum circuits can be achieved implementing only constant depth layers, i.e. only contanining a fixed number of contiguos operations as defined in Definition 2, regardless of the system size n. Notably, combining the latter approach with the level spectroscopy technique on small clusters, makes the problem suitable for quantum computers, even in the

near-term. In later Sections, we explore this possibility through numerical simulations of ideal quantum circuits, focusing on the variational approximation of the excited states of the J_1-J_2 Heisenberg model. However, our approach can be suitably extended to handle other Hamiltonians, and it results particularly valuable to treat those that conserve total spin or momentum.

3.3 Variational ansatz

Following the VQA framework, we wish to approximate the eigenstates of $H_{J_1-J_2}$ by exploiting the variational principle, which involves minimizing the variational energy

$$E_{\boldsymbol{\theta}} = \langle \Psi_{\boldsymbol{\theta}} | H_{J_1-J_2} | \Psi_{\boldsymbol{\theta}} \rangle .$$

Here, $|\Psi_{\boldsymbol{\theta}}\rangle$ constitutes a variational state depending on $\boldsymbol{\theta}$, a vector of parameters. The ansatz is represented as a quantum circuit identified by a unitary transformation $U_{\boldsymbol{\theta}}$ acting on an initial state $|\phi\rangle$, such that $|\Psi_{\boldsymbol{\theta}}\rangle = U_{\boldsymbol{\theta}} |\phi\rangle$. In this Section, we utilize a quantum circuit based on the Hamiltonian Variational ansatz (HVA) [15, 120, 83, 124, 29], which proved to be effective for approximating quantum many-body eigenstates [124, 69, 4, 122, 44, 101, 32]. Furthermore, we develop a technique for *a posteriori symmetrization*, suitable to enforce translational symmetry to the resulting quantum circuit.

3.3.1 Hamiltonian Variational ansatz for $H_{J_1-J_2}$

Given a Hamiltonian H, the Hamiltonian Variational Ansatz (HVA) approach involves introducing a set of auxiliary Hamiltonians $H_1, H_2, \ldots, H_M$ such that $[H_m, H_{m'}] \neq 0 \ \forall m \neq m'$. The variational state is then expressed as:

$$|\psi_{\boldsymbol{\theta}}\rangle = \prod_{l=1}^{L} e^{-i\theta_M^l H_M} \ldots e^{-i\theta_2^l H_2} e^{-i\theta_1^l H_1} |\phi\rangle \ , \tag{3.2}$$

where the initial state $|\phi\rangle$ is typically identified as a low-energy eigenstate of one of the auxiliary Hamiltonians H_m ($m > 1$). The number of variational parameters in the HVA approach is $M \cdot L$ and the accuracy of the variational state is mainly determined by the number of layers L.

In the HVA framework, the standard practice is to require that $\sum_{m=1}^{M} H_m = H$. However, since our focus lies mostly on the symmetry properties of the auxiliary Hamiltonians rather than its sum, in this Section we relax this constraint while still maintaining a connection between $\sum_{m=1}^{M} H_m$ and H. For the interested reader, we provide an HVA ansatz satisfying this constraint in Appendix A.2.

The simplest way to define a non-trivial variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ is to consider a set of $M = 2$ auxiliary Hamiltonians H_1 and H_2. As a result, the unitary operator takes the following form:

$$U_{\boldsymbol{\theta}} = \prod_{l=1}^{L} e^{i\theta_2^l H_2} e^{i\theta_1^l H_1} \ . \tag{3.3}$$

For the J_1-J_2 Heisenberg model in Eq. (3.1), one way to define H_1 and H_2 is to observe that the nearest-neighbor term $\sum_r \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+1}$ can be decomposed into the sum of two contributions

$$H_{\text{even}} = \sum_{r=1}^{n/2} \boldsymbol{S}_{2r-1} \cdot \boldsymbol{S}_{2r} \ \text{ and } \ H_{\text{odd}} = \sum_{r=1}^{n/2} \boldsymbol{S}_{2r} \cdot \boldsymbol{S}_{2r+1} \, ,$$

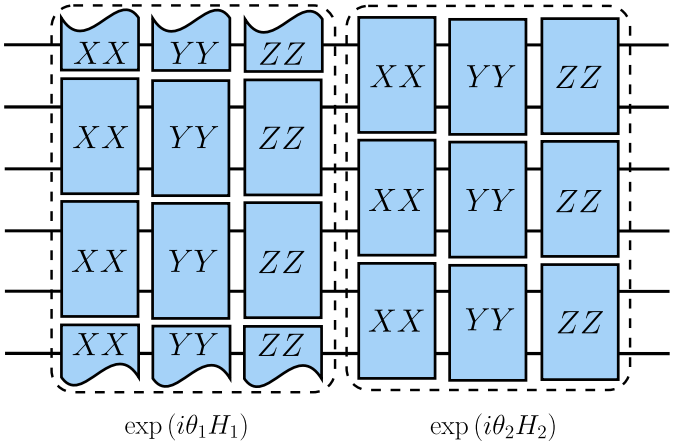

FIGURE 3.1: Implementation of a single layer of the quantum circuit defined in Eq. (3.5). This circuit consists of 2-local rotations, such as XX, which denotes the unitary $\exp\left(i\theta S^x_r \otimes S^x_{r+1}\right)$ and similarly for YY and ZZ.

where each term in the sums acts on a different pair of qubits, starting from an even and odd indexed site respectively. Additionally, the ground state of H_{even} is a product state of singlet pairs, and its excited states can be constructed by replacing a singlet pair with a triplet one, as described in the next Section. Therefore, we set $H_2 = H_{\text{even}}$, and choose the state $|\phi\rangle$ to be an appropriate eigenstate of H_{even}. Subsequently, the most natural choice for the other term is $H_1 = H_{\text{odd}}$, since $[H_{\text{even}}, H_{\text{odd}}] \neq 0$, and as a consequence we get a non-trivial circuit. Since the Hamiltonians H_1 and H_2 are defined as sums of commuting terms, the unitary operator $U_{\boldsymbol{\theta}}$ in Eq. (3.3) becomes

$$U_{\boldsymbol{\theta}} = \prod_{l=1}^{L} \prod_{r=1}^{n/2} e^{i\theta_2^l \boldsymbol{S}_{2r-1}\cdot\boldsymbol{S}_{2r}} \prod_{q=1}^{n/2} e^{i\theta_1^l \boldsymbol{S}_{2q}\cdot\boldsymbol{S}_{2q+1}} \,. \tag{3.4}$$

Furthermore, considering the definition $\boldsymbol{S}_{2r-1}\cdot\boldsymbol{S}_{2r} = \sum_\alpha S^\alpha_{2r-1} S^\alpha_{2r}$ where $\alpha = x, y, z$, it is worth noting that pairs of equal Paulis commute, namely

$$[S^\alpha_{2r-1} S^\alpha_{2r}, S^\beta_{2r-1} S^\beta_{2r}] = 0 \quad \forall \alpha, \beta \in \{x, y, z\},$$

and similarly for $\boldsymbol{S}_{2r}\cdot\boldsymbol{S}_{2r+1}$. Consequently, we can rewrite Eq. (3.4) without any approximation as

$$U_{\boldsymbol{\theta}} = \prod_{l=1}^{L} \prod_{r=1}^{n/2} \prod_{\alpha} e^{i\theta_2^l S^\alpha_{2r-1} S^\alpha_{2r}} \prod_{q=1}^{n/2} \prod_{\beta} e^{i\theta_1^l S^\beta_{2q} S^\beta_{2q+1}} \,. \tag{3.5}$$

The last step is crucial as it enables the implementation of each layer of $U_{\boldsymbol{\theta}}$ in constant depth relative to the number of qubits n. In fact, $U_{\boldsymbol{\theta}}$ in Eq. (3.5) is the composition of 2-local gates, many of which can be executed simultaneously, as shown in Fig. 3.1.

3.3.2 Enforcing symmetries on the ansatz

Our strategy for approximating excited states involves constructing variational states with definite quantum numbers, thereby energy minimizations are performed within specific symmetry sectors. In order to detect the phase transition point in the J_1-J_2 Heisenberg model, we need to fix both the momentum k and the total spin S. As a result $|\Psi_{\boldsymbol{\theta}}\rangle$ should be

invariant with respect to the operators $\boldsymbol{S^2}$ and T, namely

$$\boldsymbol{S^2}\,|\Psi_{\boldsymbol{\theta}}\rangle = S(S+1)\,|\Psi_{\boldsymbol{\theta}}\rangle \;\text{ and }\; T\,|\Psi_{\boldsymbol{\theta}}\rangle = e^{ik}\,|\Psi_{\boldsymbol{\theta}}\rangle \;\; \forall \boldsymbol{\theta} \in \Theta, \tag{3.6}$$

Given the construction $|\Psi_{\boldsymbol{\theta}}\rangle = U_{\boldsymbol{\theta}}\,|\phi\rangle$, this condition can be reduced to the *equivariance* of the circuit, namely

$$\left[U_{\boldsymbol{\theta}}, \boldsymbol{S^2}\right] = 0 \;\text{ and }\; [U_{\boldsymbol{\theta}}, T] = 0 \;\forall \boldsymbol{\theta} \in \Theta\,. \tag{3.7}$$

This way, if the initial state $|\phi\rangle$ lies in a definite symmetry sector, then the resulting variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ will automatically satisfy Eq. (3.6), and can therefore effectively estimate low-energy excitations.

In the rest of this Section, we study how to enforce this constraint to the HVA ansatz of Section 3.3.1.

Total spin conservation

Regarding the total spin, each rotation $e^{i\theta \boldsymbol{S}_j \cdot \boldsymbol{S}_k}$ in Eq. (3.4) is equivariant under spin symmetry by construction, namely

$$[\boldsymbol{S^2}, e^{i\theta \boldsymbol{S}_j \cdot \boldsymbol{S}_k}] = 0 \;\; \forall j,k \;\forall \theta \in \mathbb{R}\,.$$

Therefore, to restrict the optimization to a specific spin sector, it suffices to choose $|\phi\rangle$ with a definite quantum number S. The singlet state ($S = 0$) is implemented as the ground state of H_2, i.e.

$$|\phi_0\rangle = \prod_{r=1}^{n/2} |s\rangle_{2r-1,2r}\;, \tag{3.8}$$

where $|s\rangle_{r,r'} = (|0\rangle_r\,|1\rangle_{r'} - |1\rangle_r\,|0\rangle_{r'})/\sqrt{2}$ is a singlet pair. Similarly, the triplet state ($S = 1$) can be constructed as the first excited state of H_2 by replacing in Eq. (3.8) a singlet pair with a triplet one $|t\rangle_{r,r'} = (|0\rangle_r\,|1\rangle_{r'} + |1\rangle_r\,|0\rangle_{r'})/\sqrt{2}$. Thus we can select as initial state either of the following $n/2$ degenerate states

$$\left|\tilde{\phi}_1^j\right\rangle = |t\rangle_{2j-1,2j} \prod_{r=1, r\neq j}^{n/2} |s\rangle_{2r-1,2r}\;, \tag{3.9}$$

where $j = 1, \ldots, n/2$.

Translational invariance

Concerning translational invariance, while the Hamiltonian $H_{J_1\text{-}J_2}$ in Eq. (3.1) preserves one-site translations, the unitary operator $U_{\boldsymbol{\theta}}$ does not, namely $[U_{\boldsymbol{\theta}}, T] \neq 0$. However, the Hamiltonians H_1 and H_2 are equivariant under translations of two lattice sites, meaning

$$[H_1, T^2] = [H_2, T^2] = 0 \;\;\Rightarrow\;\; [e^{i\theta_1 H_1} e^{i\theta_2 H_2}, T^2] = 0 \;\forall \theta_1, \theta_2 \in \mathbb{R}\,.$$

Consequently, by sharing the variational parameters across different qubits as explicitely indicated in Eq. (3.5), we easily achieve $[U_{\boldsymbol{\theta}}, T^2] = 0$. While preserving two-site translations is not sufficient for our purposes, the latter condition leads us to consider a simple way to restore the one-site symmetry, namely *a posteriori* symmetrization.

As a first step in this procedure, it is necessary to define initial states $|\phi\rangle$ that possess translational invariance over two lattice sites, namely $T^2\,|\phi\rangle = |\phi\rangle$. Notably, the state $|\phi_0\rangle$ in Eq. (3.8) already exhibits this symmetry, but the states in Eq. (3.9) lack it. This issues can be

fixed by preparing a superposition of the form

$$|\phi_1\rangle = \frac{1}{\sqrt{n/2}} \sum_{j=1}^{n/2} \left|\tilde{\phi}_1^j\right\rangle , \tag{3.10}$$

implying $T^2 |\phi_1\rangle = |\phi_1\rangle$. Preparation of such superposition with a quantum circuit is generally a challenging task. Here, we provide an efficient solution to this problem based on W-states preparation circuits in Appendix A.1.

Putting everything together, our variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ satisfies $T^2 |\Psi_{\boldsymbol{\theta}}\rangle = |\Psi_{\boldsymbol{\theta}}\rangle$. In order to achieve the correct symmetry, we wish to implement the following linear combination to represent normalized, translationally invariant states with momentum $k = 0$ or $k = \pi$ as:

$$\left|\Psi_{\boldsymbol{\theta}}^k\right\rangle = \frac{|\Psi_{\boldsymbol{\theta}}\rangle + e^{ikT} |\Psi_{\boldsymbol{\theta}}\rangle}{|| \, |\Psi_{\boldsymbol{\theta}}\rangle + e^{ikT} |\Psi_{\boldsymbol{\theta}}\rangle \, ||} . \tag{3.11}$$

Indeed, starting from the definition of $\left|\Psi_{\boldsymbol{\theta}}^k\right\rangle$ it is easy to show that

$$T \left|\Psi_{\boldsymbol{\theta}}^{k=0}\right\rangle = \left|\Psi_{\boldsymbol{\theta}}^{k=0}\right\rangle \;\text{ and }\; T \left|\Psi_{\boldsymbol{\theta}}^{k=\pi}\right\rangle = - \left|\Psi_{\boldsymbol{\theta}}^{k=\pi}\right\rangle . \tag{3.12}$$

We call this procedure *a posteriori* symmetrization of the variational state. In the following Section, we show how to implement such procedure on a quantum computer for general symmetries, and subsequently apply it to prepare the symmetrized state in Eq. (3.11).

3.4 A posteriori symmetrization

While the transformation highlighted in Eq. (3.11) is in general non-unitary, in this Section we exploit a technique named Linear Combination of Unitaries (LCU) [1], which allows to implement arbitrary linear transformations on quantum computers, to enforce general symmetries.

3.4.1 Linear Combination of Unitaries in a nutshell

Linear Combination of Unitaries (LCU) is a method that allows the preparation of normalized states in the form $|\Phi\rangle \propto \Gamma |\Phi_0\rangle$, given a normalized initial state $|\Phi_0\rangle$ and a linear combination of unitaries

$$\Gamma = \sum_{a=0}^{A-1} c_a U_a .$$

Since Γ is in general non unitary, the procedure succeeds only with a certain probability, which is related to the normalization constant of $\Gamma |\Phi_0\rangle$. To achieve this goal, it is necessary to *unitarize* Γ, a process that involves introducing $\lceil \log_2(A) \rceil$ ancillary qubits, followed by a *projection*. The first step to achieve this involves acting on such qubits to produce the superposition

$$R |0\rangle = \sum_{a=0}^{A-1} \sqrt{\frac{c_a}{\|c\|_1}} |a\rangle , \tag{3.13}$$

where $\{|a\rangle\}_{a=1}^{\lceil \log_2(A) \rceil}$ is a basis of the ancillary register. The operation R is often referred to as a *prepare* step. Note that, while Eq. (3.13) does not univocally define R, all operations satifing this requirement equally work in the following steps. After applying R, the procedure requires applying the composition of all unitaries U_a *controlled* on the corresponding basis

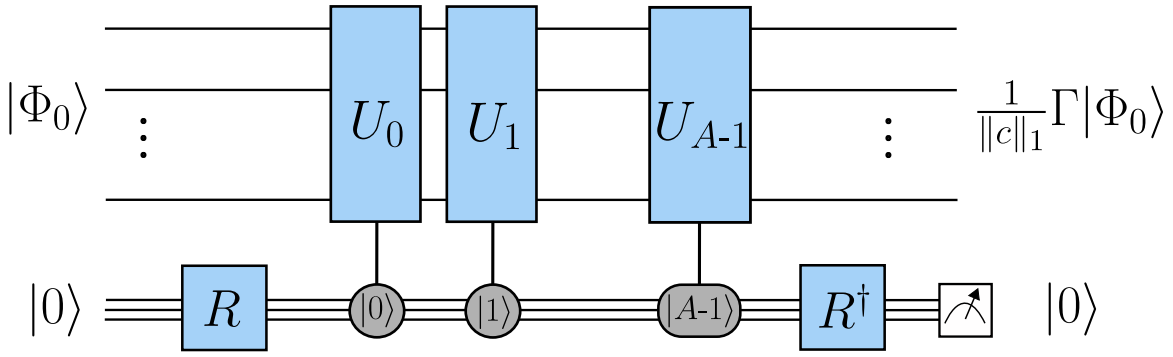

FIGURE 3.2: General circuit to implement symmetrization by LCU. To control a unitary U_a on a specific basis state $|a\rangle$ implies the realization of the unitary $|a\rangle\langle a| \otimes U_a + (\mathbb{1} - |a\rangle\langle a|) \otimes \mathbb{1}$. This notation reduces to ordinary controlled gates if $A = 2$, i.e. when only one ancillary qubit is required.

element $|a\rangle$, namely

$$\prod_{a=0}^{A-1} (|a\rangle\langle a| \otimes U_a + (\mathbb{1} - |a\rangle\langle a|) \otimes \mathbb{1}) = \sum_{a=0}^{A-1} |a\rangle\langle a| \otimes U_a. \tag{3.14}$$

The above result can be obtained by expanding the left hand side as

$$\prod_{a=0}^{A-1} (|a\rangle\langle a| \otimes U_a + (\mathbb{1} - |a\rangle\langle a|) \otimes \mathbb{1}) = \sum_{\boldsymbol{b}\in\{1,0\}^A} \prod_{b_k\in\boldsymbol{b}} (|k\rangle\langle k|)^{b_k} (\mathbb{1} - |k\rangle\langle k|)^{1-b_k} \otimes U_k^{b_k} \tag{3.15}$$

and noticing how the product $\prod_{b_k\in\boldsymbol{b}} (|k\rangle\langle k|)^{b_k} (\mathbb{1} - |k\rangle\langle k|)^{1-b_k}$ is only non-vanishing when the string $\boldsymbol{b}$ contains 1 in exactly one spot. If we let $k = a$ denote the position of the sole 1, the product reduces to the projector $|a\rangle\langle a|$, and the right hand side of Eq. (3.15) reduces to the desired result.

This procedure is usually referred to as the *select* step. Combining the two, we can unitarily prepare the superposition

$$|0\rangle_a |\Phi_0\rangle \to \sum_{a=0}^{A-1} \sqrt{\frac{c_a}{\|c\|_1}} |a\rangle U_a |\Phi_0\rangle . \tag{3.16}$$

The key point now is the observation that our desired result is contained in Eq. (3.16), and in it can be prepared by applying the projection $R|0\rangle\langle 0|R^\dagger \otimes \mathbb{1}$. In particular we get

$$\sum_{a=0}^{A-1} \sqrt{\frac{c_a}{\|c\|_1}} R|0\rangle\langle 0|R^\dagger |a\rangle U_a |\Phi_0\rangle = R|0\rangle \sum_{a=0}^{A-1} \frac{c_a}{\|c\|_1} U_a |\Phi_0\rangle = R|0\rangle \frac{1}{\|c\|_1} \Gamma |\Phi_0\rangle , \tag{3.17}$$

where we used the relation $\langle a|R|0\rangle = \sqrt{c_a/\|c\|_1}$, derived from Eq. (3.13). Practically, such a projection is realized by applying an inverse prepare step, namely $R^\dagger$, and measuring the ancillary register. We obtain the desired result when the outcome of the measurement is $|0\rangle$, which happens with a probability of success $p_s = \langle\Phi_0|\Gamma^\dagger\Gamma|\Phi_0\rangle / \|c\|_1^2$, namely the norm of the projected state in Eq. (3.17). This is pictorially represented in Fig. 3.2.

3.4.2 The procedure

We now illustrate how to apply the LCU technique to enforce any discrete, finite abelian group symmetry G, with elements $g_0, g_1, \ldots, g_{A-1}$. In general, given a character χ specifying

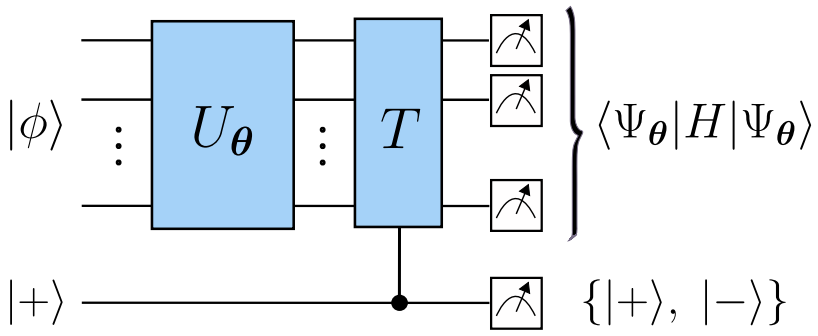

FIGURE 3.3: Circuit used to perform symmetrization by LCU, where a controlled T is added to $U_{\boldsymbol{\theta}}$ to symmetrize $|\Psi_{\boldsymbol{\theta}}\rangle$.

the symmetry sector, this is achieved by applying $\Gamma_G = \sum_{a=0}^{A-1} \chi_a^* g_a$ where χ_a is the character at element a and g_a is the unitary representation of the corresponding group element.

The symmetrization procedure requires the implementation of $A-1$ controlled operations[1]. In particular, the prepare unitary R will take the form

$$R|0\rangle = \sum_{a=0}^{A-1} \sqrt{\frac{\chi_a^*}{\sum_{a'}|\chi_{a'}|}}|a\rangle \ . \tag{3.18}$$

and the select step will employ the controlled unitary representation g_a. As before, the procedure is successful if the ancilla is found in the state $|0\rangle$.

For our purposes, consider $G = \mathbb{Z}_A$, namely the *multiplicative group of integers modulo A*. It admits a simple representation given by qubit translations, i.e. $g_a = T^a$ and $\chi_a = e^{-ika}$ for $a = 0, \ldots, A-1$ and for a chosen k. This symmetry is especially useful for the variational state of Section 3.3.2 and Appendix A.2, which is already invariant under translations of A sites, and we wish to implement full translational symmetry.

In particular, for the ansatz of Section 3.3.2, we can implement Eq. (3.11) by choosing $A = 2$, $|\Phi_0\rangle = U_{\boldsymbol{\theta}}|\phi\rangle$ and $\Gamma = \mathbb{1} \pm T$. Consequently, only one ancilla is required, regardless of the system size n. The corresponding circuit is schematically illustrated in Fig. 3.3. First, we prepare the ancilla in the $|+\rangle$ state. Then, we apply a *controlled* version of T. Finally, since the prepare unitary is very simple, the final projection is performed by measuring the ancilla in the $\{|+\rangle, |-\rangle\}$ basis directly. Depending on the measurement outcome, the computational register is prepared in the $|\Psi_\theta^{k=0}\rangle$ or $|\Psi_\theta^{k=\pi}\rangle$ states respectively. Thus, by post-selecting the appropriate measurement results from the ancillary qubit, both momenta ($k = 0$ or $k = \pi$) can be obtained using just one circuit[2]. Additionally, we mention that other symmetrization approaches, based on classical post-processing, are possible but may require a higher number of circuit evaluations [101].

While successful, the use of a posteriori symmetrization requires an increased amount of quantum resources. In the following section, we quantify them, both in terms of additional gates, ans in terms of nummber of shots required.

3.4.3 Computational costs

In this Section, we consider the asymptotic increase in quantum resources due to a posteriori symmetrization as a function of the number n of qubits. This analysis is conducted

[1] Note that, although G has A elements, the element g_0 is always mapped to the identity $\mathbb{1}$, and hence needs not to be implemented.

[2] In Section 3.4.3 we discuss both the computational cost of LCU to restore translational symmetry on a two-site translationally invariant state and to enforce a generic discrete symmetry.

in terms of:

1. *Gates overhead*. We determine the scaling of quantum operations not included in the state preparation $U_{\boldsymbol{\theta}}$.

2. *Samples overhead*. We analyze the success probability p_{s} of the algorithm. Indeed, achieving (on average) a number of accepted samples of n_s requires performing a *larger* number of experiments n_{sym} such that $p_{\mathrm{s}}\, n_{\mathrm{sym}} = n_s$. This gives a sampling overhead inversely proportional to p_{s}.

Here, we explicitly compute these overheads, both in our use case and in general.

Restoring full translational invariance given two-site invariance

As show in Section 3.4.2, LCU can be used to restore full translational symmetry in a variational state $|\Psi_{\boldsymbol{\theta}}\rangle$ by applying the circuit summarized in Fig. 3.3. In particular, this holds provided that the initial state is already invariant under two-site translations, i.e., $T^2 |\Psi_{\boldsymbol{\theta}}\rangle = |\Psi_{\boldsymbol{\theta}}\rangle$. The overheads in this procedure are listed below.

1. *Gates overhead*. Additional operations include preparation of the $\{|+\rangle, |-\rangle\}$ states in the ancilla requiring $O(1)$ operations, and implementation of T, which requires $n-1$ SWAP gates, i.e. $O(n)$ operations. Note that in the controlled version, instead of SWAPs, one needs to implement Fredkin gates (controlled SWAPs), which although more expensive, still require $O(n)$ elementary operations overall, yielding a total cost of $O(n)$ gates.

2. *Samples overhead*. Consider the preparation of variational states of momentum $k = 0$, namely we post-select samples where the ancilla is found in the state $|+\rangle$. In this case, $p_{\mathrm{s},k=0}$ can be computed using standard LCU theory, namely

$$p_{\mathrm{s},k=0} = \frac{\langle \Phi_0 | \Gamma^\dagger \Gamma | \Phi_0 \rangle}{\|\boldsymbol{c}\|_1^2} \tag{3.19}$$

where $|\Phi_0\rangle = |\Psi_{\boldsymbol{\theta}}\rangle$, $\Gamma = \mathbb{1} + T$ and $\boldsymbol{c} = (1,1)^T$. After straightforward manipulations, we obtain

$$p_{\mathrm{s},k=0}(\boldsymbol{\theta}) = \frac{1}{2} + \frac{\Re \langle \Psi_{\boldsymbol{\theta}} | T | \Psi_{\boldsymbol{\theta}} \rangle}{2}. \tag{3.20}$$

Given $T = T^\dagger T^2$ and the invariance under two-site translations of $|\Psi_{\boldsymbol{\theta}}\rangle$, it immediately follows that $\langle \Psi_{\boldsymbol{\theta}} | T | \Psi_{\boldsymbol{\theta}} \rangle = \langle \Psi_{\boldsymbol{\theta}} | T^\dagger | \Psi_{\boldsymbol{\theta}} \rangle \in \mathbb{R}$. Since only two outcomes are possible when measuring a single qubit, it implies $p_{\mathrm{s},k=\pi} = 1 - p_{\mathrm{s},k=0}$. As a result

$$p_{\mathrm{s},k}(\boldsymbol{\theta}) = \frac{1}{2} + e^{ik} \frac{\langle \Psi_{\boldsymbol{\theta}} | T | \Psi_{\boldsymbol{\theta}} \rangle}{2} \quad , \quad k = 0, \pi. \tag{3.21}$$

Note that $p_{\mathrm{s},k}$ depends on the variational parameters, and hence will vary depending on initialization and during the energy minimization.

General method for a posteriori symmetrization

The same type of analysis can be also done for the more general case, as function of not only n, but also of the properties of the considered symmetry group G.

1. *Gate overhead*. In the general case, it is not easy to estimate the number of additional operations as a function of n and A. In particular, it depends on the number of gates N_a and

N_R to implement each controlled g_a and R, respectively.

2. *Samples overhead*. A general result can be derived from Eq. (3.19), namely

$$p_s = \frac{\sum_a |\chi_a|^2 + 2\sum_{b>a} \Re \chi_a \chi_b^* \langle \Psi_{\boldsymbol{\theta}} | g_a^\dagger g_b | \Psi_{\boldsymbol{\theta}} \rangle}{\left(\sum_a |\chi_a|\right)^2} , \tag{3.22}$$

which follows directly by the definition of Γ_G and generalizes Eq. (3.20). However, due to the dependence on variational parameters $\boldsymbol{\theta}$, it is not trivial to estimate the scaling of p_s. Nevertheless, the average scaling can be computed. In particular, if we assume $U_{\boldsymbol{\theta}}$ forms a unitary 1-design over the initialization probability[3], we can use the relation

$$\mathbb{E}_{\boldsymbol{\theta}} \{ \langle \Psi_{\boldsymbol{\theta}} | O | \Psi_{\boldsymbol{\theta}} \rangle \} = \frac{1}{2^n} \mathrm{Tr}\{O\} , \tag{3.23}$$

which holds for any operator O. This allows to compute the average probability of success $\mathbb{E}_{\boldsymbol{\theta}}\{p_s\} = \bar{p}_s$. Employing Eq. (3.23) we get

$$\bar{p}_s = \frac{\sum_a |\chi_a|^2 + 2^{-N} \sum_{b>a} 2\Re \chi_a \chi_b^* \mathrm{Tr}\{g_a^\dagger g_b\}}{\left(\sum_a |\chi_a|\right)^2} , \tag{3.24}$$

which now only depends on n, A and the representation of G.

As an example, consider as before $G = \mathbb{Z}_A$, and its representation given by qubit translations, i.e. $g_a = T^a$ and $\chi_a = e^{-ika}$ for $a = 0, \ldots, A-1$. Assuming the implementation of R does not scale with n, the number of additional operations can be estimated to scale as $O(A \cdot N)$. Regarding sampling overhead, applying Eq. (3.24) yields

$$\bar{p}_{s,k} = \frac{1}{A} + \frac{\sum_{b>a} \cos[ik(a-b)] \mathrm{Tr}\{T^{b-a}\}}{2^{N-1} A^2} . \tag{3.25}$$

Furthermore, by noting that $0 \leq \mathrm{Tr}\{T^{b-a}\} \leq 2^A$ we can bound the magnitude of the second term:

$$\bar{p}_{s,k} = \frac{1}{A} + O(2^{A-n}) . \tag{3.26}$$

This is especially useful in cases where $n \gg A$, since it ensures a sizable initial probability. Setting $A = 2$, we get the use-case discussed in this Chapter, thus formally proving a sizable success probability in the symmetrization step regardless of the problem's size. To strengthen our claims, we numerically show in Fig. 3.4 that the average probability of success $\bar{p}_s$ indeed approaches 1/2 exponentially fast as the system size n increases in our use-case, as predicted by Eq. (3.26).

3.4.4 Interaction with noise

While effective in the ideal case, *a posteriori* symmetrization procedures (like those discussed in this Section) are notably sensitive to errors arising from noise [14, 19]. Furthermore, since the circuit depth scales *linearly* with the system size n, scaling our approach in the presence of noise may be challenging. Nevertheless, energy level crossings are expected to be effective at relatively moderate system sizes. Therefore, by combining small sized circuits with error mitigation techniques, this approach could still effectively detect the phase

[3] We refer to Chapter 5 for a formal defintion of unitary designs.

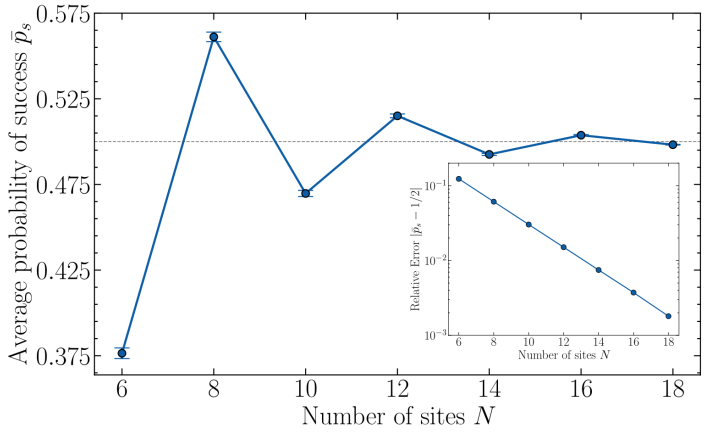

FIGURE 3.4: Scaling of the average probability of success $\bar{p}_s$ as a function of the system size n over the initialization distribution for the ansatz in Section 3.3.1, assuming $S = 1$ and $k = \pi$. Inset: Relative error $|\bar{p}_s - 1/2|$ as a function of the system size n.

transition point, even when using imperfect devices. For this reason, in the following, we investigate the impact of small perturbations on the quantum circuit.

3.5 Noise and mitigation

Accurately modeling incoherent errors is challenging [35, 23, 71]. Here, we define a simple noise model that captures only local incoherent errors, allowing to examine how small perturbations affect the symmetries of the variational state. This noise model is implemented as quantum channels $\mathcal{E}_\tau$, which generally depend on the gate time τ. They are used to approximate the dominant local errors typically occurring on real devices during the gate execution. The most common models include depolarization $\mathcal{E}_\tau^D$ and thermal relaxation $\mathcal{E}_\tau^R$, which are applied after each gate in the quantum circuit [77, 9]. Both maps induce single-qubit decoherence. However, the fixed points of the two maps differ, thus introducing competing effects. On the one hand, single-qubit depolarization tends reduce coherence and to bring the state towards $\mathbb{1}/2$, namely the maximally mixed one. On the other hand, focusing on superconducting circuit implementations [47], thermal relaxation tends to bring qubits that are in the excited state $|1\rangle$ to the ground state $|0\rangle$, thus having the pure state $|0\rangle\langle 0|$ as a fixed point. (see Appendix A.3 for a formal definition and a discussion of the noise channels). Here, we choose to combine the two as $\mathcal{E}_\tau = \mathcal{E}_\tau^D \circ \mathcal{E}_\tau^R$. Furthermore, we neglect cross talks and correlated noises [99] implying that the total channel associated to m-qubit gates is $\mathcal{E}_\tau^{\otimes m}$, i.e., the tensor product of the single-qubit one. While such effects generally are present in realistic hardware, single-qubit contributions to noise are typically dominant [106], hence providing a good approximation.

Together with errors arising from the gate execution, readout errors are also present during the measurement procedure at the end of computation [74]. However, we neglect this error source, since highly effective techniques such as T-REX [109], capable of mitigating these errors, are already accessible on current quantum devices. Moreover, since such quantum devices can only implement a specific set of universal gates (see Appendix A.3), to conduct numerical simulations that are more faithful to the real hardware, we transpiled our algorithm accordingly.

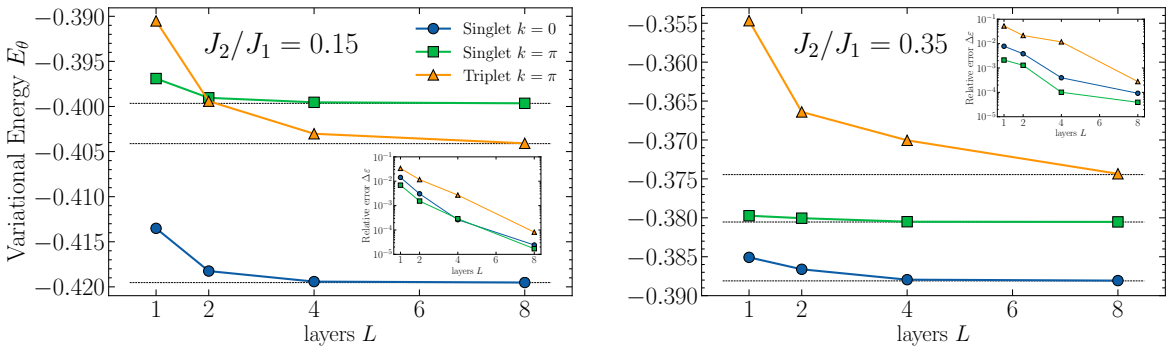

FIGURE 3.5: The variational energy for a cluster of $n = 16$ sites as a function of the number of layer $L = 1, \ldots, n/2$ for $J_2/J_1 = 0.15$ (left panel) and $J_2/J_1 = 0.35$ (right panel). Both plots share the same color code. The exact energies are also reported as dotted lines in both panels. The corresponding relative error $\Delta\varepsilon = |E_{\boldsymbol{\theta}} - E_{\text{ex}}|/E_{\text{ex}}$ with respect to the exact energies is reported as insets as a function of L.

In the noise model outlined previously, the quantum channels $\mathcal{E}_\tau$ depend on the gate time τ and the error probabilities, associated to each channel, increase with τ (see Appendix A.3). This parameter can be theoretically adjusted to control the noise level, a manipulation that can also be realized experimentally through various techniques [106, 58, 36]. This allows to perform Zero Noise Extrapolation (ZNE) [106], an error mitigation strategy suitable for expectation value estimations. In this approach, the expectation value is computed with increasing noise levels (i.e., the gate time τ), in order to extrapolate the ideal result in the zero-noise limit (i.e., $\tau \to 0$).

Generally, the introduction of noise explicitly breaks the symmetries of the ansatz [108], implying possibly wrong estimations of the low-lying excited energies. To counteract this effect we modify the problem's cost function with a penalty term which favours variational states lying in the correct symmetry sector [60]. By exploiting the optimizer to mitigate the noise-induced effects on the symmetries of the ansatz, in combination with ZNE, we accurately recover the zero-noise limit.

3.6 Numerical Results

In this Section we perform both noisy and ideal simulations of our proposed algorithm, to evaluate its performance. In particular, noiseless simulations are used to assess the required depth and accuracy of the excited state energies achievable with our ansatz, while noisy simulations are performed in conjunction with quantum error mitigation to approximate the transition point by measuring the level crossing in a realistic scenario.

3.6.1 Ideal setting

Here, we examine the results obtained in the noiseless scenario by performing numerical state-vector simulations (PennyLane's `lightning.qubit` backend [113]). Specifically, we investigate how the accuracy of the variational state can be systematically enhanced by increasing the number of layers L in the circuit across the different symmetry sectors.

We focus on a cluster of $n = 16$ qubits for two frustration ratios: $J_2/J_1 = 0.15$ (gapless phase) and $J_2/J_1 = 0.35$ (gapped phase). In Fig. 3.5 we show the dependence of the variational energies with respect to the number of layers L of the circuit. The optimizations are carried out fixing the quantum numbers of total spin S (singlet or triplet) and momentum

($k = 0, \pi$). For both frustration ratios, the ground state is a singlet ($S = 0$) with $k = 0$. The first excited state before the phase transition ($J_2/J_1 = 0.15$, left panel) is a triplet state ($S = 1$) with momentum $k = \pi$. Then, after the transition ($J_2/J_1 = 0.35$, right panel), it becomes a singlet state ($S = 0$) with momentum $k = \pi$. The situation is reversed for the second excited state.

For both frustration ratios, the relative error of the states compared to the exact ones [56, 98] is of order $\Delta\varepsilon \approx 0.01\%$ for a number of layers $L = n/2$ [44] (see insets in Fig. 3.5). We observe that the convergence of the triplet state is slower compared to the singlet case (see Fig. 3.5). This behavior can be attributed to the larger symmetry sector of the triplet state ($S = 1$), which includes $S^z = -1, 0, 1$, in contrast to the singlet excited state $S = 0$, which only includes $S^z = 0$. Consequently, more layers are required to adequately span the entire symmetry subspace [60]. Nevertheless, by reaching $L = n/2$, the accuracy is of the same order of magnitude for both cases. Additionally, we emphasize that using a highly symmetric ansatz, as we do here, generally reduces the subspace that needs to be explored during optimization. This approach helps mitigate trainability challenges in variational circuits, such as barren plateaus, introduced in Chapter 5. Furthermore, we employ a Gaussian initialization of the parameters, with variance scaling as $\sim 1/L$, which was shown to be effective to mitigate barren plateaus in the general setting [50]. The accurate results obtained before and after the transition suggest that the variational state accurately captures the transition point through the crossing of excited states.

3.6.2 Noisy setting

The introduction of noise in the circuit modifies the performance of the ansatz. In this Section we investigate the impact of the noise model (see Section 3.5) on the symmetries of the variational state. We show that it is possible to mitigate errors through ZNE combined with the introduction of a suitable penalty term in the cost function. Here, we focus on a system of $n = 4$ sites using a circuit with $L = 1$ layer. Indeed, for such a small cluster one layer is sufficient to get accurate estimations in the noiseless limit. Using as reference current IBM superconducting devices, the single-qubit gate time is $t_g \approx 3.5 \times 10^{-8}$ s. For this reason, variational optimizations are performed setting $\tau = t_g$ (see Appendix A.3). Regarding mitigation, we dub *ZNE region* the gate-time interval attainable to perform ZNE on current quantum hardware (i.e. from $\tau = t_g$ up to $\tau \approx 3t_g$). All numerical simulations are performed by employing a density matrix simulator (PennyLane's `default.mixed` backend [113]).

Breaking of Equivariance due to Noise

In Fig. 3.6 we show the results at $J_2/J_1 = 0.15$ for the singlet excited state ($S = 0$) at $k = \pi$. We start describing what happens when changing continuously the gate time τ in the variational circuit. First, we optimize the variational state at $\tau = t_g$ (empty green diamond). Then, fixing the optimal parameters, the noise level is changed shifting τ in the interval $\tau/t_g \in [10^{-7}, 10^2]$ (dashed green curve). We point out that for small values of the gate time ($\tau/t_g \sim 10^{-7}$) the variational result approaches the exact energy (red star).

However, in the *ZNE region* (shaded interval) the variational energy is lower with respect to the exact one in the selected symmetry sector[4]. As a result, due to the non-monotonic behaviour of the noisy energy curve, performing the ZNE in the *ZNE region* gets a off-target result, committing an error of 45% with respect to the exact energy. This suggests that the variational state has no definite momentum, implying a breaking in the equivariance of the circuit due to the presence of noise [108].

[4] We remark that the variational energy is consistently higher than the ground state energy. However, if the symmetries of the variational state are not preserved, its energy may be lower than that of the excited state we are approximating.

To better understand this behaviour, we measure, during the energy optimization at $\tau = t_g$, the probability $p_s(\boldsymbol{\theta}) = (1 + e^{ik} \langle \Psi_{\boldsymbol{\theta}} | T | \Psi_{\boldsymbol{\theta}} \rangle)/2$, which quantifies the success in performing the LCU symmetrization (green dashed curve in the inset (1) of Fig. 3.6). This probability decreases during the optimization and at the end is quite low ($p_s \approx 15\%$). We point out that in a noiseless scenario, the definition of a state with a definite momentum is achieved irrespective of the value of p_s. However, when noise is present, the probability of success p_s in executing the LCU symmetrization becomes relevant to the effective generation of translationally invariant states. In general for both scenarios, low values of p_s imply that the one-site translations are primarly restored by LCU, indicating that components in the wrong symmetry sectors are relevant before the LCU application. Conversely, when p_s is high, most of the symmetrization is effectively performed by optimizing the parameters within the variational circuit $U_{\boldsymbol{\theta}}$, relegating a minor role to LCU. Since the latter is the most susceptible component to noise in the quantum circuit [14, 19], its contribution to equivariance loss is expected to be predominant. Consequently, we identify the decay of p_s during training as the main indicator of noise-induced equivariance breaking within the circuit.

Noise mitigation and symmetry restoration

In this Section, we devise a strategy to mitigate the effect of noise on the symmetries of the ansatz. As discussed in the previous Section, in order to reduce the role of the LCU in the construction of a translationally invariant state we aim at increasing the probability p_s. This can be achieved by adding a penalty term in the loss function

$$P(\lambda, \boldsymbol{\theta}) = \lambda \left[1 - p_s(\boldsymbol{\theta})\right]^2 . \tag{3.27}$$

Here, λ is an hyperparameter which controls the intensity of the penalty term. Performing energy minimizations at different values of λ, fixing the gate time $\tau = t_g$, allows to determine its optimal value. In the inset (2) of Fig. 3.6 we show a sharp transition at $\lambda = 0.05$ from a regime of low ($p_s \approx 15\%$) to high ($p_s \approx 90\%$) probability measured at the end of the training. In addition, in the inset (1) we show how the behaviour of the probability p_s, during the energy optimization, is modified by introducing a penalty term (solid green curve). As a result, by setting $\lambda = 0.05$, the energy of the optimized state (empty green square in Fig. 3.6) results in a reliable approximation of the energy of the singlet excited state at $k = \pi$ (with a relative error $\Delta\varepsilon \approx 5\%$).

At this stage, maintaining the optimal variational parameters obtained at $\tau = t_g$, the noise level is changed shifting τ in the interval $\tau/t_g \in [10^{-7}, 10^2]$ (green solid curve in Fig. 3.6). Here, a monotonic behaviour emerges, facilitating the implementation of Zero Noise Extrapolation within the *ZNE region* (see below).

Mitigated energy level crossing

Finally, in Fig. 3.7 we estimate the variational energies, at gate time $\tau = t_g$, for the different symmetry sectors varying the values of the frustration ratio in the interval $J_2/J_1 \in [0.15, 0.35]$[5]. As shown in the left panel, the variational energies are shifted with respect to the exact ones (marked by dashed lines) due to noise. The mitigated results with ZNE technique performed in the *ZNE region* are depicted in the right panel.

In the inset, for $J_2/J_1 = 0.15$, the empty points represent the noisy energies (also depicted on the left panel), while the crosses denote the expectation values obtained with increasing gate time. The filled points indicate the extrapolated values in the zero-noise limit (also displayed in the right panel) after fitting the data by linear regression.

[5] In particular, an appropriate value of λ is chosen for each simulation.

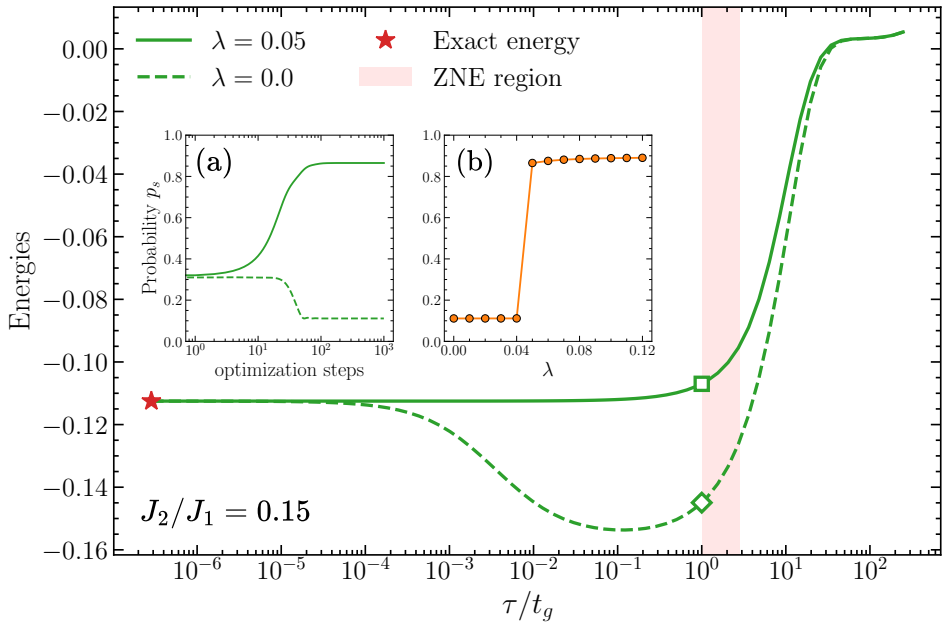

FIGURE 3.6: Variational energies for $S = 0$ and $k = \pi$ on a $n = 4$ site cluster at $J_1/J_2 = 0.15$, plotted as a function of τ/t_g, with $\lambda = 0.0$ (dashed green curve) and $\lambda = 0.05$ (solid green curve). The empty green square (diamond) indicates the energy at $\tau/t_g = 1$ for $\lambda = 0.05$ ($\lambda = 0.0$). The shaded red region marks the *ZNE region* spanning $\tau/t_g \in [1,3]$. The exact energy value is also shown for comparison (red star). The inset (1) displays the evolution of the probability p_s over optimization steps for the gate time $\tau = t_g$, with $\lambda = 0.0$ (dashed green curve) and $\lambda = 0.05$ (solid green curve). The inset (2) shows the values of the probability p_s, at the end of the optimization, as a function of λ.

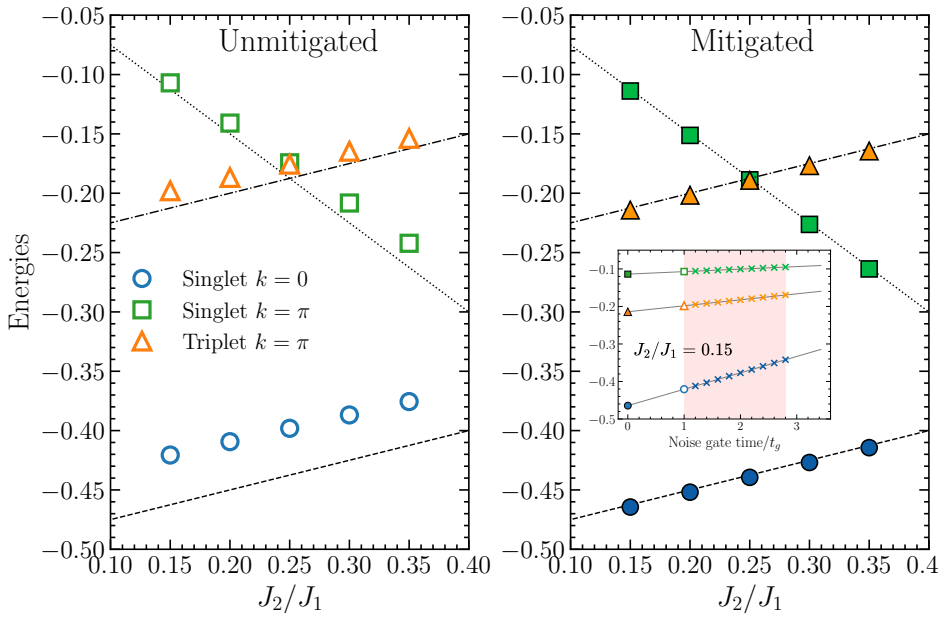

FIGURE 3.7: Variational energies for the different symmetry sectors in the interval $J_2/J_1 \in [0.15, 0.35]$. Unmitigated results denoted by empty markers are obtained at $\tau = t_g$ (left panel), while the mitigated energies obtained via ZNE and denoted by filled markers are shown in the right panel. The inset shows ZNE extrapolations performed within the *ZNE region* ($\tau/t_g \in [1,3]$) for $J_1/J_2 = 0.15$.

We point out that even in the unmitigated scenario (left panel), the crossing point is adequately captured, despite slight energy shifts. However, through mitigation, we not only identify the energy level crossing accurately but also approximate the exact energies with an $\Delta\varepsilon \approx 1\%$ error (right panel).

3.7 Discussion

We have introduced an ansatz inspired by HVA to investigate the excited states of an interacting spin model on a lattice. Specifically, we have discussed the possibility to carry out optimizations in specific symmetry sectors by fixing the quantum numbers in the variational state. While HVA states are not yet competitive with classical methods, such as Tensor Networks and NQS, for the approximation of the low-energy states of quantum spin models, this Chapter aims to emphasize that the total-spin symmetry, generally challenging to incorporate in classical approaches [112, 111, 105, 100], can be more straightforwardly implemented in quantum circuit-based ansätze. Although highly accurate results can often be achieved with classical states even without enforcing exact total-spin conservation, we believe that having a simple method to construct $SU(2)$-symmetric states presents an interesting application of HVA quantum circuits. It is worth noting that, by employing level spectroscopy, the quantum phase transition point in the one-dimensional J_1-J_2 Heisenberg model is estimated at $J_2/J_1 = 0.25$ on a cluster of $n = 4$ sites (see Fig. 3.7), which is exact for the finite size, and has an error of just 3.6% compared to the thermodynamic limit result $(J_2/J_1)_c = 0.24117(6)$, highlighting the small size effects of this technique. For this reason, it can be properly generalized to study other models, specifically the application of the same approach to systems with long-range interactions or two-dimensional systems, where the presence of phase transitions is still under debate.

Furthermore, we discussed how noise, which explicitly breaks state symmetries, can be mitigated by standard techniques such as ZNE, with the addition of an appropriate penalty term which helps in finding symmetric solutions. From this perspective, the implementation of this ansatz on current quantum devices represents the next step in verifying whether the mitigation techniques employed in simulations are still effective. Moreover, this implementation will involve assessing also the impact of finite samples on the estimation of the expectation values [7].

Chapter 4

Variational correction of coherent errors

Stabilizer states are a central resource in quantum information processing, underpinning a wide range of applications. While they can be efficiently generated via Clifford circuits, the presence of coherent errors, such as small-angle miscalibrations in native gate implementations, can significantly impact their quality. In this Chapter, we introduce a method that employs the stabilizer formalism to suppress coherent errors through variational optimization of gate parameters. This method is provably robust, remaining largely unaffected by incoherent noise, and enabling pre-compensation of coherent errors prior to the application of standard error mitigation techniques. Beyond theoretical results, we further support our claims through numerical simulations.

4.1 Coherent errors on stabilizer states

Consider $\mathbf{P}_n$ the Pauli group on n qubits. An n-qubit state $|\psi_S\rangle$ is called a stabilizer state if there exists a subgroup $\mathbf{S}_n \subset \mathbf{P}_n$ of dimension 2^n and such that

$$S_i |\psi_S\rangle = |\psi_S\rangle \ , \tag{4.1}$$

for every generator S_i of $\mathbf{S}_n$ with $i = 1, \ldots, n$ [6]. The S_i operators are called *stabilizers* for the state $|\psi_S\rangle$ and they have eigenvalues ± 1, each with degeneracy 2^{n-1}, namely

$$S_i|\psi_{ik}^{(+)}\rangle = |\psi_{ik}^{(+)}\rangle \ \forall k = 1, \ldots, 2^{n-1} \ \text{ and } \ S_i|\psi_{ik}^{(-)}\rangle = -|\psi_{ik}^{(-)}\rangle \ \forall k = 1, \ldots, 2^{n-1}, \tag{4.2}$$

where the eigenvectors $|\psi_{ik}^{(\pm)}\rangle$ are orthonormal $\langle\psi_{ik}^{(\pm)}|\psi_{ik'}^{(\pm)}\rangle = \delta_{kk'}$. Being a stabilizer state, it is always possible [77] to identify a Clifford circuit U_S such that $|\psi_S\rangle = U_S\,|0\rangle^{\otimes n}$. However, in the presence of even a small amount of coherent errors, the state $|\psi_S\rangle$ becomes non-stabilized, and Eq. (4.1) does not hold.

Coherent errors can be described as unwanted small extra rotations in the native gates of a quantum device resulting from gates miscalibration. In what follows, we assume that each gate in U_S has its own small coherent error which is random but stable, i.e. constant over the preparation time of $|\psi_S\rangle$. Our aim is to correct for such errors.

Towards that aim, note that any stabilizer state can be expanded as a superposition of the +1 eigenvectors for each S_i

$$|\psi_S\rangle = \sum_{k=1}^{2^{n-1}} a_{ik}|\psi_{ik}^{(+)}\rangle, \tag{4.3}$$

where a_{ik} are generic amplitudes, while, by definiton, the -1 eigenvectors do not play any role. However, in the presence of coherent errors the state is modified, and can lose its *stabilizerness*. In particular, if we take $\boldsymbol{\epsilon}$ as the vector collecting each gate's calibration error, then

in the presence of coherent noise, we have $|\psi_S(\boldsymbol{\epsilon})\rangle = U_S(\boldsymbol{\epsilon})\,|0\rangle^{\otimes n}$ and $U_S(\boldsymbol{\epsilon})$ is in general non-Clifford. Contrarily to Eq. (4.3), the -1 eigenvectors of S_i are now needed in the expansion, and the modified state can be written as

$$|\psi_S(\boldsymbol{\epsilon})\rangle = \sum_{k=1}^{2^{n-1}} a_{ik}(\boldsymbol{\epsilon})|\psi_{ik}^{(+)}\rangle + b_{ik}(\boldsymbol{\epsilon})|\psi_{ik}^{(-)}\rangle \ , \tag{4.4}$$

where $\sum_k |a_{ik}(\boldsymbol{\epsilon})|^2 + |b_{ik}(\boldsymbol{\epsilon})|^2 = 1$, with $b_{ik}(\boldsymbol{\epsilon})$ generally different from zero. In what follows, we will assume that the circuit is simple enough that, when a small perturbation is introduced into the circuit parameters, the state $|\psi_S\rangle$ is also perturbed, i.e. the effects of the perturbation do not cancel out. More precisely, this idea can be expressed in terms of Eq. (4.4) saying that, whenever $\|\boldsymbol{\epsilon}\| \ll 1$ but $\boldsymbol{\epsilon} \neq 0$, then at least on b_{ik} coefficient satisfies $|b_{ik}(\boldsymbol{\epsilon})| > 0$. While apparently trivial, this assumption is key for the convexity claims made in the following Sections.

4.2 Variational coherent error mitigation

The main idea behind variational correction of coherent errors is that one can introduce a vector of variational parameters $\boldsymbol{\theta}$ to the circuit $U_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})$ in order to compensate the miscalibrations. More precisely, we wish to transpile the circuit into the native gates of a given quantum device and parametrize their rotation angles to identify the over-rotations. The resulting state $|\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\rangle = U_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\,|0\rangle^{\otimes n}$ can again be expanded analogously to Eq. (4.4) as

$$|\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\rangle = \sum_{k=1}^{2^{n-1}} a_{ik}(\boldsymbol{\theta}+\boldsymbol{\epsilon})|\psi_{ik}^{(+)}\rangle + b_{ik}(\boldsymbol{\theta}+\boldsymbol{\epsilon})|\psi_{ik}^{(-)}\rangle\,, \tag{4.5}$$

and our aim can be restated in terms of the amplitudes b_{ik} as a simultaneous minimization of their modulus, namely

$$\min_{\boldsymbol{\theta}} \ \sum_{k=1}^{2^{n-1}} \sum_{i=1}^{n} |b_{ik}(\boldsymbol{\theta}+\boldsymbol{\epsilon})|^2, \tag{4.6}$$

which attains its minimum when $|\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\rangle$ is a the correct stabilizer state. More practically, we can define the following cost function

$$C(\boldsymbol{\theta}) = \sum_{i=1}^{n} C_i(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \langle\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})|S_i|\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\rangle\,, \tag{4.7}$$

i.e. the sum of the expectation values of the stabilizers S_i, which has the same optimal parameters. Indeed

$$\begin{aligned} C_i(\boldsymbol{\theta}) &= -\langle\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})|S_i|\psi_S(\boldsymbol{\theta}+\boldsymbol{\epsilon})\rangle = -1 + 2\sum_k |b_{ik}(\boldsymbol{\theta}+\boldsymbol{\epsilon})|^2 \geq -1\,, \\ &\Rightarrow C(\boldsymbol{\theta}) = -n + 2\sum_{k=1}^{2^{n-1}} \sum_{i=1}^{n} |b_{ik}(\boldsymbol{\theta}+\boldsymbol{\epsilon})|^2. \end{aligned} \tag{4.8}$$

In the same way as before, the global minimum of the cost function $-n$ is exactly obtained at the correct stabilizer state. By finding the minimum of the cost function, the resulting optimal parameters $\boldsymbol{\theta}^*$ are such that $\boldsymbol{\theta}^* = -\boldsymbol{\epsilon}$ and coherent errors are corrected. When $\boldsymbol{\theta}^*$ is obtained, one runs the corrected circuit $U_S(\boldsymbol{\theta}^*+\boldsymbol{\epsilon}) = U_S$. We name this technique Variational Coherent Error Mitigation (VCEM). A pictorial representation of the VCEM circuit is shown in Fig. 4.1.

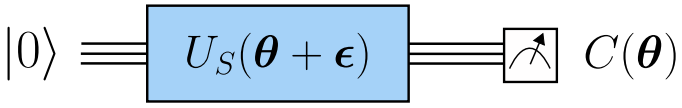

FIGURE 4.1: Parametrized circuit to evaluate the VCEM cost function.

On a practical note, assuming that $\|\boldsymbol{\epsilon}\| \ll 1$, one can start the optimization in $\boldsymbol{\theta} = 0$, that is equivalent to optimize the cost function in a region $\mathcal{B}(-\boldsymbol{\epsilon}, \delta)$ for $\delta \ll 1$, and such that $\boldsymbol{\theta}^* \in \mathcal{B}$. Furthermore, we argue that, under the simplicity assumptions of Section 4.1, the cost function is convex in $\mathcal{B}$ as in that case, $\boldsymbol{\theta} = -\boldsymbol{\epsilon}$ is the sole global minimum in $\mathcal{B}$. This can be formally stated by saying that the Hessian $\mathbf{H} > 0$ computed at $\boldsymbol{\theta} = -\boldsymbol{\epsilon}$ is positive definite. This excludes trainability issues such as barren plateaus in the ideal unitary case, consistently with the literature [90]. In the following Sections, we show how the same claim continues to hold also in the presence of incoherent errors.

The natural application of this method is the correction of coherent errors arising in the preparation of a stabilizer state. However, the method has broader applicability, as the error mitigation is performed at the level of native gates acting on physical qubits. This enables its use as a general-purpose technique, independent of the specific stabilizer state initially considered. The choice of stabilizer state can be tailored to the target application: for instance, in simulations of many-body quantum systems, one might select a stabilizer state reflecting the system's topology, such as spin lattices or molecular structures. Another practical choice is the graph state associated with the connectivity of the specific quantum device.

4.3 VCEM in presence of incoherent errors

In the previous section we assumed the presence of coherent errors only, here we show that under certain assumptions VCEM can still be effective in correcting coherent errors also in presence of incoherent ones. Thus from this point on we work with density matrices instead of state vectors. We define $\rho_S = |\psi_S\rangle\langle\psi_S| = \mathcal{U}_S(\rho_0)$, $\rho_0 = (|0\rangle\langle 0|)^{\otimes n}$ and to ease the notation we omit the explicit dependence on variational parameters $\boldsymbol{\theta}$ and coherent errors $\boldsymbol{\epsilon}$. Moreover when needed, we divide the circuit $\mathcal{U}_S$ in M moments, namely

$$\mathcal{U}_S = \prod_{q=1}^{M} \mathcal{U}_q = \mathcal{U}_M \mathcal{U}_{M-1} \dots \mathcal{U}_2 \mathcal{U}_1,$$

where each moment $\mathcal{U}_q$ is a sub-circuit consisting of native gates acting simultaneously. In the following Section, we will study how the cost function changes under different noise models, and its consequences in terms of the efficacy of VCEM.

4.3.1 From general errors to Pauli maps

In the following Sections, we model noisy quantum channels as Pauli maps, i.e. consistent with the following Definition.

Definition 3 (Pauli map)**.** A quantum channel $\mathcal{P}$ acting on n qubits is a Pauli map if it can be decomposed as

$$\mathcal{P}(\rho) = \sum_{j=0}^{4^n-1} p_j P_j \rho P_j, \tag{4.9}$$

where $p_j \geq 0$, $\sum_j p_j = 1$; P_j are Pauli strings and ρ is a generic density matrix of n qubits.

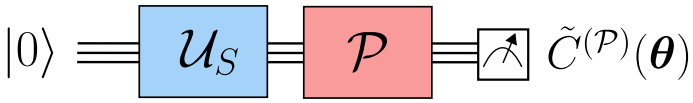

FIGURE 4.2: Parametrized circuit to evaluate the modified VCEM cost function with a Pauli noise channel after $\mathcal{U}_S$.

While theoretically convenient, noise maps affecting current quantum devices are more complex, and are more closely represented by the generic map

$$\mathcal{E}_q(\rho) = \sum_{j=0}^{4^n-1} E_{q,j}\rho E_{q,j} \tag{4.10}$$

for some Kraus operators $E_{q,j}$. In general $\mathcal{E}_q$ is not a Pauli map. However, this fact does not constitute a fundamental limitation for the following analysis. Indeed, by exploiting the Pauli twirling technique [25, 117, 39] Eq. (4.10) can be reduced to a Pauli map. Pauli twirling was already implemented efficiently on superconducting devices [110] in order to modify the noise behaviour, justifying our assumptions. Thus by combining VCEM with Pauli twirling, the conclusions drawn in this Section will apply under generic noise.

4.3.2 Analysis in the presence of Pauli noise

In the following, we study the properties of different noise models encompassing Pauli maps, going from the simplest model to a more realistic one.

General Pauli map at the end of the circuit

Within the first model, we assume that the effect of incoherent errors is entirely described by a Pauli map acting at the end of the circuit $\mathcal{U}_S$. This choice is depicted in Fig. 4.2. Such a class of noise models is especially convenient, as we have the following Lemma.

Lemma 1 (Action of Pauli maps on Pauli strings)**.** Given a Pauli map $\mathcal{P}$ and a Pauli string S, the action $\mathcal{P}^\dagger(S)$ is given by

$$\mathcal{P}^\dagger(S) = \chi S, \tag{4.11}$$

for some $\chi \in \mathbb{R}$, $|\chi| \leq 1$. In particular, $\chi = 1 - 2\Gamma$, where $\Gamma = \sum_{j:\{S,P_j\}=0} p_j$.

We refer the reafer to Appendix B.1 for a formal proof. Since the stabilizers S_i are Pauli strings, we can apply Lemma 1 to each term $C_i(\theta)$ in Eq. (4.8), getting

$$\tilde{C}_i^{(\mathcal{P})}(\boldsymbol{\theta}) = -\operatorname{Tr}\{S_i\mathcal{P}(\rho_S)\} = -\chi_i \operatorname{Tr}\{S_i\rho_S\} = \chi_i C_i(\boldsymbol{\theta}). \tag{4.12}$$

The resulting modified cost function is given by

$$\tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \chi_i C_i(\boldsymbol{\theta})\,. \tag{4.13}$$

This is especially convenient, as it allows to show that $\boldsymbol{\theta}^* = -\boldsymbol{\epsilon}$ is still a stationary point of $\tilde{C}^{(\mathcal{P})}$. This comes immediately from the fact that, for each C_i, $-\boldsymbol{\epsilon}$ is a global minimum, and hence

$$\partial_{\theta_k} C_i|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}} = 0, \;\Rightarrow\; \partial_{\theta_k} \tilde{C}^{(\mathcal{P})}|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}} = 0.$$

Furthermore, when dealing with Pauli maps, it is reasonable to assume that the probability associated to the identity operator $P_0 = \mathbb{1}$ is $p_0 > 1/2$. Indeed, the opposite describes a

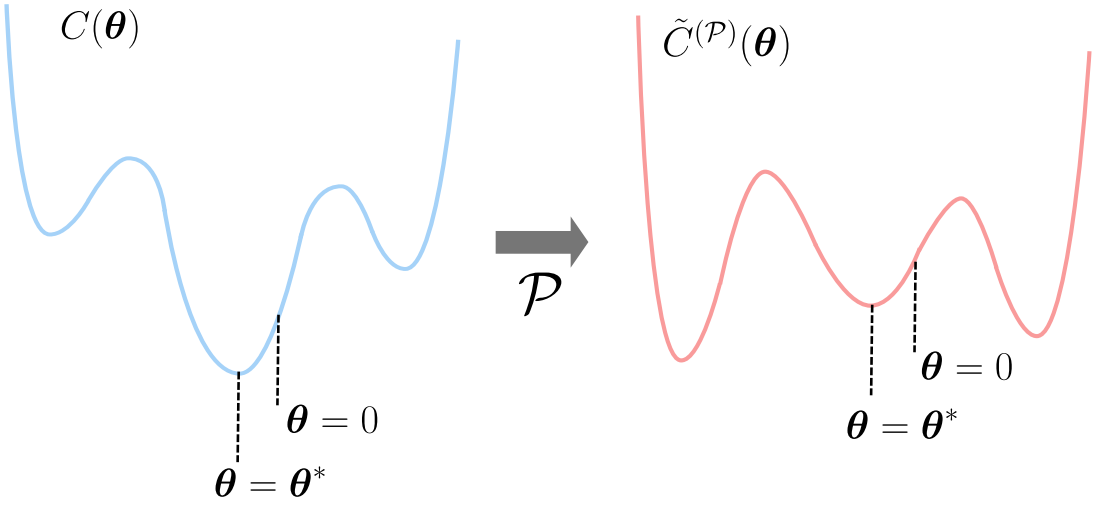

FIGURE 4.3: Sketch of how the landscape of the cost function could be modified by the Pauli incoherent noise $\mathcal{P}$ acting at the end of the circuit. The value of $\boldsymbol{\theta}$ in black is the initial choice of the parameters. While $\boldsymbol{\theta}^*$ is no longer the global minimum, it is always ensured to be a local minimum.

channel where probability of having errors would be higher than that of the ideal evolution, which is very far from a *weak* perturbation. Given this assumption, since Γ_i does not contain p_0, then $\Gamma_i < 1/2$ which implies that all terms $\chi_i > 0$. Thus, if $\tilde{\mathbf{H}}^{(\mathcal{P})}$ is the Hessian matrix of $\tilde{C}^{(\mathcal{P})}$, namely

$$\tilde{\mathbf{H}}^{(\mathcal{P})}_{kl} = \partial_{\theta_k}\partial_{\theta_l}\tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}} ,$$

and we let $\mathbf{H}_i$ denote the Hessian matrices of $C_i(\boldsymbol{\theta})$, we have that

$$\tilde{\mathbf{H}}^{(\mathcal{P})} = \sum_i \chi_i \mathbf{H}_i = \chi_{min}\mathbf{H} + \sum_i (\chi_i - \chi_{min})\mathbf{H}_i \,. \tag{4.14}$$

Each term in the last summation is positive semi-definite, since by definition $(\chi_i - \chi_{\min}) \geq 0$ and $\mathbf{H}_i \geq 0$. Furthermore, since $\mathbf{H} > 0$ is positive definite by construction, and $\chi_{\min} > 0$, it follows that $\tilde{\mathbf{H}}^{(\mathcal{P})} \geq \chi_{min}\mathbf{H} > 0$ is also positive definite. This ensures the convexity of the modified cost function in the region $\mathcal{B}(-\boldsymbol{\epsilon}, \delta)$. However, it's worth noticing that despite the fact that $\boldsymbol{\theta}^*$ is still a local minimum, is not guaranteed to be a *global* minimum. This is pictorially depicted in Fig. 4.3.

General Pauli map after each circuit moment

We now move to a more realistic scenario, where a general Pauli map is assumed to act not only at end, but also *in between* the circuit, i.e. after each circuit moment. Such maps will be identified by the moment index q, and denoted by $\mathcal{P}_q$. The corresponding noisy circuit reads

$$\mathcal{N} \equiv \prod_{q=1}^{M} \mathcal{P}_q \mathcal{U}_q = \mathcal{P}_M \mathcal{U}_M \ldots \mathcal{P}_2 \mathcal{U}_2 \mathcal{P}_1 \mathcal{U}_1 \ . \tag{4.15}$$

In order to proceed in this setting, we make the assumption that all the native gates $\mathcal{G}_w$ in the circuit are such that

$$\mathcal{G}_w = e^{-i(\phi_w + \theta_w + \epsilon_w)\mathcal{H}_w} = e^{-i(\theta_w + \epsilon_w)\mathcal{H}_w} e^{-i\phi_w \mathcal{H}_w} \equiv \mathcal{G}_w^{(\mathrm{N})} \mathcal{G}_w^{(\mathrm{C})} \ , \tag{4.16}$$

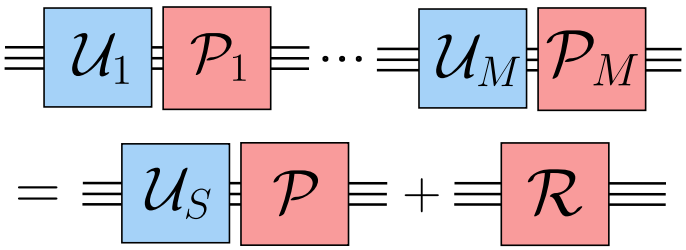

FIGURE 4.4: The action of local Pauli maps after each moment is equivalent to an effective Pauli map after $\mathcal{U}_S$ plus a reminder $\mathcal{R}$.

where each perturbation $\theta_w + \epsilon_w$ gives rise to the non-Clifford transformation $\mathcal{G}_w^{(\mathrm{N})}$ associated to coherent noise and each ϕ_w is chosen such that $\mathcal{G}_w^{(\mathrm{C})}$ is the ideal Clifford transformation. We now try to reduce the study for the new cost function, namely

$$\tilde{C}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \mathrm{Tr}\{S\mathcal{N}(\rho_0)\}$$

to the simpler case of Section 4.3.2, where stationarity of the desired solution and local convexity were already proven. We begin by considering the decomposition

$$\mathcal{N}(\rho_0) = \mathcal{P}\mathcal{U}_S(\rho_0) + \mathcal{R}(\rho_0) = \mathcal{P}(\rho_S) + \mathcal{R}(\rho_0)\,. \tag{4.17}$$

where $\mathcal{P}$ is an effective Pauli map acting at the end of the circuit and $\mathcal{R}$ denotes a remainder term. Assuming Eq. (4.16), an explicit expression for both terms can can be found in Appendix B.2, while a pictorial representation is given in Fig. 4.4. Employing Eq. (4.17), the modified cost function reads

$$\tilde{C}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \tilde{C}_i(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \mathrm{Tr}\{S_i\mathcal{N}(\rho_0)\} = \tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta}) - \sum_{i=1}^{n} \mathrm{Tr}\{S_i\mathcal{R}(\rho_0)\}\ , \tag{4.18}$$

where $\tilde{C}^{(\mathcal{P})}$ is the cost function in Eq. (4.13). In particular $\tilde{C}$ differs from $\tilde{C}^{(\mathcal{P})}$ only by a quantity $\Delta\tilde{C}(\boldsymbol{\theta})$ depending solely on the remainder term, namely

$$\Delta\tilde{C}(\boldsymbol{\theta}) \equiv \tilde{C}(\boldsymbol{\theta}) - \tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \mathrm{Tr}\{S_i\mathcal{R}(\rho_0)\}\,. \tag{4.19}$$

Given that the derivative of $\tilde{C}^{(\mathcal{P})}$ already vanishes at $\boldsymbol{\theta} = -\boldsymbol{\epsilon}$, stationarity of the desired solution can be recovered by studying the derivative of $\Delta\tilde{C}$. This is given in the following Theorem.

Theorem 4.3.1 (Stationarity of the solution)**.** Let $\mathcal{N}$ be a noisy quantum circuit affect by Pauli noise after each circuit moment as defined in Eq. (4.15). Furthermore, assume that each native gate in $\mathcal{U}$ has a parameterization consistent with Eq. (4.16). Then

$$\partial_{\theta_k}\Delta\tilde{C}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}} = 0 \ \ \forall k\,. \tag{4.20}$$

regardless of the noise strength.

A formal proof of Theorem 4.3.1 is given in Appendix B.3. Informally, the core of this result can be understood by saying that, while the action of noise can in principle completely disrupt our quantum state, in the Pauli noise setting, the searched for solution $\boldsymbol{\theta}^*$ will remain

stationary, which greatly simplifies the experimental realization of the VCEM protocol.

Another interesting property to estimate in this setting, are second derivatives of $\Delta\tilde{C}(\boldsymbol{\theta})$ in Eq. (4.19), which are related to the convexity of the cost function $\tilde{C}(\boldsymbol{\theta})$. In particular, in this Section we give some estimates of the Hessian matrix $\tilde{\mathbf{H}}^{(\mathcal{R})}$ computed at the stationary point $\boldsymbol{\theta} = -\boldsymbol{\epsilon}$, namely

$$\tilde{\mathbf{H}}^{(\mathcal{R})}_{kl} = \partial_{\theta_k}\partial_{\theta_l}\Delta\tilde{C}_{\theta}|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}} = \sum_{i=1}^{n} \partial_{\theta_k}\partial_{\theta_l} \operatorname{Tr}\{S_i\mathcal{R}(\rho_0)\}\Big|_{\boldsymbol{\theta}=-\boldsymbol{\epsilon}}. \tag{4.21}$$

Proving positive semi-definiteness of $\tilde{\mathbf{H}}^{(\mathcal{R})}$ in general is a complicated task. Instead of going through this path, we focus on weak Pauli noise, introducing the parameter $\eta \ll 1$ to measure its strength. In particular, we can define the parameter η by writing all Pauli maps inside $\mathcal{R}$ as

$$\mathcal{P}_q = e^{\eta\mathcal{L}_q}, \text{ where } \mathcal{L}_q(\rho) = \sum_k \mu_{qk}(P_{qk}\rho P_{qk} - \rho) \tag{4.22}$$

where μ_{qk} are adimensional numbers and $\mathcal{L}_q$ is the generator of the Pauli map. From this construction, it is clear that, as a function of η, the Hessian $\tilde{\mathbf{H}}^{(\mathcal{R})}$ is a continuous function, and as result, so are its eigenvalues. By defining $\mathbf{H}$ and $\tilde{\mathbf{H}}$ respectively the Hessian of the cost function C in Eq. (4.7) and of $\tilde{C}$ in Eq. (4.18) and by recalling that $\tilde{\mathbf{H}}^{(\mathcal{P})}$ is the Hessian of $\tilde{C}^{(\mathcal{P})}$ in Eq. (4.13), we can now make the following observation.

Observation 2 (Zero-noise limit)**.** In the limit of zero noise $\eta \to 0$, we have that $\tilde{\mathbf{H}}^{(\mathcal{P})} \to \mathbf{H}$, $\Delta\tilde{C} \to 0$ and hence $\tilde{\mathbf{H}}^{(\mathcal{R})} \to 0$, implying that in a neighborhood of $\eta = 0$, i.e. for *weak* noise, $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}^{(\mathcal{P})} + \tilde{\mathbf{H}}^{(\mathcal{P})} > 0$ is positive definite.

This is because at $\eta = 0$, all noise maps are trivial, i.e. $\mathcal{P}_q = \mathbb{1}$, then $\tilde{C}^{(\mathcal{P})} = C$ and $\Delta\tilde{C} = 0$. Since $\mathbf{H}$ is positive definite, all its eigenvalues $\lambda_k(\eta = 0)$ are strictly larger than zero. The argument is concluded by noting that, by continuity, eigenvalues $\lambda_k(\eta)$ of $\tilde{\mathbf{H}}$ are also strictly positive for small enough η. In other words, if the Pauli noise is weak enough, $\tilde{C}$ is locally convex near its stationary point, which hence is a minimum.

Approximation error upper bound

Since $\tilde{C}(\boldsymbol{\theta})$ and $\tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta})$ share some key properties for VCEM, it is natural to ask to what degree we can use one to approximate the other. Indeed, the latter is much simpler and easy to manipulate, and it would be much more advantageous to use in practical simulations than the former. This question is quantitatively answered in the following Theorem, by studying the scaling of $\Delta\tilde{C}$.

Theorem 4.3.2 (Upper bound on $\Delta\tilde{C}$)**.** Given a cost function $\tilde{C}(\boldsymbol{\theta})$ as defined in Eq. (4.15), and $\tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta})$ chosen such that $\mathcal{P}$ matches the first term in Eq. (4.17). Then the approximation error is upper bounded by

$$|\Delta C(\boldsymbol{\theta})| \in O(\epsilon^2 n^3 M^3), \tag{4.23}$$

where ϵ is the order of magnitude of each coherent error ϵ_w, and M denotes the number of moments in the circuit.

While this scaling is very unfavorable in terms of the number of qubits n, we can refine our approximation error upper bound by considering only *local* native gates and fixed number of moments $M \in O(1)$. With this very reasonable assumptions, Theorem 4.3.2 simplifies in the following Corollary.

Corollary 4.3.1 (Upper bound for fixed depth circuits)**.** Assuming only *local* native gates and fixed number of moments $M \in O(1)$, the scaling of Theorem 4.3.2 can be improved to

$$|\Delta C(\boldsymbol{\theta})| \leq O(\epsilon^2 n) \tag{4.24}$$

thus $|\Delta C(\boldsymbol{\theta})|$ scales linearly with n in this case. A formal proof of both Theorem 4.3.2 and Corollary 4.3.1 is given in Appendix B.4. On a final note, we observe that, as both $\tilde{C}(\boldsymbol{\theta})$ and $\tilde{C}^{(\mathcal{P})}(\boldsymbol{\theta})$ are composed as sum of n terms, then $|\tilde{C}(\boldsymbol{\theta})| \sim n$ for fixed noise strength. Combined with Corollary 4.3.1, this gives a constant upper bound on the relative error

$$\left| \frac{\Delta C(\boldsymbol{\theta})}{\tilde{C}(\boldsymbol{\theta})} \right| \in O(\epsilon^2),$$

which does not depend on n. It follows that the case of general Pauli maps after each circuit moment can be essentially reduced to the one considered in Section 4.3.2 even for approximate circuit simulations when coherent errors are small enough.

This can be useful considering that two of the main quantum algorithm subroutines based on the preparation of stabilizer states are the preparation of graph states [42, 41, 103] and GHZ states [20, 22], which generally employ fixed depth circuits. For example, 2D/3D graph states are a foundamental resource in the framework of Measurement Based Quantum Computation (MBQC) [11, 89]. Respectively for 2D/3D square/cubic graph states, the optimal number of the moments in which the circuit can be divided, without transpiling the latter into native gates, is $M' = 4$ and $M' = 6$. These are constant numbers independent from the number of nodes, i.e. of qubits, and this is a general property of 2D/3D graph states and GHZ states. Furthermore, when the circuit is transpiled into native gates the resulting number of moments M might increase as $M = \nu M'$, by only by a constant amount, i.e. ν is again a constant. These, are practical examples in which the scaling in Corollary 4.3.1 is valid.

4.4 Numerical results

In the following we show numerically VCEM effectiveness in realistic scenarios. All the simulations are performed with the software library Pennylane [113]. We focus in particular on graph states preparation.

Given a graph with n nodes and by calling $n(j)$ the set of all neighbors of node j, the graph state stabilizers are given by

$$G_i = X_i \otimes \left(\bigotimes_{j \in n(i)} Z_j \right) \; i = 1, \ldots, n \; . \tag{4.25}$$

The stabilizer state associated to Eq. (4.25), which is named graph state, is $|\psi_G\rangle = U_G |0\rangle^{\otimes n}$ with the Clifford circuit U_G given by

$$U_G = \left(\prod_{<i,j> \in E} CZ_{ij} \right) \mathrm{H}^{\otimes n} \; , \tag{4.26}$$

where E is the set of edges of the graph, CZ and H are respectively controlled-Z and Hadamard gates. Here we consider a 2D rectangular graph with $n = 10$ and sides of length 2 and 5. We

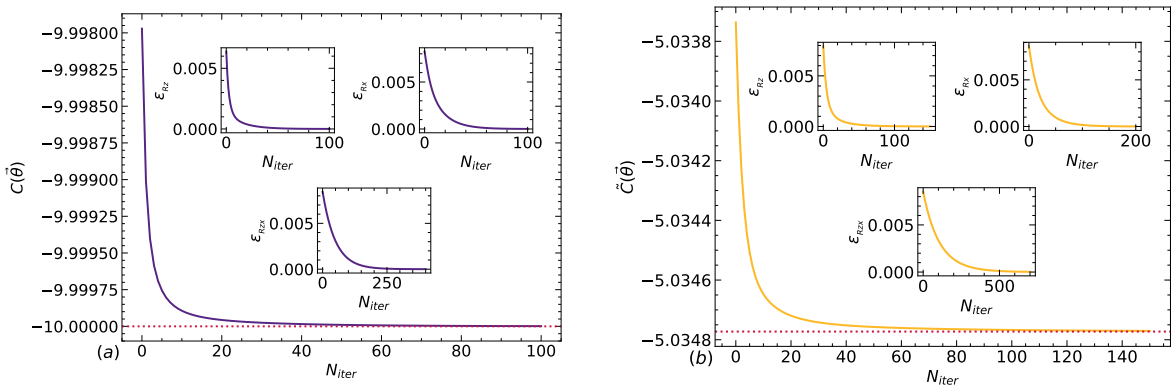

FIGURE 4.5: 10 qubits 2D-rectangular graph state cost function optimization. The left plot shows the value of the cost function $C(\boldsymbol{\theta})$ with coherent errors only as a function of the number of optimization steps N_{iter}. The red dashed line is the minimal value $C(\boldsymbol{\theta}_{opt}) = -10$. The insets represent the behaviour of ε_{R_z}, ε_{R_x} and $\varepsilon_{R_{zx}}$ defined in the main text. The right plot shows the cost function $\tilde{C}(\boldsymbol{\theta})$ in presence of local Pauli incoherent errors after each circuit momentum and the red dashed line is the minimal value $\tilde{C}(\boldsymbol{\theta}_{opt}) = -5.0347723$. The insets in panel (b) have the same meaning as for panel (a).

choose the following native gates basis

$$R_x(\phi) = e^{-i\frac{\phi}{2}X}, \quad R_z(\phi) = e^{-i\frac{\phi}{2}Z}, \quad \text{and} \quad R_{zx}(\phi) = e^{-i\frac{\phi}{2}Z\otimes X}; \tag{4.27}$$

where X, Z are respectively the x, z Pauli gates. In the following simulations we assume that given a qubit, the same single-qubit native gate acting on it has the same random coherent error and correspondingly the same parameter. Analogously, given a couple of qubits each R_{zx} gate acting on it has the same coherent error and parameter. A pictorial representation of the transpilation of the Hadamard and the CZ gates with coherent errors and parameters is provided in Appendix B.5.

As a first test, we optimize the cost function in the case of coherent errors only. We sample random coherent errors in the interval $[-0.01, 0.01]$. On the left plot of Fig. 4.5, we show the value of $C(\boldsymbol{\theta})$ as a function of the number of optimization steps N_{iter}. The red dashed line is the minimal value $C(\boldsymbol{\theta}^*) = -10$. The insets represent the behaviour of $\varepsilon_l = \|\boldsymbol{\theta}_l + \boldsymbol{\epsilon}_l\|_2$ with $l \in \{R_z, R_x, R_{zx}\}$. Here, $\boldsymbol{\theta}_l$ denotes the vector containing all the parameters associated with the gate l, while $\boldsymbol{\epsilon}_l$ collects the corresponding coherent errors. For the single-qubit rotations R_z and R_x, these vectors include parameters and errors defined per qubit, whereas for the two-qubit gate R_{zx} they are defined over pairs of qubits. As shown, each ε_l converges to zero with increasing number of optimization steps.

Secondly we study the modification introduced by incoherent errors by applying m-local Pauli maps after each circuit momentum with $m = 1, 2$. Analogously, we sample random incoherent errors in the interval $[-0.01, 0.01]$, while coherent ones are the same of the previous simulation. On the right plot of Fig. 4.5 we show the behaviour of $\tilde{C}(\boldsymbol{\theta})$ and the red dashed line is the minimal value $\tilde{C}(\boldsymbol{\theta}^*) = -5.0347723$. The insets have the same meaning in both panels.

Moreover, our simulations confirm the scaling of the upper bound in Eq. (4.23) as we show in Fig. 4.6. We simulated a 1D-linear graph state with 10 qubits and computed $\Delta\tilde{C}(0)$ by varying ϵ. Then, we performed a linear fit using a log-log scale which resulted in a slope $a = 1.97 \approx 2$. This indicates that $\Delta\tilde{C}$ grows quadratically with ϵ as expected. Additionally we computed $\Delta\tilde{C}(0)$ for the same graph state by increasing the number of qubits n from 9 to 12. The data follow a linear trend, confirming the expected linear dependence on n.

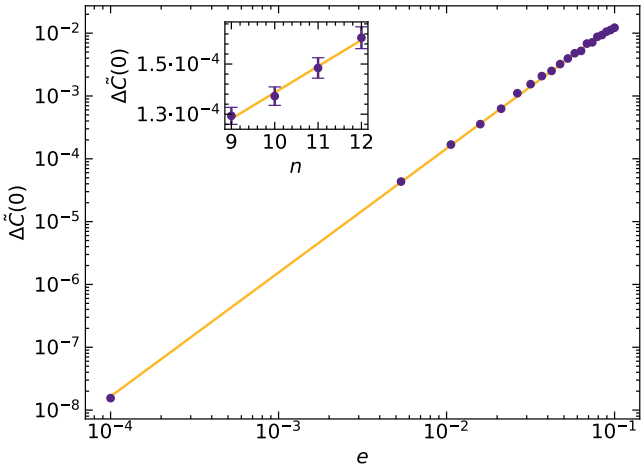

FIGURE 4.6: $\Delta\tilde{C}\boldsymbol{\theta}$) evaluated in $\boldsymbol{\theta} = 0$ for 10 qubits 1D-linear graph state. The main plot shows the linear fit of $\Delta\tilde{C}(0)$ as a function of ϵ in loglog scale. The angular coefficient of the yellow line is $a = 1.97 \simeq 2$ confirming the quadratic scaling of $\Delta\tilde{C}$ in terms of ϵ. The inset shows the linear fit of $\Delta\tilde{C}(0)$ as a function of the number of qubits n. The linear behaviour is again satisfied.

4.5 Discussion

This Chapter introduced the Variational Coherent Error Mitigation (VCEM) framework, which effectively suppresses coherent errors through stabilizer-based variational optimization. VCEM operates directly on native gate parameters and remains robust even under realistic incoherent noise. Numerical results confirm its scalability and accuracy on current quantum hardware with reasonable noise assumptions. Further exploration of these kind of techniques, and their practical implementation, may open promising avenues to improve VCEM effectiveness enabling robust coherent error mitigation even with large-scale noisy stabilizer states.

The techincal points raised in this Chapter are also very interesting, as they suggest that, for certain classes of variational ansätze, the *location* of the minimum $\boldsymbol{\theta}^*$ rather than the optimal value of the cost function $C(\boldsymbol{\theta})$ could be better exploited, as it is an intrinsically *noise-resistant* property of the model. Future work will include exploring whether similar results can be achieved for other classes of circuits, and their potential for other domains of application, such as quantum machine learning [10].

Chapter 5

Scaling Variational Quantum Algorithms

Variational quantum computing offers a powerful framework with applications across diverse fields such as quantum chemistry, machine learning, and optimization. However, its scalability is hindered by the exponential concentration of the loss function in the number n of qubits, known as the barren plateau (BP) problem [67, 65]. More precisely, in the presence of BP, we have an exponentially vanishing probability of being able to efficiently find a loss-minimizing direction after parameter initialization, making the optimization loop typical of VQAs unfeasible. In this Chapter, we formally introduce this effect and give an overview of the main results in the literature on the subject.

5.1 The cost concentration phenomenon

Given a method for computing gradients directly on quantum hardware, the next natural question concerns the *required precision* of these estimations. In classical optimization, gradient-based methods are typically robust to small relative errors in the derivatives. However, in the quantum setting, the situation is more delicate: since all observables must be estimated statistically from a finite number of measurements (shots), it is the *absolute precision* that plays the dominant role in determining optimization performance. Indeed, as discussed in Chapter 1, when expectation values such as the cost fuction

$$C(\boldsymbol{\theta}) = \mathrm{Tr}\{\mathcal{U}_{\boldsymbol{\theta}}(\rho)H\} \tag{5.1}$$

are estimated via N_s independent projective measurements, the central limit theorem guarantees that the standard error of this estimator scales as $\epsilon_{\text{stat}} \propto 1/\sqrt{N_s}$. Hence, only a *polynomial precision* can be obtained with polynomial resources (i.e. shots). This scaling directly limits the precision attainable in gradient estimation, even assuming ideal quantum operations.

In this setting, a fundamental limitation emerges as the ansatz structure becomes increasingly expressive or deep. Indeed, it has been widely observed that the gradients of variational cost functions tend to *vanish exponentially* with system size or circuit depth, a phenomenon known as the *barren plateau* (BP) problem, pictorially represented in Fig. 5.1.

This effect can be intuitively understood as a manifestation of the *curse of dimensionality*, where the exponential growth of the Hilbert space $\mathcal{H}$ turns from a computational advantage into a severe obstacle. To illustrate this intuition, consider a typical variational cost function of the form of Eq. (5.1). Exploiting the tools form Chapter 1, it can be seen as a scalar product between the exponentially large vectors H and $\rho_{\boldsymbol{\theta}} = \mathcal{U}_{\boldsymbol{\theta}}(\rho)$, namely the Hilbert-Schmidt scalar product

$$\langle \rho_{\boldsymbol{\theta}}, H\rangle_{HS} = \mathrm{Tr}\{\mathcal{U}_{\boldsymbol{\theta}}(\rho)H\} \tag{5.2}$$

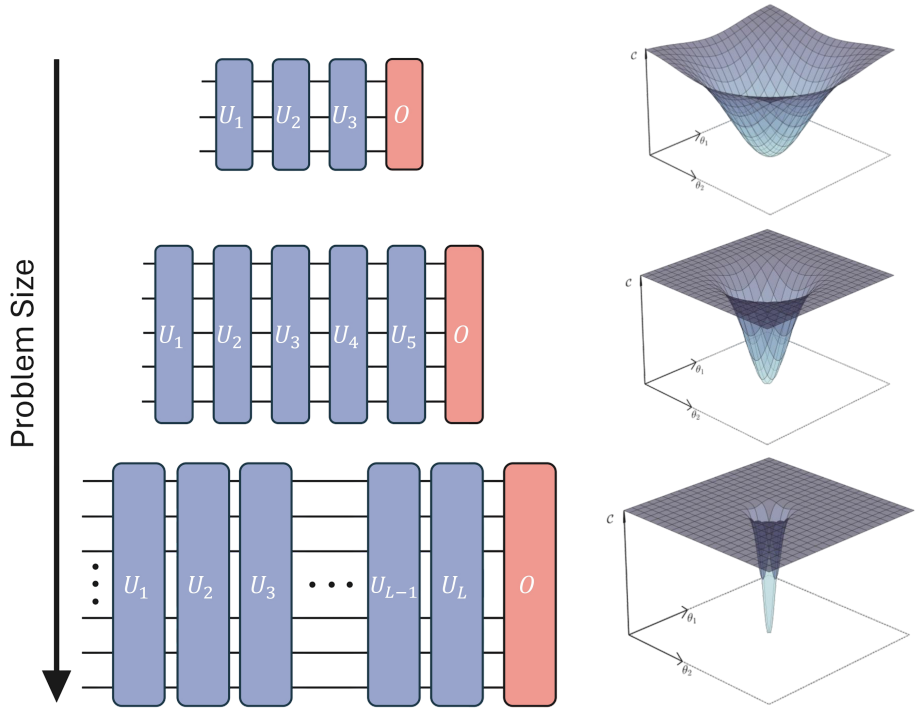

FIGURE 5.1: Pictorial representation of the barren plateau issue. As the the number of qubits becomes larger and the ansatz becomes deeper, a general flattening of the cost function appears.

defined over the space $\mathcal{B}$ of linear bounded operators acting on $\mathcal{H}$.

If we now assume that, by randomly sampling the parameters $\boldsymbol{\theta}$, the resulting states $\rho_{\boldsymbol{\theta}}$ will distribute approximately uniformly in the space of density matrices, then the typical overlap between $\rho_{\boldsymbol{\theta}}$ and H will be exponentially small in the number of qubits, scaling inversely proportional to the dimension of $\mathcal{B}$. This is because the large dimension of this vector space makes it very hard to *guess* the sole preferred direction by the problem, which is the one corresponding to H, and consequently, by uniformly guessing a trial state, we are bound to fail most of the time. In this case, the cost function values concentrate exponentially close to their mean, and the gradients, being first derivatives of such concentrated quantities, also exhibit exponentially small fluctuations around zero.

More formally, barren plateaus can be characterized by an *exponential decay of the gradient variance*. In particular we use the following Definition.

Definition 5.1.1 (Barren plateau)**.** Given an observable H and an initial state ρ, the ansatz $\mathcal{U}_{\boldsymbol{\theta}}$ is said to suffer from a barren plateau if the variance $\mathbb{V}_{\rho,H}$ of the derivative of its cost function satisfy

$$\mathbb{E}_{\boldsymbol{\theta}}\left\{(\partial_{\theta_i} C)^2\right\} - \mathbb{E}_{\boldsymbol{\theta}}\left\{\partial_{\theta_i} C\right\}^2 = \mathbb{V}_{\rho,H}\left\{\partial_{\theta_i} C\right\} \in O\left(e^{-\alpha n}\right) \;\; \forall \theta_i \in \boldsymbol{\theta}, \tag{5.3}$$

where n is the number of qubits and $\alpha > 0$ is constant.

This exponential suppression of gradient variance directly translates to an optimization landscape that is effectively flat for all practical purposes. To make this explicit, let's consider the class of ansätze that satisfy $\mathbb{E}_{\boldsymbol{\theta}}\left\{\partial_{\theta_i} C\right\} = 0$. This is indeed a very general assumption, since simply using uncorrelated Pauli rotation gates, which are ubiquitous in VQA ansätze

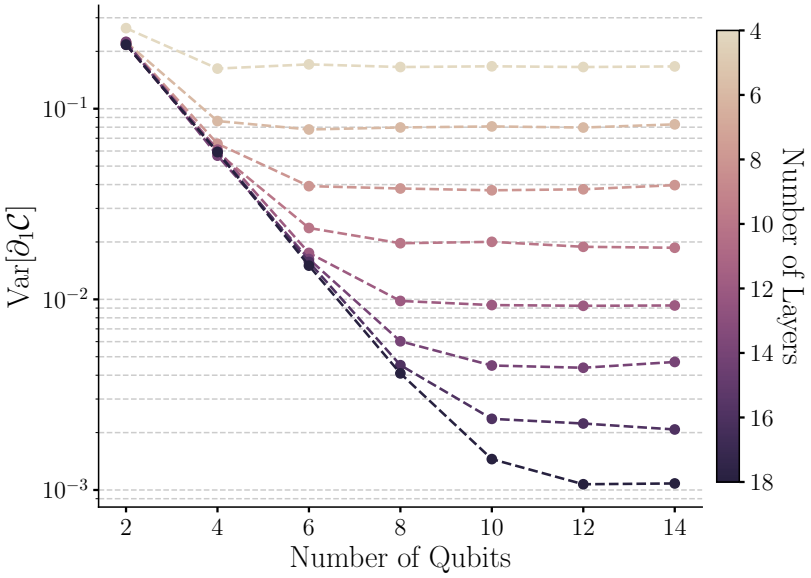

FIGURE 5.2: Example of the exponential decay of the gradients with increasing number of qubits and layers.

due to their practicality, ensures this. Now, considering Chebyshev inequality, we can rigorously bound the probability of obtaining a sizeable gradient. In particular, we have

$$\Pr\left\{|\partial_{\theta_i} C| \geq k\sigma_i\right\} \leq \frac{1}{k^2}. \tag{5.4}$$

where $\sigma_i^2 = \mathbb{V}_{\rho,H}\left\{\partial_{\theta_i} C\right\} \sim e^{-\alpha n}$. This implies that, gradient magnitudes lower-bounded by a polynomially vanishing quantity $k\sigma_i$, require $k^2 \sim e^{\alpha n}$, meaning that the probability of such an event becomes exponentially small. Then, with overwhelming probability, any polynomially small estimation error ϵ_{stat} introduced by shot noise will dominate the exponentially vanishing gradient signal. As a result, even with a perfect gradient estimation method, the optimizer receives effectively no information about the descent direction, rendering the optimization process intractable. This effect can be numerically estimated by measuring the gradient variance for an increasing number of qubits, as shown in Fig. 5.2.

It is worth noting that the definition above, centered on the mean and variance of the gradients, can be equivalently formulated substituting the gradients with cost function itself. In such cases, we will more precisely speak of *cost concentration*.

Definition 5.1.2 (Cost concentration)**.** Given an observable H, an initial state ρ, and an $\mathcal{U}_{\boldsymbol{\theta}}$, its cost function of Eq. (5.1) is said to be exponentially concentrated if the variance $\mathbb{V}_{\rho,H}$ of its cost function satisfies

$$\mathbb{E}_{\boldsymbol{\theta}}\left\{C^2\right\} - \mathbb{E}_{\boldsymbol{\theta}}\left\{C\right\}^2 = \mathbb{V}_{\rho,H}\left\{C\right\} \in O\left(e^{-\alpha n}\right), \tag{5.5}$$

where n is the number of qubits and $\alpha > 0$ is constant.

Indeed, several works have formally established the correspondence between the concentration of the cost function values and the concentration of its gradients [46]. In the following Sections, we will lose this distinction, and exploit this equivalence to analyze expressivity and optimization properties within specific variational frameworks.

Determining whether a given ansatz will exhibit such concentration behavior is a highly nontrivial problem. The presence or absence of barren plateaus has been shown to depend

on multiple interrelated factors, including the expressivity of the ansatz [57, 81, 95, 46, 84, 17], the choice of initial state [107], circuit depth [16], entanglement structure [46], and even the locality of the cost function [16]. Recently, these diverse phenomena have been unified under a *Lie-algebraic framework* [88, 33], which characterizes the expressivity and trainability of variational ansatzes in terms of the algebra generated by their parameterized operators. This framework, which we shall briefly introduce in the next Section, provides a rigorous means to predict whether gradient concentration will occur in the unitary setting.

Furthermore, noise introduces an additional source of concentration. Since most physically relevant noise channels are strictly contractive, they tend to suppress distinguishability between quantum states, effectively flattening the optimization landscape as well. This *noise-induced barren plateau* phenomenon [82, 119, 62] will also be discussed in later Sections of this Chapter.

Finally, to simplify the notation, we will drop the explicit cost function dependence of the variance, and use $\mathbb{V}_{\rho,H}$ instead.

5.2 Unitary Barren Plateaus

While the emergence of cost concentration in VQAs can be linked to many factors, recent works [88, 33] have underlined that the fundamental notions of *expressivity* and *alignment* suffice to explain most effects. In particular, the Lie-algebraic framework they introduce, provides a natural and mathematically precise setting in which these concepts can be formalized, offering a unified description of when and why concentration phenomena arise in the unitary setting.

Consider a layered ansatz composed of L layers, where each layer is made of the composition of unitary gates of the form

$$U_l(\theta_l) = \prod_{k=1}^{K} e^{-i\theta_{l,k}H_k}, \tag{5.6}$$

where $\{H_k\}_{k=1}^K$ are fixed Hermitian generators acting on n qubits. Since in general they do not commute, the generator of the product $U_l(\theta_l)$ cannot be simply described by a linear combination of H_k. Instead, composing these gates gives rise to a series of nested commutators according to the Baker–Campbell–Hausdorff formula, i.e.

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] - \frac{1}{12}[B,[A,B]] + \cdots\right). \tag{5.7}$$

This implies that the effective generator of the circuit requires an intricate combination of commutators between the H_k as well. To capture this structure, it is convenient to introduce the notion of the dynamical Lie algebra (DLA), denoted by $\mathfrak{g}$. It is formally defined as the *Lie closure* of the set of the generators, namely

$$\mathfrak{g} = \langle \{iH_k\}_{k=1}^K \rangle_{\text{Lie}} = \text{span}\left\{iH_k, [iH_k, iH_{k'}], [iH_k, [iH_{k'}, iH_{k''}]], \ldots\right\}. \tag{5.8}$$

With this construction, the Lie group $G = e^{\mathfrak{g}}$ associated to the DLA, which collects all unitaries that can be generated by exponentiating elements of $\mathfrak{g}$, precisely characterizes the unitaries implementable by the ansatz, i.e. $U_{\boldsymbol{\theta}} \in G \ \forall \boldsymbol{\theta}$. In other words, we have the inclusion

$$G_\Theta = \{U_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\} \subseteq G. \tag{5.9}$$

In this sense, the *dimension* of $\mathfrak{g}$ provides a natural quantitative measure of expressivity: a large DLA allows exploration of a larger portion of the Hilbert space, up to full $\mathfrak{su}(2^n)$ for

maximally expressive ansätze, and vice versa a small DLA severely restricts the unitaries implementable by the model.

A very useful concept in this framework is that of a t-design over G.

Definition 5.2.1 (t-design over G)**.** Let $G \subseteq \mathbb{U}(2^n)$ be a Lie group, and let $G_\Theta \subseteq G$ be an ensable of unitaries parameterized by some parameter vector $\boldsymbol{\theta} \in \Theta$. Then G_Θ is said to form a t-design over G if

$$\int_\Theta d\boldsymbol{\theta}\, U_{\boldsymbol{\theta}}^{\otimes t} \otimes U_{\boldsymbol{\theta}}^{*\otimes t} = \int_{V\in G} d\mu(V)\, V^{\otimes t} \otimes V^{*\otimes t} \tag{5.10}$$

where μ denotes the uniform, or *Haar* measure over the group G.

When the t-design property holds, all polynomials of matrix elements up to degree t have the same expectation value when averaged with both measures. This capture the idea of the Haar measure on G and integration over G_Θ as being indistinguishable for the first t moments. When the group G used is the full unitary group $\mathbb{U}(2^n)$, we call this ensemble a *unitary* t-design.

Indeed, as one composes many layers, i.e., as L increases, it can be shown that the distribution of unitaries $U_{\boldsymbol{\theta}}$ induced by random parameter choices becomes increasingly uniform over the group G. More precisely, G_Θ approaches a 2-design over G.

Theorem 5.2.1 (Convergence to a 2-design over G [88])**.** Moments of the uniform distribution over the ensemble G_Θ of Eq. (5.9) converge to Haar values exponentially fast in the number of layers, namely

$$\left|\mathbb{E}_{\boldsymbol{\theta}}\left\{\mathrm{Tr}\{\mathcal{U}_{\boldsymbol{\theta}}(\rho)\}^t H\right\} - \mathbb{E}_\mu\left\{\mathrm{Tr}\{\mathcal{V}(\rho)\}^t H\right\}\right| \in O(e^{-\beta L}) \tag{5.11}$$

for $t \leq 2$, and for any ρ and H.

In this regime, often referred to as the *deep circuit limit*, expectation values over random parameters converge to those computed under the Haar measure, allowing for analytical calculation of the limiting variance $\mathbb{V}^\infty_{\rho,H}$ using tools from Weingarten calculus. In particular, this can be done by decomposing the Lie algebra into a direct sum of mutually commuting ideals,

$$\mathfrak{g} = \bigoplus_{\lambda=0}^{k} \mathfrak{g}_\lambda, \tag{5.12}$$

where $\mathfrak{g}_0$ is the center of $\mathfrak{g}$, and $\mathfrak{g}_{\lambda>0}$ are simple Lie algebras [88]. Given this decomposition, any operator $A \in \mathcal{B}$ acting on system can be decomposed as $A = \sum_\lambda A_\lambda$, where A_λ denotes the component of A supported on the subalgebra $\mathfrak{g}_\lambda$. This decomposition allows one to define a *purity measure* $\mathcal{P}_\lambda(A)$ for the observable A, which can be formally, defined as the 2-norm of A_λ. When applied to the initial state ρ and the observable H, the quantities $\mathcal{P}_\lambda(\rho)$ and $\mathcal{P}_\lambda(H)$ quantify how much of the initial state and the observable, respectively, lie within each subalgebra component.

Using these definitions, one can prove the following Theorem.

Theorem 5.2.2 (Deep circuit variance [88, 33])**.** The deep circuit variance $\mathbb{V}^\infty_{\rho,H}$ of a layered ansatz of type Eq. (5.6) is given by

$$\mathbb{V}^\infty_{\rho,H} = \sum_{\lambda>0} \frac{\mathcal{P}_\lambda(\rho)\,\mathcal{P}_\lambda(H)}{d_\lambda}, \tag{5.13}$$

where $d_\lambda = \dim(\mathfrak{g}_\lambda)$.

Equation Eq. (5.13) makes explicit the interplay between *expressivity* (through the subalgebra dimensions) and *alignment* (through the purities). Large expressivity, corresponding

to large d_λ, tends to suppress the variance and thus induces cost concentration. However, even if the expressivity is limited, concentration may still occur if the observable and the initial state are poorly aligned, i.e., if they share little overlap within the same subalgebra components, leading to small purities products.

Finally, while this Lie-algebraic model provides a rigorous and general framework to quantify cost concentration in the deep circuit limit, it relies heavily on the group-theoretic properties of G. Such a structure is no longer available when noise acts on the system, as quantum channels are generally non-unitary and do not form a compact group. In the following Section, we extend this analysis to noisy settings, showing how contractivity can lead in the worst case to an analogous suppression of variances.

5.3 Noise-induced Barren Plateaus

In the presence of noise, cost concentration is still expected to occur, but it takes a qualitatively different form compared to the purely unitary case. While the unitary barren plateau phenomenon manifests as a *probabilistic concentration*, i.e., gradients and costs become exponentially unlikely to deviate from their mean, noise-induced barren plateaus (NIBPs) often exhibits a *deterministic* character. In this regime, one can bound the magnitude of the cost function directly, regardless of the parameter values. This stronger bound originates from the *contractivity properties* of the quantum channels used to describe the noise, as we describe in the following.

5.3.1 Local Pauli Noise

We illustrate the main idea behind the emergence of noise-induced concentration for unital noise using the example of *local Pauli noise*, a widely studied and analytically tractable class of quantum channels. In particular, we assume to have a noisy layered ansatz composed of L layers, and whose noise channel are given by

$$\mathcal{E} = \bigotimes_{i=1}^{n} \mathcal{E}_i \quad \text{where} \quad \mathcal{E}_i(\rho) = \sum_{P \in \{\mathbb{1}, X, Y, Z\}} p_{P_i} P_i \rho P_i \tag{5.14}$$

are Pauli channels acting on the i-th qubit and the probabilities $\{p_P\}$ form a normalized distribution. If all nontrivial Pauli errors have strictly positive probability, i.e. $p_{P_i} > 0$ for all $P_i \neq \mathbb{1}$, the map is both unital and strictly contractive. This follows immediately by noting that, given a local Pauli Q_i, then $\mathcal{E}_i(Q_i) = \eta_{Q_i} Q_i$, where

$$\eta_{Q_i} = p_{\mathbb{1}_i} + p_{Q_i} - \sum_{P \notin \{\mathbb{1}, Q\}} p_{Q_i} < 1$$

by normalization. Working in the 2-norm, the channel's induced operator norm satisfies

$$\|\mathcal{E}_i\|_{2\to 2} = \max_{P \neq \mathbb{1}} |\eta_{P_i}| < 1, \tag{5.15}$$

and consequently, $\|\mathcal{E}\|_{2\to 2} = \max_i \|\mathcal{E}_i\|_{2\to 2} < 1$. No assumption on the unitary layers are needed, as all unitary trasformation preserve this norm, i.e. $\|\mathcal{U}\|_{2\to 2} = 1 \ \forall \mathcal{U}$. This allows direct application of Hölder's inequality bound the cost function absolute value to an exponentially decaying quantity in the number L of layers. In particular we have

$$|C(\boldsymbol{\theta})| = |\mathrm{Tr}\{H\, \mathcal{E}\mathcal{U}_L \cdots \mathcal{E}\mathcal{U}_1(\rho)\}| \leq \|H\|_2 \, \|\rho\|_2 \, \|\mathcal{E}\|_{2\to 2}^{L}, \tag{5.16}$$

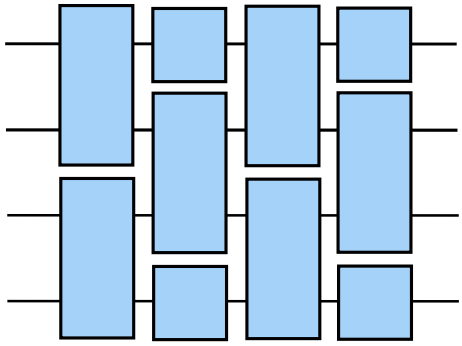

FIGURE 5.3: Structure of the brickwall circuit. Each blue box represent a unitary 2-design acting on two qubits.

where we used that unitary layers preserve the 2-norm. Since $\|\mathcal{E}\|_{2\to 2} < 1$, the right-hand side decays exponentially with the number of noisy layers L, namely

$$|C(\boldsymbol{\theta})| \in O(e^{-\beta L}) \tag{5.17}$$

If the circuit depth L scales linearly with the number of qubits, the cost function is thus exponentially suppressed in system size, leading to a deterministic form of concentration even without any probabilistic averaging over parameters.

This phenomenon can be intuitively understood by noting that unital and strictly contractive noise maps drive all states toward the maximally mixed state, which is invariant under both the noisy and unitary layers. Consequently, the composite dynamics has the maximally mixed state as a global fixed point, and after sufficiently many layers, the system converges to it. In this limit, the output state becomes independent of the parameters, rendering the cost function effectively constant.

In contrast, when the noise is *non-unital*, the channel admits nontrivial fixed points, and such uniform convergence no longer occurs. This leads to qualitatively different behavior, as we discuss next Section.

5.3.2 Non-Unital Noise

In the presence of non-unital noise, the situation becomes more intricate. Non-unital channels do not preserve the maximally mixed state, and as a result, deterministic concentration is generally lost. Indeed, certain constructions exist that exploit non-unital noise to counteract decoherence effects. For instance, in the so-called *quantum refrigerator* schemes [8], non-unital channels can be used to implement error correcting circuits. As a consquence, in this setting, noisy cost functions can sustain large values, just as in purely unitary settings.

Quantitative estimates of cost variance in this regime are difficult to obtain, but progress has been made for specific architectures. In particular, for *brickwall circuits*, in which local unitary 2-design blocks are interleaved with single-qubit non-unital noise, it has been shown that the variance of the cost function may remain bounded away from zero. An example of the brickwall structure is pictorially represented in Fig. 5.3.

To formalize this, consider the Pauli decomposition of H, namely $H = \sum_P h_P P$. Furthermore, let the noise channel take the form

$$\mathcal{E} = \mathcal{E}_1^{\otimes n}, \tag{5.18}$$

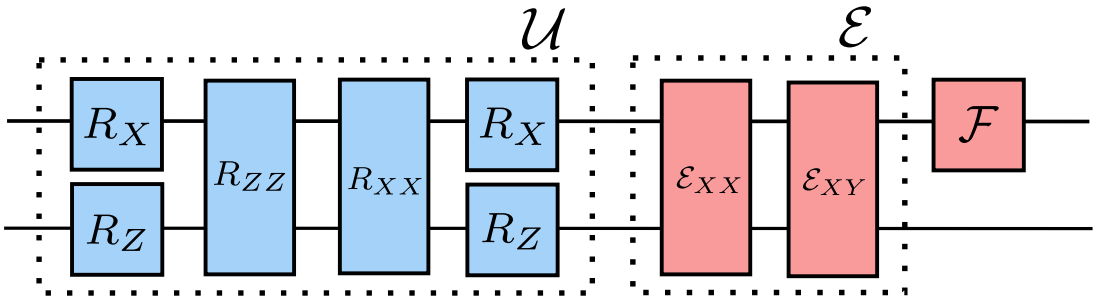

FIGURE 5.4: Single layer structure used in the numerical examples. The circuit acts on a two-qubit register and is divided into three blocks, each representing either unitary gates ($\mathcal{U}$) or noise channels ($\mathcal{E}$ and $\mathcal{F}$).

where $\mathcal{E}_1$ is a single qubit, non-unital noise map Then we have the following Theorem.

Theorem 5.3.1 (Non-unital, single qubit noise [5])**.** Given a noisy brickwall circuit, composed of L layers, intial state ρ and observable H, where each noise map satisfy Eq. (5.18). Then the variance $\mathbb{V}^L_{\rho,H}$ reads

$$\mathbb{V}^L_{\rho,H} \in \sum_P h_P \, e^{-\boldsymbol{\theta}(|P|)}, \tag{5.19}$$

where $|P|$ denotes the Hamming weight of P.

Crucially, this result implies that for local observables, i.e. those involving only few-body operators, the terms with small Hamming weight dominate the sum, allowing a non-vanishing variance even in the deep circuit limit. Thus, while global observables may still produce cost functions prone to concentration, local observables can be used to retain sensitivity to the parameters and avoid barren plateaus.

It should be emphasized, however, that this preservation of variance relies on two key conditions, namely the presence of *single-qubit* non-unital noise and the specific *brickwall* structure of the circuit. Indeed, when either of these conditions is relaxed, e.g., by introducing correlated or multi-qubit noise channels, more intricate phenomena can arise.

5.4 Restricted noise models are not enough

Although certain results in the literature may be misconstrued as implying that unital, non-unitary channels inevitably induce deterministic concentration in deep circuits [82, 119], this interpretation does not hold in general without strict contractivity. Indeed, here we provide two examples where unital, but not strictly contractive, noise is either neutral or beneficial for the scaling of the variance.

Commuting noise and unitary layers

A simple yet compelling counterexample involves commuting noise and unitary layers, i.e., when $\mathcal{E}_l \circ \mathcal{U}_{l'}(\rho) = \mathcal{U}_{l'} \circ \mathcal{E}_l(\rho)$ for all ρ and all l, l'. In such cases, one can commute all noisy operations to the beginning of the computation, effectively reducing the noisy circuit to a fully unitary one, provided the initial state is redefined as $\tilde{\rho} = \mathcal{E}_L \circ \cdots \circ \mathcal{E}_1(\rho)$. If ρ is a steady state of all intermediate channels $\mathcal{E}_l$, then $\tilde{\rho} = \rho$, and the entire circuit becomes noise-free. Consequently, it cannot exhibit NIBP, although BP may still occur.

An explicit example is given by the circuit depicted in Fig. 5.4, which features a unitary layer $\mathcal{U}$, combined with a fixed noise channel $\mathcal{E}_l = \mathcal{E} = \mathcal{E}_{XX} \circ \mathcal{E}_{XY}$ $\forall l$. Specifically, $\mathcal{E}$ is defined

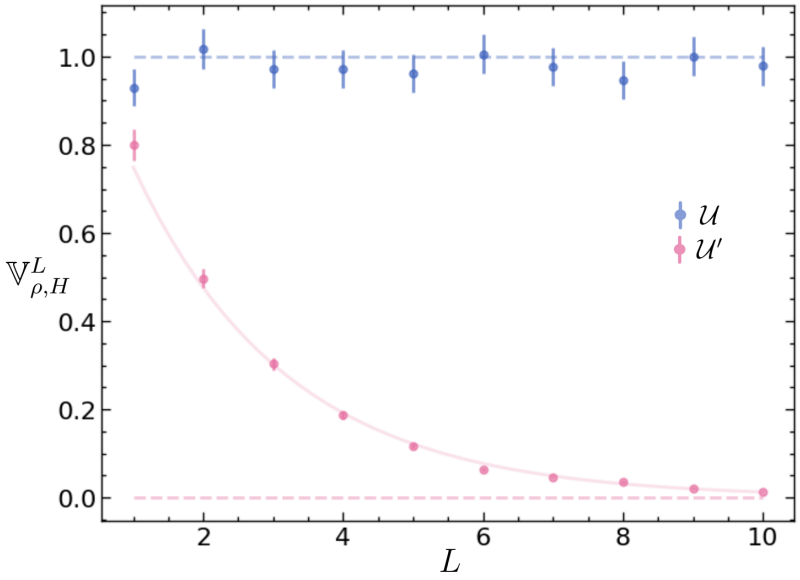

FIGURE 5.5: Example of NIBP-free unital, non-unitary channel. Dotted lines represent the theoretical deep-circuit limiting values for the variance in the case of layers defined by $\mathcal{U} \circ \mathcal{E}$ (blue) and $\mathcal{U}' \circ \mathcal{E}$ (pink), where $\mathcal{U}'$ is obtained from $\mathcal{U}$ in Fig. 5.4 by substituting the first RX gate with an RY gate. The solid pink line is an exponential fit of the convergence to the limiting value.

as the composition of two dephasing-like channels on the XX and XY bases respectively, namely $\mathcal{E}_{XX}(\rho) = (1-p)\rho + pXX\rho XX$ and $\mathcal{E}_{XY}(\rho) = (1-p)\rho + pXY\rho XY$.

By analyzing the generators of the parameterized gates R_X and R_{ZZ} (i.e., X and ZZ), one can verify that they commute with both the other rotations appearing in $\mathcal{U}$ (i.e., R_Z and R_{XX}) and with the noise channel for any choice of parameters. Consequently, their action can be moved to the end of the circuit, leaving a mixed channel at the beginning. Notably, the state $\rho = |00\rangle\langle 00| + |11\rangle\langle 11|$ is a fixed point of this mixed channel, as it is singularly as fixed point of each of the remainig operations. As a result, by choosing ρ as the initial state, the circuit is effectively equivalent to its purely unitary counterpart and thus free of NIBP.

Furthermore, by substituting $\mathcal{U}$ with $\mathcal{U}'$, obtained by replacing the first RX gate with an RY gate, Fig. 5.5 shows how the NIBP phenomenon reappears, highlighting the crucial role in the interplay between noise and circuit structure.

Non-uniform noise channels

Even in the non-commuting case, the emergence of concentration is not guaranteed. In fact, the variance can even increase compared to the purely unitary circuit. This counterintuitive behavior can result from noise-induced symmetry breaking [108, 21], a phenomenon typically studied in non-unital channels but also possible in unital ones. In both cases, this effect can lead to an enhanced variance.

As an example for this second case, we keep the previous circuit structure, but we change initial state, setting it to $\rho = \mathbb{1} \otimes |0\rangle\langle 0|$, while keeping H. By virtue of this choice, the variance of the unitary circuit is now identically zero, as shown in Fig. 5.6. Here, we consider a modified noise channel $\mathcal{E}' = \mathcal{E} \circ \mathcal{F}$, where $\mathcal{F}$ is a single-qubit channel applied to the first

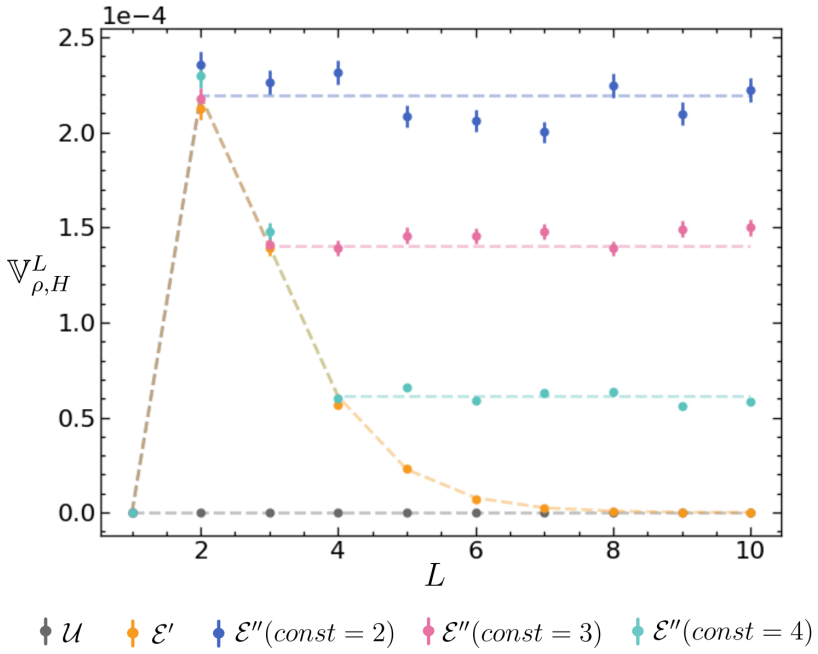

FIGURE 5.6: Variance improvment with unital channels. Dotted lines represent the theoretical prediction given by Eq. (6.10), while points indicate the numerically estimated circuit variances.

qubit, defined as $\mathcal{F}(\rho) = (1-p)\rho + p\mathcal{A}(\rho)$, and

$$\begin{aligned}\mathcal{A}(\rho) = &\frac{q}{\sqrt{2}}\operatorname{Tr}\{X\rho\}X + \frac{q}{\sqrt{2}}\operatorname{Tr}\{X\rho\}Y \\ &+ (1-q)\operatorname{Tr}\{Y\rho\}Y \\ &+ \frac{q}{\sqrt{2}}\operatorname{Tr}\{Z\rho\}Z + \frac{q}{\sqrt{2}}\operatorname{Tr}\{Z\rho\}Y\end{aligned} \tag{5.20}$$

By computing its Choi matrix, one can confirm that $\mathcal{F}$ is completely positive and trace-preserving (CPTP), and therefore constitutes a valid quantum channel whenever $0 \leq p \leq 1$ and $0 \leq q \leq 1/2$. Although $\mathcal{E}'$ is not strictly contractive, it still induces the NIBP effect, as numerically shown in Fig. 5.6. Interestingly however, the behavior of $\mathbb{V}_{\rho,H}$ under this channel is not monotonic with respect to the circuit depth: it initially increases before eventually decaying to zero in the deep-circuit limit, thus temporarily enhancing the variance. This observation motivates the consideration of a non-uniform noise profile, where $\mathcal{E}_l \neq \mathcal{E}_{l'}$ for some l, l', defined as:

$$\mathcal{E}''_l = \begin{cases}\mathcal{E}' & \text{if } l \leq const \\ \mathcal{E} & \text{otherwise}\end{cases} \tag{5.21}$$

In such a setup, the circuit benefits from the absence of concentration observed in the first example, as well as the transient variance enhancement from the second, converging to non-zero value dependent on the constant rather than the number of layers. This hybrid effect is clearly demonstrated in Fig. 5.6. The mechanism at play here is at the core of the *absorption* effects shown in the next Chapter. Altogether, the above discussion highlights that, beyond the non-trivial effects observed in non-unital channels, the broader class of contractive, but not *strictly* contractive channels gives rise to a rich variety of largely unexplored variance behaviors. In the following Chapter we will characterize these scenarios for general channels, encompassing the previously discussed cases as specific instances.

Chapter 6

Cost concentration for general noise channels

In this Chapter, we introduce a novel analytical framework based on non-negative matrix theory that enables the description of the variance in layered noisy quantum circuits with arbitrary noise channels. This approach enables the derivation of exact expressions in the deep-circuit regime, uncovering the complex interplay between unitary layers and noise. Notably, we identify a noise-induced absorption mechanism, a phenomenon absent in purely unitary dynamics, which provides new insight into how noise shapes circuit behavior. We further present a controlled convergence analysis, establishing general lower bounds on the variance of both deep and shallow circuits. This leads to a principled connection between noise resilience and the expressive capacity of parameterized quantum circuits, particularly under smart initialization strategies. Our theoretical results are supported by numerical simulations and illustrative applications. We refer to Fig. 6.1 for a schematic summary of the main results of this Chapter.

6.1 The theoretical model

In the following, we consider n-qubit quantum systems with a Hilbert space $\mathcal{H} = \bigotimes_{m=1}^{M} \mathcal{H}_m$ of dimension d, divided into M subsystems, each of dimension d_m, representing either single qubits or groups of qubits. More specifically, we study the problem where a quantum state ρ is evolved using a parameterized quantum channel $\Phi_{\boldsymbol{\theta}}$, whose parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Theta$ are optimized by minimizing the cost function

$$C(\boldsymbol{\theta}) = \mathrm{Tr}\{\Phi_{\boldsymbol{\theta}}(\rho) H\}, \tag{6.1}$$

where H is an observable of the system. Within this framework, we focus on the case of noisy *layered* quantum channels, namely

$$\Phi_{\boldsymbol{\theta}} = \mathcal{U}_{\theta_{L+1}} \mathcal{E}_L \mathcal{U}_{\theta_L} \cdots \mathcal{E}_1 \mathcal{U}_{\theta_1} \tag{6.2}$$

where each layer is composed of a unitary part $\mathcal{U}_{\theta_l}$ and an arbitrary quantum channel $\mathcal{E}_l$. If we denote by $\mathcal{B}$ the space of bounded operators acting on $\mathcal{H}$, both such components can be regarded as linear, completely positive functions mapping $\mathcal{B}$ to itself.

6.1.1 The Locality Transfer Matrix formalism

To accurately characterize the effects of general noise maps, it is essential to first analyze the structural transformations induced on the computational space by the unitary layers. In particular, it worth noticing that the subdivision of $\mathcal{H}$ into local subsystems induces also a partition of $\mathcal{B}$. More specifically, if we denote by $\kappa \in \{0,1\}^M$ a binary string of length M, we

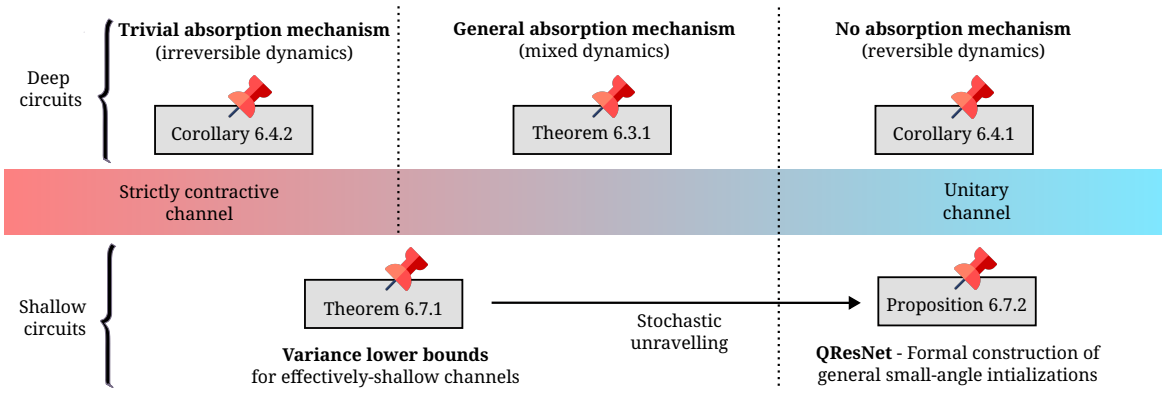

FIGURE 6.1: Cost concentration in variational quantum computing. The analytical formulation proposed here employs non-negative matrix theory to describe the interplay between local 2-designs and noise. This allows for precise calculation of cost variances for generic noise maps, from strictly contractive to unitary channels, as illustrated by the coloured band in the Figure.

can split $\mathcal{B}$ into *local* subspaces $\mathcal{B}_\kappa \subset \mathcal{B}$, each spanned by the *traceless* operators acting non trivially on $\mathcal{H}_m$ if and only if $\kappa_m = 1$. Indeed, we can partition the whole space as

$$\mathcal{B} = \bigoplus_{\kappa\in\{0,1\}^M} \mathcal{B}_\kappa \tag{6.3}$$

with $d_\kappa = \dim(\mathcal{B}_\kappa) = \prod_m (d_m^2 - 1)^{\kappa_m}$. Clearly, if $\kappa = 0$, $\mathcal{B}_0 = \operatorname{span}\mathbb{1}$. In this Chapter, such decomposition, is delineated as bringing the notion of *locality* from $\mathcal{H}$ to $\mathcal{B}$. Specifically, we say that $A \in \mathcal{B}$ is κ-local if $A \in \mathcal{B}_\kappa$, and we associate to A a locality vector $\ell_A \in \mathbb{R}^{2^M}$ defined element-wise by

$$(\ell_A)_\kappa = \sum_{j=1}^{d_\kappa} \operatorname{Tr}\{B_j A\}^2 \tag{6.4}$$

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of $\mathcal{B}_\kappa$. Clearly, from the definition, a κ-local operator A has locality vector $(\ell_A)_\lambda = \delta_{\kappa,\lambda}\|A\|_2^2$, where $\|\cdot\|_2$ is the Hilbert-Schmidt norm. We remark that similar quantities are not new in the context of PQCs, and in fact the vector ℓ_A is analogous to the purity measures defined in Chapter 5. Based on this, a derived notion of *locality preservation* can be introduced for linear maps acting on $\mathcal{B}$. In particular, given a map $\Lambda : \mathcal{B} \to \mathcal{B}$ and two subspaces $\mathcal{B}_\kappa$ and $\mathcal{B}_\lambda$, we can measure the degree to which Λ is able to put them in communication. The more subspaces are connected, the less locality preserving the map Λ will be. This idea is captured formally in the following definition of a locality transfer matrix (LTM).

Definition 6.1.1 (Locality transfer matrix)**.** Given a linear map $\Lambda : \mathcal{B} \to \mathcal{B}$, its locality transfer matrix T is defined elementwise as

$$T_{\kappa,\lambda} = \frac{1}{d_\kappa}\sum_{j=1}^{d_\kappa} (\ell_{\Lambda(B_j)})_\lambda \tag{6.5}$$

for some Hermitian, orthonormal basis $\{B_j\}_{j=1}^{d_\kappa}$ of $\mathcal{B}_\kappa$.

In this formalism, *locality preserving* transformations reflect all maps whose LTM coincides with the identity, i.e. $T_{\kappa,\lambda} = \delta_{\kappa,\lambda}$. Trivially, unitary maps, separable with respect to the partition, namely $\mathcal{U} : \rho \mapsto U\rho U^\dagger$, where $U = \bigotimes_m U_m$, are locality preserving.

With this description in mind, we assume that $\mathcal{U}_{\boldsymbol{\theta}}$ is locality preserving, and hence describes an ideal operation limited to the local subsystems $\mathcal{H}_m$, while each channel $\mathcal{E}_l$ encodes both the operations which entangle the subsystems as well as any interaction between the system and the environment. This formally captures the idea of a variational quantum algorithm running on a real, noisy, device, where quantum computation can be realised very precisely within each subsystem, but is still inaccurate when dealing with more complex entangling operations. Furthermore, we assume that, within each subsystem and for all layers l, $\mathcal{U}_{\theta_l}$ is deep enough to form a 2-design over the local unitary groups $\mathbb{U}(d_m)$. More precisely, we ask that it satisfy the following definition for at least $t = 2$.

Definition 6.1.2. (Local t-design) Given a unitary ensemble $\{U_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ with a given probability distribution over the parameter space Θ, we say it forms a local t-design for the system if each element is factorized with respect to the partition, i.e. $U_{\boldsymbol{\theta}} = \bigotimes_m U^{(m)}_{\theta_m}$ each acting solely on $\mathcal{H}_m$, and additionally

$$\int_\Theta d\boldsymbol{\theta} U^{(m)\otimes t}_{\boldsymbol{\theta}} \otimes U^{(m)*\otimes t}_{\boldsymbol{\theta}} = \int_{V\in U(d_m)} d\mu(V) V^{(m)\otimes t} \otimes V^{(m)*\otimes t} \tag{6.6}$$

where the second integral is performed with respect to the Haar measure.

This assumption is justified by the fact that such local operations are relatively inexpensive, and that the necessary depth can be very small even for large systems, since in general it depends on d_m rather than d. Putting everything together, we can give a formal definition of the family of PQCs covered by our analysis, which we call *layered, locally random channels*.

Definition 6.1.3 (Layered, Locally Random Channel)**.** Given a n-qubit quantum system described by the Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, and a decomposition into M local subsystems $\mathcal{H}_m$ such that $\mathcal{H} = \bigotimes_{m=1}^M \mathcal{H}_m$, a parameterized quantum channel $\Phi_{\boldsymbol{\theta}} : \mathcal{H} \to \mathcal{H}$ defined as

$$\Phi_{\boldsymbol{\theta}} = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}, \tag{6.7}$$

where $\mathcal{E}_l$ are arbitrary channels, is said to be *layered, locally random* if each $\mathcal{U}_{\theta_l}$ is a unitary, parametrized channel forming a local 2-design for the partition $\mathcal{H} = \bigotimes_{m=1}^M \mathcal{H}_m$.

6.1.2 Main properties of Locality Tranfer Matrices

When dealing with an LTM T, the first thing to observe is that, by construction, T is a non-negative matrix, namely $T_{\kappa,\lambda} \geq 0 \ \forall \kappa, \lambda$. This means it can always be expressed in canonical, block-upper triangular form, where each of the diagonal blocks is irreducible [102]. Throughout this Chapter, irreducible components of T will be regarded as *essential* if they cannot lead outside the block, and will be denoted by T_z. Otherwise, an irreducible block will be deemed *inessential*, and their collection will be denoted by Q. A useful pictorial representation of these possibilities is shown in Fig. 6.2, together with the matrix canonical form. A foundational reference for this definition can be found in Ref. [102], while a more detailed discussion is proposed in Appendix C.1.

Another crucial property in the following analysis is the relation between the LTMs of the channel $\mathcal{E}_l$ and its dual $\mathcal{E}_l^\dagger$ with respect to the Hilbert-Schmidt scalar product, namely T and $T^\dagger$. Such relation can be characterized for generic linear maps $\Lambda : \mathcal{B} \to \mathcal{B}$ as a direct consequence of Definition 6.1.1, and in particular we have

$$TD = (T^\dagger D)^t,$$

with $D_{\kappa,\lambda} = d_\kappa \delta_{\kappa,\lambda}$. For sake of readability, here we introduce a shorthand notation for the scalar product $(\cdot,\cdot)$ in $\mathbb{R}^{2^M}$ such that T and $T^\dagger$ are *also* Hermitian adjoint of one another, i.e.

$$(a, b) = a^t D^{-1} b = \sum_\kappa \frac{a_\kappa b_\kappa}{d_\kappa}. \tag{6.8}$$

We refer the reader to Appendix C.2 for a more detailed discussion on the fundamental properties of LTMs.

6.2 Cost variance calculation

The overarching goal of this Chapter is to characterise the properties of the variance $\mathbb{V}^L_{\rho,H}$ as a function of the number L of layers. To this end, in this Section, we provide a formal expression for general layered locally random channels. In particular, we compute the variance $\mathbb{V}^L_{\rho,H}$ in terms of the properties of ρ, H and the intermediate channels, showing its relevance and main domains of applicability.

As a first step, we consider the expectation value of the cost function. A useful property of local 2-designs is that they automatically form a *global* 1-design, meaning we can easily compute the expectation values of locally random circuits by integration over the Haar measure. This property can be derived by considering that each $\mathcal{U}_\theta$ also forms a local 1-design by definition, and composition of local 1-designs gives a global 1-design. Thanks to this property, we have

$$\mathbb{E}_\theta \left\{ \mathrm{Tr}\{\Phi_\theta(\rho) H\} \right\} = \frac{\mathrm{Tr}\{H\}\,\mathrm{Tr}\{\rho\}}{d} = \frac{\mathrm{Tr}\{H\}}{d}. \tag{6.9}$$

A formal proof is given in Appendix C.3. Concerning the second moment, we have the following proposition.

Proposition 6.2.1 (General formula)**.** Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition 6.1.3. Then, we have

$$\mathbb{E}_\theta \left\{ \mathrm{Tr}\{\Phi_\theta(\rho) H\}^2 \right\} = \left(\ell_\rho, \prod_{l=1}^{L} T_l \, \ell_H \right) \tag{6.10}$$

where $(\cdot, \cdot)$ is the scalar product defined in Eq. (6.8), and each T_l is the LTM associated to the respective $\mathcal{E}_l^\dagger$.

Again, a formal proof is given in Appendix C.3. Putting everything together, this allows a direct calculation of $\mathbb{V}^L_{\rho,H}$ as a function of the LTMs of the intermediate channels. In particular, from the defintion of variance we have

$$\begin{aligned} \mathbb{V}^L_{\rho,H} &= \mathbb{E}_\theta \left\{ \left(\mathrm{Tr}\{\Phi_\theta(\rho) H\} - \mathbb{E}_{\theta'} \left\{ \mathrm{Tr}\{\Phi_{\theta'}(\rho) H\} \right\} \right)^2 \right\} \\ &= \mathbb{E}_\theta \left\{ \mathrm{Tr}\{\Phi_\theta(\rho) H\}^2 \right\} - \mathbb{E}_\theta \left\{ \mathrm{Tr}\{\Phi_\theta(\rho) H\} \right\}^2 \end{aligned} \tag{6.11}$$

which immediately gives

$$\mathbb{V}^L_{\rho,H} = \left(\ell_\rho, \prod_{l=1}^{L} T_l \, \ell_H \right) - \frac{\mathrm{Tr}\{H\}^2}{d^2}. \tag{6.12}$$

Note that this formulation provides an exact formula, which in principle gives access to the study of cost function concentration for any L. Indeed, this is the case when there are few subsystems, since in this case one can easily estimate and manipulate the matrices T_l. An example of this will be given in Section 6.3.1. However, this approach becomes rather cumbersome in a general setting, for large systems, since the resources needed to represent the

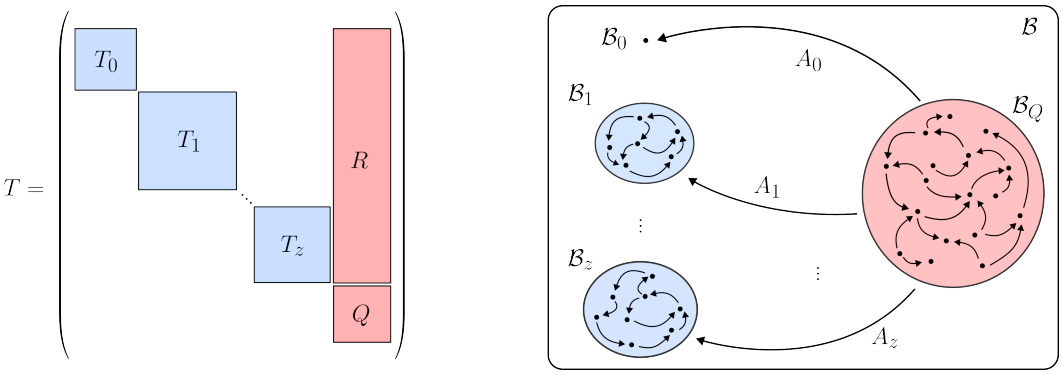

FIGURE 6.2: Graphical representations of the stochastic process describing $\mathbb{V}^L_{\rho,H}$. On the left, we show the structure of the general locality transfer matrix (LTM), highlighting the decomposition into irreducible components [102]. Light blue blocks represent irreducible, essential components of T, while red blocks are related to inessential ones. In particular, Q represents the collection of all irreducible, inessential components of T and R their relation with the essential components T_z. On the right, the same process is represented graphically, in terms of the local subspaces $\mathcal{B}_k$. Here, each dot represents a single subspace, while the arrows represent the adjoint action of the channel $\mathcal{E}$. Essential and inessential components share the same colour code of T.

LTM can grow exponentially. Nevertheless, as shown in the following Sections, Proposition 6.2.1 can still be profitably used as a theoretical tool, as it allows characterizing $\mathbb{V}^L_{\rho,H}$ in both the deep and shallow circuit limit.

6.3 Deep circuit limit

As the circuits become deeper, it is natural to expect the contribution of the leading eigenvectors of T_l to become dominant, as repeatedly multiplying them gives rise to a process analogous to power iteration methods. The structure of such eigenvectors captures several interesting properties of the interaction between the unitary and non-unitary parts of $\Phi_{\boldsymbol{\theta}}$, particularly *absorption*, which can only arise in this picture.

For simplicity, we will focus for the remainder of this Chapter on *homogeneous* channels, i.e. we fix $\mathcal{E}_l = \mathcal{E}\ \forall l$, and consider, without loss of generality, a traceless observable H. In this case, Eq. (6.12) becomes

$$\mathbb{V}^L_{\rho,H} = \left(\ell_\rho, T^L \ell_H\right), \tag{6.13}$$

greatly simplifying the calculations. In what follows, we will use this construction to understand convergence properties, limitig value and their physiscal implications for our noisy circuit model.

6.3.1 Convergence of the limit

While the convergence of $\mathbb{V}^L_{\rho,H}$ to a well defined value in the deep circuit limit is often taken for granted, when using general entangling channels as in the case of Definition 6.1.3, convergence is not always guaranteed. More precisely, the limit

$$\mathbb{V}^\infty_{\rho,H} = \lim_{L\to\infty} \mathbb{V}^L_{\rho,H} \tag{6.14}$$

need not to exists in general. Indeed, while the variance is upper bounded by Hölder inequality, i.e. $\mathbb{V}^L_{\rho,H} \leq \|H\|^2_\infty\ \forall L$, different circuit sub-sequences for increasing L can in general

lead to different limits. As discussed in Appendix C.4, this property is related to the presence of *cycles* in T, i.e. the presence of periodic irreducible blocks T_z with period $p > 1$. The notion of periodicity of a non-negative matrix such as T_z is rigorously introduced in Appendix C.1, however, to help the intuition, we can informally think of p as the number of distinct eigenvalues $\lambda_1, \cdots \lambda_p$ such that their modulus $|\lambda_i|$ is the same, and it is simultaneously maximal across the spectrum of T_z. Intuitively, since there are multiple dominant eigenvalues, we also have a multitude of potential limiting values, which ultimately cause the lack of convergence.

Indeed, in the variance calculation setting, this can happen for instance when the entangling operation is not chosen carefully with respect to the partition of $\mathcal{H}$. In those cases, the value of $\mathbb{V}^L_{\rho,H}$ will depend on which equivalence class of integers modulo p the depth L belongs to. This idea is put into an effective example in this Section.

Consider the circuit depicted in Fig. 6.3. The simple structure of this circuit allows to explicitly compute $\mathbb{V}^L_{\rho,H}$ as function of L. In particular, we have

$$\mathbb{V}^L_{\rho,H} = \begin{cases} \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{3} & \text{if } L \text{ is even} \\ 0 & \text{if } L \text{ is odd} \end{cases} \tag{6.15}$$

from which it is clear that the deep circuit limit does not converge. Moreover, thanks to the contained dimension of the system, it is possible to compute and represent T:

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \quad T_0 = T_2 = \begin{pmatrix} 1 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{6.16}$$

Here we can see that T_1 has in fact period 2, as its eigenvalues are $\lambda_1 = 1$, $\lambda_2 = -1$, both of modulus 1. This implies the presence of 2 distinct limiting values whenever both ρ and H have a component belonging to T_1, and hence the limit of Eq. (6.14) does not exist. In such cases, we can still make sense of the deep circuit limit in terms of the *Cesàro* average of $\mathbb{V}^L_{\rho,H}$. The Cesàro average is a method of assigning a value to certain non-convergent series by averaging their partial sums: given a series $\sum_{k=1}^{\infty} a_k$ with partial sums $s_n = \sum_{k=}^{n} a_k$, its Cesàro sum is defined as the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} s_k,$$

if this limit exists. Concerning the above example, the Cesàro average of the variance is indeed well-defined, and in this case we get

$$\mathbb{V}^{\infty}_{\rho,H} = \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \mathbb{V}^l_{\rho,H} = \frac{(\|\rho\|_2^2 - 1/2)\|H\|_2^2}{6}, \tag{6.17}$$

which can be recovered from direct calculation. As we will see in the next Section, this is not unique to this example, and indeed the Cesàro average in this context is perfectly well defined and closely related to the structure of T.

6.3.2 Limiting variance

We now use the general properties of LTMs to compute $\mathbb{V}^{\infty}_{\rho,H}$. All results presented here are formalized a rigorously proven in Appendix C.4.

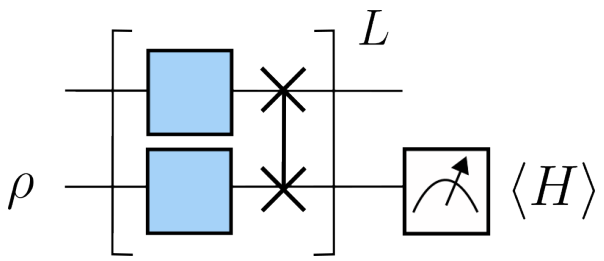

FIGURE 6.3: Simple 2-qubit circuit, designed to have a non-convergent variance. In this construction, the channel $\mathcal{E}$ is chosen to be unitary and in particular to be a SWAP gate. Moreover, both the initial state ρ and the observable H are chosen to be non-trivial only on the second qubit, and are therefore represented as single qubit operators.

Recalling the structure of T (see Fig. 6.2), we can decompose $\mathbb{V}^L_{\rho,H}$ into a sum of terms, coming from each irreducible component T_z and Q, namely

$$\mathbb{V}^L_{\rho,H} = (\ell_\rho, T^L \ell_H) = \sum_z \left[(\ell_\rho, T_z^L \ell_H) + (\ell_\rho, A_z^{(L)} \ell_H) \right] + (\ell_\rho, Q^L \ell_H), \tag{6.18}$$

where $A_z^{(L)} = \sum_{l=0}^{L-1} T_z^l R Q^{L-1-l}$. While the presence of T_z^L and Q^L is clear from the multiplication of the diagonal blocks of T, the structure of $A_z^{(L)}$ is more elusive, and its calculation requires a proof by induction, provided in Lemma C.4.1. Given this structure, it will be particularly useful to denote by $\mathcal{B}_z = \bigoplus_{\kappa \in T_z} \mathcal{B}_\kappa$ the union of subspaces put in communication within T_z, by d_z their total dimension, and given $A \in \mathcal{B}$, by $(\ell_A)_z = \sum_{\kappa \in T_z} (\ell_A)_\kappa$ the corresponding locality.

Regardless of the specific channel $\mathcal{E}$ used, some general properties of the blocks T_z can be identified. For instance, due to trace preservation, the map $\mathcal{E}^\dagger$ is unital, and consequently the trivial subspace $\mathcal{B}_0$ is always mapped into itself, thus forming an essential component of T, which we denote by T_0. Similarly, all other essential blocks T_z must satisfy $\mathcal{E}^\dagger(\mathcal{B}_z) \subset \mathcal{B}_z$.

Concerning the spectral properties of T, we have the following Proposition.

Proposition 6.3.1 (Spectral radius of T)**.** Let T be the LTM of a CPU, and let Q represent the collection of all irreducible, inessential components of T as in Fig. 6.2. Then T is contractive in the sense of the spectral radius, i.e. $\rho(T) \leq 1$. Moreover, the block Q is strictly contractive, namely $\rho(Q) < 1$.

By inclusion, this implies the same bound also on $\rho(T_z)$. When the bound is saturated, namely $\rho(T_z) = 1$, the celebrated *Perron-Frobenius* theorem guarantees that there exists one, positive, non-degenerate eigenavlue r_z, the so called *Perron* eigenvalue, such that $r_z = \rho(T_z) = 1$. This holds even in the presence of cycles, where multiple eigenvalues λ_k with modulus $|\lambda_k| = r_k$ might be present. In particular, for LTMs, the left eigenvector of the Perron eigenvalue in this setting can be analytically computed, as given by the following Proposition.

Proposition 6.3.2. Let T be a LTM of a CPU map and T_z be an irreducible block, then $\rho(T_z) = 1 \Leftrightarrow \sum_\kappa (T_z)_{\kappa,\lambda} = 1$, or equivalently $v_z^t T_z = v_z^t$, where $(v_z)_\kappa = 1 \forall \kappa$ is the left eigenvector of the dominant eigenvalue.

Furthermore, the Perron-Frobenius theorem also relates the left and right eigenvectors v_z and w_z, to the Cesàro average of T_z^L. In particular, the limit exists and we have the relation

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} T^l = w_z v_z^t, \tag{6.19}$$

suggesting that each term $(\ell_\rho, T_z^L \ell_H)$ in Eq. (6.18) contributes in the limit as

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} (\ell_\rho, T_z^L \ell_H) = (\ell_\rho, w_z v_z^t \ell_H) = (\ell_\rho, w_z)(\ell_H)_z \,. \tag{6.20}$$

In contrast, when the spectral radius is strictly less than unity, the corresponding variance contributions vanish under repeated applications of the channel by Gelfand's formula. This is given explicitly by the following Proposition.

Proposition 6.3.3 (Vanishing terms)**.** Let T be a LTM, and let T_z be an irreducible block with $r_z < 1$. Then, as $L \to \infty$, $\|T_z^L\| \to 0$ and $\|A_z^{(L)}\| \to 0$ exponentially fast for any matrix norm $\|\cdot\|$. Similarly, also $\|Q^L\| \to 0$.

This suggests that the corresponding terms in Eq. (6.18) vanish in the limit, and thus the general form for the variance should read

$$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z\,|\,\rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)_z + (\ell_\rho, w_z)(A\ell_H)_z, \tag{6.21}$$

where A is a matrix of the same shape as R in Fig. 6.2, which collects contributions coming from all $A_z^{(L)}$ in Eq. (6.18) in the limit of large L. Such reasoning is formally encapsulated in the following Theorem.

Theorem 6.3.1 (Deep circuit variance)**.** Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ a be layered, locally random quantum channel as in Definition 6.1.3. Then the Cesàro average of $\mathbb{V}_{\rho,H}^{L}$ converges to Eq. (6.21), and we have

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}_{\rho,H}^{l} - \mathbb{V}_{\rho,H}^{\infty} \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \tag{6.22}$$

for some constant $\beta > 0$. Additionally, if all essential blocks are aperiodic (i.e. with period $p = 1$), then $\mathbb{V}_{\rho,H}^{L}$ is convergent, and we have

$$\left| \mathbb{V}_{\rho,H}^{L} - \mathbb{V}_{\rho,H}^{\infty} \right| \in O\left(e^{-\beta L} \|H\|_2^2\right), \tag{6.23}$$

where the absorption matrix is given by $A = R(\mathbb{1} - Q)^{-1}$.

As a special case, if the intermediate channel can be reduced unitarily to a tensor product of single qubit channels, then also w_z of Eq. (6.21) can be explicitly computed, yielding the following Corollary.

Corollary 6.3.1. If the intermediate channel takes the form $\mathcal{E} = \mathcal{N} \circ \mathcal{W}$, where the noise channel $\mathcal{N} = \bigotimes_m \mathcal{N}_m$ is the composition of single qubit channels and $\mathcal{W}$ is unitary, then

$$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z\,|\,\rho(T_z)=1} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \tag{6.24}$$

The formal proof of Theorem 6.3.1 and Corollary 6.3.1 can be found in Appendix C.4, where we complete our analysis computing w_z also for global Pauli noise, yielding the same limiting value of Eq. (6.24). While these results hold in general, for the rest of this Chapter, we will assume *aperiodicity* for each irreducible block, as this will considerably simplify the notation while keeping the fundamental insights from the Theorem.

We start the analysis of Theorem 6.3.1 by unpacking Eq. (6.24). Here, $\mathbb{V}_{\rho,H}^{\infty}$ is shown to depend on four important quantities, namely the invariant subspaces $\mathcal{B}_z$ of the noisy circuit's dual channel $\Phi_{\boldsymbol{\theta}}^{\dagger}$, the respective locality of ρ and H, together with the matrix A. As a

reminder, in order for this structure to arise, the invariant subspaces of $\mathcal{U}_\theta^\dagger$ and $\mathcal{E}^\dagger$ have to be *well-aligned*, so that non-trivial subspaces $\mathcal{B}_z \subset \mathcal{B}$ are present. This extends the notion of alignment already introduced in the literature [65] for ρ, H and unitary circuits. Moreover, we observe that such subspaces act as *attractors* for the variance, as each summand not only depends on the local components of ρ and H, but also on the components of H belonging to $\mathcal{E}(\mathcal{B}_z) \cap \mathcal{B}_Q$. In this sense, A can be interpreted as an absorption matrix, which quantifies the extent of the contribution of such terms. We remark that this contribution is always non-negative, and it is intimately related to the contractivity properties of $\mathcal{E}^\dagger$. Overall, this suggests that appropriate non-unitary layers will alleviate the concentration typical of unitary circuits by a mechanism that allows to bring the contribution of the components of H belonging to strictly contractive, high dimensional subspaces, to non-contractive, smaller dimensional ones.

In the following Sections, we show how the general formula introduced in Theorem 6.3.1 reduces to ordinary barren plateau results when $\mathcal{E}$ is unitary, and ordinary noise-induced barren plateau when $\mathcal{E}$ is unital strictly contractive. We denote these limits as the *extreme cases*. Furthermore, we analyze the scaling of the absorption terms, giving theoretical estimates for the noise contributions.

6.4 Extreme cases of the deep circuit formula

Starting from the analysis of Section 6.2, we show how we can use Eq. (6.24) to recover previously known results by considering specific classes of quantum channels $\mathcal{E}$. Formal proofs of these statements can be found in Appendix C.4.

6.4.1 Unitary circuits

Concerning the case where the entangling map $\mathcal{E}$ is unitary, we point out that, being an absorption term, the last term in Eq. (6.24) vanishes for unitary dynamics, which is reversible by definition. In particular, in this case, reversibility implies that the matrix T is block diagonal, and the block Q vanishes. Formally, we have the following Corollary of Theorem 6.3.1.

Corollary 6.4.1 (Deep, unitary circuits)**.** Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition 6.1.3, where $\mathcal{E}(\cdot) = W \cdot W^\dagger$, $W \in \mathbb{U}(d)$ is an arbitrary unitary transformation. Then we have

$$\mathbb{V}^\infty_{\rho,H} = \sum_{z>0} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}, \tag{6.25}$$

where $\mathcal{B}_z$ are invariant subspaces of $\mathcal{E}$ which can be expressed as the direct sum of $\mathcal{B}_\kappa$.

This Corollary contains many interesting properties of the variance in the deep circuit limit. First, it captures the necessity of *alignment* between ρ, H and Φ_θ in order to achieve a substantial variance, i.e. ρ and H need both to have non-negligible components on the same invariant subspace $\mathcal{B}_z$. Due to the structure of the channel, this idea is extended to the components $\mathcal{U}_\theta$ and $\mathcal{E}$, whose invariant subspaces need to align in order to keep the dimension d_z of $\mathcal{B}_z$ from being exponentially large. Indeed, while $\mathcal{B}_0$ is always an invariant subspace satisfying Corollary 6.4.1, misalignment of the local and entangling parts of the circuit could result in the entirety of the remaining space falling under a single, irreducible component of dimension $d^2 - 1$. In such cases we get

$$\mathbb{V}^\infty_{\rho,H} = \frac{\left(\|\rho\|_2^2 - 1/d\right)\left(\|H\|_2^2 - \mathrm{Tr}\{H\}^2/d\right)}{d^2 - 1}, \tag{6.26}$$

which implies the presence of BP regardless of ρ and H as long as $\|H\|_2 < O(d)$. Indeed, one can interpret the misalignment of $\mathcal{U}_{\boldsymbol{\theta}}$ and $\mathcal{E}$ as introducing an excess of expressibility, which is known to lead to exponential concentration [46].

6.4.2 Strictly contractive circuits

As a complementary remark, we point out that, conversely to the above, the first term in Eq. (6.24) pertains to non-contractive subspaces, and as such, vanishes if $\mathcal{E}^\dagger$ is strictly-contractive in, at least, one direction in each $\mathcal{B}_z$. This is formalized in the following Corollary of Theorem 6.3.1.

Corollary 6.4.2 (Deep, noisy circuits)**.** Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3, where $\mathcal{E}$ is such that $\|\mathcal{E}(A)\|_2 < \|A\|_2$, for at least one $A \in \mathcal{B}_\kappa \subset \mathcal{B}_z \; \forall z > 0$. Then, we have

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}. \tag{6.27}$$

In particular, if the channel is unital, $\mathbb{V}^{\infty}_{\rho,H} = 0$.

Note that, even if Eq. (6.27) is inversely proportional to $d = 2^n$, $\mathbb{V}^{\infty}_{\rho,H}$ is not necessarily exponentially suppressed, as in general the contribution of the observable increases with the same speed, i.e. $\|H\|_2^2 \sim d$. As before, this Corollary captures the main features of noise-induced barren plateaus (NIBP). In fact, it is clear that strictly contractive channels with a unique fixed point, fall into the assumptions of Corollary 6.4.2, and therefore exhibit some form of concentration. However, Corollary 6.4.2 is not limited to them, extending the noise-induced concentration to a wider class of noise maps, which crucially depend on the structure of the unitary part of the channel $\mathcal{U}_{\boldsymbol{\theta}}$. According to the method introduced here, this can clearly be interpreted as a consequence of the interaction between noise and unitary layers, which effectively "spread" the contractive effect of $\mathcal{E}$ to the whole irreducible component.

Unital channels will suffer most severely from NIBP, since in that case the absorption term in Eq. (6.27) vanishes. Contrarily, as pointed out in the literature [5, 82], non-unital channels may avoid the exponential concentration. Here, we show how previous results can be expressed as the contribution to the absorption term of $\mathcal{B}_0$, which always forms a norm-preserving subspace due to trace preservation of $\Phi_{\boldsymbol{\theta}}$. A precise analysis is provided in Section 6.5, where we analytically estimate the scaling of absorption coefficients.

However, it is crucial to remark that this specific contribution is not due to the retention of any computational power to the PQC, since the dependence on the initial state is completely lost, but rather to the competing effects between the drive of $\mathcal{U}_{\boldsymbol{\theta}}$ and $\mathcal{E}$ towards the respective, different fixed points. An example of this phenomenon is provided in Section 6.6, where we study how extensive entanglement can further exacerbate the issue. This is in stark contrast with the general case discussed above, i.e. the absorption terms of $\mathcal{B}_z$, $z > 0$, as they keep a non-trivial dependence on the initial state, and as such it can be said to genuinely avoid concentration if d_z scales appropriately. As an example, this regime can be realized within the single-qubit noise model of Corollary 6.3.1 by employing a structured entangling gate and a mixture of clean and noisy qubits, as precisely shown in Section 6.5.

6.5 Estimates of the absorption coefficients

In this section, we estimate the absorption coefficients in the general setting of Corollary 6.3.1, where the entangling channel admits a decomposition of the form

$$\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger), \quad \mathcal{N} = \bigotimes_m \mathcal{N}_m \tag{6.28}$$

that is, a generic unitary entangling gate followed by local single-qubit noise. In the first part, we provide theoretical lower bounds for all absorption coefficients, along with scaling guarantees for the absorption into the trivial subspace, thereby demonstrating consistency with prior literature [5, 82]. In the second part, we present numerical evidence of absorption into non-trivial subspaces, revealing qualitatively distinct variance behaviors depending on the choice of initial state. This effect goes beyond the scope of existing analyses and suggests the emergence of a previously uncharacterized phenomenon.

6.5.1 Theoretical estimates

As a first remark, we give a general lower bound for the absorption coefficients in the setting of Corollary 6.3.1. Such lower bounds can be obtained exploiting the structure of A in Theorem 6.3.1, which in the aperiodic case reads $A = R(\mathbb{1} - Q)^{-1}$, noticing that since $Q_{\kappa,\lambda} \geq 0$, then

$$A = R(\mathbb{1} - Q)^{-1} = R\sum_{l=0}^{\infty} Q^l \geq R \tag{6.29}$$

entrywise. Estimation of R in the general case yields the following Proposition.

Proposition 6.5.1. Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3, and let the intermediate channel take the form of Eq. (6.28). Then, for all essential T_z of the LTM T with unit spectral radius, we have:

$$\frac{(A\ell_H)_z}{d} \geq \sum_{j \,|\, P_j \in \mathcal{B}_z} (\Lambda_j, \ell_H) \tag{6.30}$$

where the operators Λ_j are defined by the action of the noise channels as

$$\Lambda_j = \mathcal{Q}(\tilde{\Lambda}_j), \quad \tilde{\Lambda}_j = \bigotimes_m \mathcal{N}_m\left(\frac{P_{j_m}}{\sqrt{2}}\right) \tag{6.31}$$

with P_j are normalized Pauli strings and $\mathcal{Q} : \mathcal{B} \to \mathcal{B}_\mathcal{Q}$ is a projection onto $\mathcal{B}_\mathcal{Q}$.

Note that, the factor $1/\sqrt{2}$ is introduced in the definition of $\tilde{\Lambda}_j$ to ensure $\mathrm{Tr}\{\tilde{\Lambda}_j\} \leq 1$, following the normalization constraints for quantum states. In this way, if we restrict to the trivial subspace $\mathcal{B}_0$, $\tilde{\Lambda}_j = \tau$ becomes a quantum state, and in particular it becomes the action of the noise channel on the maximally mixed state $\tau = \mathcal{N}(\rho_m)$, $\rho_m = \mathbb{1}/d$. As a special case of Proposition 6.5.1, we have that

$$\frac{(A\ell_H)_0}{d} \geq (\ell_{\mathcal{Q}(\tau)}, \ell_H), \tag{6.32}$$

which hints at noise contribution arising from the absorption term to $\mathcal{B}_0$. If the map is unital, $\tau = \mathcal{N}(\mathbb{1}/d) = \mathbb{1}/d$ and $\ell_{\mathcal{Q}(\tau)} = 0$, so the bound becomes vacous in agreement with Corollary 6.4.2.

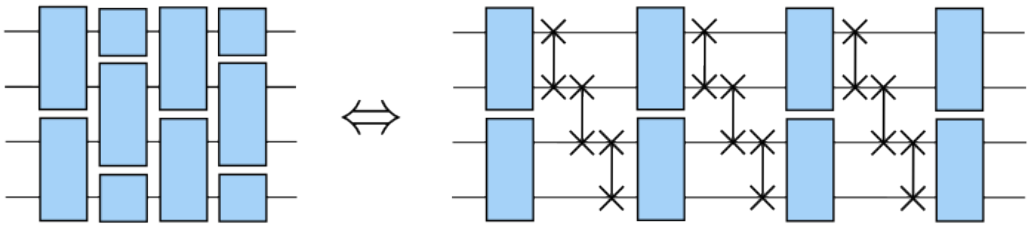

FIGURE 6.4: Relation between the circuit structure used in [5, 30] and the present work. The interleaved local two-design architecture is a special case of Eq. (6.2), with a *SWAP* cascade as entangling layer.

As a second theoretical point, we show how the general scaling for the variance obtained in the literature can be also recovered in this formalism, by considering more closely the absorption to the trivial subspace only.

Proposition 6.5.2. If the noise channel in Eq. (C.57), is further simplified to $\mathcal{N} = \bar{\mathcal{N}}^{\otimes n}$, where $\bar{\mathcal{N}}$ is a single-qubit, non-unital channel, then the trivial block T_0 is the only one with unit spectral radius, and its corresponding absorption coefficient scales as

$$\frac{(A\ell_H)_0}{d} \in \sum_i h_i^2 e^{\Theta(|P_i|)} \tag{6.33}$$

where $H = \sum_i h_i P_i$ denotes the Pauli decomposition of H and $|P_i|$ denotes the number of non-trivial terms in P_i.

This includes as a special special case Theorem 6 in Ref. [5], as their model can be expressed in our formalism by considering two-qubit local subsystems $\mathcal{H}_m = (\mathbb{C}^2)^{\otimes 2}$ in the defintion of Section 6.1, and a SWAP-cascade entangling layer as illustrated in Fig. 6.4.

6.5.2 Numerical validation

While, as shown in Proposition 6.5.2, for noise models composed of repeated applications of the same single-qubit non-unital channel, $\mathcal{B}_0$ remains the only relevant subspace in the deep circuit limit, our analysis indicates that when unital and non-unital channels are mixed, additional subspaces can also contribute non-trivially. This suggests the emergence of a novel absorption phenomenon not captured by prior models. We numerically validate this claim below with a pedagogical example.

Consider a circuit composed of 4 qubits, with single-qubit local subsystems $\mathcal{H}_m = \mathbb{C}^2$. Let the entangling gate be of the form Eq. (C.57), where W and $\mathcal{N} = \mathcal{N}_0 \otimes \mathcal{N}_2$ are described in Fig. 6.5. In particular, $\mathcal{N}_0$ is a *depolarizing* channel defined as

$$\mathcal{N}_0(\rho) = \left(1 - \frac{3\gamma_0}{4}\right)\rho + \frac{\gamma_0}{4} X\rho X + \frac{\gamma_0}{4} Y\rho Y + \frac{\gamma_0}{4} Z\rho Z, \tag{6.34}$$

with $\gamma_0 = 5 \times 10^{-2}$ and $\mathcal{N}_2$ is an *amplitude damping* channel, defined as

$$\begin{gathered} \mathcal{N}_2(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger, \\ K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma_2} \end{pmatrix} \quad K_0 = \begin{pmatrix} 0 & \sqrt{\gamma_2} \\ 0 & 0 \end{pmatrix} \end{gathered} \tag{6.35}$$

with $\gamma_2 = 2 \times 10^{-1}$.

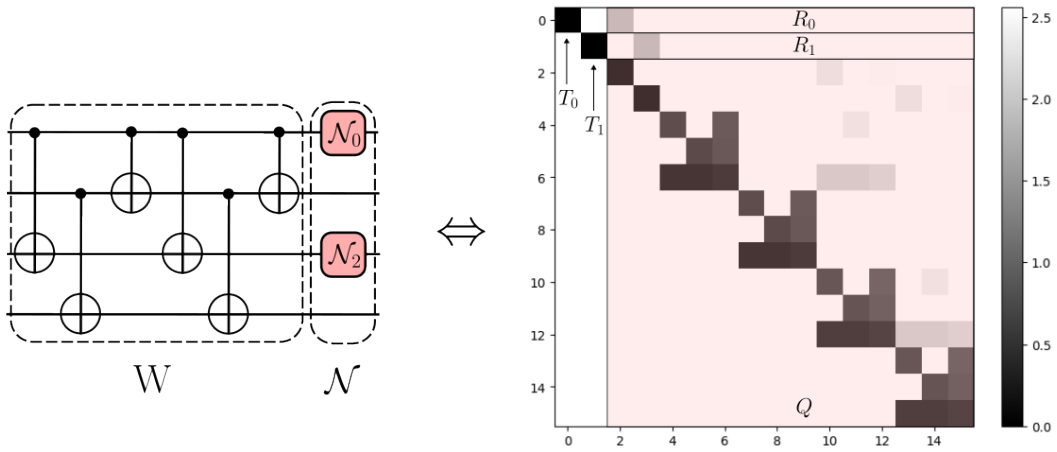

FIGURE 6.5: Intermediate channel for the numerical experiments and corresponding LTM. On the left, the circuit used in the numerical study, including the structure of the entangling unitary W, and noise channel $\mathcal{N} = \mathcal{N}_0 \otimes \mathcal{N}_2$, where $\mathcal{N}_0$ is a depolarizing channel and $\mathcal{N}_2$ is an amplitude damping channel. On the right, the heatmap of the LTM entries in logarithmic scale, where strictly contractive component Q, and the absorption terms R_0 and R_1 are highlighted in red.

Since the dimension of the system is small, we can block diagonalize T, showing the presence of *two* distinct norm-preserving, strongly connected components, namely the trivial component T_0 corresponding to $\mathcal{B}_0$, and T_1, corresponding to $\mathcal{B}_1 = \mathcal{B}_{\kappa=(0100)}$. Their respective dimensions are $d_0 = 1$ and $d_1 = d_{\kappa=(0100)} = d_{m=1}^2 - 1 = 3$. We now estimate the magnitude of the absorption effects using Proposition 6.5.1. We start by computing the action of $\mathcal{N}_2$ on the Pauli basis. Applying the definition, we have:

$$\begin{aligned} \mathcal{N}_2(\mathbb{1}) &= \mathbb{1} + \gamma_2 Z, \;\; \mathcal{N}_2(X) = \sqrt{1-\gamma_2}X \\ \mathcal{N}_2(Y) &= \sqrt{1-\gamma_2}Y, \;\; \mathcal{N}_2(Z) = (1-\gamma_2)Z. \end{aligned} \tag{6.36}$$

Using these relations, we can now easily compute the operators τ, and $\tilde{\Lambda}_j$, $j \in \mathcal{B}_{\kappa=(0100)}$. In particular we have:

$$\begin{aligned} \tau &= \mathcal{N}_0(\mathbb{1}) \otimes \mathbb{1} \otimes \mathcal{N}_2(\mathbb{1}) \otimes \mathbb{1}/d \\ &= \mathbb{1} \otimes \mathbb{1} \otimes (\mathbb{1} + \gamma_2 Z) \otimes \mathbb{1} \\ &= \mathbb{1}/d + \gamma_2 \mathbb{1}\mathbb{1}Z\mathbb{1}/d, \\ \\ \tilde{\Lambda}_j &= \mathcal{N}_0(\mathbb{1}) \otimes P_{j_1} \otimes \mathcal{N}_2(\mathbb{1}) \otimes \mathbb{1}/d \\ &= \mathbb{1} \otimes P_{j_1} \otimes (\mathbb{1} + \gamma_2 Z) \otimes \mathbb{1} \\ &= \mathbb{1}P_{j_1}\mathbb{1}\mathbb{1}/d + \gamma_2 \mathbb{1}P_{j_1}Z\mathbb{1}/d, \end{aligned} \tag{6.37}$$

which yields the operators $\Lambda_0 = \mathcal{Q}(\tau) = \gamma_2 \mathbb{1}\mathbb{1}Z\mathbb{1}/d$ and $\Lambda_j = \mathcal{Q}(\tilde{\Lambda}_j) = \gamma_2 \mathbb{1}P_{j_1}Z\mathbb{1}/d$ to estimate the absorption coefficients. In particular, this leads to the locality vectors

$$(\ell_{\Lambda_0})_\kappa = \gamma_2^2 \delta_{\kappa,(0010)} \;\text{ and }\; \sum_j (\ell_{\Lambda_j})_\kappa = 3\gamma_2^2 \delta_{\kappa,(0110)}. \tag{6.38}$$

This implies that, by choosing $H_0 \in \mathcal{B}_{\kappa=(0010)}$ and $H_1 \in \mathcal{B}_{\kappa=(0110)}$, we are guaranteed to have

a non-vanishing absorption term to $\mathcal{B}_0$ and $\mathcal{B}_1$ respectively. Let's fix them to $H_0 = \mathbb{1}\mathbb{1}X\mathbb{1}$ and $H_1 = \mathbb{1}XX\mathbb{1}$. Then we have

$$(A\ell_{H_0})_0 \geq d\gamma_2^2/3 \text{ and } (A\ell_{H_1})_1 \geq d\gamma_2^2/3, \tag{6.39}$$

where the factors of 3 appear in accordance with the dimensions $d_{\kappa=(0010)} = 3$ and $d_{\kappa=(0110)} = 9$. The last step involves putting the lower bounds of Eq. (6.39) back into the formula for $\mathbb{V}^{\infty}_{\rho,H}$ of Eq. (6.24), namely estimating the quantity

$$\frac{(\ell_\rho)_z (A\ell_H)_d}{d_z}. \tag{6.40}$$

Since all quantum states satisfy $\mathrm{Tr}\{\rho\} = 1$, $(\ell_\rho)_0 = 1/d$, and the first absorption term is always lower bounded by $\gamma_2^2/3$, regardless of the initial state.

On the contrary, the other absorption terms depends on the locality of the initial state on $\mathcal{B}_{z=1}$, which is not guaranteed to be non-zero in general. As an example, the state $\rho_A = |0000\rangle\langle 0000|$ has non-trivial component here, namely $(\ell_{\rho_0})_{z=1} = 1/d$, but the mixed state $\rho_B \propto |0000\rangle\langle 0000| + |0110\rangle\langle 0110|$ does not, and $(\ell_{\rho_B})_{z=1} = 0$. Without resorting to mixed states, also the entangled GHZ state ρ_{GHZ}, can be shown to similarly satisfy $(\ell_{\rho_{GHZ}})_{z=1} = 0$. Indeed, the lower bound established by Proposition 6.5.1, only holds for the H-term in $\mathbb{V}^{\infty}_{\rho,H}$ of Eq. (6.24), but the contribution could still result in an exponentially decaying variance if the corresponding ρ-term vanishes. Putting this into Eq. (6.40), we get the lower-bounds

$$\frac{(\ell_\rho)_0 (A\ell_{H_0})_0}{d_0} \geq \frac{\gamma_2^2}{3} \ \forall \rho \tag{6.41}$$

and

$$\frac{(\ell_{\rho_A})_1 (A\ell_{H_1})_1}{d_1} \geq \frac{\gamma_2^2}{9}, \ \frac{(\ell_{\rho_B})_1 (A\ell_{H_1})_1}{d_1} = 0 \tag{6.42}$$

This means that, using the same circuit and observable, we can get a noise-induced barren plateau (NIBP) concentration effect or not, depending on the initial state that we select. Moreover, since H_0 and H_1 are orthogonal Pauli strings, $\ell_{H_0+H_1} = \ell_{H_0} + \ell_{H_1}$, suggesting that, by using the observable $H_{tot} = H_0 + H_1$ we are guaranteed to observe both contributions when using ρ_A as an initial state.

In the rest of this section, we numerically validate the above claims. In particular, we study the behavior of $\mathbb{V}^{L}_{\rho,H}$ as for an increasing amount of layers L, changing observable H and initial state ρ to isolate different kind of contributions. Recalling Eq. (6.18), we can decompose the variance contributions like

$$\mathbb{V}^{L}_{\rho,H} = (\ell_\rho, T^L \ell_H) = \sum_z \Big[(\ell_\rho, T^L_z \ell_H) + (\ell_\rho, A^{(L)}_z \ell_H) \Big] + (\ell_\rho, Q^L \ell_H), \tag{6.43}$$

where the sum includes the non-vanishing terms in the limit of large L, while $(\ell_\rho, Q^L \ell_H)$ is the exponentially decaying contribution typical of the NIBP phenomenon, which vanishes in the limit. In what follows, we'll use the term "NIBP terms" to refer to this quantity. Thanks the choice of ρ_A and ρ_B, we have that $(\ell_{\rho_A}, Q^L \ell_{H_1}) = (\ell_{\rho_B}, Q^L \ell_{H_1})$, thus removing the ambiguity when both initial states are present in the same plot.

Furthermore, since for both $H = H_0$ and $H = H_1$, $H \in \mathcal{B}_Q$, the norm-preserving terms vanish, i.e. $T^L_z \ell_H = 0 \ \forall L$, leaving only the absorption contributions. In what follows, we'll use the term "Absorption terms to $\mathcal{B}_z$" to refer to the quantity $(\ell_\rho, A^{(L)}_z \ell_H)$. As rigorously proven in Appendix C.4, these transient quantities are related to the above analysis by the

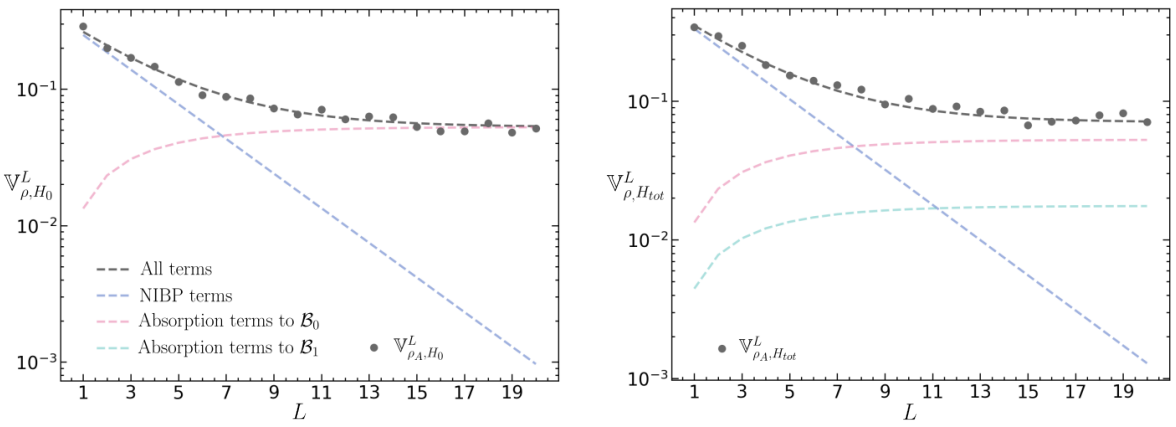

FIGURE 6.6: Example of the absorption effect in a system comprising a mixture of clean and noisy qubits, as described in Fig. 6.5. On the left, we set $H = H_0$ and $\rho = \rho_A$. For this choice of H, the absorption terms to $\mathcal{B}_{z=1}$ vanish, enabling us to isolate and verify the model's ability to accurately reproduce the non-vanishing variance contributions analogous to those discussed in the literature, but within a more general framework. On the right, we set $H = H_{\text{tot}}$ and $\rho = \rho_A$. In this case, both absorption terms are non-zero, illustrating how, in the more general setting, a combination of the two effects—initial-state-dependent and initial-state-independent—can occur. The total variance matches the sum of these contributions, in agreement with the prediction of Eq. (6.24) in the main text.

limit

$$\lim_{L\to\infty}(\ell_\rho, A_z^{(L)}\ell_H) = \frac{(\ell_\rho)_z(A\ell_H)_z}{d_z}. \tag{6.44}$$

After computing each of these terms for the current model, we estimate the variance numerically by randomly sampling $R = 200$ circuit parameters and computing the sample variance of the cost function defined in Eq. (6.1). Simulations are performed using the Pennylane software framework [113].

The results are presented in both Fig. 6.7 and Fig. 6.6. In Fig. 6.7, we numerically demonstrate the emergence of an absorption term in the variance that depends on the circuit's initial state. Specifically, we show that an NIBP effect can be suppressed by appropriately choosing the initial state ($\rho = \rho_A$) and can reappear otherwise ($\rho = \rho_B$). The agreement with the theoretical predictions indicates that this behavior is explained by absorption into a non-trivial, norm-preserving invariant subspace $\mathcal{B}_{z=1}$. This stands in stark contrast to previous descriptions of cost concentration under both unital and non-unital noise, in which the initial state played no role in determining the limiting variance. Furthermore, as shown in Fig. 6.6, we also identify an initial-state-independent contribution from absorption into the trivial subspace $\mathcal{B}_0$ (left) as well as a regime combining both effects (right).

Finally, we note that while the numerical simulations in this example are feasible due to the small size of the system, the theoretical tools from Section 6.5 used to motivate this example do not share this limitation and are, in principle, applicable to much larger systems. Furthermore, if additional details about the structure and entangling power of the intermediate channel $\mathcal{E}$ are known, absorption coefficients can be computed analytically, as shown in the following Section.

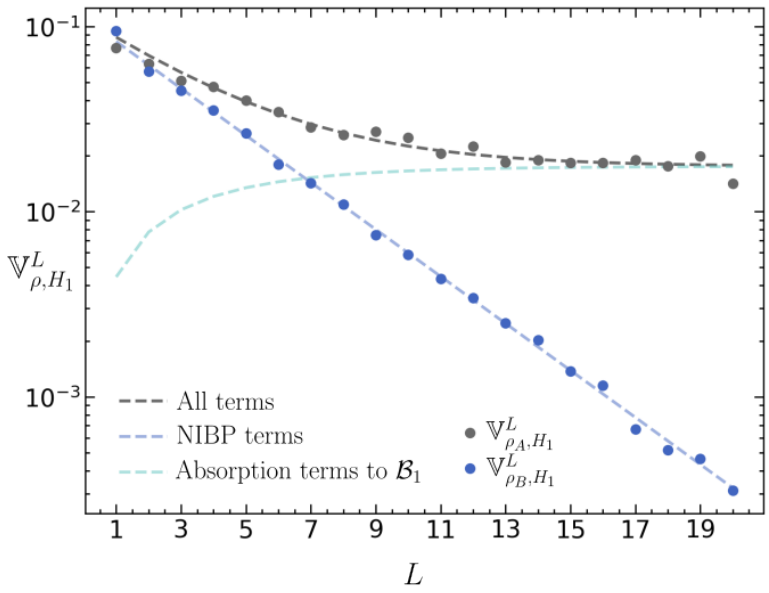

FIGURE 6.7: Example of the absorption effect in a system comprising a mixture of clean and noisy qubits. Dotted lines indicate theoretical predictions for each contribution to the variance. As predicted, $(\ell_{\rho_B})_{z=1} = 0$ in this case, so absorption occurs only for ρ_A, resulting in qualitatively distinct variance behaviors depending on the initial state.

6.6 Non-unital noise and entanglement

Exploiting the knowledge about the structure of the absorption matrix A, derived from Theorem 6.3.1, it is possible to study the scaling of the variance $\mathbb{V}^{\infty}_{\rho,H}$ as a function of the noise strength and entangling power of the unitary circuit. To do so, let us consider a non-unital map of the form

$$\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}, \tag{6.45}$$

where $\tilde{\rho} \neq \mathbb{1}/d$ is an arbitrary quantum state, $\mathcal{E}$ is a unitary channel representing the entangling operation, and p is the error probability associated to $\mathcal{E}_c$. Intuitively, we can think of the resulting channel $\Phi_{\boldsymbol{\theta}}$ as the repetition of L layers, each made up of the composition of $\tilde{\Phi}(\rho) = (1-p)\rho + p\tilde{\rho}$ and of $\mathcal{E}\circ\mathcal{U}_{\theta_l}$. Clearly $\tilde{\rho}$ is the unique fixed-point of $\tilde{\Phi}$, while since the local unitaries are assumed to form 2-designs, the only fixed point for the unital part, valid for all parameters, is the maximally mixed state $\mathbb{1}/d$. This causes the emergence of competing effects, which are the ultimate origin of the variance in such models. However, depending on the relative strength of the two effects, the behaviour of $\mathbb{V}^{\infty}_{\rho,H}$ as a function of p may vary greatly.

Assume that $\tilde{\Phi}$ reaches the deep-circuit limit rapidly, i.e., already after only a few layers. In terms of Theorem 6.3.1, this corresponds to $\beta \gg 1$, since $e^{-\beta L}$ controls the maximal deviation from the asymptotic variance. In this regime, the mixing of local subspaces induced by $\mathcal{E}$ occurs much faster than the accumulation of noise, leaving limited time for noise-induced effects to develop. We refer to this regime as *rapidly entangling*. Conversely, when $\tilde{\Phi}$ reaches the deep-circuit limit slowly, i.e., $\beta \ll 1$, the mixing of local subspaces is significantly slower than the action of noise, allowing noise-induced effects to accumulate over many layers. We refer to this regime as *slowly entangling*.

Theorem 6.3.1 allows to quantify the variance scaling in the limit of a rapidly entangling and slowly entangling channel, i.e. $\beta \to \infty$ and $\beta \to 0$ respectively. In particular, for both limits, the LTM approaches a projection, namely the dominant eigenprojection in the former

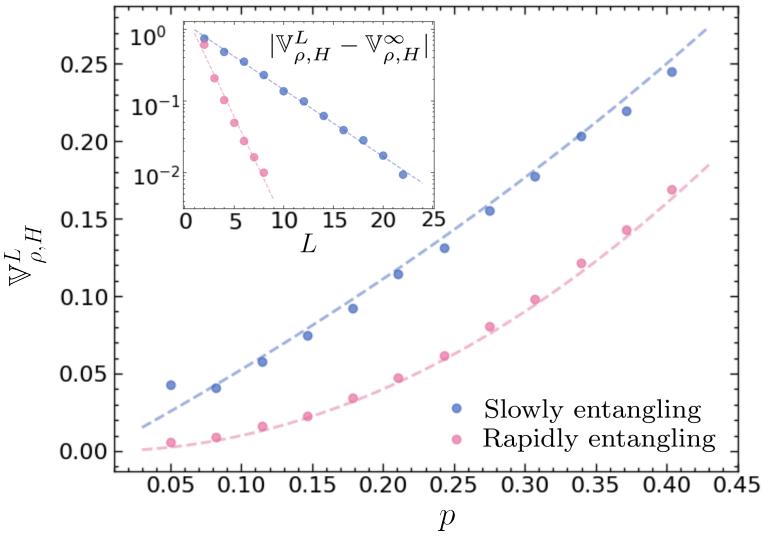

FIGURE 6.8: Scaling of $\mathbb{V}^L_{\rho,H}$ as a function of the noise strength and the entangling power of the intermediate channel. The main figure illustrates the scaling of $\mathbb{V}^\infty_{\rho,H}$ with noise strength p for both rapidly entangling (pink) and slowly entangling (light blue) channels, using $L = 8$ and $L = 20$, respectively. The dotted lines represent the theoretical predictions of Eq. (6.49) and Eq. (6.48) The inset verifies the exponential convergence of $\mathbb{V}^L_{\rho,H}$ to $\mathbb{V}^\infty_{\rho,H}$ at $p = 0.1$, justifying the chosen number of layers. The dotted lines represent an exponential fit to the numerical data. All plots are obtained using a $n = 10$ qubit system.

and the identity in the latter. In these cases, $\mathbb{V}^\infty_{\rho,H}$ is given by the following Lemma.

Lemma 6.6.1. Let $\mathcal{E}_c$ be a quantum channel of type Eq. (6.45), with $0 < p \leq 1$ and let T be the LTM of $\mathcal{E}^\dagger$. Then

$$\mathbb{V}^\infty_{\rho,H} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1} - (1-p)^2 T)^{-1} \ell_H). \tag{6.46}$$

In particular, if T is a projection, then

$$\mathbb{V}^\infty_{\rho,H} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \tag{6.47}$$

The first thing to notice is that the dependence on the initial state of $\mathbb{V}^\infty_{\rho,H}$ is completely lost: the component associated to it decays exponentially fast in the number of layers when $p \in \Theta(1)$, and as a consequence, vanishes in the limit. The remaining terms, instead, pertain to the fixed point of the noise channel $\tilde{\rho}$, and therefore still appear. In particular, the last term pertains to the very last layer, while the first collects the contribution of all preceding layers. Clearly, since the decay of these contributions is exponential, only layers where $L - l \lesssim -2 \log p$ will contribute sensibly to the variance.

Lemma 6.6.1 allows to compute the scaling as a function of p in the two opposite limits. Starting from the slowly entangling case, i.e. $T \approx \mathbb{1}$, the scaling is approximately *linear* in p. More precisely, we have

$$\mathbb{V}^\infty_{\rho,H} = \frac{p}{2-p}(\ell_{\tilde{\rho}}, \ell_H). \tag{6.48}$$

The result shows that first term in Eq. (6.47) dominates, suggesting that $\mathbb{V}^\infty_{\rho,H}$ emerges from the contribution of the last $O(-\log p)$ layers. In this sense, for fixed noise rates, only the last portions of the channels are relevant for VQAs [5]. Conversely, in the opposite limit, the dependence on p is more complex, as it now depends on the irreducible components T_z. For

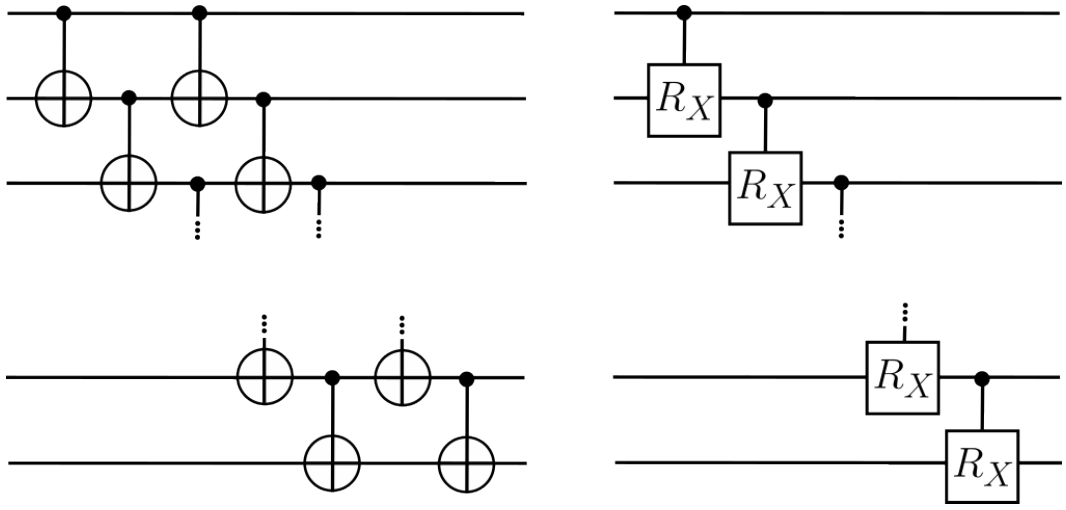

FIGURE 6.9: Entangling unitaries used in the examples. On the left, the rapidly entangling configuration is composed of a double cascade of CNOT gates, while on the right, the slowly entangling one is composed of a single cascade of controlled RX gates, where $RX(\boldsymbol{\theta}) = e^{i\boldsymbol{\theta}X/2}$ and X is the Pauli X gate. In particular, we fixed $\boldsymbol{\theta} = \pi/20$.

simplicity, if we consider the case of a highly expressive ansatz, we may take T to have only one irreducible component T_1. In this case, the last term in Eq. (6.47) dominates, and we get a *quadratic* dependence on p up to an exponentially vanishing correction, namely

$$\mathbb{V}^{\infty}_{\rho,H} = p^2(\ell_{\tilde{\rho}}, \ell_H) + O\left(4^{-n}\right). \tag{6.49}$$

This worsens the concentration, suggesting that now only the very last layer is able to produce a sizable effect, hence giving an effectively constant depth circuit.

We provide numerical evidence for the application discussed here, derived by the formalism introduced in this Chapter. We utilize Pennylane [113] to construct and optimize the PQC described in the following.

Regarding the initial state, we use $\rho = (|0\rangle\langle 0|)^{\otimes n}$ for simplicity. As for the channel, we consider maps $\mathcal{E}_c$ of the family $\mathcal{E}_c = \mathcal{N} \circ \mathcal{E}$, $\mathcal{E}_c(\rho) = (1-p)\mathcal{E}(\rho) + p\tilde{\rho}$. In particular, $\mathcal{E}$ is a unitary, entangling channel depicted in Fig. 6.9 and $p \in (0,1]$ represents the noise strength of the noise map $\mathcal{N}(\rho) = (1-p)\rho + p\tilde{\rho}$ with fixed point $\tilde{\rho}$. Specifically, we fix $\tilde{\rho}$ to be a highly entangled, pure state, i.e. the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$.

Concerning H, we fix $H = h\sum_{k=1}^{n} 2^{n/2} Z_k \otimes Z_{k+1}$, as it represents the simplest observable involving all n qubits while having a non-vanishing normalization factor $(\ell_{\tilde{\rho}}, \ell_H)$. In particular, as shown in Appendix C.6.2, if we fix $h = 9/n$, we have $(\ell_{\tilde{\rho}}, \ell_H) = 1$, which makes the scaling of $\mathbb{V}^{\infty}_{\rho,H}$ especially easy to check. Finally, the entangling part is chosen according to Fig. 6.9. While the number of layers L needed to reach convergence to $\mathbb{V}^{\infty}_{\rho,H}$ is logarithmic, as shown in Theorem 6.3.1 and numerically assessed in the inset of Fig. 6.8, the mixing speed varies depending on the entangling part. For this reason, $L = 8$ layers are sufficient in the rapidly entangling case, but $L = 20$ are necessary in the slowly entangling case. Finally, all simulations are performed using $n = 10$ qubits. The results are shown in Fig. 6.8. The plots show both the quadratic and linear scaling with p predicted by our model, as well as an exponential decay in the difference $|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}|$ for fixed noise rates. A slight deviation from the predicted scaling is observed in the slowly entangling setting for $p \approx 0$. This effect can be explained by the finite amount of layers used: the slow speed of convergence imposed by the condition $T \approx \mathbb{1}$ prevents the variance to reach the asymptotic limit $\mathbb{V}^{\infty}_{\rho,H}$, while the action of the noise is still too weak to erase the contribution of the first layers, hence deviating from

the predicted behaviour. This same phenomenon is not observed in the rapidly entangling case, as entanglement production already exponentially suppresses those contributions.

6.7 Avoiding concentrated cost functions with smart initializations

The results showed so far were based on the analysis of the dominant eigenvectors of T. When L is not deep enough to reach convergence to its asymptotic limit, we need to characterize better the behaviour of $\mathbb{V}^L_{\rho,H}$. This is the subject of the following Section.

6.7.1 Lower bounds on "shallow" circuits

While for shallow circuits we cannot rely on the spectral properties of T to determine $\mathbb{V}^L_{\rho,H}$, we can still use knowledge of the convergence speed to $\mathbb{V}^\infty_{\rho,H}$ to set general lower bounds. Intuitively, such bounds can be obtained by preventing the variance to reach its stationary state, which can be done only if the exponential upper bounds appearing in Theorem 6.3.1 are sufficiently loose, namely $\beta L \in O(\log n)$. This implies either that the circuit is *shallow*, i.e. there are not enough layers to reach the asymptotic value $\mathbb{V}^\infty_{\rho,H}$, or *effectively shallow*, i.e. the mixing speed β of T is slowed down according to $\beta \in O(\log n/L)$ so that $\mathbb{V}^\infty_{\rho,H}$ is never reached, regardless of L. As this speed is related to the amount of correlations with respect to the partition introduced by the channel, we study this scenario in the limit of $\mathcal{E}^\dagger$ being close to locality preserving. This idea is formally captured in the following Theorem, which extends its application to generally non-homogeneous layered, locally random channels.

Theorem 6.7.1 (General lower bound)**.** Let $\rho, H \in \mathcal{B}$ and consider a sequence of quantum channels $\{\mathcal{E}_l\}_{l=1}^L$, and let $\{T_l\}_{l=1}^L$ be the respective LTMs. Finally let $K \subset \{0,1\}^M$ denote a subset of indices, and by $\alpha_l = \min_{\kappa\in K}(T_l)_{\kappa,\kappa}$. Then

$$\mathbb{V}^L_{\rho,H} \geq \alpha^L(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{6.50}$$

where $\mathcal{K} : \mathcal{B} \to \mathcal{B}_K$ is the projector onto $\mathcal{B}_K = \oplus_{\kappa\in K}\mathcal{B}_\kappa$ and α is the geometric mean of α_l.

Depending on the scaling of α and the dimensions d_m of the subsystems, Eq. (6.50) can provide a meaningful lower bound. For instance, focusing on the case of subsystems with constant dimension, we have the following Corollary.

Corollary 6.7.1 (Lower bound examples)**.** Let $\mathcal{H} = \otimes_{m=1}^M \mathcal{H}_m$, $d_m \in \Theta(1)$. If either of the conditions

(a) $\alpha > 0$, $\alpha \in \Omega(1)$ and $L \in O(\log n)$,

(b) $\alpha = 1 - f(n, L)$, $f \in O(\log n/L)$ and $L \in \Omega(\log^{1+\epsilon} n)$ for some arbitrary $\epsilon > 0$

is satisfied, then

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{6.51}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

The conditions of Corollary 6.7.1 reflect the aforementioned scenarios; in particular, condition (a) ensures the absence of concentration for *shallow* circuits, both unitary and noisy. Specifically, this holds true whenever $0 < \alpha \in \Omega(1)$, indicating that the intermediate channel does not become increasingly rapidly entangling, as the problem size grows. As a notable example, this condition is satisfied for brickwall circuits, equipped with local noise and a

local observable [5]. Similarly, condition (b) reflects the absence of concentration for *effectively shallow* quantum channels. A significant example is that of deep finite local-depth circuits (FLDCs) [128]. We can interpret condition (b) as a limit to the mixing speed of T by noting that, in the homogeneous case, it is equivalent to the more explicit relation $|1-\lambda| \in O(\log n/L)$ for all the eigenvalues λ of T by Gershgorin circles theorem [70], which directly implies $\beta \in O(\log n/L)$.

Interestingly, since these results have been obtained by imposing $T \approx \mathbb{1}$, they have a strong resemblance with small angle initialization strategies [126, 50], which similarly hinge on identity manipulation. In fact, while the primary concern of Theorem 6.7.1 is noise, it can be seen as a theoretical foundation of such initialization strategies. In the following Section we discuss in details the strong relation between smart initializations for noiseless circuits and the properties of noisy layers. In doing so, we analytically estimate the scaling of α for a broad class of circuits, hence providing rigorous lower bounds on $\mathbb{V}^L_{\rho,H}$.

6.7.2 Connection with small angle initializations

Here we expand on the concept of small angle initialization introduced in Refs. [50, 126]. In particular, we establish a general relationship between the insights gained from controlling cost concentration in noisy circuits (as presented in Theorem 6.7.1) and BP mitigation strategies that typically only apply to ideal, unitary circuits.

The main idea behind small angle initialization strategies considers a layered quantum circuit $U_{\boldsymbol{\theta}} = \prod_l U_{\theta_l}$ where absence of concentration for a given initialization distribution is shown. This is typically very peaked around 0, with variance scaling inversely to the number of layers: $\sigma^2 \in O(1/L)$. The core idea of all such strategies relies on *identity manipulation*, i.e. on choosing initialization distributions such that $U_{\boldsymbol{\theta}} \approx \mathbb{1}$ with high probability. This introduces a sizable variance to the circuit, at the price of having a large bias towards the identity in the quantum model.

A similar structure can be defined in our framework by considering quantum channels $\Phi_{\boldsymbol{\theta},\boldsymbol{\phi}}$, as in Definition 6.1.3, where the intermediate channels $\mathcal{E}_{\phi_l}$ are now parameterized. This differs from the main idea of small angle initialization, as the local components $\mathcal{U}_{\theta_l}$ of the channel remain Haar random, and instead it is the allowed channels $\mathcal{E}_\phi$ that get restricted. Intuitively, this will lead to a different model bias for small angles, i.e. $\Phi_{\boldsymbol{\theta},\boldsymbol{\phi}} \approx \mathcal{U}_{\boldsymbol{\theta}}$. We name this model Quantum Residual Network (QResNet), as we interpret the large identity component of $\mathcal{E}_\phi$ as a *skip-connection*, in analogy to classical Residual Networks [40]. Indeed, this structure is enough to avoid concentration when a small angle initialization strategy is used on the parameter ϕ. To see this, we remark that while a constant channel was used to derive Proposition 6.2.1, it can be readily generalized to parameterized channels $\mathcal{E}_\phi$, as long as the parameters ϕ_l are independent. In that case it is sufficient to use $\mathbb{E}_\phi\{T_\phi\}$ in place of T, where T_ϕ is the LTM of $\mathcal{E}^\dagger_\phi$. Indeed, if each parameter in ϕ is sampled independently of the other parameters, including $\boldsymbol{\theta}$, then by linearity we have

$$\begin{aligned}\mathbb{V}^L_{\rho,H} &= \mathbb{E}_\phi \mathbb{E}_{\boldsymbol{\theta}} \left\{ \mathrm{Tr}\left\{ \Phi_{\boldsymbol{\theta},\phi}(\rho) H \right\}^2 \right\} \\ &= \mathbb{E}_\phi \left\{ (\ell_\rho, T_\phi \ell_H) \right\} = (\ell_\rho, \mathbb{E}_\phi\{T_\phi\} \ell_H).\end{aligned} \tag{6.52}$$

Exploiting this, we can derive the following Proposition.

Proposition 6.7.1 (QResNet)**.** Let $\mathcal{E}_\phi(\cdot) = e^{i\phi G} \cdot e^{-i\phi G}$ be a unitary entangling gate with normalized generator, i.e. $\|G\|_2^2/d = 1$. Furthermore, let μ, σ^2 be the mean and variance of the initialization distribution of ϕ. Then, if $\mu = 0$, $\sigma^2 \in O(\log n/L)$, and $L \in \Omega(\log^{1+\epsilon} n)$ we have

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H), \tag{6.53}$$

where $F(n) \in \Omega(1/\text{poly}(n))$.

This result represents an application of Corollary 6.7.1 (b) in the case of unitary, parameterized intermediate channels.

Note however that Theorem 6.7.1, which is the backbone of Proposition 6.7.1, is not limited to unitary circuits, but is applicable to generic quantum channels. Indeed, if we take $\mathcal{E}$ to be a noise model, Corollary 6.7.1 (b) may be analogously interpreted as a condition on the noise rates to avoid concentration. This showcases the connection between BP-free QResNets and noise models escaping NIBP, which can be interpreted as follows: if a noise map satisfies Corollary 6.7.1, then there exists a QResNet associated to it that is able to avoid BP. As an example, assume that the channel $\mathcal{E} = e^{\Delta t \mathcal{L}}$ is obtained as a solution of the Lindblad equation at time $\Delta t \ll 1$, where $\mathcal{L}(\rho) = \sum_i L_i \rho L_i - 1/2\{L_i^2, \rho\}$, $L_i = L_i^\dagger \, \forall i$, i.e. we consider weak Lindbladian noise [23]. In this case, there always exist a unitary, stochastic unravelling $\{U_\phi\}_\phi$, such that $\mathcal{E}(\rho) = \mathbb{E}_\phi\{U_\phi \rho U_\phi^\dagger\}$ [2]. Then, the ensemble corresponding to the choice $\mathcal{E}_\phi(\rho) = U_\phi \rho U_\phi^\dagger$ will avoid BP if Δt is small enough. Formally, we have the following result.

Proposition 6.7.2 (General QResNets from a quantum channel)**.** Let $\mathcal{E}(\rho) = e^{t\mathcal{L}}(\rho)$ be a quantum channel generated by the Lindbladian

$$\mathcal{L}(\rho) = \sum_k \gamma_k \left(L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\} \right) \tag{6.54}$$

with $\|L_k\|_2^2/d = 1$. Then, when $L \in \Omega(\log^{1+\epsilon} n)$, the ensemble generated by $\mathcal{E}_\phi = E_\phi \cdot E_\phi^\dagger$,

$$E_\phi = \prod_k e^{i\phi_k L_k + \sigma_k^2 (L_k^2 - L_k^\dagger L_k)/2} \tag{6.55}$$

and ϕ_k normally distributed with $\mu = 0$, $\sigma_k^2 \in O\left(\frac{\gamma_k}{\sum_{k'} \gamma_{k'}} \frac{\log n}{L}\right) \forall k$ satisfies

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho, \ell_H) \tag{6.56}$$

where $F(n) \in \Omega(1/\text{poly}(n))$.

If $L_k = L_k^\dagger$, we get a unitary QResNet as in the previous discussion. Interestingly, Proposition 6.7.2 can be equally applied if the ensemble is not unitary, which can happen if the noise channel $\mathcal{E}$ is non-unital (e.g. $L_k \neq L_k^\dagger$). This effectively extends the framework of small angle initialization to non-unitary quantum models, e.g. those based on linear combination of unitaries (LCU) or analogous techniques [48]. Finally, it is important to note that, while a *specific* unraveling was used to derive Proposition 6.7.2, the stochastic unraveling of a given channel is not unique, allowing multiple QResNets (unitary or non-unitary) to be derived from the same channel. Additionally, different choices among such models may lead to distinct variances [125], all bounded from below by the variance of $\mathcal{E}$. Indeed, any stochastic unraveling is connected to the physical channel solely through their first moment, while higher-order moments are relevant only in the context of the associated QResNet, as they do not represent otherwise measurable quantities.

6.8 Discussion

The study of cost function concentration is a central topic in variational quantum computing. While the description of this effect in the absence of noise has been recently formulated using Lie-algebraic theory [88, 33, 73], this approach inevitably fails in the general

setting of non-unitary circuits, where the group structure description is lost. In this Chapter, we employ non-negative matrix theory to derive a general expression for $\mathbb{V}^L_{\rho,H}$ in quantum circuits composed of local 2-designs interleaved with arbitrary quantum channels. This circuit architecture is chosen for its combination of analytical accessibility and physical relevance: it is rich enough to capture qualitatively novel variance behavior, while structured enough to perform analytical calculations.

In particular, within this framework, the variance is formulated in terms of the channel's locality transfer matrix (LTM), which in turn enables the precise computation of $\mathbb{V}^\infty_{\rho,H}$ in the deep-circuit limit. Specifically, the observed structure of $\mathbb{V}^\infty_{\rho,H}$ brings out a new mechanism, which we call *absorption*, whereby components of H pertaining to strictly contractive subspaces of $\mathcal{B}_Q$ can augment the variance of the model by coupling with non-contractive ones through the non-reversible action of the quantum noise channel. Such result is supported by numerical simulations, provided in Section 6.5, Section 6.6 and Section 5.4.

This indicates that a mixed configuration of ideal and noisy qubits, as well as a partial realization of error correction, could potentially outperform purely ideal or purely noisy systems in terms of variance, and highlights novel properties of $\mathbb{V}^\infty_{\rho,H}$ in this setting, e.g. the initial state dependence of noise-induced concentration, which is crucial for algorithm engineering. This is especially important in the early stages of fault-tolerant quantum computing, where a large quantity of noisy qubits can be utilized, but the availability of ideal qubits is still constrained [85, 51]. This may also be beneficial in characterizing variational quantum algorithms executed on hardware, where different qubits may experience varying error rates [23]. In this context, $\mathbb{V}^\infty_{\rho,H}$ can be used to approximate $\mathbb{V}^L_{\rho,H}$ in regimes where L is sufficiently large to significantly impact some qubits but not yet others.

Furthermore, this approach clarifies the role of unitality in noise-induced barren plateaus, shifting the focus on the more relevant aspect of *contractivity* and *alignment* between invariant subspaces of the unitary layers $\mathcal{U}_\theta$ and strictly contractive subspaces of $\mathcal{E}$.

Additionally, we introduced a general lower bound on the variance of noisy circuits. This bound is derived by restricting the mixing speed of $\mathbb{V}^L_{\rho,H}$ to prevent it from reaching its asymptotic limit. We subsequently employ this approach to introduce QResNets, an initialization strategy analogous to small angle initializations [50, 126, 83, 18] that can effectively mitigate the occurrence of barren plateaus. Moreover, we demonstrate that an analogous procedure can be applied to noise maps, establishing a formal connection between weak noise and QResNets. This enables us to derive BP-free QResNets as stochastic unravellings of sufficiently weak noise maps. Notably, since these unravellings are not necessarily unitary, this approach can yield BP-free architectures beyond unitary circuits, encompassing more complex models such as those employing linear combinations of unitaries (LCU) [48].

Future research may extend our findings by relaxing the local 2-design property of the unitary layer. Such a generalization would broaden the domain of applicability of our results beyond hardware-specific designs, but it would primarily offer a technical refinement without altering the conceptual insights gained in this Chapter. Furthermore, recent studies have established a connection between the absence of concentration and classical simulability for both ideal [61] and noisy [5] quantum circuits. While these results often focus on strictly contractive noise models, our work suggests a potential avenue for combining these concepts, extending their validity to more complex noise environments. A representation of our contributions is depicted in Fig. 6.1: the non negative matrix approach offers a new and timely answers to several open questions related to concentration phenomena in quantum circuits and provides valuable insights into the optimal utilization of near-term and early fault-tolerant quantum devices, thus guiding the community towards the effective application of the variational quantum computation framework.

Chapter 7

Conclusions and outlook

This Thesis has explored the interplay between noise, trainability, and scalability in variational quantum algorithms, combining concrete applications on near-term devices with a general theoretical analysis of their fundamental limitations. The unifying theme throughout has been the role of structure, both of noise and of the circuit, in determining whether variational approaches can remain effective as system size and circuit depth increase.

In the first part of this work, we focused on practical applications of variational quantum algorithms in noisy settings. Following a brief introduction to noisy quantum computation in Chapter 1 and Chapter 2, variational methods are applied in Chapter 3 to detect quantum phase transitions through level spectroscopy in the one-dimensional J_1–J_2 Heisenberg model. By explicitly exploiting symmetries of the Hamiltonian and restricting the optimization to selected symmetry sectors, we demonstrated that low-lying excited states can be reliably approximated with shallow, structured ansätze. Crucially, this approach allowed the identification of a Berezinskii–Kosterlitz–Thouless transition using relatively small system sizes, highlighting how physically motivated constraints can dramatically enhance the efficiency of variational algorithms. The incorporation of error mitigation techniques, such as Zero-Noise Extrapolation and a posteriori symmetrization, further showed that meaningful physical information can be extracted even in the presence of realistic noise.

Chapter 4 addressed a complementary, hardware-oriented application: the variational correction of coherent calibration errors. By leveraging the structure of stabilizer states, we isolated coherent errors from other noise sources and constructed a variational procedure capable of compensating miscalibrations in a robust manner. This study emphasized that variational algorithms need not be limited to solving abstract optimization problems, but can also be employed as adaptive tools for characterizing and correcting imperfections in quantum hardware. Importantly, the analysis showed that the effectiveness of such methods persists even when incoherent noise is present.

While these applications demonstrate the promise of variational methods on near-term devices, the second part of the Thesis addressed their most severe obstacle to scalability: the cost concentration, or barren plateau, phenomenon. Chapter 5 and Chapter 6 developed a comprehensive theoretical framework to analyze cost concentration under arbitrary noise processes. By introducing the Locality Transfer Matrix (LTM) formalism, we provided a unified description that encompasses both unitary and noisy circuits, extending and generalizing existing results in the literature. This framework made it possible to identify the precise mechanisms through which noise, depth, and circuit architecture conspire to suppress cost-function variance and gradients.

A key insight emerging from this analysis is that cost concentration is not an unavoidable consequence of increasing system size, but rather a structural property of the chosen ansatz and noise model. In particular, the distinction between strictly contractive and non-strictly contractive dynamics, as well as the role of non-unital noise, proved crucial in determining the asymptotic behavior of the cost variance. These results clarify why many highly expressive, hardware-efficient ansätze suffer from severe trainability issues, especially in noisy

settings, and why naive increases in depth or expressivity can be counterproductive.

Within this general framework, we identified a class of architectures termed *Quantum Residual Networks* (QResNets), which are inspired by residual connections in classical deep learning. By effectively limiting the accumulation of contractive effects across layers, QResNets were shown to preserve meaningful cost-function variance and gradients even at large depths. From a theoretical perspective, these architectures represent explicit counterexamples to the notion that deep variational circuits must necessarily exhibit barren plateaus. From a practical standpoint, they provide concrete design principles for constructing scalable variational ansätze that remain trainable in the presence of noise.

Taken together, the results of this Thesis suggest a coherent picture bridging practical algorithm design and fundamental scalability limits. The successful applications discussed in the first part can be reinterpreted, in light of the theoretical analysis, as instances where structure—symmetry constraints, shallow effective depth, or restricted entanglement growth—naturally avoids the most severe forms of cost concentration. Conversely, the general theory developed in the second part explains why such structure is not merely beneficial, but essential, for scalability.

Looking forward, these insights have several implications for the design of future quantum algorithms. First, variational algorithms are likely to remain relevant well beyond the near-term era, provided that their architectures are informed by rigorous trainability analyses. The QResNet paradigm points toward a class of hybrid quantum-classical models where depth can be increased without sacrificing optimization viability, opening the door to more expressive yet controllable quantum circuits. Second, error mitigation strategies should not be viewed as purely post-processing tools, but as ingredients that can be co-designed with the ansatz itself, shaping the effective noise model seen by the optimization landscape. Finally, the mathematical framework introduced here offers a foundation for systematically evaluating new algorithmic proposals, allowing one to anticipate scalability bottlenecks before they arise in practice.

Appendix A

Additional material on the variational estimation of phase transitions

A.1 Initial states

Singlet state $|\phi_0\rangle$

The initial state representing factorized singlet pairs, as given in Eq. (3.8), is easily prepared on a quantum computer. Simultaneously for each singlet pair $2r-1, 2r$, we prepare the state $|1,+\rangle$ using Pauli X and Hadamard H gates, followed by a CNOT gate (refer to Fig. A.1). In general, we dub U_s the unitary transformation responsible for generating the product of singlet state on n qubits.

Triplet state $|\phi_1\rangle$

The triplet states in Eq. (3.9) are not two-site translation invariant. To get the invariant state $|\phi_1\rangle$ in Eq. (3.10) we use a particular class of entangled states called W states [24]. The W state for n qubits is the superposition of all the possible n-qubit states with just a single qubit in the state $|1\rangle$

$$\left|W^{(n)}\right\rangle = \frac{1}{\sqrt{n}}\big(\,|10\ldots0\rangle + |01\ldots0\rangle + \cdots + |00\ldots1\rangle\,\big). \tag{A.1}$$

As an example for the preparation of $|\phi_1\rangle$, we consider the case of $n = 4$ qubits. We set qubits 2 and 4 in the state $|00\rangle$ and qubits 1 and 3 in the W state, that reads

$$\left|W^{(2)}\right\rangle_{13} = \frac{1}{\sqrt{2}}\big(\,|10\rangle_{13} + |01\rangle_{13}\,\big). \tag{A.2}$$

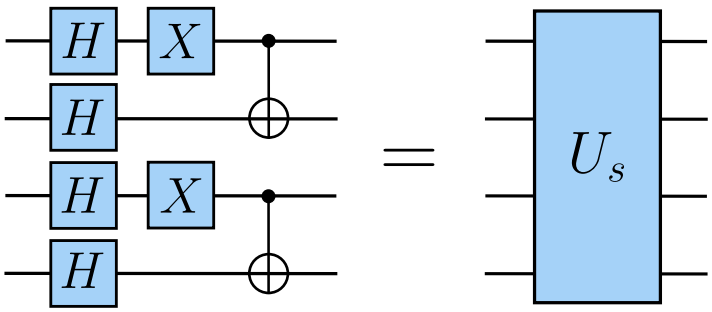

FIGURE A.1: Circuit U_s for preparing the $|\phi_0\rangle$ state on 4 qubits.

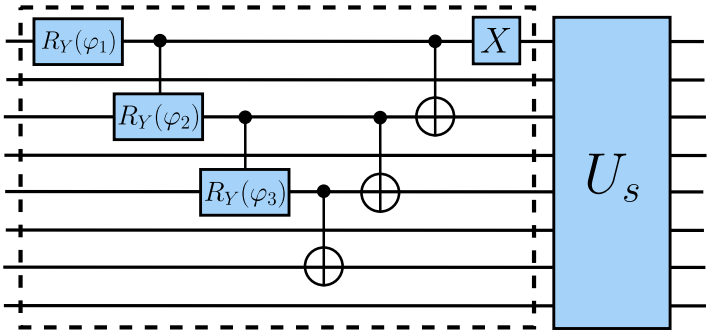

FIGURE A.2: Circuit for preparing the $|\phi_1\rangle$ state on 8 qubits.

Then, the total state is expressed as

$$\left|W^{(2)}\right\rangle_{13} |00\rangle_{24} = \frac{1}{\sqrt{2}}\big(|1000\rangle + |0010\rangle \big) , \tag{A.3}$$

where in the right-hand side of Eq. (A.3) the states of single qubits are indexed in ascending order. By applying the unitary U_s to the state in Eq. (A.3) we get

$$\frac{1}{\sqrt{2}}\big(|t\rangle_{12}|s\rangle_{34} + |s\rangle_{12}|t\rangle_{34}\big) = \frac{1}{\sqrt{2}}\Big(\big|\tilde{\phi}_1^1\big\rangle + \big|\tilde{\phi}_1^2\big\rangle\Big), \tag{A.4}$$

where the right-hand side of Eq. (A.4) is $|\phi_1\rangle$ for $n = 4$ qubits. Thus in general, to get $|\phi_1\rangle$ all odd indexed qubits are prepared in the W state and this transformation is followed by the singlet preparation U_s on all qubits. An example of the resulting circuit for eight qubits is shown in Fig. A.2, where the preparation of the W state is performed efficiently with the techniques described in Ref. [68]. The angles φ_i in Fig. A.2 are defined as $\varphi_i = 2\arccos\big(1/\sqrt{n-i+1}\big)$ where i is the index of the qubit.

A.2 Alternative choice of H_1 and H_2

As detailed in Section 3.3.1, we choose the Hamiltonians H_1 and H_2 to define the variational circuit in Eq. (3.3) as H_{odd} and H_{even}, respectively. Together, they constitute the nearest-neighbor contribution in the J_1-J_2 Heisenberg model. However, an alternative choice for H_1 and H_2 is feasible, incorporating contributions from the next-nearest neighbor term $\sum_{r=1}^{n} \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+2}$. For $J_2/J_1 = 0.5$, the Hamiltonian in Eq. (3.1) corresponds to the exactly solvable Majumdar–Ghosh model [63], denoted here as H_{MG}. In this model, the ground state is represented by a product of singlet pairs, and the first excited state can be obtained by replacing one singlet pair with a triplet one, analogously to the Hamiltonian H_{even}. Consequently, the J_1-J_2 Hamiltonian can be expressed as

$$H_{J_1\text{-}J_2} = H_{MG} + \left(J_2 - \frac{J_1}{2}\right) \sum_{r=1}^{n} \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+2} \, . \tag{A.5}$$

Accordingly, we can define a variational quantum circuit in the form of Eq. (3.3) setting $H_2 = H_{\text{MG}}$ and $H_1 = \sum_{r=1}^{n} \boldsymbol{S}_r \cdot \boldsymbol{S}_{r+2}$.

In Fig. A.3, we illustrate a comparison of the accuracy of the variational states with the two different choices of H_2 and H_1. Specifically, we plot the relative error of the ground state energy for a cluster of $n = 10$ sites as a function of the optimization steps. The left panel represents the results for a small value of the frustration ratio $J_2/J_1 = 0.1$, while the right panel corresponds to a large value $J_2/J_1 = 0.8$. For each frustration ratio, multiple

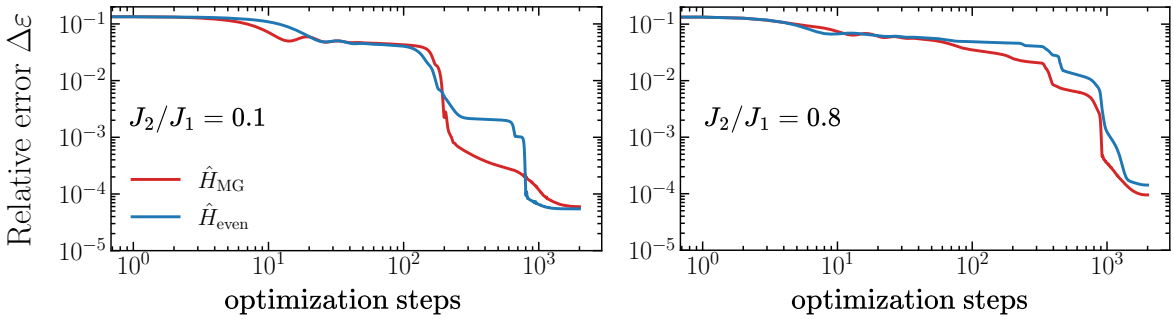

FIGURE A.3: Comparison of the accuracy of variational energies, relative to exact values, for the states defined by $H_{\rm even}$ and $H_{\rm MG}$ as a function of the optimization steps. Optimizations are performed on a cluster of $n = 10$ sites for $J_2/J_1 = 0.1$ (left panel) and $J_2/J_1 = 0.8$ (right panel). The mean behaviours over multiple optimizations are represented by solid red and blue lines for the variational states related to $H_{\rm MG}$ and $H_{\rm even}$, respectively. Orange and light-blue intervals indicate the corresponding standard deviations.

optimizations are performed using $L = n/2$ layers for $J_2/J_1 = 0.1$ and $L = N$ layers for $J_2/J_1 = 0.8$. The relative error of the two variational states is of the same order of magnitude ($\Delta\varepsilon \approx 0.01\%$) at the end of the optimizations.

Given that the variational state defined using $H_{\rm MG}$ includes gates that directly entangle next-nearest neighbors qubits, we might expect that it is more accurate for large values of the frustration ratio ($J_2/J_1 > 0.5$) compared to the one defined in Section 3.3.1. However, even for $J_2/J_1 = 0.8$, the two variational states yield similar results. This could be attributed to the fact that increasing the number of layers L in the circuit also the state defined through $H_{\rm even}$ entangles (indirectly) next-nearest neighbor sites. As a result, the two unitaries produce comparable variational results.

Moreover, by parameter sharing among different qubits the circuit corresponding to the Ansatz with $H_{\rm MG}$ is translational invariant of four lattice sites. Consequently, constructing a translational invariant state with defined momentum $k = 0$ or $k = \pi$ requires a more computationally expensive LCU, since now we must use $A = 4$ in the discussion of Section 3.4.3. Therefore, given the similar accuracy of the two states, the numerical calculations in this work were carried out using the the circuit defined in the main text.

A.3 Details on the noise models

We report the Kraus maps used for the noisy simulations in the main text. The Kraus map of single-qubit depolarization reads [9, 77]

$$\mathcal{E}^D_\tau(\rho) = \Big(1 - \frac{3}{4}p_\tau\Big)\rho + \frac{p_\tau}{4}X\rho X + \frac{p_\tau}{4}Y\rho Y + \frac{p_\tau}{4}Z\rho Z\,, \tag{A.6}$$

where ρ is the density matrix of the single-qubit, X, Y, Z are the Pauli matrices and p_τ quantifies the probability of having a bit flip, a phase flip or a bit-phase flip of the states of the computational basis. We assume a behaviour in time of the form $p_\tau = (1 - e^{-\gamma_d \tau})$ for a characteristic time $T_d = 1/\gamma_d$ [64]. The Kraus map of the single-qubit thermal relaxation is given by [9, 77]

$$\mathcal{E}^R_\tau(\rho) = K\rho K + p^{(1)}_\tau \sigma^- \rho \sigma^+ + p^{(z)}_\tau Z\rho Z + p^{(1)}_\tau \mathcal{P}_0 \rho \mathcal{P}_0 \tag{A.7}$$

where we define the operators $K = \sqrt{1 - p^{(1)}_\tau - p^{(z)}_\tau}\,\mathbb{1}$, $\mathcal{P}_0 = |0\rangle\langle 0|$ and σ^-, σ^+ are the lowering and raising Pauli operators, respectively. Here, $p^{(1)}_\tau = (1 - e^{-\tau/T_1})$ is the probability of reset

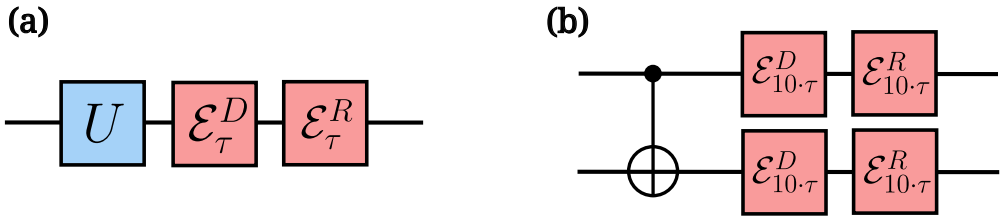

FIGURE A.4: Schematic depiction of the noise model. In panel (a) we show quantum channels associated to a generic single-qubit gate U and in panel (b) those associated to CNOT gates.

to $|0\rangle$ and T_1 the relaxation time. Additionally, we introduce $p_\tau^{(z)} = (1 - p_\tau^{(1)}) \cdot p_\tau^{(pd)}$, where $p_\tau^{(pd)} = (1 - e^{-\tau/T_{pd}})$ is the probability of pure dephasing with $T_{pd} = T_1 T_2/(2T_1 - T_2)$ and T_2 the decoherence time. The time scales T_1 and T_2 are related as $T_2 \leq 2T_1$ [64].

The circuits used in our simulations are transpiled into the native gate set of IBM devices, i.e., $\{RZ(\alpha), X, \sqrt{X}, \text{CNOT}\}$, where $\alpha \in [-\pi, \pi]$ is a rotation angle. Typically the duration of the execution of CNOT gate is 10 times larger with respect to the single-qubit gate time. We set the noise time scales as $T_1 = T_2 = T_d \approx 10^{-4}$ s, compatible with the average values of current IBM devices (see Fig. A.4).

Appendix B

Additional material on the variational correction of coherent errors

B.1 Useful properties of Pauli maps and intermediate results

We start the section by introducing the concept of Pauli map, and later show some of their useful properties. One of the main theoretical advantages of Pauli maps, is that they are self-adjoint with respect to the Hilbert-Schimdt scalar product, i.e. $\mathcal{P} = \mathcal{P}^\dagger$. Additionally, they act on Pauli strings simply by rescaling them. This idea is formalized in the following Lemma.

Lemma 2 (Action of Pauli maps on Pauli strings)**.** Given a Pauli map $\mathcal{P}$ and a Pauli string S, the action $\mathcal{P}^\dagger(S)$ is given by

$$\mathcal{P}(S) = \mathcal{P}^\dagger(S) = \chi S, \tag{B.1}$$

for some $\chi \in \mathbb{R}$, $|\chi| \leq 1$. In particular, $\chi = 1 - 2\sum_{j:\{S,P_j\}=0} p_j$.

Proof. The proof is based on the observation that any pair of Pauli string either commutes or anti-commutes with each other. More specifically, given two Pauli strings P_j and S, then if $[P_j, S] \neq 0 \Rightarrow \{P_j, S\} = 0$ and vice versa, as it can be directly verified using the commutation relations of Pauli matrices. Consequently it follows that

$$P_j S P_j = \begin{cases} S & \text{if} \quad [S, P_j] = 0 \\ -S & \text{if} \quad \{S, P_j\} = 0 \end{cases} \tag{B.2}$$

By following Definition 3, we have

$$\mathcal{P}(S) = \sum_{j=0}^{4^n-1} p_j P_j S P_j = \left(\sum_{j:[S,P_j]=0} p_j - \sum_{j:\{S,P_j\}=0} p_j \right) S = \left(1 - 2 \sum_{j:\{S,P_j\}=0} p_j \right) S = \chi S, \tag{B.3}$$

where the condition $|\chi| \leq 1$ is ensured by the relation

$$0 \leq \sum_{i:\{S,P_j\}=0} p_j \leq 1. \tag{B.4}$$

The first equality directly follows from the relation $\mathcal{P} = \mathcal{P}^\dagger$. □

Another useful property of Pauli maps, is that they preserve their structure under conjugation with Clifford unitaries. More specifically, we have the following Lemma.

Lemma 3 (Pauli and unitary Clifford maps)**.** Given a Pauli map $\mathcal{P}$ and a Clifford unitary $\mathcal{U}$, then we have

$$\mathcal{U}\mathcal{P}\mathcal{U}^\dagger = \mathcal{P}', \text{ or equivalently } \mathcal{U}\mathcal{P} = \mathcal{P}'\mathcal{U}, \tag{B.5}$$

where the new map $\mathcal{P}'$ is still of the form of a diagonal Pauli map.

Proof. We prove the statement by showing that $\mathcal{U}\mathcal{P}\mathcal{U}^\dagger(\rho) = \mathcal{P}'(\rho)\ \forall\rho$. In particular, we use the definitions of both $\mathcal{P}$ and $\mathcal{U}$ to achieve the following chain of equalities

$$\begin{aligned}\mathcal{U}\mathcal{P}\mathcal{U}^\dagger(\rho) &= U\left(\sum_{j=0}^{4^n-1} p_j P_j \left(U^\dagger\rho U\right) P_j\right)U^\dagger = \sum_{j=0}^{4^n-1} p_j \left(UP_jU^\dagger\right)\rho\left(UP_jU^\dagger\right)\\ &= \sum_{j=0}^{4^n-1} p_j P_j'\rho P_j' = \sum_{j=0}^{4^n-1} p_j' P_j\rho P_j = \mathcal{P}'(\rho),\end{aligned} \tag{B.6}$$

where we used the property that, for any Clifford unitary U and Pauli string P_j, $UP_jU^\dagger = P_j'$ is still a Pauli string. □

An similar property also holds for the generators of unitary superoperators.

Lemma 4. Given a generic unitary super-operator $\mathcal{U}$ and a generic generator $\mathcal{H}$ defined by the Hamiltonian H, as $\mathcal{H}(\rho) = [H,\rho]$, we have

$$\mathcal{U}\mathcal{H}\mathcal{U}^\dagger = \mathcal{H}' \text{ or equivalently } \mathcal{U}\mathcal{H} = \mathcal{H}'\mathcal{U}, \tag{B.7}$$

where $\mathcal{H}'$ is defined as $\mathcal{H}'(\rho) = [\mathcal{U}(H),\rho]$.

Proof. This Lemma follows immediately from the definitons of $\mathcal{U}$ and $\mathcal{H}$ by the following chain of equalities:

$$\begin{aligned}\mathcal{U}\mathcal{H}\mathcal{U}^\dagger(\rho) &= U[H,U^\dagger\rho U]U^\dagger = UHU^\dagger\rho UU^\dagger - UU^\dagger\rho UHU^\dagger\\ &= UHU^\dagger\rho - \rho UHU^\dagger = [UHU^\dagger,\rho] = [\mathcal{U}(H),\rho] = \mathcal{H}'(\rho)\end{aligned} \tag{B.8}$$

where $\mathcal{H}'$ is defined as $\mathcal{H}'(\rho) = [\mathcal{U}(H),\rho]$. □

We now show some useful upper-bounds to compute derivatives in the later sections.

Lemma 5. Given a quantum state ρ, a Pauli string S, a diagonal Pauli map $\mathcal{P}$, and a super-operator $\mathcal{H}$, defined by the Hamiltonian H, as $\mathcal{H}(\rho) = [H,\rho]$, we have the following upper bound

$$|\mathrm{Tr}\{S\mathcal{H}\mathcal{P}(\rho)\}| \le \|H\|_2\|[S,\rho]\|_2\,, \tag{B.9}$$

where $\|\cdot\|_2$ is the operator 2-norm. In particular, if $[S,\rho] = 0$, then $\mathrm{Tr}\{S\mathcal{H}\mathcal{P}(\rho)\} = 0$.

Proof. We begin our proof by expanding the Hamiltonian H in the Pauli basis, namely $H = \sum_k h_k P_k$. Using this decomposition and since $\mathrm{Tr}\{S\mathcal{H}(\cdot)\} = \mathrm{Tr}\{\mathcal{H}^\dagger(S)\cdot\} = \mathrm{Tr}\{[S,H]\cdot\}$ we get

$$\mathrm{Tr}\{[S,H]\mathcal{P}(\rho)\} = \sum_k h_k\,\mathrm{Tr}\{[S,P_k]\mathcal{P}(\rho)\}. \tag{B.10}$$

Note that, since both S and P_k are Pauli strings, also their commutator has the same structure. This can be easiliy derived from the defining relations of Pauli matrices. By Lemma 2, we have that $\mathcal{P}^\dagger([S,P_k]) = \chi_k[S,P_k]$. In particular, this gives

$$\begin{aligned}\sum_k h_k\,\mathrm{Tr}\{[S,P_k]\mathcal{P}(\rho)\} &= \sum_k \chi_k h_k\,\mathrm{Tr}\{[S,P_k]\rho\} = \sum_k \chi_k h_k\,\mathrm{Tr}\{P_k[S,\rho]\}\\ &= \mathrm{Tr}\left\{\left(\sum_k \chi_k h_k P_l\right)[S,\rho]\right\} = \mathrm{Tr}\{H'[S,\rho]\},\end{aligned} \tag{B.11}$$

where we used cyclicity and linearity properties of the trace, and used the shorthand notation $H' = \sum_k h'_k P_k$, with $h'_k = \chi_k h_k$. Recall that, by definiton of the 2-norm, we have the handy relation $\|H\|_2^2 = d \sum_k h_k^2$, where d is the dimension of the system H acts on. Together with Lemma 2, which ensures that $\chi_k^2 \leq 1$, we can show that $h_k'^2 = h_k^2 \chi_k^2 \leq h_k^2$, which implies $\|H'\|_2^2 \leq \|H\|_2^2$. Finally, invoking Cauchy-Schwarz inequality we have

$$\left|\mathrm{Tr}\{H'[S,\rho]\}\right| < \|H'\|_2 \|[S,\rho]\|_2 \leq \|H\|_2 \|[S,\rho]\|_2, \tag{B.12}$$

which concludes the proof. □

We can easily extend Lemma 5 to a more generic superoperator $\alpha = \sum_{l=1}^{L} \mathcal{H}_l$, getting the immediate upper bound

$$|\mathrm{Tr}\{S\alpha\mathcal{P}(\rho)\}| \leq L \max_l \left(\|H_l\|_2\right) \|[S,\rho]\|_2. \tag{B.13}$$

Another useful upper bound involving super-operators of the form of α is given in the following Lemma.

Lemma 6. Given a quantum state ρ, a Pauli string S, two diagonal Pauli maps $\mathcal{P}_1$ and $\mathcal{P}_2$, let's define two super-operators α_1 and α_2 by

$$\alpha_q = \sum_{l=1}^{L_q} \mathcal{H}_{ql}, \quad \text{where} \quad \mathcal{H}_{ql}(\rho) = [H_{ql}, \rho]\,. \tag{B.14}$$

Further assume that each S and H_{ql} acts non-trivially on m qubits, i.e. they are m-local, for both $q = 1, 2$, each H_{ql} acts on a different subset of qubits when varying l. Furthermore, assume that $\|H_{ql}\|_2^2 \in O(d)$, where $d = 2^m$. Then we have the following upper bound

$$|\mathrm{Tr}\{S\alpha_1\mathcal{P}_1\alpha_2\mathcal{P}_2(\rho)\}| \leq \mathrm{Y} N_c \in O(4^{2m-1})\,. \tag{B.15}$$

where $\mathrm{Y} = \max_{ql} \|H_{ql}\|_2^2/d$ is a constant, and $N_c \in O(4^{2m-1})$ is an integer number.

Proof. We start by noting that, decomposing each Hamiltonian H_{ql} into its Pauli decomposition we get $H_{ql} = \sum_k h_{qlk} P_k$, depending on both k and H. Collecting terms which share the same P_k, we can obtain the following Pauli decomposition for the action of α_q, i.e.

$$\alpha_q(\rho) = \sum_k \left(\sum_l h_{qlk}\right) [P_k, \rho] = \sum_k \left(\sum_l h_{qlk}\right) \mathcal{H}_k^{(P)}(\rho),$$

where $\mathcal{H}_k^{(P)}$ defined by the corresponding P_k. Since we assume that each H_{ql} act non-trivially on distinct qubits, then all terms in $\sum_l h_{qlk}$ must vanish except when k matches the support of H_{ql}, which can only happen for one l. As there is no ambiguity in the choice of H_{ql}, we hence can drop the index l entirely, and simply write h_{qk} instead.

Using the more explicit form $\mathrm{Tr}\{S\mathcal{H}(\cdot)\} = \mathrm{Tr}\{\mathcal{H}^\dagger(S)\cdot\} = \mathrm{Tr}\{[S,H]\cdot\}$ we get

$$
\begin{aligned}
|\mathrm{Tr}\{S\alpha_1\mathcal{P}_1\alpha_2\mathcal{P}_2(\rho)\}| &= \left|\sum_{kk'} h_{1k}h_{2k'}\,\mathrm{Tr}\left\{S\mathcal{H}_k^{(P)}\mathcal{P}_1\mathcal{H}_{k'}^{(P)}\mathcal{P}_2(\rho)\right\}\right| \\
&\leq \sum_{kk'} |h_{1k}|\cdot|h_{2k'}|\cdot\left|\mathrm{Tr}\left\{[S,P_k]\mathcal{P}_1\mathcal{H}_{k'}^{(P)}\mathcal{P}_2(\rho)\right\}\right| \\
&\leq \sum_{kk'} |h_{1k}|\cdot|h_{2k'}|\cdot|\mathrm{Tr}\{[[S,P_k],P_{k'}]\mathcal{P}_2(\rho)\}| \\
&\leq \sum_{kk'} |h_{1k}|\cdot|h_{2k'}|\cdot|\mathrm{Tr}\{[[S,P_k],P_{k'}]\rho\}| \\
&\leq \max_k(|h_{1k}|)\cdot\max_{k'}(|h_{2k'}|)\cdot\sum_{kk'}|\mathrm{Tr}\{[[S,P_k],P_{k'}]\rho\}|\,,
\end{aligned}
\tag{B.16}
$$

where in the third term of the first line and in the first term of the second line we used Lemma 2 for $\mathcal{P}_1^\dagger([S,P_k])$ and $\mathcal{P}_2^\dagger([[S,P_k],P_{k'}])$ by upper bounding the corresponding constants to 1.

Recalling the relation $\|H_{ql}\|_2^2 = d\sum_k h_{qlk}^2$, where d denotes the dimension of the system, we can note that for any given a k, $h_{qk}^2 \leq \|H_{ql}\|_2^2$. As a consequence, it is immediate to see that $d\max_k|h_{qk}| \leq \max_l\|H_{ql}\|_2$, $\forall q$. Thanks to this observation, we can finally write the bound for the product

$$
\max_k(|h_{1k}|)\cdot\max_{k'}(|h_{2k'}|) \leq \max_{ql}\left(\frac{\|H_{ql}\|_2^2}{d^2}\right) = \Upsilon \in O(1)\,. \tag{B.17}
$$

Furthermore, S, P_k and $P_{k'}$ are m-local by virtue of the locality of H_{ql}. This implies that only a constant number N_c of nested commutators $[[S,P_k],P_{k'}]$ are non-vanishing. Indeed, each operator S_i commutes with all Pauli strings that act trivially on the same qubits and for all the others Pauli strings P_k, since they act non-trivially on at most a constant number m of qubits, the overlap between the supports of S_i and P_k is bounded by 4^{2m-1}, yielding only a constant number of terms in the sum.

Using Hölder inequality, we have $|\mathrm{Tr}\{[[S,P_k],P_{k'}]\rho\}| \leq \|[[S,P_k],P_{k'}]\|_\infty\|\rho\|_1 = 1$, giving the final upper bound

$$
|\mathrm{Tr}\{S\alpha_1\mathcal{P}_1\alpha_2\mathcal{P}_2(\rho)\}| \leq \Upsilon N_c \in O(4^{2m-1})\,. \tag{B.18}
$$

□

B.2 Explicit form of the decomposition of $\tilde{C}$

We can start showing that, for each moment of a given imperfect stabilizer circuit, we can separate the Clifford and non-Clifford contributions. In order to do that, we assume that all the native gates $\mathcal{G}_w$ in the circuit are such that

$$
\mathcal{G}_w = e^{-i(\phi_w+\theta_w+\epsilon_w)\mathcal{H}_w} = e^{-i(\theta_w+\epsilon_w)\mathcal{H}_w}e^{-i\phi_w\mathcal{H}_w} \equiv \mathcal{G}_w^{(\mathrm{N})}\mathcal{G}_w^{(\mathrm{C})} \tag{B.19}
$$

where each $\theta_w+\epsilon_w$ gives rise to the non-Clifford transformation $\mathcal{G}_w^{(\mathrm{N})}$ associated to coherent noise and each ϕ_w is chosen such that $\mathcal{G}_w^{(\mathrm{C})}$ is Clifford.

Moreover, given a moment $\mathcal{U}_q = \bigotimes_{l=1}^{L_q}\mathcal{G}_{ql}$ since $\mathcal{G}_{ql}$ act on different qubits, they commute with each other inside the same moment. This allows us to write

$$
\mathcal{U}_q = \bigotimes_{l=1}^{L_q}\mathcal{G}_{ql}^{(\mathrm{N})}\mathcal{G}_{ql}^{(\mathrm{C})} = \left(\bigotimes_{l=1}^{L_q}\mathcal{G}_{ql}^{(\mathrm{N})}\right)\left(\bigotimes_{l=1}^{L_q}\mathcal{G}_{ql}^{(\mathrm{C})}\right) \equiv \mathcal{U}_q^{(\mathrm{N})}\mathcal{U}_q^{(\mathrm{C})}, \tag{B.20}
$$

where $\mathcal{U}_q^{(\mathrm{N})}$ and $\mathcal{U}_q^{(\mathrm{C})}$ are respectively the non-Clifford and Clifford parts. Such operations, are still of the form of exponentials, except now they are generated by the more complex super-operator

$$\bar{\alpha}_q = \sum_{l=1}^{L_q} (\theta_{ql} + \epsilon_{ql}) \mathcal{H}_{ql}, \quad \text{i.e. } \mathcal{U}_q^{(\mathrm{N})} = e^{-i\bar{\alpha}_q}. \tag{B.21}$$

Similar structure also holds for $\mathcal{U}_q^{(\mathrm{C})}$, but we'll not need it.

Lemma 7. (Circuit decomposition and shape of the remainder). The noisy quantum circuit $\mathcal{N}$ in Eq. (4.17) can be decomposed as

$$\mathcal{N} = \mathcal{P}\mathcal{U}_S + \mathcal{R} \tag{B.22}$$

where $\mathcal{P} \equiv \left(\prod_{q=1}^{M} \mathcal{P}'_q\right)$ is an effective Pauli map and the remainder $\mathcal{R}$ is expressed as

$$\mathcal{R} = \sum_{p=1}^{M-1} \sum_{q>p}^{M} \mathcal{T}_{qp}, \quad \mathcal{T}_{qp} = \left(\prod_{s=p+1}^{M} \mathcal{P}'_s\right) \left(\prod_{r=q+1}^{M} \mathcal{U}_r^{(\mathrm{N})} \mathcal{U}_r^{(\mathrm{C})}\right) \Delta_{qp} \mathcal{U}_q^{(\mathrm{C})} \left(\prod_{r'=p}^{q-1} \mathcal{U}_{r'}^{(\mathrm{N})} \mathcal{U}_{r'}^{(\mathrm{C})}\right) \left(\prod_{t=1}^{p-1} \mathcal{P}_t \mathcal{U}_t^{(\mathrm{N})} \mathcal{U}_t^{(\mathrm{C})}\right) \tag{B.23}$$

where $\Delta_{qp} \equiv [\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{pq}]$ is the commutator between coherent and incoherent noise maps, and $\mathcal{P}'_{qp}$ are diagonal Pauli maps obtained by conjugation of $\mathcal{P}_p$ with the Clifford moments between $p+1$ and q, namely

$$\mathcal{P}'_{qp} = \left(\prod_{v=p+1}^{q} \mathcal{U}_v^{(\mathrm{C})}\right) \mathcal{P}_p \left(\prod_{v=p+1}^{q} \mathcal{U}_v^{(\mathrm{C})}\right)^{\dagger}. \tag{B.24}$$

Finally, for the last layer, we use the shorthand notation $\mathcal{P}'_s = \mathcal{P}'_{Ms}$.

Proof. The Lemma is proven giving a constructive approach. In particular, we will commute each $\mathcal{P}_q$ to the left of the expression one step at the time, keeping track of the remainders. Each term of the double sum in Eq. (B.23) represents exactly this. We now give the blueprint on how to explicitly build this construction, using as an example the first few p and q. Let's start from $\mathcal{N}$ where we highlight the last moments, namely

$$\mathcal{N} = \mathcal{P}_M \mathcal{U}_M^{(\mathrm{N})} \mathcal{U}_M^{(\mathrm{C})} \mathcal{P}_{M-1} \mathcal{U}_{M-1}^{(\mathrm{N})} \mathcal{U}_{M-1}^{(\mathrm{C})} \prod_{q=1}^{M-2} \mathcal{P}_q \mathcal{U}_q^{(\mathrm{N})} \mathcal{U}_q^{(\mathrm{C})} \tag{B.25}$$

According to Lemma 3, we can swap $\mathcal{U}_M^{(\mathrm{C})}$ and $\mathcal{P}_{M-1}$ at the price of having a new Pauli map $\mathcal{P}'_{M-1}$, obtained by conjugation, i.e. $\mathcal{U}_M^{(\mathrm{C})} \mathcal{P}_{M-1} = \mathcal{P}'_{M-1} \mathcal{U}_M^{(\mathrm{C})}$. However we cannot do this for the non-Clifford moment $\mathcal{U}_M^{(\mathrm{N})}$. In this case we swap them and keep track of the error $\Delta_{M,M-1}$, i.e. we use the relation

$$\mathcal{U}_M^{(\mathrm{N})} \mathcal{P}'_{M-1} = \mathcal{P}'_{M-1} \mathcal{U}_M^{(\mathrm{N})} + [\mathcal{U}_M^{(\mathrm{N})}, \mathcal{P}'_{M-1}] = \mathcal{P}_{M-1} \mathcal{U}_M^{(\mathrm{N})} + \Delta_{M,M-1}. \tag{B.26}$$

In this way, we get the first decomposition

$$\begin{aligned} \mathcal{N} = {} & \mathcal{P}_M \mathcal{P}'_{M-1} \mathcal{U}_M^{(\mathrm{N})} \mathcal{U}_M^{(\mathrm{C})} \mathcal{U}_{M-1}^{(\mathrm{N})} \mathcal{U}_{M-1}^{(\mathrm{C})} \prod_{q=1}^{M-2} \mathcal{P}_q \mathcal{U}_q^{(\mathrm{N})} \mathcal{U}_q^{(\mathrm{C})} \\ & + \mathcal{P}_M \Delta_{M,M-1} \mathcal{U}_M^{(\mathrm{C})} \mathcal{U}_{M-1}^{(\mathrm{N})} \mathcal{U}_{M-1}^{(\mathrm{C})} \prod_{q=1}^{M-2} \mathcal{P}_q \mathcal{U}_q^{(\mathrm{N})} \mathcal{U}_q^{(\mathrm{C})} \end{aligned} \tag{B.27}$$

Note that the first term is one step closer to look like the first term in Eq. (B.22), while the second term corresponds exactly to $\mathcal{T}_{M\,M-1}$. By iterating this process on first term, we can obtain all subsequent $\mathcal{T}_{qp}$ terms. □

Beyond the technicalities involved in keeping track of all the conjugations appearing in $\mathcal{P}'_{qp}$, the real power of Lemma 7 is to provide the general structure of the remainder $\mathcal{R}$ in terms of a sequence of Clifford, non-Clifford and Pauli maps, which will be key for the subsequent proofs.

B.3 Proof of stationarity of the solution

In this section we show that, despite the presence of Pauli noise in the system, the optimal parameters that we aim to find, namely $\boldsymbol{\theta} = -\boldsymbol{\epsilon}$ remains a stationary point. In order to do so we prove Theorem 4.3.1 of the main text.

Proof. For simplicity now we define $\gamma_k = \theta_k + \epsilon_k$ in the native gates inside the non-Clifford moments $\mathcal{U}_q^{(\mathrm{N})}$. In this manner, it is sufficient to compute the derivative $\partial_{\gamma_k}\mathcal{U}_q^{(\mathrm{N})}$ and evaluate it in $\boldsymbol{\gamma} = \boldsymbol{\theta} + \boldsymbol{\epsilon} = 0$ to achieve the desired result of Eq. (4.20).

By employing the expression of $\mathcal{R}$ in Eq. (B.23) of Lemma 7 we have

$$\partial_{\gamma_k}\Delta\tilde{C}(\boldsymbol{\theta})|_{\boldsymbol{\gamma}=0} = \sum_{i=1}^{n}\partial_{\gamma_k}\operatorname{Tr}\{S\mathcal{R}(\rho_0)\}\big|_{\boldsymbol{\gamma}=0} = \sum_{i=1}^{n}\sum_{p=1}^{M-1}\sum_{q>p}^{M}\partial_{\gamma_k}\operatorname{Tr}\big\{S\mathcal{T}_{qp}(\rho_0)\big\}\big|_{\boldsymbol{\gamma}=0}. \tag{B.28}$$

At this point we focus on two distinct cases:

Observation 3 (Selection of the remainder terms)**.** Given a parameter γ_k, for each moment $\mathcal{U}_q^{(\mathrm{N})}$ that does not depend on γ_k, then

$$\partial_{\gamma_k}\mathcal{T}_{qp}|_{\boldsymbol{\gamma}=0} \equiv 0 \tag{B.29}$$

This follows from the definition of $\mathcal{T}_{qp}$, and in particular on the structure of $\Delta_{qp} = [\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}]$ arising from Lemma 7. Indeed, we assume that γ_k is not contained in a moment $\mathcal{U}_q^{(\mathrm{N})}$, and without loss of generality to ease the notation, we assume that instead, it is contained in the moment $\mathcal{U}_1^{(\mathrm{N})}$. Then we have

$$\begin{aligned}\partial_{\gamma_k}\mathcal{T}_{qp}|_{\boldsymbol{\gamma}=0} &= \left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{N})}\mathcal{U}_r^{(\mathrm{C})}\right)\Delta_{qp}\mathcal{U}_q^{(\mathrm{C})}\left(\prod_{r'=p}^{q-1}\mathcal{U}_{r'}^{(\mathrm{N})}\mathcal{U}_{r'}^{(\mathrm{C})}\right)\left(\prod_{t=2}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(\mathrm{N})}\mathcal{U}_t^{(\mathrm{C})}\right)\mathcal{P}_1\left(\partial_{\gamma_k}\mathcal{U}_1^{(\mathrm{N})}\right)\mathcal{U}_1^{(\mathrm{C})}\Big|_{\boldsymbol{\gamma}=0}\\ &= -i\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{N})}\mathcal{U}_r^{(\mathrm{C})}\right)\Delta_{qp}\mathcal{U}_q^{(\mathrm{C})}\left(\prod_{r'=p}^{q-1}\mathcal{U}_{r'}^{(\mathrm{N})}\mathcal{U}_{r'}^{(\mathrm{C})}\right)\left(\prod_{t=2}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(\mathrm{N})}\mathcal{U}_t^{(\mathrm{C})}\right)\mathcal{P}_1\left(\mathcal{H}_k\mathcal{U}_1^{(\mathrm{N})}\right)\mathcal{U}_1^{(\mathrm{C})}\Big|_{\boldsymbol{\gamma}=0}\end{aligned} \tag{B.30}$$

where we used the definition $\mathcal{U}_q^{(\mathrm{N})} = e^{-i\bar{\alpha}_q}$, and $\bar{\alpha}_q = \sum_k \gamma_k\mathcal{H}_k$. Then, noting that, according to the same definition, $\bar{\alpha}_q = 0$ when evaluated at $\boldsymbol{\gamma} = 0$, we get the simple result that $\mathcal{U}_q^{(\mathrm{N})} = e^{-i\bar{\alpha}_q} = \mathbb{1}\ \forall q$.

In particular this implies that, at $\boldsymbol{\gamma} = 0$, $\Delta_{qp} \equiv 0$, since in that case $\Delta_{qp} = [\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}] = [\mathbb{1}, \mathcal{P}'_{qp}] \equiv 0$, killing the whole product. Clearly from this reasoning, all other terms where the moment does not depend on γ_k will similarly vanish.

The same however cannot be said when $\mathcal{U}_q^{(\mathrm{N})}$ depends on γ_k. Indeed, in that case the γ_k dependence is included in Δ_{qp}, and hence we need to compute $\partial_{\gamma_k}\Delta_{qp}$. In particular, we get

$$\partial_{\gamma_k}\Delta_{qp}|_{\gamma=0} = \partial_{\gamma_k}[\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}]\Big|_{\gamma=0} = [\partial_{\gamma_k}\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}]\Big|_{\gamma=0} = -i[\mathcal{H}_k\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}]\Big|_{\gamma=0} = -i[\mathcal{H}_k, \mathcal{P}'_{qp}], \quad \text{(B.31)}$$

which is in general non-vanishing. However, we will now prove that, when evaluated in the trace, such contributions do indeed vanish. As a first step, we make the following simple observation.

Observation 4 (Structure of the derivative)**.** For any non-vanishing $\partial_{\gamma_k}\mathcal{T}_{qp}$, we can write its gradient contribution as

$$\partial_{\gamma_k}\operatorname{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\}|_{\gamma=0} = -i\chi'_i\operatorname{Tr}\left\{S_i\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)[\mathcal{H}_k, \mathcal{P}'_{qp}]\mathcal{P}''_{qp}\left(\prod_{r'=1}^{q}\mathcal{U}_{r'}^{(\mathrm{C})}\right)(\rho_0)\right\} \quad \text{(B.32)}$$

where $\mathcal{P}''_p$ is a Pauli map including the contribution of the first p Pauli noise maps.

To show this, we recall the definition of $\mathcal{T}_{qp}$ in Eq. (B.23) and the previous result that when $\gamma = 0$, all non-Clifford moments become the identity. Using this we get

$$\begin{aligned}\partial_{\gamma_k}\mathcal{T}_{qp}|_{\gamma=0} &= \left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{N})}\mathcal{U}_r^{(\mathrm{C})}\right)\partial_{\gamma_k}\Delta_{qp}\mathcal{U}_q^{(\mathrm{C})}\left(\prod_{r'=p}^{q-1}\mathcal{U}_{r'}^{(\mathrm{N})}\mathcal{U}_{r'}^{(\mathrm{C})}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(\mathrm{N})}\mathcal{U}_t^{(\mathrm{C})}\right)\Big|_{\gamma=0} \\ &= \left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)\partial_{\gamma_k}\Delta_{qp}\big|_{\gamma=0}\mathcal{U}_q^{(\mathrm{C})}\left(\prod_{r'=p}^{q-1}\mathcal{U}_{r'}^{(\mathrm{C})}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(\mathrm{C})}\right).\end{aligned} \quad \text{(B.33)}$$

Inserting the explicit form of Δ_{qp}, using Lemma 2 to the leftmost Pauli maps and Eq. (B.31), we can obtain

$$\partial_{\gamma_k}\operatorname{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\}|_{\gamma=0} = -i\chi'_i\operatorname{Tr}\left\{S_i\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)[\mathcal{H}_k, \mathcal{P}'_{qp}]\mathcal{U}_q^{(\mathrm{C})}\left(\prod_{r'=p}^{q-1}\mathcal{U}_{r'}^{(\mathrm{C})}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(\mathrm{C})}\right)(\rho_0)\right\}. \quad \text{(B.34)}$$

Finally, the last step is achieved by noting that, since we only have Clifford moments, we can swap all Pauli maps next to the commutator, at the price of conjugation with Clifford moments as in Lemma 3.

While it would be possible to keep track exactly of the form of those Pauli maps, it suffices for our purposes to note that they will remain Pauli maps. With this in mind, we are ready to conclude the proof by computing for simplicity the modulus of the derivative (without using the modulus the derivation is more involved). In particular, we have

$$\begin{aligned}\left|\partial_{\gamma_k}\operatorname{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\}|_{\gamma=0}\right| &= \left|\chi'_i\operatorname{Tr}\left\{S_i\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)[\mathcal{H}_k, \mathcal{P}'_{qp}]\mathcal{P}''_{qp}\left(\prod_{r'=1}^{q}\mathcal{U}_{r'}^{(\mathrm{C})}\right)(\rho_0)\right\}\right| \\ &\le \left|\operatorname{Tr}\left\{S_i\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)\mathcal{H}_k\mathcal{P}'_{qp}\mathcal{P}''_{qp}\left(\prod_{r'=1}^{q}\mathcal{U}_{r'}^{(\mathrm{C})}\right)(\rho_0)\right\}\right| \\ &+ \left|\operatorname{Tr}\left\{S_i\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(\mathrm{C})}\right)\mathcal{P}'_{qp}\mathcal{H}_k\mathcal{P}''_{qp}\left(\prod_{r'=1}^{q}\mathcal{U}_{r'}^{(\mathrm{C})}\right)(\rho_0)\right\}\right|,\end{aligned} \quad \text{(B.35)}$$

where we used $|\chi_i| \leq 1 \; \forall i$ coming from Lemma 2 and the triangle inequality to expand the commutator outside the modulus. If we now introduce the notations

$$S_i^{(q)} \equiv \left(\prod_{r=q+1}^{M} \mathcal{U}_r^{(\mathrm{C})} \right)^\dagger S_i \;\text{ and }\; \rho_q \equiv \left(\prod_{r'=1}^{q} \mathcal{U}_{r'}^{(\mathrm{C})} \right)(\rho_0), \tag{B.36}$$

using the ciclicity of the trace we have the final bound

$$\begin{aligned} \left| \partial_{\gamma_k} \mathrm{Tr}\{S_i \mathcal{T}_{qp}\} \right|_{\gamma=0} \Big| &\leq \left| \mathrm{Tr}\left\{ S_i^{(q)} \mathcal{H}_k \mathcal{P}'_{qp} \mathcal{P}''_{qp}(\rho_q) \right\} \right| + \left| \mathrm{Tr}\left\{ S_i^{(q)} \mathcal{P}'_{qp} \mathcal{H}_k \mathcal{P}''_{qp}(\rho_q) \right\} \right| \\ &\leq \left| \mathrm{Tr}\left\{ S_i^{(q)} \mathcal{H}_k \mathcal{P}'_{qp} \mathcal{P}''_{qp}(\rho_q) \right\} \right| + \left| \mathrm{Tr}\left\{ S_i^{(q)} \mathcal{H}_k \mathcal{P}''_{qp}(\rho_q) \right\} \right| \end{aligned} \tag{B.37}$$

where again we used Lemma 2 applied to $\mathcal{P}'_{qp}$ in the second term of the first line. By Lemma 5, we can now write

$$\left| \partial_{\gamma_k} \mathrm{Tr}\{S_i \mathcal{T}_{qp}(\rho_0)\} \right|_{\gamma=0} \Big| \leq 2 \|H_k\|_2 \|[S_i^{(q)}, \rho_q]\|_2 = 0, \tag{B.38}$$

where the equality comes from the fact that $S_i^{(q)}$ is a stabilizer for ρ_q by construction, and hence they commute. □

B.4 Upper bound on $\Delta\tilde{C}$

In this section we prove Theorem 4.3.2 by computing an analytical upper bound on the error $\Delta\tilde{C}(\boldsymbol{\theta})$ that we make when approximating the circuit $\mathcal{N}$ in Eq. (4.15) with the circuit $\mathcal{P}\mathcal{U}_S$, where all noise maps have been commuted to the end giving rise to an effective Pauli map $\mathcal{P}$ as in Eq. (B.22).

Proof. We evaluate the expression

$$\left| \Delta\tilde{C}(\boldsymbol{\theta}) \right| = \left| \sum_{i=1}^{n} \mathrm{Tr}\{S_i \mathcal{R}(\rho_0)\} \right| \leq \sum_{i=1}^{n} \sum_{p=1}^{M-1} \sum_{q>p}^{M} \left| \mathrm{Tr}\left\{ S_i \mathcal{T}_{qp}(\rho_0) \right\} \right|, \tag{B.39}$$

where as shown in the main text $\mathcal{R} = \mathcal{N} - \mathcal{P}\mathcal{U}_S$ and we inserted the explicit expression of $\mathcal{R}$ in Eq. (B.23).

As next steps, we assume that the coherent errors in each native gate have the same order of magnitude $\epsilon \ll 1$. This allows to find the general scaling of $\mathcal{T}_{qp}$ in terms of ϵ, which will give the final upper bound. To do so, lets consider the scalings of Δ_{qp} and $\mathcal{U}_q^{(\mathrm{N})}$ in Eq. (B.23), namely

$$\mathcal{U}_q^{(\mathrm{N})} \sim \mathbb{1} - i\epsilon\alpha_q - \frac{\epsilon^2}{2}\alpha_q^2 + O(\epsilon^3) \tag{B.40}$$

where we used the shorthand notation $\alpha_q = \sum_{l=1}^{L_q} \mathcal{H}_{ql}$ which represents the sum over generators for all gates in the same moment q, and

$$\Delta_{qp} = [\mathcal{U}_q^{(\mathrm{N})}, \mathcal{P}'_{qp}] \sim -i\epsilon[\alpha_q, \mathcal{P}'_{qp}] - \frac{\epsilon^2}{2}[\alpha_q^2, \mathcal{P}'_{qp}] + O(\epsilon^3), \tag{B.41}$$

where $\mathcal{P}'_{qp}$ is defined in Eq. (B.24). Substituting Eqs. (B.40) and (B.41) into the definition of $\mathcal{T}_{qp}$, and keeping terms up to order ϵ^2, we get

$$\begin{aligned}\mathcal{T}_{qp} = &-i\epsilon\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(C)}\right)[\alpha_q,\mathcal{P}'_{qp}]\left(\prod_{r'=p}^{q}\mathcal{U}_{r'}^{(C)}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(C)}\right)\\
&-\epsilon^2\sum_{q<u}\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{b=u+1}^{M}\mathcal{U}_b^{(C)}\right)\alpha_u\left(\prod_{b'=q+1}^{u}\mathcal{U}_{b'}^{(C)}\right)[\alpha_q,\mathcal{P}'_{qp}]\left(\prod_{r'=p}^{q}\mathcal{U}_{r'}^{(C)}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(C)}\right)\\
&-\epsilon^2\sum_{p\le u<q}\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(C)}\right)[\alpha_q,\mathcal{P}'_{qp}]\left(\prod_{b=u+1}^{q}\mathcal{U}_b^{(C)}\right)\alpha_u\left(\prod_{b'=p}^{u}\mathcal{U}_{b'}^{(C)}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(C)}\right)\\
&-\epsilon^2\sum_{u<p}\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(C)}\right)[\alpha_q,\mathcal{P}'_{qp}]\left(\prod_{r'=p}^{q}\mathcal{U}_{r'}^{(C)}\right)\left(\prod_{b=u+1}^{p-1}\mathcal{P}_b\mathcal{U}_b^{(C)}\right)\mathcal{P}_u\alpha_u\left(\prod_{b'=1}^{u}\mathcal{P}_{b'}\mathcal{U}_{b'}^{(C)}\right)\\
&-\frac{\epsilon^2}{2}\left(\prod_{s=p+1}^{M}\mathcal{P}'_s\right)\left(\prod_{r=q+1}^{M}\mathcal{U}_r^{(C)}\right)[\alpha_q^2,\mathcal{P}'_{qp}]\left(\prod_{r'=p}^{q}\mathcal{U}_{r'}^{(C)}\right)\left(\prod_{t=1}^{p-1}\mathcal{P}_t\mathcal{U}_t^{(C)}\right).\end{aligned} \tag{B.42}$$

This decomposition has a lot of structure, which can be exploited to greatly simplify our calculations. In particular, we can make a few observations.

Observation 5 (No linear terms in ϵ)**.** The term proportional to ϵ vanish in the expectation value with S_i and ρ_0, i.e.

$$\mathrm{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\} \in O(\epsilon^2) \tag{B.43}$$

Indeed one can notice that each of the terms arising from $\alpha_q = \sum_{l=1}^{L_q}\mathcal{H}_{ql}$ in the first line of Eq. (B.42) is proportional to the expression $\partial_{\gamma_k}\mathrm{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\}|_{\gamma=0}$ in Eq. (B.32), which was already proven to vanish in Section B.3.

Observation 6 (All non-vanishing terms have the same structure)**.** If we expand the commutators $[\alpha_q,\mathcal{P}'_{qp}] = \alpha_q\mathcal{P}'_{qp} - \mathcal{P}'_{qp}\alpha_q$ and $[\alpha_q^2,\mathcal{P}'_{qp}] = \alpha_q^2\mathcal{P}'_{qp} - \mathcal{P}'_{qp}\alpha_q^2$, the terms proportional to ϵ^2 in Eq. (B.42) have the same underlying structure, namely

$$\epsilon^2\,\mathcal{P}_3\mathcal{U}_3\alpha_2\mathcal{P}_2\alpha_1\mathcal{P}_1\mathcal{U}_0 \tag{B.44}$$

where the explicit expression of the maps in Eq. (B.44) depends on the specific term considered in Eq. (B.42).

Consider as an example the second term in Eq. (B.42) for some specific $u > q$, then we set

$$\mathcal{P}_3 = \prod_{s=p+1}^{M}\mathcal{P}'_s\,. \tag{B.45}$$

Concerning $\mathcal{U}_3$ and α_2, we can iteratively apply Lemma 4 to move all unitary operation indexed by b and b' next to each other collecting them in a single unitary. By naming α'_u the result of this conjugation, we are left with

$$\mathcal{U}_3 = \left(\prod_{b=u+1}^{M}\mathcal{U}_b^{(C)}\right)\left(\prod_{b'=q+1}^{u}\mathcal{U}_{b'}^{(C)}\right) = \prod_{r=q+1}^{M}\mathcal{U}_r^{(C)} \text{ and } \alpha_2 = \alpha'_u = \left(\prod_{b'=q+1}^{u}\mathcal{U}_{b'}^{(C)}\right)^{\dagger}\alpha_u\left(\prod_{b'=q+1}^{u}\mathcal{U}_{b'}^{(C)}\right). \tag{B.46}$$

Continuing, $\mathcal{P}_2$ is equal either the identity or to $\mathcal{P}'_{qp}$ depending on which branch of the commutator we're considering, $\alpha_1 = \alpha_q$, and finally, $\mathcal{P}_1$ and $\mathcal{U}_0$ can be obtained by Lemma 3

moving all remaining Pauli maps next to the right of α_q, getting

$$\mathcal{P}_1 = \prod_{t=1}^{p-1} \mathcal{P}'_{qt} \text{ and } \mathcal{U}_0 = \prod_{t=1}^{q} \mathcal{U}_t^{(C)}, \tag{B.47}$$

where $\mathcal{P}'_{qt}$ has the structure in Eq. (B.24).

With similar arguments, we can rewrite all terms arising from Eq. (B.42) in this form. It then suffices to upper bound the equation

$$|\mathrm{Tr}\{S_i\mathcal{P}_3\mathcal{U}_3\alpha_2\mathcal{P}_2\alpha_1\mathcal{P}_1\mathcal{U}_0(\rho_0)\}| \leq |\mathrm{Tr}\{S_i\mathcal{U}_3\alpha_2\mathcal{P}_2\alpha_1\mathcal{P}_1\mathcal{U}_0(\rho_0)\}| = \left|\mathrm{Tr}\left\{S'_i\alpha_2\mathcal{P}_2\alpha_1\mathcal{P}_1(\rho')\right\}\right|, \tag{B.48}$$

where we used Lemma 2 on $\mathcal{P}_3^\dagger(S_i)$ and we defined $S'_i \equiv \mathcal{U}_3^\dagger(S_i)$ and $\rho' \equiv \mathcal{U}_0(\rho_0)$. Assuming that S'_i and all generators in α_2, α_1 are m-local, we can apply Lemma 6, and get

$$\left|\mathrm{Tr}\left\{S'_i\alpha_2\mathcal{P}_2\alpha_1\mathcal{P}_1(\rho')\right\}\right| \leq \mathrm{Y}N_c \in O(4^{2m-1}). \tag{B.49}$$

Counting all the terms of this form appearing in Eq. (B.42), we get a total of M contributions of this kind for each $\mathrm{Tr}\{S_i\mathcal{T}_{qp}(\rho_0)\}$, giving the final bound

$$|\Delta C(\boldsymbol{\theta})| \leq \epsilon^2 n M^2 (M-1)\mathrm{Y}N_c. \tag{B.50}$$

We remark that, the bound obtained scales with $\mathrm{Y}N_c \in O(4^{2m-1})$, which is constant only if m does not scale with n. This is indeed satisfied for locally-connected graph state circuits, where the stabilizers S_i can be chosen to be local, and the depth M is also constant, meaning that $m \leq 2^M$ is bounded by the maximum dimension of the lightcone of each stabilizer. □

B.5 Transpilation of Hadamard and CZ gates

Here in Fig. B.1 we show the transpilation of Hadamard and CZ gates into native gates used in section 4.4.

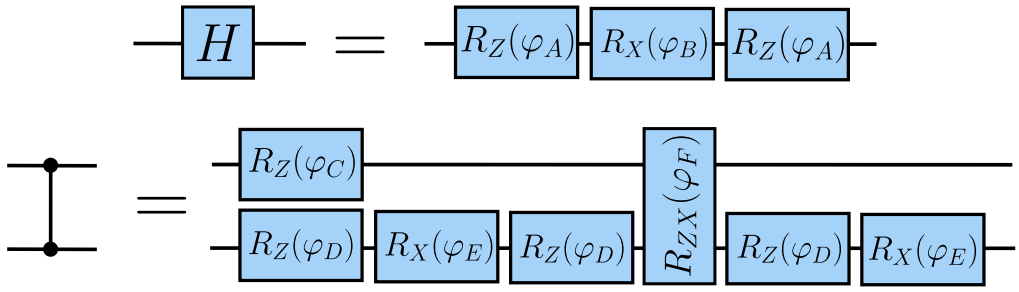

FIGURE B.1: Transpile of Hadamard and CZ gates into Rz, Rx and Rzx native gates. The angles are defined in the following way: $\varphi_A = \frac{\pi}{2} + \theta_1 + \epsilon_1$, $\varphi_B = \frac{\pi}{2} + \theta_2 + \epsilon_2$, $\varphi_C = -\frac{3}{2}\pi + \theta_1 + \epsilon_1$, $\varphi_D = -\frac{\pi}{2} + \theta_3 + \epsilon_3$, $\varphi_E = \frac{\pi}{2} + \theta_4 + \epsilon_4$, $\varphi_F = \frac{\pi}{2} + \theta_5 + \epsilon_5$.

,

Appendix C

Additional material on cost concentration

C.1 Elements of non-negative matrix theory

In this Section, we briefly recap on the main results on non-negative matrix theory useful in the proof of Theorem 6.3.1 of the main text. For a complete discussion and proofs of the cited results, we refer the interested reader to Refs. [102, 70]. Let's start by the definition of non-negative matrix.

Definition C.1.1 (Non-negative matrix)**.** A $n \times n$ matrix T is said to be non-negative if each entry $(T)_{ij} \geq 0$.

The general behaviour of non negative matrices can vary greatly, but there is a class of matrices, called *irreducible*, which have very informative spectral properties.

Definition C.1.2 (Irreducible matrix)**.** A $n \times n$ non-negative matrix T is said to be *irreducible* if for two arbitrary indices $i, j = 1, ...n$, there exist $l = l(i,j) \in \mathbb{N}$ such that $(T^l)_{ij} > 0$. Moreover, we will say that T has period p, where p is the greatest common divisor of all $l(i,i)$ that satisfy $(T^l)_{ii} > 0 \ \forall i$.

Equivalently, if we introduce the graph $\mathcal{G}_T$ whose adjacency matrix is T, then it can be shown that T is irreducible if and only if $\mathcal{G}_T$ is strongly connected, and that the period p reduces to the great common divisor of the lengths of all closed directed paths in $\mathcal{G}_T$ [102]. Furthermore, it will be useful in the following to distinguish two classes of irreducible matrices, namely *cyclic* (or *periodic*) and *primitive* (or *aperiodic*), which are characterized as having period $p > 1$ and $p = 1$ respectively.

One of the main results involving irreducible matrices is the celebrated *Perron-Frobenius* theorem, which characterizes the spectral properties of this class. We recall it here for convenience.

Theorem C.1.1 (Perron-Frobenius)**.** Let T be a $n \times n$ non-negative, irreducible matrix. Then there exists an eigenvalue r of T, with corresponding right and left eigenvectors v, w such that:

(a) $r \in \mathbb{R}, r > 0$ and is a simple root of the characteristic polynomial,

(b) both w,v are the only eigenvectors that have strictly positive components, i.e. $v_i, w_i > 0 \ \forall i = 1, ...n$,

(c) v and w are unique up to a scalar multiple, and hence can be taken to be normalized, i.e. $v^t w = 1$,

(d) $r \geq |\lambda|$, for all eigenvalue λ of T,

where r is called the Perron-Frobenius eigenvalue and $P = wv^t$ the Perron projector. Moreover, if T is also aperiodic, then we have the more restrictive

(d') $r > |\lambda|$, for all eigenvalue $\lambda \neq r$ of T,

Another important result is the so called subinvariance theorem, which is a useful tool to bound the value of r for a given irreducible matrix.

Theorem C.1.2 (Subinvariance Theorem)**.** Let T be a $n \times n$ non-negative, irreducible matrix, $s > 0$ and y be a n-dimensional row vector such that each component $y_i \geq 0$ and satisfying

$$y^t T \leq s y^t \tag{C.1}$$

component-wise. Then $y_i > 0 \forall i$, and $s \geq r$. Moreover, equality holds if and only if $s = r$.

Finally, the last result allows us to exploit the knowledge of the dominant eigenvalue to determine the asymptotic properties of T^L.

Theorem C.1.3 (Asymptotic behaviour of irreducible matrices)**.** Let T be a $n \times n$ non-negative, irreducible matrix. Then the Cesàro average of T converges, and we have

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} T^l / r^l = P \tag{C.2}$$

Moreover, if T is also aperiodic, then limit of T^L/r^L converges, and we have

$$\lim_{L\to\infty} T^L / r^L = P \tag{C.3}$$

where r and P are the Perron-Frobenius eigenvalue and Perron projector respectively.

Despite being less structured, it is a well known fact that general non-negative matrices can be cast to a canonical block upper triangular form, where all blocks in the diagonal are *irreducible* simply by means of a permutation matrix, i.e. by a relabelling of the basis elements. In particular, concerning the diagonal, irreducible blocks appearing in such decomposition, we will use the term *essential* when referring to the blocks such that all $(T)_{i,j} = 0$ for all columns apart from the block itself, and *inessential* otherwise. In terms of the graph $\mathcal{G}_T$, this distinction is readily understood. As discussed above, irreducible blocks correspond to strongly connected components, and consequently essential blocks are strongly connected components which do not have edges connecting vertices in it to vertices pertaining to other components. More simply, we can describe essential components as those whose edges "do not lead outside". In what follows, we name by T_z all essential, irreducible components, and we group into a single block Q all inessential components. The blocks R_z, appearing on top of the block Q, represent the collection of edges coming from inessential components and leading to essential ones. A graphical summary is depicted in Eq. (C.4).

$$T = \begin{pmatrix} T_0 & & & & R_0 \\ & T_1 & & & R_1 \\ & & \ddots & & \vdots \\ & & & T_z & R_z \\ & & & & Q \end{pmatrix} \tag{C.4}$$

This in particular allows us to apply the results of this Section also to more generic matrices, such as LTMs, which in general are not irreducible.

C.2 Locality and locality transfer matrix properties

In order to show the main properties of locality vectors and locality transfer matrices (LTM), it is convenient, for each subsystem $\mathcal{H}_m$, to fix an orthonormal basis $\{P_j^{(m)}\}_{j=1}^{d_m-1}$ of $\mathcal{B}_m$ composed of traceless, Hermitian operators together with the identity operator, each normalized with respect to the Hilbert-Schmidt norm, namely $P_0^{(m)} = \mathbb{1}/\sqrt{d_m}$, $P_{j_m}^{(m)} = P_{j_m}^{(m)\dagger}\ \forall j_m$,$\mathrm{Tr}\left\{P_{j_m}^{(m)} P_{k_m}^{(m)}\right\} = \delta_{j_m k_m}$. Starting from these, we can build an orthonormal basis $\{P_j\}_j$ for the whole space by means of tensor products. Each basis element will be labelled by the multi-index $j = (j_1, ..., j_M)$, where the entries j_m refer to an element of the local bases, and hence $j_m \in \{0, ..., d_m^2 - 1\}$. Such a basis will be dubbed a *local* basis. As an example, if each $\mathcal{H}_m = \mathbb{C}^2$ is a qubit, the normalized Pauli strings form a local basis for $\mathcal{B}$. Given a local basis $\{P_j\}_j$, it is possible to group the elements in disjoint sets. In particular, for any given binary string $\kappa = \{0,1\}^M$, we can collect in the set S_κ all basis elements acting non trivially on $\mathcal{H}_m$ if and only if $\kappa_m = 1$. For practical reasons, we introduce the indicator function $\tilde{\delta}_{i,\kappa}$ for the set S_κ, defined by

$$\tilde{\delta}_{i,\kappa} = \prod_{m=0}^{M} \tilde{\delta}_{i_m,\kappa_m} = \begin{cases} 1 & \text{if } \kappa_m = 0 \Leftrightarrow i_m = 0\ \forall m \\ 0 & \text{otherwise} \end{cases} \tag{C.5}$$

From the definition, we can derive some simple properties of this function.

Lemma C.2.1. The indicator function $\tilde{\delta}_i b$ has the following properties:

$$\sum_i \tilde{\delta}_{i,\kappa} = d_\kappa, \quad \sum_\kappa \tilde{\delta}_{i,\kappa} = 1, \quad \sum_{\kappa\in K} \tilde{\delta}_{i,\kappa}\tilde{\delta}_{i,\lambda} = \tilde{\delta}_{i,\lambda} \sum_{\kappa\in K} \tilde{\delta}_{\kappa,\lambda} \tag{C.6}$$

where $d_\kappa = \prod_{m=0}^{M} (d_m^2 - 1)^{\kappa_m}$, $K \subset \{0,1\}^M$, and $\delta_{\kappa,\lambda}$ is the usual Kronecker delta.

Proof. All the results follow directly from the definition. □

Using this notation, we can express the *locality* ℓ_A of some operator $A \in \mathcal{B}$ and *locality transfer matrix* T of a linear map $\Lambda : \mathcal{B} \to \mathcal{B}$ defined in the main text as

$$(\ell_A)_\kappa = \sum_j \mathrm{Tr}\left\{P_j A\right\}^2 \tilde{\delta}_j \kappa \tag{C.7}$$

and

$$T_{\kappa,\lambda} = \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\left\{P_i \Lambda(P_j)\right\}^2 \tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\lambda} \tag{C.8}$$

respectively. It is immediately realized that both these quantities are basis-independent.

Lemma C.2.2. Given a bounded operator $A \in \mathcal{B}$ and a partition into subsystems, the locality vector ℓ_A is uniquely defined, i.e. it does not depend on the choice of local basis. Similarly, given a quantum channel $\mathcal{E} : \mathcal{B} \to \mathcal{B}$. the corresponding locality transfer matrix T is uniquely defined.

Proof. Let $\{P_j\}$ and $\{B_i\}$ be two local bases for a given subsystem. Then

$$\begin{aligned}
&\sum_j \mathrm{Tr}\{P_j A\}^2 \tilde{\delta}_{j,\kappa} = \sum_j \mathrm{Tr}\left\{\sum_i \mathrm{Tr}\{B_i P_j\} B_i A\right\}^2 \tilde{\delta}_{j,\kappa} \\
&= \sum_j \sum_{i,i'} \mathrm{Tr}\{B_i P_j\}\, \mathrm{Tr}\{B_{i'} P_j\}\, \mathrm{Tr}\{B_i A\}\, \mathrm{Tr}\{B_{i'} A\} \tilde{\delta}_{j,\kappa} \\
&= \sum_{i,i'} \left(\sum_j \mathrm{Tr}\{B_i P_j\}\, \mathrm{Tr}\{B_{i'} P_j\} \tilde{\delta}_{j,\kappa} \right) \mathrm{Tr}\{B_i A\}\, \mathrm{Tr}\{B_{i'} A\} = \sum_i \mathrm{Tr}\{B_i A\}^2 \tilde{\delta}_{i,\kappa}
\end{aligned} \tag{C.9}$$

where the last equality is due to

$$\begin{aligned}
\sum_j \mathrm{Tr}\{B_i P_j\}\, \mathrm{Tr}\{B_{i'} P_j\} \tilde{\delta}_{j,\kappa} &= \prod_{m=1}^{M} \left(\sum_{j_m=1}^{d_m} \mathrm{Tr}\{B_{i'_m} P_{j_m}\}\, \mathrm{Tr}\{B_{i_m} P_{j_m}\} \tilde{\delta}_{j_m,\kappa_m} \right) \\
&= \prod_{m=1}^{M} \left(\mathrm{Tr}\{B_{i'_m} B_{i_m}\} \tilde{\delta}_{i_m,\kappa_m} \right) = \delta_{i,i'} \tilde{\delta}_{i,\kappa}
\end{aligned} \tag{C.10}$$

which holds since both $\{P_{j_m}\}$ and $\{B_{i_m}\}$ are orthonormal bases of $\mathcal{B}_m$ by definition of local basis. A totally analogous calculation yields the same result for the locality transfer matrix T. □

Thanks to the formulation of Eq. (C.8), the relation between the LTM of a map Λ and the Hermitian adjoint $\Lambda^\dagger$ with respect to the Hilbert-Schmidt scalar product can be seen. In particular, we have

$$d_\lambda T_{\kappa,\lambda} = \sum_{i,j} \mathrm{Tr}\{P_i \Lambda(P_j)\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = \sum_{i,j} \mathrm{Tr}\{\Lambda^\dagger(P_i) P_j\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = d_\kappa T^\dagger_{\lambda,\kappa}, \tag{C.11}$$

which can be compactly written in matrix form as $TD = (T^\dagger D)^t$, with $D_{\kappa,\lambda} = d_\kappa \delta_{\kappa,\lambda}$. For sake of readability, here we introduce a shorthand notation for the scalar product $(\cdot,\cdot)$ in $\mathbb{R}^{2^M}$ such that T and $T^\dagger$ are *also* Hermitian adjoint of one another, i.e.

$$(a,b) = a^t D^{-1} b = \sum_\kappa \frac{a_\kappa b_\kappa}{d_\kappa}. \tag{C.12}$$

This follows from the chain $(a, Tb) = a^t D^{-1} T b = a^t D^{-1} D (T^\dagger)^t D^{-1} b = (T^\dagger a)^t D^{-1} b = (T^\dagger a, b)$.

C.3 Proof of the general variance formula

In this Section we provide a proof for the building blocks of the main results of this Chapter. Note that, what follows hinge on the structure of the circuit $\Phi_{\boldsymbol{\theta}}$ provided in the main text, which we recall is composed of L layers of interleaved unitary and noise channels

$$\Phi_{\boldsymbol{\theta}} = \mathcal{U}_{\theta_{L+1}} \circ \mathcal{E}_L \circ \mathcal{U}_{\theta_{L-1}} \cdots \circ \mathcal{E}_1 \circ \mathcal{U}_{\theta_1}. \tag{C.13}$$

In particular, we assume that $\mathcal{U}_{\boldsymbol{\theta}} : \rho \mapsto U_{\boldsymbol{\theta}} \rho U^\dagger_{\boldsymbol{\theta}}$, where $U_{\boldsymbol{\theta}} = \otimes_m U^{(m)}_{\theta_m}$ is a *local 2-design* for the system. This statement is more precisely captured here.

Definition C.3.1. (Local design) Given a unitary ensemble $\{U_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ with a given probability distribution over the parameter space Θ, we say it forms a local t-design for the system if each element is factorized with respect to the partition, i.e. $U_{\boldsymbol{\theta}} = \otimes_m U^{(m)}_{\theta_m}$ each acting solely

on $\mathcal{H}_m$, and additionally

$$\int_\Theta d\boldsymbol{\theta} U_{\boldsymbol{\theta}}^{(m)\otimes t} \otimes U_{\boldsymbol{\theta}}^{(m)*\otimes t} = \int_{V\in\mathbb{U}(d_m)} d\mu(V) V^{(m)\otimes t} \otimes V^{(m)*\otimes t} \tag{C.14}$$

where the second integral is performed with respect to the Haar measure.

With this notation in place, we are ready to start. The first Lemma provides a formula for the expectation value of a circuit in the aforementioned class, showing that they form global 1-designs.

Lemma C.3.1 (Global 1-design)**.** Let $A, B \in \mathcal{B}$, and let $\{U_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ be a unitary ensemble forming a local 1-design. Then

$$\mathbb{E}_{\boldsymbol{\theta}} \left\{ \operatorname{Tr}\left\{ A U_{\boldsymbol{\theta}}^\dagger B U_{\boldsymbol{\theta}} \right\} \right\} = \frac{\operatorname{Tr}\{A\} \operatorname{Tr}\{B\}}{d} \tag{C.15}$$

Proof. Let $\{P_j\}_j$ be a local basis for the system, and consider the respective decompositions of A and B, namely $A = \sum_i a_i P_i$ and $B = \sum_j b_j P_j$. Note that, each component a_i is defined as $a_i = \operatorname{Tr}\{P_i A\}$, and consequently $a_0 = \frac{1}{\sqrt{d}} \operatorname{Tr}\{A\}$ (respectively for B). Then we have the following chain of equalities:

$$\begin{aligned}
\int_\Theta \prod_{m=1}^M d\boldsymbol{\theta} \operatorname{Tr}\left\{ A U_{\boldsymbol{\theta}}^\dagger B U_{\boldsymbol{\theta}} \right\} &= \sum_{i,j} a_i b_j \int_\Theta \prod_{m=1}^M d\boldsymbol{\theta} \operatorname{Tr}\left\{ P_i U_{\boldsymbol{\theta}}^\dagger P_j U_{\boldsymbol{\theta}} \right\} \\
&= \sum_{i,j} a_i b_j \prod_{m=1}^M \int_{U\in\mathbb{U}(d_m)} d\mu(U) \operatorname{Tr}\left\{ P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U \right\} \\
&= \sum_{i,j} a_i b_j \prod_{m=1}^M \frac{1}{d_m} \operatorname{Tr}\left\{ P_{i_m}^{(m)} \right\} \operatorname{Tr}\left\{ P_{j_m}^{(m)} \right\} = \sum_{i,j} a_i b_j \prod_{m=1}^M \delta_{0,i_m} \delta_{0,j_m} = a_0 b_0
\end{aligned} \tag{C.16}$$

where Tr_m denotes the partial trace over the m-th subsystem. □

Regarding the second moment, it can be computed using the following Lemma.

Lemma C.3.2. Let $\{P_j\}$ be a local basis and let $\{U_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ be a unitary ensemble forming a local 2-design. Then

$$\mathbb{E}_{\boldsymbol{\theta}} \left\{ \operatorname{Tr}\left\{ P_i U_{\boldsymbol{\theta}}^\dagger P_j U_{\boldsymbol{\theta}} \right\} \operatorname{Tr}\left\{ P_k U_{\boldsymbol{\theta}}^\dagger P_l U_{\boldsymbol{\theta}} \right\} \right\} = \delta_{ij}\delta_{kl} \prod_{m=1}^M \left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m}) \frac{1}{d_m^2 - 1} \right) \tag{C.17}$$

Proof. To prove this, we make use of the following result of Weingarten Calculus

$$\begin{aligned}
&\int_{U\in\mathbb{U}(d)} d\mu(U) \operatorname{Tr}\left\{ A U^\dagger B U \right\} \operatorname{Tr}\left\{ C U^\dagger D U \right\} \\
&= \frac{1}{d^2-1} \left(\operatorname{Tr}\{A\} \operatorname{Tr}\{B\} \operatorname{Tr}\{C\} \operatorname{Tr}\{D\} + \operatorname{Tr}\{AC\} \operatorname{Tr}\{BD\} \right) \\
&+ \frac{1}{d(d^2-1)} \left(\operatorname{Tr}\{AC\} \operatorname{Tr}\{C\} \operatorname{Tr}\{D\} + \operatorname{Tr}\{A\} \operatorname{Tr}\{C\} \operatorname{Tr}\{BD\} \right)
\end{aligned} \tag{C.18}$$

Based on Eq. (C.18), the result follows from direct integration:

$$
\begin{aligned}
&\prod_{m=1}^{M} \int_{U\in\mathbb{U}(d_m)} d\mu(U)\, \mathrm{Tr}\Big\{P_{i_m}^{(m)} U^\dagger P_{j_m}^{(m)} U\Big\}\, \mathrm{Tr}\Big\{P_{k_m}^{(m)} U^\dagger P_{l_m}^{(m)} U\Big\} \\
&= \prod_{m=1}^{M} \frac{1}{d_m^2-1}\Big(\mathrm{Tr}\Big\{P_{i_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{j_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{k_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{l_m}^{(m)}\Big\} + \mathrm{Tr}\Big\{P_{i_m}^{(m)}P_{k_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{j_m}^{(m)}P_{l_m}^{(m)}\Big\}\Big) \\
&\quad - \frac{1}{d_m(d_m^2-1)}\Big(\mathrm{Tr}\Big\{P_{i_m}^{(m)}P_{k_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{j_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{l_m}^{(m)}\Big\} + \mathrm{Tr}\Big\{P_{i_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{k_m}^{(m)}\Big\}\mathrm{Tr}\Big\{P_{j_m}^{(m)}P_{l_m}^{(m)}\Big\}\Big) \\
&= \prod_{m=1}^{M} \frac{1}{d_m^2-1}\left(d_m^2\,\delta_{0i_m}\delta_{0j_m}\delta_{0k_m}\delta_{0l_m} + \delta_{i_mk_m}\delta_{j_ml_m} - \delta_{0i_m}\delta_{0k_m}\delta_{j_ml_m} - \delta_{i_mk_m}\delta_{0j_m}\delta_{0l_m}\right) \\
&= \prod_{m=1}^{M} \delta_{0i_m}\delta_{0j_m}\delta_{0k_m}\delta_{0l_m} + \frac{1}{d_m^2-1}(\delta_{i_mk_m}-\delta_{0i_m}\delta_{0k_m})(\delta_{j_ml_m}-\delta_{0j_m}\delta_{0l_m}) \\
&= \delta_{ik}\delta_{jl}\prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right)
\end{aligned}
\tag{C.19}
$$

□

In particular, Lemma C.3.2 can be used to compute the variance of cost functions computed as expectation values $C(\boldsymbol{\theta}) = \mathrm{Tr}\{\mathcal{U}_{\boldsymbol{\theta}}(A)B\}$, i.e. in the absence of intermediate channels $\mathcal{E}_l$.

Proposition C.3.1. Let $A, B \in \mathcal{B}$, and let $\{U_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ be a unitary ensemble forming a local 2-design. Then

$$
\mathbb{E}_{\boldsymbol{\theta}}\left\{\mathrm{Tr}\left\{AU_{\boldsymbol{\theta}}^\dagger BU_{\boldsymbol{\theta}}\right\}^2\right\} = (\ell_A, \ell_B)\,, \tag{C.20}
$$

where $(\cdot,\cdot)$ is the scalar product defined in Eq. (6.8).

Proof. Let $\{P_j\}$ be a local basis for the system, and consider the respective decompositions of A and B. By Lemma C.3.2 we have

$$
\mathbb{E}_{\boldsymbol{\theta}}\left\{\mathrm{Tr}\left\{AU_{\boldsymbol{\theta}}^\dagger BU_{\boldsymbol{\theta}}\right\}^2\right\} = \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0j_m})\frac{1}{d_m^2-1}\right) \tag{C.21}
$$

In the following, it will be convenient to recast the product on the right-hand side of Eq. (C.21) into the equivalent formulation

$$
\prod_{m=1}^{M}\left(\delta_{0i_m}\delta_{0j_m} + (1-\delta_{0i_m})(1-\delta_{0i_m})\frac{1}{d_m^2-1}\right) = \sum_{\kappa\in\{0,1\}^M}\frac{1}{d_\kappa}\prod_{m=1}^{M}(\delta_{0i_m}\delta_{0j_m})^{1-\kappa_m}(1-\delta_{0i_m})^{\kappa_m}(1-\delta_{0j_m})^{\kappa_m} \tag{C.22}
$$

where the binary vectors $\kappa \in \{0,1\}^M$ identify all possible sets S_κ introduced in Appendix C.2 and $d_\kappa = \prod_{m=1}^{M}(d_m^2-1)^{\kappa_m}$. Putting it back into Eq. (C.21) we get

$$
\begin{aligned}
&\sum_{\kappa\in\{0,1\}^M} \frac{1}{d_\kappa} \sum_{i,j} a_i^2 b_j^2 \prod_{m=1}^{M} (\delta_{0i_m}\delta_{0j_m})^{1-\kappa_m} (1-\delta_{0i_m})^{\kappa_m} (1-\delta_{0j_m})^{\kappa_m} \\
&= \sum_{\kappa\in\{0,1\}^M} \frac{1}{d_\kappa} \left(\sum_i a_i^2 \prod_{m=1}^{M} \delta_{0i_m}^{1-\kappa_m} (1-\delta_{0i_m})^{\kappa_m} \right) \left(\sum_j b_j^2 \prod_{m=1}^{M} \delta_{0j_m}^{1-\kappa_m} (1-\delta_{0j_m})^{\kappa_m} \right) \\
&= \sum_{\kappa\in\{0,1\}^M} \frac{1}{d_\kappa} \left(\sum_i a_i^2 \tilde{\delta}_{i,\kappa} \right) \left(\sum_j b_j^2 \tilde{\delta}_{j,\kappa} \right) \\
&= \sum_{\kappa\in\{0,1\}^M} \frac{(\ell_A)_\kappa (\ell_B)_\kappa}{d_\kappa} = (\ell_A, \ell_B)
\end{aligned}
\tag{C.23}
$$

from which the Proposition follows.

□

This can be extended to the general case introducing the action of the intermediate channels $\mathcal{E}_l$, and in particular we get the following Proposition.

Proposition C.3.2. Let $A, B \in \mathcal{B}$ and $\Lambda : \mathcal{B} \to \mathcal{B}$ be a linear map. Furthermore, let $\{U_{\theta_1}\}_{\theta_1\in\Theta}$ and $\{V_{\theta_2}\}_{\theta_2\in\Theta}$ be independent, unitary ensembles each forming a local 2-design. Then

$$
\mathbb{E}_{\theta_1,\theta_2} \left\{ \mathrm{Tr}\left\{ A U_{\theta_1}^\dagger \Lambda(V_{\theta_2} B V_{\theta_2}^\dagger) U_{\theta_1} \right\}^2 \right\} = (\ell_A, T\ell_B) \tag{C.24}
$$

where T is the locality transfer matrix associated to Λ.

Proof. Let $\tilde{B}_{\theta_2} = \Lambda(V_{\theta_2} B V_{\theta_2})$. By Proposition C.3.2, we have

$$
\mathbb{E}_{\theta_2}\mathbb{E}_{\theta_1} \left\{ \mathrm{Tr}\left\{ A U_{\theta_1}^\dagger B_{\theta_2} U_{\theta_1} \right\}^2 \right\} = \mathbb{E}_{\theta_2} \left\{ (\ell_A, \ell_{\tilde{B}_{\theta_2}}) \right\} = \sum_{\kappa\in\{0,1\}^M} \frac{1}{d_\kappa} (\ell_A)_\kappa \, \mathbb{E}_{\theta_2} \left\{ (\ell_{\tilde{B}_{\theta_2}})_\kappa \right\} \tag{C.25}
$$

Expanding the definition on the last term with respect to the local basis $\{P_j\}_j$, ad applying again Proposition C.3.2 we get

$$
\begin{aligned}
\mathbb{E}_{\theta_2} \left\{ (\ell_{\tilde{B}_{\theta_2}})_\kappa \right\} &= \sum_i \mathbb{E}_{\theta_2} \left\{ \mathrm{Tr}\left\{ P_i \tilde{B}_{\theta_2} \right\}^2 \right\} \tilde{\delta}_{i,\kappa} \\
&= \sum_i \mathbb{E}_{\theta_2} \left\{ \mathrm{Tr}\left\{ \Lambda^\dagger(P_i) V_{\theta_2}^\dagger B V_{\theta_2} \right\}^2 \right\} \tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda\in\{0,1\}^M} \frac{1}{d_\lambda} \sum_{i,j,k} \mathrm{Tr}\left\{ P_i \Lambda(P_j) \right\}^2 \mathrm{Tr}\{P_k B\}^2 \tilde{\delta}_{j,\lambda} \tilde{\delta}_{k,\lambda} \tilde{\delta}_{i,\kappa} \\
&= \sum_{\lambda\in\{0,1\}^M} \left(\frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\left\{ P_i \Lambda(P_j) \right\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \right) \left(\sum_k \mathrm{Tr}\{P_k A\}^2 \tilde{\delta}_{k,\lambda} \right) \\
&= \sum_{\lambda\in\{0,1\}^M} T_{\kappa,\lambda} (\ell_B)_\lambda = (T\ell_B)_\kappa
\end{aligned}
\tag{C.26}
$$

where $\Lambda^\dagger$ is the Hermitian adjoint of Λ with respect to the Hilbert-Schmidt scalar product. □

Iterated application of Proposition C.3.2 for an initial state ρ, an observable H, and a general intermediate quantum channel $\mathcal{E}$, yields Eq. (6.10) of the main text.

C.4 Proof of the deep circuit formula

The proof of Theorem 6.3.1 is based on the characterization of the general LTM for the Hermitian adjoint $\mathcal{E}^\dagger$ of arbitrary quantum channel. To do so, several aspects of non-negative matrix theory, as well as the contractivity properties of $\mathcal{E}^\dagger$ are key. For sake of clarity all the main concepts are introduced in Appendix C.1.

The proof is divided into two parts: in the first, we characterize LTMs of general CPU maps, and subsequently we use their properties to prove the deep circuit variance formula.

Characterization of LTMs of general CPU maps

We now study the structure of the LTM of a general CPU map. Thanks to this analysis, we will be able to compute the limiting value $\mathbb{V}^\infty_{\rho,H}$ by describing the quantum circuit in the Heisenberg picture. Denoting by T the resulting LTM, we start by computing the general form of integer powers T^L of T.

Lemma C.4.1 (Limiting form of T)**.** Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T and T^L take the form

$$T = \begin{pmatrix} T_0 & & & & R_0 \\ & T_1 & & & R_1 \\ & & \ddots & & \vdots \\ & & & T_z & R_z \\ & & & & Q \end{pmatrix}, \quad T^L = \begin{pmatrix} T_0^L & & & & A_0^{(L)} \\ & T_1^L & & & A_1^{(L)} \\ & & \ddots & & \vdots \\ & & & T_z^L & A_z^{(L)} \\ & & & & Q^L \end{pmatrix} \tag{C.27}$$

up to a basis state index permutation, where each T_z is an irreducible matrix, and $A_z^{(L)} = \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l}$.

Note that here, we split the contributions R_z corresponding to each component T_z for convenience. This is related to the convention of Fig. 6.2 by the relation $R = \sum_z R_z$ with clear meaning of the symbols.

Proof. We start by putting T into the canonical form of Eq. (C.4). In this form, the powers of the diagonal blocks T_z^L are trivially the diagonal blocks of T^L. Instead, the result about $A^{(L)}$ follows by induction. In fact, both the base case and the inductive one follow from matrix multiplication rules, of $T \cdot T$ and $T^{L-1} \cdot T$ respectively. In particular we have

$$\begin{aligned} A_z^{(2)} &= T_z R_z + R_z Q \\ A_z^{(L)} = T_z^{L-1} R_z + A^{(L-1)} Q &= T_z^{L-1} R_z + \sum_{l=0}^{L-2} T_z^l R_z Q^{L-1-l} = \sum_{l=0}^{L-1} T_z^l R_z Q^{L-1-l} \end{aligned} \tag{C.28}$$

which gives the proposition. □

As an immediate consequence of Hölder inequality and contractivity of CPTP maps, the value of the variance is upper bounded by $\|H\|_\infty^2$. Since by Proposition 6.2.1, this quantity is linked to $(\ell_\rho, T^L \ell_H)$, it is expected that the spectral radius $\rho(T)$ of T is upper bounded by 1. More specifically, we can prove the following Proposition.

Proposition C.4.1 (Spectral radius of T)**.** Let $\Lambda : \mathcal{B} \to \mathcal{B}$ be a CPU map and T be the corresponding LTM. Then T is contractive in the sense of the spectral radius, i.e. $\rho(T) \leq 1$. Moreover, the component Q is strictly contractive, namely $\rho(Q) < 1$.

Proof. The statement follows as a consequence of the Kadison-Schwarz inequality and the Subinvariance theorem.

First note that, for a generic Λ, the structure of Q is not as well-behaved as T_z, as Q is not necessarily *irreducible*. However, as any non-negative matrix, also Q can be cast in canonical block upper triangular form by means of a basis permutation, where each diagonal block is irreducible.

$$Q = \begin{pmatrix} \boxed{Q_1} & \boxed{*} & \cdots & \boxed{*} \\ & \boxed{Q_2} & & \boxed{*} \\ & & \ddots & \vdots \\ & & & \boxed{Q_k} \end{pmatrix} \tag{C.29}$$

With this in mind, we can study the spectral radius of T and Q in terms of the spectral radii of each block T_z and Q_k, i.e. the corresponding Perron eigenvalues, since $\rho(T) = \max\{\max_z r_z, \max_k r_{Q_k}\}$ and $\rho(Q) = \max_k r_{Q_k}$.

By Lemma C.2.2, we are free to choose the basis used to express the matrix T. In particular, we choose the normalized Pauli basis $\{P_i\}$, which, besides being a local basis for $\mathcal{B}$, is also unitary up to normalization, namely $P_i^2 = \mathbb{1}/d \; \forall i$. Exploiting Eq. (C.8), we can compute the column sum of T as

$$\begin{aligned} \sum_\kappa T_{\kappa,\lambda} &= \sum_\kappa \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\big\{P_i \Lambda(P_j)\big\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\big\{P_i \Lambda(P_j)\big\}^2 \tilde{\delta}_{j,\lambda} \left(\sum_\kappa \tilde{\delta}_{i,\kappa} \right) \\ &= \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\big\{P_i \Lambda(P_j)\big\}^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\big\{\Lambda(P_j)^2\big\} \tilde{\delta}_{j,\lambda} \\ &\leq \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\Big\{\Lambda(P_j^2)\Big\} \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \tilde{\delta}_{j,\lambda} = 1 \end{aligned} \tag{C.30}$$

where the third and last equality follow from Lemma C.2.1, the inequality is Kadison-Schwarz and the second to last equality is the unitary property of the basis. If Λ is unitary, the inequality is saturated, and T becomes a *stochastic* matrix. In general, Eq. (C.30) show *sub-stochasticity* of T. Indeed, this condition can be recast in vector form as $v^t T \leq v^t$, where $v_\kappa = 1 \forall \kappa$. In particular, this holds for all irreducible blocks in the diagonal, which by the Subinvariance theorem implies $r_z \leq 1$ and $r_{Q_k} \leq 1$, giving $\rho(T) \leq 1$. Focusing on Q, we observe that, by definition, each irreducible block Q_k is *inessential*, i.e. is connected to some other block. In terms of the matrix T, this means that, considering the columns involving Q_k, there is always an index λ in the support of Q_k such that

$$\sum_\kappa (T - Q_k)_{\kappa,\lambda} > 0, \tag{C.31}$$

which implies that $\exists \lambda$ s.t. $\sum_\kappa (Q_k)_{\kappa,\lambda} < 1$. Written in matrix form, this reads $v^t Q_k \leq v^t$, $v^t Q_k \neq v^t$, which by the Subinvariance theorem, implies $r_{Q_k} \neq 1$. Putting everything together, one gets $\rho(Q) < 1$.

□

When analysing the single irreducible components T_z, we can be more specific, and find an equivalence between the value of the column-sum of the block and the value of the corresponding spectral radius. This is especially useful in the computation of the dominant eigenvectors, which is explicitly stated in the following Corollary.

Corollary C.4.1. Let T be a LTM and T_z be an irreducible block, then $\rho(T_z) = 1 \Leftrightarrow \sum_\kappa (T_z)_{\kappa,\lambda} = 1$, or equivalently $v_z^t T_z = v_z^t$, where $(v_z)_\kappa = 1 \forall \kappa$ is the left eigenvector of the dominant eigenvalue.

Proof. The result follows from the same proof strategy as above, and is a direct consequence of the Subinvariance theorem. □

Intuitively, the blocks which are out of such hypothesis won't contribute to the large L limit, and indeed the contribution of the Q, T_z and $A_z^{(L)}$ is bounded to decay exponentially in the number of layers.

Proposition C.4.2. Let T be a LTM, and let T_z be an irreducible block with $r_z < 1$. Then, as $L \to \infty$, $\|T_z^L\| \to 0$ and $\|A_z^{(L)}\| \to 0$ exponentially fast for any matrix norm $\|\cdot\|$. Similarly, also $\|Q^L\| \to 0$.

Proof. The proposition can be proven using Gelfand's formula. In particular since $\lim_{L\to\infty} \|T_z^L\|^{1/L} = r_z$, we can always bound $\|T_z^L\| \leq K\tau^L$, for some constant $K > 0$ and $\tau = r_z + \epsilon < 1$ for an arbitrarily small ϵ. In the same way, by Proposition C.4.1 a similar result can be obtained for Q. Finally, the absorption term $A_z^{(L)}$ can also be bounded using Lemma C.4.1. In that case we have

$$\|A_z^{(L)}\| \leq \sum_{l=0}^{L-1} \|T_z^L\| \|R_z\| \|Q^{L-1-l}\| \leq \|R_z\| K_T K_Q \tau^l \kappa^{L-1-l} \leq K_A \alpha^L \tag{C.32}$$

with some constant $K_A > 0$ and $\alpha = \max\{\kappa, \tau\} < 1$. This can be obtained again using Gelfand's formula on both T_z and Q, and by sub-additivity and sub-multiplicativity of the matrix norm $\|\cdot\|$. □

Proof of the theorem

As stated in the main text, the limiting value of $\mathbb{V}_{\rho,H}^L$ for layered, locally random channels as in Definition 6.1.3 is obtained by studying the spectral properties of the LTM of the intermediate channel in the Heisenberg picture. In particular, the previous discussion suggests a limiting value for the variance of the form

$$\mathbb{V}_{\rho,H}^\infty = \sum_{z\,|\,\rho(T_z)=1} (\ell_\rho, w_z)(\ell_H)_z + (\ell_\rho, w_z)(A\ell_H)_z \tag{C.33}$$

with some normalized, strictly positive vector w_z, and an absorption matrix A given by the limit[1]

$$A = \lim_{L\to\infty} \sum_z A_z^{(L)}. \tag{C.34}$$

Indeed, the following shows that this is the case, and provides an analytical form for A.

[1] This is especially clear in light of Eq. (6.18) of the main text, and the relation $(A\ell_H)_z = A_z \ell_H$, coming from the definition of locality and Eq. (C.34).

Theorem C.4.1 (Deep circuit variance)**.** Let $\rho, H \in \mathcal{B}$ and let Φ_θ a be layered, locally random channel as in Definition 6.1.3. Then the Cesàro average of $\mathbb{V}^L_{\rho,H}$ converges to Eq. (C.33), and we have

$$\left|\frac{1}{L}\sum_{l=0}^{L}\mathbb{V}^l_{\rho,H} - \mathbb{V}^\infty_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right), \tag{C.35}$$

for some constant $\beta > 0$. Additionally, if all essential blocks are aperiodic, then $\mathbb{V}^L_{\rho,H}$ is convergent, and we have

$$\left|\mathbb{V}^L_{\rho,H} - \mathbb{V}^\infty_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right), \tag{C.36}$$

where the right eigenvector w_z of T_z is a strictly positive vector, i.e. $(w_z)_\kappa > 0\ \forall \kappa$, and $A = R(\mathbb{1} - Q)^{-1}$ are the absorption coefficients of each essential block.

Proof. Thanks to Proposition C.4.2, only irreducible components with $\rho(T_z) = 1$ will contribute to the limit, so we can restrict our analysis to those alone. Consider then an irreducible block T_z with unit spectral radius, and of period d. While the full version of Perron-Frobenius theorem does not directly apply to T_z, it is a well-known result of non-negative matrix theory [102, 70] that the matrix T_z^d can be cast to a block diagonal form by a permutation, with irreducible and *aperiodic* blocks, for which we can apply it. However, it is crucial to notice that while $\lim_{N\to\infty} T_z^{dN} = T_z^{(d\infty)}$ exists, this does not imply that $\lim_{L\to\infty} T_z^L$ does. In fact, different sub-sequences might have different limiting values, and in particular $\lim_{N\to\infty} T_z^{dN+m} = T_z^{(d\infty)}T^m$, which is different for all $m = 0, \ldots d-1$. In the periodic scenario then, T_z^L does not have a limit, and the only convergent quantity is the *Cesàro average*, i.e.

$$P_z = \lim_{L\to\infty}\frac{1}{L}\sum_{l=1}^{L}T_z^l = T_z^{(d\infty)}\frac{1}{d}\sum_{m=0}^{d-1}T_z^m \tag{C.37}$$

where $P_z = w_z v_z^t$ can be shown to be the Perron projector associated to T_z [102, 70]. Despite more cumbersome, a totally analogous approach allows determining the limiting values of $A^{(dN+m)}$ as well, as shown in the following Proposition.

Proposition C.4.3. Given a LTM with a periodic irreducible block T_z of period d, then

$$\lim_{N\to\infty} A^{(dN+m)} = T^{(d\infty)}A^{(m)} + A^{(d\infty)}Q^m \tag{C.38}$$

where $A^{(d\infty)} = \sum_{m=0}^{d-1} T^{(d\infty)}A^{(d)}(\mathbb{1} - Q^d)^{-1}$ and $A^{(0)} = 0$.

Proof. Starting from the definition of $A^{(l)}$, we can first find the limiting value $A^{(d\infty)}$ of $A^{(dN)}$:

$$\begin{aligned} A^{(dN)} &= \sum_{l=0}^{Nd-1} T_z^l R_z Q^{dN-1-l} = \sum_{m=0}^{d-1}\sum_{n=0}^{N-1} T_z^{nd+m} R_z Q^{(N-1)d-nd+d-1-m} \\ &= \sum_{m=0}^{d-1}\sum_{n=0}^{N-1} T_z^{nd} T_z^m R_z Q^{d-1-m}(Q^d)^{N-n} = \sum_{n=0}^{N-1} T_z^{nd} A^{(d)}(Q^d)^{N-n} \\ &= T_z^{(d\infty)}A^{(d)}\sum_{n=0}^{N-1}(Q^d)^{N-n} + \sum_{n=0}^{N-1}\Delta_z^{(n)}A^{(d)}(Q^d)^{N-n} \end{aligned} \tag{C.39}$$

where $\Delta_z^{(N)} = T_z^{dN} - T_z^{(d\infty)}$. Since T_z^d is block diagonal, and each of the d blocks T_{zm} is irreducible and aperiodic, we have that $T_{zm}^N \to P_{zm}$ exponentially fast, i.e. $\|T_{zm}^N - P_{zm}\| = \|\Delta_{zm}^{(N)}\| < K_{zm}\tau_{zm}^N$, for some $K > 0$ and $\tau < 1$. Then, $\|\Delta_z^{(N)}\| \le \sum_{m=0}^{d}\|\Delta_{zm}^{(N)}\| \le K_z\tau_z^N$, where $\tau = \max_m\{\tau_{zm}\} < 1$. Together with Proposition C.4.1, this implies that, the last term in

Eq. (C.39) approaches zero with the same exponential speed, similarly to what happens in Proposition C.4.2. Putting everything together, and considering that $\sum_{n=0}^{N-1} X^n \to (\mathbb{1} - X)^{-1}$ $\forall X$ such that $\rho(X) < 1$, we have

$$A^{(d\infty)} = T^{(d\infty)} A^{(d)} (\mathbb{1} - Q^d)^{-1} \tag{C.40}$$

At this point, by induction similarly to Lemma C.4.1, one can easily show that

$$A^{(dN+m)} = T^{(dN)} A^{(m)} + A^{(dN)} Q^m \tag{C.41}$$

which yields the proposition. □

In particular, the Cesàro average converges and we have the expression

$$A_z = \lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} A^{(l)} = \frac{1}{d} \sum_{m=0}^{d-1} T^{(d\infty)} A^{(m)} + A^{(d\infty)} Q^m. \tag{C.42}$$

Recalling that the right eigenvector v_z can be explicitly calculated when $\rho(T_z) = 1$ (see Corollary C.4.1), this allows to obtain the final form of $\mathbb{V}^{\infty}_{\rho,H}$ by ordinary matrix vector multiplication. As a special case, if all relevant blocks T_z are aperiodic, then the limits of T_z^L and $A_z^{(L)}$ converge, and we have $P_z = T_z^{(1\infty)}$ and $A_z = A_z^{(1\infty)} = P_z R_z (\mathbb{1} - Q)^{-1}$. Finally, the exponential upper bound in Eq. (C.35) and Eq. (C.36) is also obtained as a consequence of the preceding analysis. In particular if we denote by T^{∞} the matrix with P_z in place of T_z, A_z in place of R_z and zero otherwise, we have

$$\left| \frac{1}{L} \sum_{l=0}^{L} \mathbb{V}^{l}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H} \right| = \left| \left(\ell_\rho, \frac{1}{L} \sum_{l=0}^{L} T^l - T^{\infty} \ell_H \right) \right| \leq \left\| \frac{1}{L} \sum_{l=0}^{L} T^l - T^{\infty} \right\| \sqrt{(\ell_\rho, \ell_\rho)(\ell_H, \ell_H)}. \tag{C.43}$$

by Cauchy-Schwarz inequality. Since all blocks converge exponentially fast from the above discussion, the matrix norm is also exponentially decaying. Moreover, by construction $(\ell_A, \ell_A) \leq \sqrt{\sum_\kappa (\ell_A)_\kappa^2} \leq \sum_\kappa (\ell_A)_\kappa = \|A\|_2^2 \; \forall A \in \mathcal{B}$, which concludes the proof. □

While for an explicit calculation of w_z one should in general rely on case-specific analyses, a general result can be derived for a subclass of channels especially useful in the context of quantum computing, namely single qubit noise.

Proposition C.4.4 (Single qubit noise)**.** Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3. Assume moreover that the intermediate channel is of the form $\mathcal{E} = \mathcal{N}(W\rho W^\dagger)$, where W is a unitary transformation and $\mathcal{N} = \mathcal{N}_1 \otimes ... \otimes \mathcal{N}_n$ is a composition of single qubit quantum channels. Then

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z \,|\, \rho(T_z)=1} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z (A\ell_H)_z}{d_z}. \tag{C.44}$$

Proof. Without loss of generality, we can consider the single qubit channels $\mathcal{N}_m$ to be in their normal form of Lemma 2.2.1. In particular, this holds due to the invariance of the LTM with respect to changes of local bases (Lemma C.2.2). In terms of the adjoint maps $\mathcal{N}_m^\dagger$, this condition reads $\mathcal{N}_m^\dagger(P_{j_m}) = t_{j_m} P_0 + \lambda_{j_m} P_j$ and can be used to compute $\mathrm{Tr}\{\mathcal{N}_m^\dagger(P_{j_m})^2\} = t_{j_m}^2 + \lambda_{j_m}^2 \leq 1$ by Lemma 2.2.1. We now show that, when $P_j \in \mathcal{B}_z$ pertains to an irreducible component of spectral radius $\rho(T_z) = 1$, then the inequality must be saturated. In particular,

thanks to Corollary C.4.1 we know that $\sum_\kappa (T_z)_{\kappa,\lambda} = 1$. By Eq. (C.30) this implies

$$\begin{aligned}\sum_\kappa (T_z)_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\big\{(W^\dagger \mathcal{N}^\dagger(P_j)W)^2\big\}\tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\big\{\mathcal{N}^\dagger(P_j)^2\big\}\tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda}\sum_j \prod_{m=1}^{M} \mathrm{Tr}\big\{\mathcal{N}_m^\dagger(P_{j_m})^2\big\}\tilde{\delta}_{j,\lambda} = 1.\end{aligned} \tag{C.45}$$

Since each term in the product is upper-bounded by 1, Eq. (C.45) implies $\mathrm{Tr}\big\{\mathcal{N}_m^\dagger(P_{j_m})^2\big\} = t_{j_m}^2 + \lambda_{j_m}^2 = 1\ \forall j_m$. This condition is only compatible with Lemma 2.2.1 if $t_{j_m}^2 = 0$ and $\lambda_{j_m}^2 = 1$. This result can now be used to show that, the adjoint of $T_z^\dagger$ of the LTM T_z, must also be column stochastic. Indeed, if we consider the expansion of $WP_jW^\dagger$ with respect to the normalized Pauli basis, we can get

$$\begin{aligned}\sum_\kappa (T_z)^\dagger_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\big\{\mathcal{N}(WP_jW^\dagger)^2\big\}\tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left\{\left(\sum_i \mathrm{Tr}\big\{P_iWP_jW^\dagger\big\}\mathcal{N}(P_i)\right)^2\right\}\tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\left\{\left(\sum_i \mathrm{Tr}\big\{P_iWP_jW^\dagger\big\}\prod_m \lambda_{i_m}P_i\right)^2\right\}\tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda}\sum_{i,j}\prod_m \lambda_{i_m}^2\,\mathrm{Tr}\big\{P_iWP_jW^\dagger\big\}^2\tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\Big\{WP_j^2W^\dagger\Big\}\tilde{\delta}_{j,\lambda} = 1.\end{aligned} \tag{C.46}$$

This allows to compute the right eigenvector w_z of the leading eigenvalue of T_z. Using Eq. (C.11), we have indeed

$$\sum_\lambda (T_z)_{\kappa,\lambda} d_\lambda = \sum_\lambda (T_z)^\dagger_{\lambda,\kappa} d_\kappa = d_\kappa \tag{C.47}$$

which implies $(w_z)_\lambda = d_\lambda/d_z$, where the normalization factor $d_z = \sum_\lambda d_\lambda$ is necessary to ensure $w_z v_z^t = P_z$ is indeed a projection, thus concluding the proof. □

As a consequence of this last result, we can compute the variance of generic unitary circuits.

Corollary C.4.2 (Unitary circuits)**.** Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3, with unitary intermediate channel $\mathcal{E}(\rho) = W\rho W^\dagger$. Then we have

$$\mathbb{V}^\infty_{\rho,H} = \sum_{z>0} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z} \tag{C.48}$$

Proof. This form of the variance is a special case of Proposition C.4.4, putting $\mathcal{N}(\rho) = \rho$, and noting that the absorption terms must vanish. In particular, this follows from Eq. (C.30), observing that unitary channels saturate Kadison-Schwarz inequality, which combined with Corollary C.4.1 imply $Q = 0$. □

A similar computation is also possible for global noise channels, like correlated Pauli noise. In particular $(w_z)_\kappa$ can be analytically computed the intermediate if channel $\mathcal{E}$ is of the form $\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger)$, with $\mathcal{N}(\rho)$ a *diagonal* Pauli channel. More precisely, we have the following Proposition.

Proposition C.4.5 (Global Pauli noise map)**.** Let $\rho, H \in \mathcal{B}$ and let Φ_θ be a layered, locally random quantum channel as in Definition 6.1.3. Assume moreover that the intermediate channel is of the form $\mathcal{E}(\rho) = \mathcal{N}(W\rho W^\dagger)$, where $\mathcal{N}(\rho) = \sum_k p_k \bar{P}_k \rho \bar{P}_k$, with $p_k \geq 0$, $\sum_k p_k = 1$ and $\bar{P}_k$ denote non-normalized Pauli string, i.e. satisfying $\bar{P}_k^2 = \mathbb{1}$. Then

$$\mathbb{V}^\infty_{\rho,H} = \sum_{z\,|\,\rho(T_z)=1} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}. \tag{C.49}$$

In particular, $\rho(T_z) = 1$ if and only if $[\bar{P}_k, P_j] = 0 \ \forall P_j \in \mathcal{B}_z \ \forall \bar{P}_k | p_k > 0$.

Proof. The proof follows the same path as Proposition C.4.4. In particular, from the definition of the intermediate channel, it follows that $\mathcal{N}$ is unital. This implies that $\mathrm{Tr}\{\mathcal{N}(P_j)^2\} \leq \mathrm{Tr}\{\mathcal{N}(P_j^2)\} = 1 \ \forall j$ by Kadison-Schwarz inequality. More specifically, we can compute the action of $\mathcal{N}$ explicitly on (normalized) Pauli strings as follows

$$\mathcal{N}(P_j) = \sum_k p_k \bar{P}_k P_j \bar{P}_k = \sum_k p_k s_{k,j} P_j = \mathcal{N}^\dagger(P_j), \tag{C.50}$$

where $s_{k,j} = 1$ if $\bar{P}_k$ and P_j commute, and $s_{i,j} = -1$ otherwise. From this it follows that

$$\mathcal{N}(P_j)^2 = \mathcal{N}^\dagger(P_j)^2 = \left(\sum_k p_k s_{k,j}\right)^2 \frac{\mathbb{1}}{d}. \tag{C.51}$$

Combining it to Corollary C.4.1 and Eq. (C.45), we get that $\rho(T_z) = 1 \Leftrightarrow \sum_\kappa (T_z)_{\kappa,\lambda} = 1 \Leftrightarrow \sum_k p_k s_{k,j} = 1$, i.e. $s_{k,j} = 1 \forall k | p_k > 0 \ \forall j \in T_z$.

Let's denote for convenience by c_j the quantity $c_j = \sum_k p_k s_{k,j}$. The previous discussion implies $|c_j| \leq 1 \ \forall j$, with equality holding for all elements of $\mathcal{B}_z$. Then we have

$$\begin{aligned}
\sum_\kappa (T_z)^\dagger_{\kappa,\lambda} &= \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\{\mathcal{N}(WP_jW^\dagger)^2\} \tilde{\delta}_{j,\lambda} \\
&= \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\left\{\left(\sum_i \mathrm{Tr}\{P_i W P_j W^\dagger\} \mathcal{N}(P_i)\right)^2\right\} \tilde{\delta}_{j,\lambda} \\
&= \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\left\{\left(\sum_i \mathrm{Tr}\{P_i W P_j W^\dagger\} c_i P_i\right)^2\right\} \tilde{\delta}_{j,\lambda} \\
&= \frac{1}{d_\lambda} \sum_{i,j} c_i^2 \,\mathrm{Tr}\{P_i W P_j W^\dagger\}^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_j \mathrm{Tr}\{W P_j^2 W^\dagger\} \tilde{\delta}_{j,\lambda} = 1.
\end{aligned} \tag{C.52}$$

By Eq. (C.47), we get again $(w_z)_\lambda = d_\lambda / d_z$.

As a last step, we show how the absorption coefficients must vanish in this case. This is a consequence of the fact that $|c_j| \leq 1$. Indeed, by computing a generic element $T_{\kappa,\lambda}$, we get

$$\begin{aligned}
T_{\kappa,\lambda} &= \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\{P_i W^\dagger \mathcal{N}(P_j) W\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda} \sum_{i,j} c_j^2 \,\mathrm{Tr}\{P_i W^\dagger P_j W\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} \\
&\leq \frac{1}{d_\lambda} \sum_{i,j} \mathrm{Tr}\{P_i W^\dagger P_j W\}^2 \tilde{\delta}_{i,\kappa} \tilde{\delta}_{j,\lambda} = T^W_{\kappa,\lambda},
\end{aligned} \tag{C.53}$$

where T^W is the LTM of the unitary channe defined by the unitary W alone. Since by Corollary C.4.2 Q (and consequently R) vanish for unitary channels, then R must also vanish here,

implying $A = 0$. □

On the opposite limit, if the noise map is strictly contractive in at least one direction in each $\mathcal{B}_z$, then the combination of noise and entanglement is strong enough to kill the variance in each of the absorbing subspaces. As a consequence, only the absorption term to $\mathcal{B}_0$ remains, since no channel can be contractive there by trace preservation.

Corollary C.4.3 (Noise-induced concentration)**.** Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3, with intermediate quantum channel $\mathcal{E}$, and let $\{P_j\}$ denote the normalized Pauli basis. If $\|\mathcal{E}^\dagger(P_j)\|_2 < 1$ for some $j \in T_z$, $\forall z$, then

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}. \tag{C.54}$$

In particular, if the channel is unital, $\mathbb{V}^{\infty}_{\rho,H} = 0$.

Proof. This form of the variance is a special case of Eq. (C.33), where all absorbing components vanish, and we have $T = Q$. In particular, this follows from Eq. (C.30), observing that the above condition implies that for $\mathcal{E}^\dagger$ Kadison-Schwarz inequality is strict, which combined with Corollary C.4.1 imply $T_z = 0\ \forall z > 0$, leaving only T_0. Finally, the corollary follows from the normalization condition $\mathrm{Tr}\{\rho\} = 1$ on ρ, which ensures $(\ell_\rho)_0 = 1/d$. □

C.5 Proofs for the estimates of the absorption coefficients

We first prove a general lower bound for the absorption coefficients in the setting of single qubit noise.

Proposition C.5.1. Let $\rho, H \in \mathcal{B}$ and let $\Phi_{\boldsymbol{\theta}}$ be a layered, locally random quantum channel as in Definition 6.1.3, and let the intermediate channel take the form of Eq. (C.57). Then, for all norm-preserving, strongly connected components T_z of the LTM T, we have:

$$\frac{(A\ell_H)_z}{d} \geq \sum_{j \,|\, P_j \in \mathcal{B}_z} (\Lambda_j, \ell_H) \tag{C.55}$$

where the operators Λ_j are defined by the action of the noise channels as

$$\Lambda_j = \mathcal{Q}(\tilde{\Lambda}_j), \quad \tilde{\Lambda}_j = \bigotimes_m \mathcal{N}_m\left(\frac{P_{j_m}}{\sqrt{2}}\right) \tag{C.56}$$

with P_j are normalized Pauli strings and $\mathcal{Q} : \mathcal{B} \to \mathcal{B}_{\mathcal{Q}}$ is a projection onto $\mathcal{B}_{\mathcal{Q}}$.

Proof. By Theorem 6.3.1, we can express the absorption matrix explicitely in term of the relevant portions of the LTM. In particular we have:

$$(A\ell_H)_z = \sum_{\zeta,\kappa,\lambda} R_{\zeta,\kappa}(\mathbb{1} - Q)^{-1}_{\kappa,\lambda}(\ell_H)_\lambda \delta_{\zeta,z}\delta_{\kappa,Q} \tag{C.57}$$

where we used the shortcut notation $\delta_{\zeta,z} = \sum_{\nu\in T_z} \delta_{\zeta,\nu}$ and $\delta_{\kappa,Q} = \sum_{\nu\in Q} \delta_{\kappa,\nu}$. Such terms are necessary to avoid counting portions of T_z and Q into the block R.

As a first step, we compute the term $\sum_\zeta R_{\zeta,\kappa}\delta_{\zeta,z}$ in this noise model. For the sake of readability, let us further denote by $\tilde{\delta}_{i,z}$ the quantity $\sum_\zeta \tilde{\delta}_{i,\zeta}\delta_{\zeta,z}$.

$$\begin{aligned}\sum_{\zeta} R_{\zeta,\kappa}\delta_{\zeta,z} &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\left\{P_i W^\dagger \mathcal{N}^\dagger(P_j) W\right\}^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,Q} \\ &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\left\{\sum_\mu w_{i,\mu} P_\mu \mathcal{N}^\dagger(P_j)\right\}^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,Q} \\ &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\left\{\sum_\mu w_{i,\mu} \mathcal{N}(P_\mu) P_j\right\}^2 \tilde{\delta}_{i,z}\tilde{\delta}_{j,\kappa}\delta_{\kappa,Q} \\ &= \frac{1}{d_\kappa}\sum_{j}\sum_{\mu,\nu}\left(\sum_i w_{i,\mu} w_{i,\nu}\tilde{\delta}_{i,z}\right)\mathrm{Tr}\left\{\mathcal{N}(P_\mu)P_j\right\}\mathrm{Tr}\left\{\mathcal{N}(P_\nu)P_j\right\}\tilde{\delta}_{j,\kappa}\delta_{\kappa,Q}\end{aligned} \tag{C.58}$$

where $w_{i,\mu} = \mathrm{Tr}\left\{W P_i W^\dagger P_\mu\right\}$. Recall that, by definition of LTM, if $P_i \in \mathcal{B}_z$, then $W P_i W^\dagger \in \mathcal{B}_z$ as well. For this reason, $w_{i,\mu} = 0\ \forall \mu\ s.t. P_\mu \notin \mathcal{B}_z$, i.e. $w_{i,\mu} = w_{i,\mu}\tilde{\delta}_{\mu,z}\tilde{\delta}_{i,z}$. Furthermore, we have that

$$\begin{aligned}&\sum_i w_{i,\mu} w_{i,\nu}\tilde{\delta}_{i,z}\tilde{\delta}_{\mu,z}\tilde{\delta}_{\nu,z} = \sum_i w_{i,\mu} w_{i,\nu} \\ &= \sum_i \mathrm{Tr}\left\{W P_i W^\dagger P_\mu\right\}\mathrm{Tr}\left\{W P_i W^\dagger P_\nu\right\} = \mathrm{Tr}\left\{\left(\sum_i \mathrm{Tr}\left\{P_i W^\dagger P_\nu W\right\}\right) P_i W^\dagger P_\mu W\right\} \\ &= \mathrm{Tr}\left\{W^\dagger P_\nu W W^\dagger P_\mu W\right\} = \mathrm{Tr}\left\{P_\nu P_\mu\right\} = \delta_{\mu,\nu} = \delta_{\mu,\nu}\tilde{\delta}_{\mu,z}\tilde{\delta}_{\nu,z}\end{aligned} \tag{C.59}$$

Consequently, we have

$$\begin{aligned}\sum_{\zeta} R_{\zeta,\kappa}\delta_{\zeta,z} &= \frac{1}{d_\kappa}\sum_{j}\sum_{\mu,\nu}\delta_{\mu,\nu}\tilde{\delta}_{\mu,z}\mathrm{Tr}\left\{\mathcal{N}(P_\mu)P_j\right\}\mathrm{Tr}\left\{\mathcal{N}(P_\nu)P_j\right\}\tilde{\delta}_{i,\kappa}\delta_{\kappa,Q} \\ &= \frac{1}{d_\kappa}\sum_{j,\mu}\mathrm{Tr}\left\{\mathcal{N}(P_\mu)P_j\right\}^2\tilde{\delta}_{j,\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,Q} = \frac{1}{d_\kappa}\sum_{j,\mu}\mathrm{Tr}\left\{\sqrt{d}\tilde{\Lambda}_\mu P_j\right\}^2\tilde{\delta}_{j,\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,Q} \\ &= d\sum_\mu \frac{(\ell_{\tilde{\Lambda}_\mu})_\kappa}{d_\kappa}\tilde{\delta}_{\mu,z}\delta_{\kappa,Q} = d\sum_\mu \frac{(\ell_{\Lambda_\mu})_\kappa}{d_\kappa}\tilde{\delta}_{\mu,z}\end{aligned} \tag{C.60}$$

Putting everything into Eq. (C.57), we get:

$$\begin{aligned}\frac{(A\ell_H)_z}{d} &= \frac{1}{d}\sum_{\zeta,\kappa,\lambda} R_{\zeta,\kappa}(\mathbb{1}-Q)^{-1}_{\kappa,\lambda}(\ell_H)_\lambda\delta_{\zeta,z} \\ &= \sum_\mu\sum_{\kappa,\lambda}\frac{(\ell_{\Lambda_\mu})_\kappa}{d_\kappa}(\mathbb{1}-Q)^{-1}_{\kappa,\lambda}(\ell_H)_\lambda\delta_{\zeta,z}\tilde{\delta}_{\mu,z} \\ &= \sum_\mu(\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}\ell_H)\tilde{\delta}_{\mu,z} \\ &= \sum_\mu(\ell_{\Lambda_\mu},\ell_H)\tilde{\delta}_{\mu,z} + \sum_\mu(\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}Q\ell_H)\tilde{\delta}_{\mu,z}\end{aligned} \tag{C.61}$$

where we used the identity $(\mathbb{1}-Q)^{-1} = \mathbb{1} - (\mathbb{1}-Q)^{-1}Q$. Expanding $(\mathbb{1}-Q)^{-1} = \sum_{l=0}^\infty Q^l$, we get $(\mathbb{1}-Q)^{-1}Q = \sum_{l=1}^\infty Q^l$, which is entrywise non-negative. This implies that $(\ell_{\Lambda_\mu},(\mathbb{1}-Q)^{-1}Q\ell_H) \geq 0\ \forall\mu$, proving the Proposition. □

As a second technical point, we show how the general scaling for the variance obtained in Theorem 6 in Ref. [5], can also be recovered in this formalism, by considering absorption to the trivial subspace only.

Proposition C.5.2. If the noise channel in Eq. (C.57), is further simplified to $\mathcal{N} = \bar{\mathcal{N}}^{\otimes n}$, where $\bar{\mathcal{N}}$ is a single-qubit, non-unital channel, then the trivial subspace $\mathcal{B}_0$ is the only norm-preserving, strictly connected component, its corresponding absorption coefficient scales as

$$\frac{(A\ell_H)_0}{d} \in \sum_i h_i^2 e^{\Theta(|P_i|)} \tag{C.62}$$

where $H = \sum_i h_i P_i$ denotes the Pauli decomposition of H and $|P_i|$ denotes the number of non-trivial terms in P_i.

Proof. First, we show that in this case, all subspaces $\mathcal{B}_\kappa$ pertain to strictly contractive components, except the trivial one. In particular, if $\sum_\kappa Q_{\kappa,\lambda} < 1, \forall \lambda$, this is the case by the Subinvariance theorem (Theorem C.1.2).

$$\begin{aligned}\sum_\kappa Q_{\kappa,\lambda} &= \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\{P_i W^\dagger \mathcal{N}^\dagger(P_j) W\}^2 \tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda}\sum_{i,j}\mathrm{Tr}\{B_i \mathcal{N}^\dagger(P_j)\}^2 \tilde{\delta}_{j,\lambda} \\ &= \frac{1}{d_\lambda}\sum_j \mathrm{Tr}\{\mathcal{N}^\dagger(P_j)^2\}\tilde{\delta}_{j,\lambda} = \frac{1}{d_\lambda}\sum_j\prod_m \mathrm{Tr}\{\mathcal{N}_m^\dagger(P_{j_m})^2\}\tilde{\delta}_{j,\lambda} \\ &= \prod_m \frac{1}{d_m}\sum_{j_m=1}^{d_m}\mathrm{Tr}\{\mathcal{N}_m^\dagger(P_{j_m})^2\} = \prod_m \frac{1}{d_m}\sum_{j_m=1}^{d_m}\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2\end{aligned} \tag{C.63}$$

where the product is carried over all partitions corresponding to the local 2-designs. By Lemma 2.2.1, $\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2 \leq \lambda_{j_m}^2 + t_{j_m}^2 \leq 1$. We claim that, when $\lambda_m = 1$, then at least one of such terms in the inner sum in Eq. (C.63), is strictly less than one. Indeed, if the equality holds for all j_m, since we are free to choose the local basis P_{j_m}, this would imply that the channel is norm-preserving, i.e. unitary, which is a contradiction since $\mathcal{N}_m$ are assumed to be non-unitary channels. In turn, this implies that the average is also strictly less than one. Contrarily, when $\lambda_m = 0$, $\|\mathcal{N}_m^\dagger(P_{0_m})\|_2^2 = 1$ by trace preservation.

Let c_m denote the average

$$c_m = \frac{1}{d_m}\sum_{j_m=1}^{d_m}\|\mathcal{N}_m^\dagger(P_{j_m})\|_2^2 < 1 \tag{C.64}$$

and let $c = \max_{m|\lambda_m=1} c_m$. Then

$$\sum_\kappa Q_{\kappa,\lambda} < \prod_m c = c^{|\lambda|} \tag{C.65}$$

where $|\lambda|$ denotes the number of non-zero entries of λ which proves the first part.

For the second part, recall that, by definiton of locality vector, if $H = \sum_i h_i P_i$, then $\ell_H = \sum_i h_i^2 \ell_{P_i}$, so the absorption coefficients can be computed as

$$\frac{(A\ell_H)_0}{d} = \sum_i h_i^2 \frac{(A\ell_{P_i})_0}{d} \tag{C.66}$$

Eq. (C.57) already gives us a lower bound:

$$\frac{(A\ell_{P_i})_0}{d} \geq (\ell_\tau, \ell_{P_i}) = \frac{(\ell_\tau)_\kappa}{d_k} \in \Omega\left(\frac{1}{d_\kappa}\right) = e^{\Omega(|\kappa|)} \tag{C.67}$$

For the upper bound, we estimate the scaling of $(\ell_\tau, (\mathbb{1}-Q)^{-1}Q\ell_{P_i})$. Let $\tilde{P}_i$ be an operator such that $\ell_{\tilde{P}_i} = Q\ell_{P_i}$. Then

$$(\ell_\tau, (\mathbb{1}-Q)^{-1}Q\ell_{P_i}) \leq (\ell_\tau, (\mathbb{1}-Q)^{-1}\ell_{\tilde{P}_i}) \leq \mathbb{V}^\infty_{\tau,\tilde{P}_i} \leq \|\tau\|_2^2 \|\tilde{P}_i\|_2^2 \tag{C.68}$$

by construction, since single qubit channels are contractive under the 2-norm [86]. By construction,we have

$$\|\tilde{P}_i\|_2^2 = \sum_\kappa (Q\ell_{P_i})_\kappa = \sum_{\kappa,\lambda} Q_{\kappa,\lambda}(\ell_{P_i})_\lambda \leq \|P_i\|_2^2 c^{|P_i|} \tag{C.69}$$

since $(\ell_{P_i})_\lambda = \tilde{\delta}_{i,\lambda}$, which shows shows the upper bound completing the proof.

□

As shown in Appendix C.5, this is a generalization of Theorem 6 in Ref. [5].

C.6 Additional material on non-unital noise and entanglment

C.6.1 Calculation of the absorption term

Explicit computation of the absorption matrix $A = R(\mathbb{1}-Q)^{-1}$ is a non-trivial task, which should be tackled on a case-by-case basis. Indeed, in the non-unital case, this term describes the complex phenomenon arising from the interaction of two competing effects, which drive the system towards different states. Analytical summation of A is however can be feasible and still give insight on the interaction of the two. Indeed, using a simplified model, we can perform this calculation and still be able to appreciate the different effects that rapidly entangling and slowly entangling circuits have on a fixed, non-unital noise as a function of its strength. In particular, here we prove Lemma 6.6.1, which we repeat for convenience.

Lemma C.6.1. Let $\mathcal{E}_c$ be a quantum channel of type Eq. (6.45), with $0 < p \leq 1$ and let T be the transition matrix of $\mathcal{E}^\dagger$. Then

$$\mathbb{V}^\infty_{\rho,H} = p^2(\ell_{\tilde{\rho}}, (\mathbb{1}-(1-p)^2T)^{-1}\ell_H). \tag{C.70}$$

In particular, if T is a projection, then

$$\mathbb{V}^\infty_{\rho,H} = \left(\frac{p}{2-p} - p^2\right)(\ell_{\tilde{\rho}}, T\ell_H) + p^2(\ell_{\tilde{\rho}}, \ell_H). \tag{C.71}$$

Proof. The first result follows directly from Theorem C.4.1, in particular Corollary C.4.3, by computation of A_0. In particular, we can explicitly compute R_0 element wise as

$$\begin{aligned}(R_0)_{0,\kappa} &= \frac{1}{d_\kappa}\sum_j \mathrm{Tr}\left\{\frac{\mathbb{1}}{\sqrt{d}}\mathcal{E}_c^\dagger(P_j)\right\}^2 \tilde{\delta}_{j,\kappa} = \frac{1}{d_\kappa}\sum_j \mathrm{Tr}\left\{\mathcal{E}_c\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right\}^2 \tilde{\delta}_{j,\kappa} \\ &= \frac{1}{d_\kappa}\sum_j \left((1-p)\,\mathrm{Tr}\left\{\Phi\left(\frac{\mathbb{1}}{\sqrt{d}}\right)P_j\right\} + p\sqrt{d}\,\mathrm{Tr}\{\tilde{\rho}P_j\}\right)^2 \tilde{\delta}_{j,\kappa} \\ &= dp^2\frac{1}{d_\kappa}\sum_j \mathrm{Tr}\{\tilde{\rho}P_j\}^2\tilde{\delta}_{j,\kappa} = dp^2(\ell_{\tilde{\rho}})_\kappa\end{aligned} \tag{C.72}$$

by unitality of Φ. By an analogous calculation, it can be shown that $Q = (1-p)^2 T$, and by trace preservation $(P_0)_{0,0} = 1$. With these elements, we can compute $A_0 = P_0 R_0 (1-Q)^{-1}$, and we get

$$(A_0)_{0,\lambda} = dp^2 (\ell_{\tilde{\rho}})_\kappa (1-(1-p)^2 T)^{-1}_{\kappa,\lambda}. \tag{C.73}$$

In particular, since for any initial state ρ, we have $(\ell_\rho)_0 = 1/d$ by normalization, we get the final result

$$(\ell_\rho, A_0 \ell_H) = p^2 (\ell_{\tilde{\rho}}, (\mathbb{1}-(1-p)^2 T)^{-1} \ell_H). \tag{C.74}$$

Using this, we can explicitly compute the right-hand side in the simplified setting where T is a projection. In that case in particular, we have that

$$\begin{aligned} (\mathbb{1}-(1-p)^2 T)^{-1} &= (\mathbb{1}-T) + (1-(1-p)^2)^{-1} T \\ &= (\mathbb{1}-T) + \frac{1}{p(2-p)} T = \mathbb{1} + \left(\frac{1}{p(2-p)} - 1 \right) T \end{aligned} \tag{C.75}$$

which gives the result. □

C.6.2 Choice of the system for the numerical example

In this Section we give an explicit construction of the noise channels used in the numerical example shown in the main text, as well as the choice and normalization procedure of the observable H used. In particular, the analysis focuses on the case of unitary entangling map Φ whose transfer matrix is well approximated by a projector, and a highly entangled, pure fixed point, namely the GHZ state $\tilde{\rho} = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})(\langle 0|^{\otimes n} + \langle 1|^{\otimes n})/2$. Thanks to Lemma 6.6.1, we know that the main contribution to $\mathbb{V}^{\infty}_{\rho,H}$ in this setting comes from the term $(\ell_{\tilde{\rho}}, \ell_H)$, so to numerically assess the results, it is useful to choose the observable H in order to normalize this factor. In particular, given the qubit structure, we can express the $\tilde{\rho}$ in terms of the *normalized* Pauli basis $\{\mathbb{1}, X, Y, Z\}^{\otimes n}$, which allows to easily compute the locality vectors. Using the spectral decomposition of the Pauli matrices, it is easy to see that

$$\begin{aligned} |0\rangle\langle 0| &= \frac{\mathbb{1}+Z}{\sqrt{2}}, \quad |0\rangle\langle 1| = \frac{X+iY}{\sqrt{2}}, \\ |1\rangle\langle 1| &= \frac{\mathbb{1}-Z}{\sqrt{2}}, \quad |1\rangle\langle 0| = \frac{X-iY}{\sqrt{2}}. \end{aligned} \tag{C.76}$$

Using this decomposition, we can find a formula for $\tilde{\rho}$ exploiting a generalization of the binomial theorem.

Theorem C.6.1. Let $A, B \in \mathbb{M}_d(\mathbb{C})$ be square matrices, and let S be the set of all permutations of n elements. Then

$$(A+\omega B)^{\otimes n} = \sum_{j=0}^{n} \omega^j \sum_{\sigma \in S} \sigma(A^{\otimes(n-j)} \otimes B^{\otimes j}) \tag{C.77}$$

where $\omega \in \mathbb{C}$ and the permutation σ is applied to the qubit ordering.

In particular, the following corollary will be the most useful in performing the computations

Corollary C.6.1. Let $A, B \in \mathbb{M}_d(\mathbb{C})$ be square matrices, and let S be the set of all permutations of n elements. Then

$$\frac{(A+\omega B)^{\otimes n} + (A-\omega B)^{\otimes n}}{2} = \sum_{j=0}^{\lfloor n/2 \rfloor} \omega^{2j} \sum_{\sigma \in S} \sigma(A^{\otimes(n-2j)} \otimes B^{\otimes 2j}) \tag{C.78}$$

where $\omega \in \mathbb{C}$ and the permutation σ is applied to the qubit ordering.

Proof. This corollary follow directly from Theorem C.6.1, and noticing that all even-indexed terms in both $(A+\omega B)^{\otimes n}$ and $(A-\omega B)^{\otimes n}$ are equal, while odd-numbered terms have opposite sign and therefore cancel out. □

Applying the Corollary to the appropriate pairs of projectors, we can get the final expression

$$\begin{aligned}\tilde{\rho} = \frac{1}{2^{n/2}} \sum_{j=0}^{\lfloor n/2 \rfloor} \sum_{\sigma\in S} &\sigma(\mathbb{1}^{\otimes(n-2j)} \otimes Z^{\otimes 2j}) \\ &+(-1)^j \sigma(X^{\otimes(n-2j)} \otimes Y^{\otimes 2j})\end{aligned} \tag{C.79}$$

As it is clear from Eq. (C.79), the fixed point of the channel has a non-vanishing component only on Pauli strings that are either non-trivial on all qubits, or non-trivial in *only* in an *even* number of qubits. Then it follows that the simplest observable involving all qubits and with non-vanishing variance is of form $H = h\sum_{k=1}^{n} 2^{n/2} Z_k \otimes Z_{k+1}$, where the factor $2^{n/2}$ accounts for normalization of Z, and cancels out with the corresponding factor in Eq. (C.79) in the calculations of $(\ell_{\tilde{\rho}}, \ell_H)$. Finally, since each term in the sum is orthogonal, it gives an independent contribution of $1/(d^2-1)^2 = 1/9$, consequently by choosing $h = 9/n$ we have that $(\ell_{\tilde{\rho}}, \ell_H) = 1$ is normalized.

C.7 Variance lower bounds

In this Section we prove the general variance lower bounds outlined in Section 6.2.

Proof of the general lower bound

Here we employ Proposition 6.2.1 to prove a general lower-bound on slowly entangling circuits. In particular, such result is based on the approximation $T_l \approx \mathbb{1}$, which holds either for shallow circuits, i.e. $L \in O(\log n)$, or deeper circuits, but with weakly entangling intermediate channels. The discussion is based on the following result.

Theorem C.7.1. Let $\rho, H \in \mathcal{B}$ and consider a sequence of quantum channels $\{\mathcal{E}_l\}_{l=1}^{L}$, and let $\{T_l\}_{l=1}^{L}$ be the respective LTMs. Finally let $K \subset \{0,1\}^M$ denote a subset of indices, and by $\alpha_l = \min_{\kappa\in K}(T_l)_{\kappa,\kappa}$. Then

$$\mathbb{V}_{\rho,H}^{L} \geq \alpha^L (\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{C.80}$$

where $\mathcal{K}(\cdot) = \sum_{\kappa\in K}\sum_j \mathrm{Tr}\{P_j \cdot\} P_j \tilde{\delta}_{j,\kappa}$ is a projector onto the space spanned by K and $\alpha = \left(\prod_{l=0}^{L}\alpha_l\right)^{1/L}$ is the geometric mean of α_l.

Proof. Consider a single circuit layer. Then, we can write

$$\begin{aligned}(\ell_\rho, T_l \ell_H) &= \sum_{\kappa,\lambda} \frac{(\ell_\rho)_\kappa T_{\kappa,\lambda} (\ell_H)_\lambda}{d_\kappa} \\ &\geq \sum_{\kappa\in K} \frac{(\ell_\rho)_\kappa T_{\kappa,\kappa} (\ell_H)_\kappa}{d_\kappa} \\ &\geq \alpha_l \sum_{\kappa\in K} \frac{(\ell_\rho)_\kappa (\ell_H)_\kappa}{d_\kappa} = \alpha_l (\ell_\rho, \ell_{\mathcal{K}(H)})\end{aligned} \tag{C.81}$$

where the inequality holds since all terms in the sum are non-negative by construction. The claim follows from repeated application of the latter. □

Despite its simplicity, Theorem C.7.1 can be used to deduce general bounds on weakly entangling circuits, which are the foundation of small angle initialization strategies. In particular, we get the following Corollary.

Corollary C.7.1. Let $\mathcal{H} = \bigotimes_{m=1}^{M} \mathcal{H}_m$, $d_m \in \Theta(1)$. If either of the conditions

(a) $\alpha > 0$, $\alpha \in \Omega(1)$ and $L \in O(\log n)$,

(b) $\alpha = 1 - f(n, L)$, $f \in O(\log n/L)$ and $L \in \Omega(\log^{1+\epsilon} n)$ for some arbitrary $\epsilon > 0$

is satisfied, then

$$\mathbb{V}^{L}_{\rho,H} \geq F(n)(\ell_\rho, \ell_{\mathcal{K}(H)}), \tag{C.82}$$

where $F(n) \in \Omega(1/\text{poly}(n))$.

Proof. Exploiting Theorem C.7.1, it suffices to show that $F(n) = \alpha^L \in \Omega(1/\text{poly}(n))$. In the first case, this follows directly from the shallow nature of the circuit, and in particular $F(n) \in \alpha^{O(\log(n))} = \Omega(1/n^{-\log(\alpha)}) \subset \Omega(1/\text{poly}(n))$. For the second case instead, it is useful to consider $\log(F(n))$:

$$\begin{aligned} -\log(F(n)) &< -L\log\left(1 - C\frac{\log(n)}{L}\right) \\ &= C\log(n)\left(1 + C\frac{\log(n)}{2L} + O\left(\frac{\log^2(n)}{L^2}\right)\right) \in O(\log(n)) \end{aligned} \tag{C.83}$$

which in turn implies $F(n) \in e^{-O(\log(n))} = \Omega(1/n^C) \subset \Omega(1/\text{poly}(n))$. □

Small-angle initializations lower bounds

In order to prove the general lower bounds on small angle initializations provided in the main text, it is useful to start from the following Proposition, as it is the fundamental building block in this type of proofs.

Proposition C.7.1. Let $\{E_\phi\}_\phi$ be an ensemble such that $\mathcal{E}(\rho) = \mathbb{E}_\phi\{E_\phi \rho E_\phi^\dagger\}$ is a quantum channel. Further, denote by T_ϕ the transfer matrix associated to each $\mathcal{E}_\phi^\dagger(\cdot) = E_\phi^\dagger \cdot E_\phi$, and by T the transfer matrix of $\mathcal{E}^\dagger$. Then we have

$$\mathbb{E}_\phi\{T_\phi\} \geq T, \tag{C.84}$$

with equality holding if and only if $\mathcal{E}$ is unitary.

Proof. Let A, B be arbitrary bounded operators, and consider

$$\begin{aligned} \text{Tr}\{\mathcal{E}(A)B\}^2 &= \text{Tr}\left\{A\mathcal{E}^\dagger(B)\right\}^2 \\ &= \mathbb{E}_\phi\left\{\text{Tr}\left\{A\mathcal{E}_\phi^\dagger(B)\right\}\right\}^2 \\ &\leq \mathbb{E}_\phi\left\{\text{Tr}\left\{A\mathcal{E}_\phi^\dagger(B)\right\}^2\right\} \quad \forall A, B \end{aligned} \tag{C.85}$$

which follows from the observation that $f(\phi) = \text{Tr}\left\{A\mathcal{E}_\phi^\dagger(B)\right\} \in \mathbb{R}$, and so $\mathbb{V}_\phi\{f\} \geq 0$. Applying this to the entries of T and T_ϕ gives the general inequality. Finally, the equality follows from $\mathbb{V}_\phi\{f\} = 0$, which is means that $f(\phi) = \text{Tr}\left\{AK^\dagger BK\right\}$ is a constant, where $K = E_\phi \ \forall \phi$. Hence, since $\mathcal{E}(\cdot) = K \cdot K^\dagger$ is CPTP, it must also be unitary. □

In order to translate this rather abstract formulation into a practical recipe, we need to identify the conditions that allow to treat the contribution of given an ensemble $\{\mathcal{E}_\phi\}$ of parameterized intermediate channels to the variance in terms of the mean LTM $\mathbb{E}_\phi\{T_\phi\}$. In particular, it is easily verified that, if ϕ is sampled independently of the other parameters, then

$$\begin{aligned}\mathbb{V}^L_{\rho,H} &= \mathbb{E}_\phi \mathbb{E}_{\boldsymbol{\theta}} \left\{ \mathrm{Tr}\{\Phi_{\boldsymbol{\theta},\phi}(\rho)H\}^2 \right\} \\ &= \mathbb{E}_\phi \left\{ (\ell_\rho, T_\phi \ell_H) \right\} = (\ell_\rho, \mathbb{E}_\phi\{T_\phi\}\ell_H).\end{aligned} \tag{C.86}$$

Combining this observation with Corollary C.7.1, we can get the QResNet lower bound of Proposition 6.7.1. First, a technical Lemma.

Lemma C.7.1. Let $A \in \mathcal{B}$, and P_i be a normalized Pauli string, i.e. $\mathrm{Tr}\{P_i^2\} = 1$. Then $\mathrm{Tr}\{P_i A^\dagger P_i A\} \in \mathbb{R}$ and

$$|\mathrm{Tr}\{P_i A^\dagger P_i A\}| \leq \mathrm{Tr}\{A^\dagger A P_i^2\} = \frac{\|A\|_2^2}{d}. \tag{C.87}$$

Proof. Consider the decomposition of A into the normalized Pauli basis $A = \sum_j a_j P_j$. The coefficients $\{a_j\}_j$ are related to the 2-norm of A by the following relation:

$$\begin{aligned}\|A\|_2^2 = \mathrm{Tr}\{A^\dagger A\} &= \sum_{j,k} a_j^* a_k \mathrm{Tr}\{P_j P_k\} \\ &= \sum_{j,k} a_j^* a_k \delta_{j,k} = \sum_j |a_j|^2\end{aligned} \tag{C.88}$$

Using this decomposition, we get

$$\begin{aligned}\mathrm{Tr}\{P_i A^\dagger P_i A\} &= \sum_{j,k} a_j^* a_k \mathrm{Tr}\{P_i P_j P_i P_k\} \\ &= \sum_{j,k} a_j^* a_k s_{i,j} \frac{\mathrm{Tr}\{P_j P_k\}}{d} \\ &= \sum_{j,k} a_j^* a_k s_{i,j} \frac{\delta_{j,k}}{d} = \frac{1}{d}\sum_j |a_j|^2 s_{i,j}\end{aligned} \tag{C.89}$$

Where $s_{i,j} = 1$ if P_i and P_j commute, and $s_{i,j} = -1$ otherwise. Since $|s_{i,j}| = 1$, we get

$$|\mathrm{Tr}\{P_i A^\dagger P_i A\}| = \left| \frac{1}{d}\sum_j |a_j|^2 s_{i,j} \right| \leq \frac{1}{d}\sum_j |a_j|^2 |s_{i,j}| = \frac{\|A\|_2^2}{d} \tag{C.90}$$

The equality comes directly from the fact that $P_i^2 = \mathbb{1}/d$ by the normalization constraint. □

Proposition C.7.2 (QResNet)**.** Let $\mathcal{E}_\phi(\cdot) = e^{i\phi G} \cdot e^{-i\phi G}$ be a unitary entangling gate, and let μ, σ^2 be the mean and variance of the initialization distribution $\mathcal{P}$ of ϕ. Then, if $\mu = 0$, $\sigma^2 \in O(\log n / \|G\|_2^2 L)$, and $L \in \Omega(\log^{1+\epsilon} n)$

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H) \tag{C.91}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

Proof. Let T be the locality transfer matrix of $\mathcal{E} = \mathbb{E}_\phi\{\mathcal{E}_\phi\}$, and consider the diagonal element $T_{\kappa,\kappa}$. Then, by definiton, we have

$$\begin{aligned} T_{\kappa,\kappa} &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\big\{\mathbb{E}_\phi\{P_i\mathcal{E}_\phi^\dagger(P_j)\}\big\}^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\kappa} \\ &\geq \frac{1}{d_\kappa}\sum_{i}\mathrm{Tr}\big\{\mathbb{E}_\phi\{P_i\mathcal{E}_\phi^\dagger(P_i)\}\big\}^2\tilde{\delta}_{i,\kappa} \\ &= \frac{1}{d_\kappa}\sum_{i}\mathrm{Tr}\big\{\mathbb{E}_\phi\{P_i e^{-i\phi G}P_i e^{i\phi G}\}\big\}^2\tilde{\delta}_{i,\kappa} \end{aligned} \tag{C.92}$$

Since $\sigma^2 \to 0$ as $n \to \infty$, to find the asymptotic behaviour of the diagonal elements of T we can expand $e^{i\phi G}$ around μ, and obtain $e^{i\phi G} = \mathbb{1} + i\phi G - \phi^2 G^2/2 + O(\phi^3\|G\|_2^3)$. Substituting this into Eq. (C.92), we get

$$\begin{aligned} T_{\kappa,\kappa} &\approx \frac{1}{d_\kappa}\sum_{i}\mathrm{Tr}\Big[\mathbb{E}_\phi\{P_i(\mathbb{1} + i\phi G - \phi^2G^2/2)P_i(\mathbb{1}+ \\ &\quad + i\phi G - \phi^2G^2/2)\}\Big]^2\tilde{\delta}_{i,\kappa} \\ &= \frac{1}{d_\kappa}\sum_{i}\big(1 - \mathbb{E}_\phi\{\phi^2\}\left(\mathrm{Tr}\{G^2P_i^2\} - \mathrm{Tr}\{P_iGP_iG\}\right)\big)^2\tilde{\delta}_{i,\kappa} \\ &\geq 1 - 4\|G\|_2^2\sigma^2 \quad \forall\kappa \in \{0,1\}^M \end{aligned} \tag{C.93}$$

Exploiting Corollary C.7.1, we get the proposition by showing $T_{\kappa,\kappa} \geq 1 - f(n,L)$, where $f(n,L) \in O(\log n/L)$. In particular this follows directly from the scaling of σ^2. The same proof ensures absence of concentration on a unitary QResNet, provided that $\mathcal{P}$ is chosen as an initialization probability by Proposition C.7.1. □

Going beyond unitary ansätze, a similar technique—combined with the stochastic representation of solutions to the Lindblad equation [2, 23]—can be employed to construct a BP-free QResNet from any noise model defined by a set of jump operators L_k.

Proposition C.7.3 (General QResNets from a quantum channel)**.** Let $\mathcal{E}(\rho) = e^{t\mathcal{L}}(\rho)$ be a quantum channel generated by the Lindbladian

$$\mathcal{L}(\rho) = \sum_k \gamma_k\left(L_k\rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k,\rho\}\right) \tag{C.94}$$

with $\|L_k\|_2^2/d = 1$. Then, when $L \in \Omega(\log^{1+\epsilon} n)$, the ensemble generated by $\mathcal{E}_\phi = E_\phi\cdot E_\phi^\dagger$, $E_\phi = \prod_k e^{i\phi_k L_k + \sigma_k^2(L_k^2 - L_k^\dagger L_k)/2}$, and ϕ_k normally distributed with $\mu = 0$, $\sigma_k^2 \in O\left(\frac{\gamma_k}{\sum_{k'}\gamma_{k'}}\frac{\log n}{L}\right)\forall k$ satisfies

$$\mathbb{V}_{\rho,H}^L \geq F(n)(\ell_\rho,\ell_H) \tag{C.95}$$

where $F(n) \in \Omega(1/\mathrm{poly}(n))$.

Proof. We prove this statement, we compute the scaling of α in Corollary C.7.1. Again, consider the diagonal elements $T_{\kappa,\kappa}$.

$$\begin{aligned} T_{\kappa,\kappa} &= \frac{1}{d_\kappa}\sum_{i,j}\mathrm{Tr}\big\{P_ie^{t\mathcal{L}}(P_j)\big\}^2\tilde{\delta}_{i,\kappa}\tilde{\delta}_{j,\kappa} \\ &\geq \frac{1}{d_\kappa}\sum_{i}\mathrm{Tr}\big\{P_ie^{t\mathcal{L}}(P_i)\big\}^2\tilde{\delta}_{i,\kappa} \end{aligned} \tag{C.96}$$

Expanding for $t \ll 1$, we get $e^{t\mathcal{L}}(P_i) \approx P_i + t\mathcal{L}(P_i)$, which can be used to lower bound $T_{\kappa,\kappa}$.

$$\begin{aligned}
&\operatorname{Tr}\{P_i e^{t\mathcal{L}}(P_i)\} \approx \operatorname{Tr}\{P_i^2\} + t\operatorname{Tr}\{P_i\mathcal{L}(P_i)\} \\
&= 1 + \sum_k t\gamma_k \left(\operatorname{Tr}\{P_i L_k P_i L_k^\dagger\} - \frac{1}{2}\operatorname{Tr}\{P_i\{L_k^\dagger L_k, P_i\}\}\right) \\
&= 1 + \sum_k t\gamma_k \left(\operatorname{Tr}\{P_i L_k P_i L_k^\dagger\} - \operatorname{Tr}\{L_k^\dagger L_k P_i^2\}\right) \\
&\geq 1 - \sum_k t\gamma_k \left(|\operatorname{Tr}\{P_i L_k P_i L_k^\dagger\}| + \operatorname{Tr}\{L_k^\dagger L_k P_i^2\}\right) \\
&\geq 1 - 2\sum_k t\gamma_k \|L_k\|_2^2/d = 1 - 2t\sum_k \gamma_k.
\end{aligned} \tag{C.97}$$

Putting it back into the definition of $T_{\kappa,\kappa}$ we get

$$\begin{aligned}
T_{\kappa,\kappa} &\geq \frac{1}{d_\kappa}\sum_i \left(1 - 2t\sum_k \gamma_k\right)^2 \tilde{\delta}_{i,\kappa} \\
&\geq 1 - 4t\sum_k \gamma_k
\end{aligned} \tag{C.98}$$

This directly implies that $\alpha \geq 1 - 4t\sum_k \gamma_k$ for small t, which satisfies Corollary C.7.1, if $t \in O\left(\frac{1}{\sum_{k'}\gamma_{k'}}\frac{\log n}{L}\right)$. Finally, the claim follows by Corollary C.7.1 unraveling the master equation in the Stratonovich formalism. In particular we can describe the evolution according to $\mathcal{L}$ by the (stochastic) differential equation:

$$\frac{d|\psi\rangle}{ds} = \sum_k \left[i\sqrt{\gamma_k}L_k\xi_k(s) + \frac{\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]|\psi\rangle \tag{C.99}$$

where $\xi_k(s)$ represents a *white noise*. Eq. (C.99) is formally solved by

$$|\psi\rangle_t = \mathcal{T}\left\{e^{\int_0^t \sum_k \left[i\sqrt{\gamma_k}L_k\xi_k(s) + \frac{\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]ds}\right\}|\psi\rangle_0 \tag{C.100}$$

where $\mathcal{T}\{\cdot\}$ denotes the time ordering operator. Since $t \to 0$ as n diverges, we can again expand this solution to first order in t to get the asymptotic scaling. In particular, this allows us to discard the time ordering, and, exploiting the defining relation of the white noise in terms of the Weiner process $\int_0^t \xi_k(s)ds = x_k$, where $x_k \sim \mathcal{N}(0,t)\,\forall k$ is normally distributed, we get, for each x, $|\psi\rangle_{x,t} = \bar{E}_x|\psi\rangle_0$, where

$$\begin{aligned}
\bar{E}_x &= e^{\sum_k \left[i\sqrt{\gamma_k}L_k x_k + \frac{t\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)\right]} \\
&\approx \prod_k e^{i\sqrt{\gamma_k}L_k x_k + \frac{t\gamma_k}{2}\left(L_k^2 - L_k^\dagger L_k\right)}.
\end{aligned} \tag{C.101}$$

This gives an approximate solution $e^{t\mathcal{L}} \approx \mathbb{E}_x\left[\bar{E}_x\rho\bar{E}_x^\dagger\right]$ Finally, we define $\phi_k = x_k\sqrt{\gamma_k}$, which is also normally distributed but with variance $\sigma_k^2 = t\gamma_k \in O\left(\frac{\gamma_k}{\sum_{k'}\gamma_{k'}}\frac{\log n}{L}\right)$. The proof is completed by making the substitution $E_\phi = \bar{E}_{x_k\sqrt{\gamma_k}}$. □

Bibliography

[1] In: *Quantum Information and Computation* 12 (Nov. 2012). ISSN: 1533-7146. DOI: 10.26421/qic12.11-12. URL: http://dx.doi.org/10.26421/QIC12.11-12.

[2] Adler, Stephen L and Bassi, Angelo. "Collapse models with non-white noises". In: *Journal of Physics A: Mathematical and Theoretical* 40.50 (2007), 15083.

[3] Tameem Albash and Daniel A. Lidar. "Adiabatic quantum computation". In: *Rev. Mod. Phys.* 90 (1 Jan. 2018), p. 015002. DOI: 10.1103/RevModPhys.90.015002. URL: https://link.aps.org/doi/10.1103/RevModPhys.90.015002.

[4] Baptiste Anselme Martin, Pascal Simon, and Marko J. Rančić. "Simulating strongly interacting Hubbard chains with the variational Hamiltonian ansatz on a quantum computer". In: *Physical Review Research* 4.2 (June 2022). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.4.023190. URL: http://dx.doi.org/10.1103/PhysRevResearch.4.023190.

[5] Antonio Anna Mele and Armando Angrisani and Soumik Ghosh and Sumeet Khatri and Jens Eisert and Daniel Stilck França and Yihui Quek. *Noise-induced shallow circuits and absence of barren plateaus*. 2024. arXiv: {2403.13927} (quant-ph). URL: {https://arxiv.org/abs/2403.13927}.

[6] Arab, Amir R. "Lecture notes on quantum entanglement: From stabilizer states to stabilizer channels". In: *Frontiers of Physics* 19.5 (2024), 51203.

[7] Nikita Astrakhantsev et al. "Phenomenological theory of variational quantum ground-state preparation". In: *Phys. Rev. Res.* 5 (3 Sept. 2023), p. 033225. DOI: 10.1103/PhysRevResearch.5.033225. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.5.033225.

[8] Michael Ben-Or, Daniel Gottesman, and Avinatan Hassidim. *Quantum Refrigerator*. 2013. arXiv: 1301.1995 [quant-ph]. URL: https://arxiv.org/abs/1301.1995.

[9] Benenti, Giuliano and Casati, Giulio and Rossini, Davide and Strini, Giuliano. *Principles of Quantum Computation and Information: A Comprehensive Textbook*. World Scientific, 2019.

[10] Biamonte, Jacob and Wittek, Peter and Pancotti, Nicola and Rebentrost, Patrick and Wiebe, Nathan and Lloyd, Seth. "Quantum machine learning". In: *Nature* 549.7671 (2017), 195–202.

[11] Briegel, Hans J and Browne, David E and Dür, Wolfgang and Raussendorf, Robert and Van den Nest, Maarten. "Measurement-based quantum computation". In: *Nature Physics* 5.1 (2009), 19–26.

[12] C. King and M. B. Ruskai. *Minimal Entropy of States Emerging from Noisy Quantum Channels*. 2000. arXiv: {quant-ph/9911079} (quant-ph). URL: {https://arxiv.org/abs/quant-ph/9911079}.

[13] Giuseppe Carleo and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks". In: *Science* 355.6325 (Feb. 2017), 602–606. ISSN: 1095-9203. DOI: 10.1126/science.aag2302. URL: http://dx.doi.org/10.1126/science.aag2302.

[14] Almudena Carrera Vazquez et al. "Well-conditioned multi-product formulas for hardware-friendly Hamiltonian simulation". In: *Quantum* 7 (July 2023), p. 1067. ISSN: 2521-327X. DOI: 10.22331/q-2023-07-25-1067. URL: http://dx.doi.org/10.22331/q-2023-07-25-1067.

[15] Cerezo, M. and Arrasmith, Andrew and Babbush, Ryan and Benjamin, Simon C. and Endo, Suguru and Fujii, Keisuke and McClean, Jarrod R. and Mitarai, Kosuke and Yuan, Xiao and Cincio, Lukasz and Coles, Patrick J. "Variational quantum algorithms". In: *Nature Reviews Physics* 3.9 (Aug. 2021), 625–644. ISSN: 2522-5820. DOI: {10.1038/s42254-021-00348-9}. URL: {http://dx.doi.org/10.1038/s42254-021-00348-9}.

[16] Cerezo, M. and Sone, Akira and Volkoff, Tyler and Cincio, Lukasz and Coles, Patrick J. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". In: *Nature Communications* 12.1 (Mar. 2021). ISSN: 2041-1723. DOI: {10.1038/s41467-021-21728-w}. URL: {http://dx.doi.org/10.1038/s41467-021-21728-w}.

[17] Cervero Martín, Enrique and Plekhanov, Kirill and Lubasch, Michael. "Barren plateaus in quantum tensor network optimization". In: *Quantum* 7 (Apr. 2023), 974. ISSN: 2521-327X. DOI: {10.22331/q-2023-04-13-974}. URL: {https://doi.org/10.22331/q-2023-04-13-974}.

[18] Chae-Yeun Park and Minhyeok Kang and Joonsuk Huh. *Hardware-efficient ansatz without barren plateaus in any depth*. 2024. arXiv: {2403.04844} (quant-ph). URL: {https://arxiv.org/abs/2403.04844}.

[19] Shantanav Chakraborty. *Implementing any Linear Combination of Unitaries on Intermediate-term Quantum Computers*. 2023. arXiv: 2302.13555 [quant-ph].

[20] Contreras-Tejada, Patricia and Palazuelos, Carlos and De Vicente, Julio I. "Resource theory of entanglement with a unique multipartite maximally entangled state". In: *Physical review letters* 122.12 (2019), 120503.

[21] Crognaletti, Giulio and Di Bartolomeo, Giovanni and Vischi, Michele and Loris Viteritti, Luciano. "Equivariant Variational Quantum Eigensolver to detect phase transitions through energy level crossings". In: *Quantum Science and Technology* 10.1 (Dec. 2024), 015048. ISSN: 2058-9565. DOI: {10.1088/2058-9565/ad9be3}. URL: {http://dx.doi.org/10.1088/2058-9565/ad9be3}.

[22] D'Hondt, Ellie and Panangaden, Prakash. "The computational power of the W and GHZ states". In: *arXiv preprint quant-ph/0412177* (2004).

[23] Di Bartolomeo, Giovanni and Vischi, Michele and Cesa, Francesco and Wixinger, Roman and Grossi, Michele and Donadi, Sandro and Bassi, Angelo. "Noisy gates for simulating quantum computers". In: *Physical Review Research* 5.4 (Dec. 2023). ISSN: 2643-1564. DOI: {10.1103/physrevresearch.5.043210}. URL: {http://dx.doi.org/10.1103/PhysRevResearch.5.043210}.

[24] W. Dür, G. Vidal, and J. I. Cirac. "Three qubits can be entangled in two inequivalent ways". In: *Phys. Rev. A* 62 (6 Nov. 2000), p. 062314. DOI: 10.1103/PhysRevA.62.062314. URL: https://link.aps.org/doi/10.1103/PhysRevA.62.062314.

[25] Dür, W. and Hein, M. and Cirac, J. I. and Briegel, H.-J. "Standard forms of noisy quantum operations via depolarization". In: *Phys. Rev. A* 72 (5 Nov. 2005), 052326. DOI: {10.1103/PhysRevA.72.052326}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.72.052326}.

[26] S. Eggert. "Numerical evidence for multiplicative logarithmic corrections from marginal operators". In: *Phys. Rev. B* 54 (14 Oct. 1996), R9612–R9615. DOI: 10.1103/PhysRevB.54.R9612. URL: https://link.aps.org/doi/10.1103/PhysRevB.54.R9612.

[27] Jens Eisert and John Preskill. *Mind the gaps: The fraught road to quantum advantage*. 2025. arXiv: 2510.19928 [quant-ph]. URL: https://arxiv.org/abs/2510.19928.

[28] Enrico Fontana and Nathan Fitzpatrick and David Muñ oz Ramo and Ross Duncan and Ivan Rungger. "Evaluating the noise resilience of variational quantum algorithms". In: *Physical Review A* 104.2 (Aug. 2021). DOI: {10.1103/physreva.104.022403}. URL: {https://doi.org/10.1103%2Fphysreva.104.022403}.

[29] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. *A Quantum Approximate Optimization Algorithm*. 2014. arXiv: 1411.4028 [quant-ph].

[30] Fefferman, Bill and Ghosh, Soumik and Gullans, Michael and Kuroiwa, Kohdai and Sharma, Kunal. "Effect of Nonunital Noise on Random-Circuit Sampling". In: *PRX Quantum* 5 (3 July 2024), 030317. DOI: {10.1103/PRXQuantum.5.030317}. URL: {https://link.aps.org/doi/10.1103/PRXQuantum.5.030317}.

[31] Francesco Ferrari and Federico Becca. "Gapless spin liquid and valence-bond solid in the J_1-J_2 Heisenberg model on the square lattice: Insights from singlet and triplet excitations". In: *Phys. Rev. B* 102 (1 July 2020), p. 014417. DOI: 10.1103/PhysRevB.102.014417. URL: https://link.aps.org/doi/10.1103/PhysRevB.102.014417.

[32] Verena Feulner and Michael J. Hartmann. "Variational quantum eigensolver ansatz for the J_1–J_2-model". In: *Phys. Rev. B* 106 (14 Oct. 2022), p. 144426. DOI: 10.1103/PhysRevB.106.144426. URL: https://link.aps.org/doi/10.1103/PhysRevB.106.144426.

[33] Fontana, Enrico and Herman, Dylan and Chakrabarti, Shouvanik and Kumar, Niraj and Yalovetzky, Romina and Heredge, Jamie and Sureshbabu, Shree Hari and Pistoia, Marco. "Characterizing barren plateaus in quantum ansätze with the adjoint representation". In: *Nature Communications* 15.1 (Aug. 2024). ISSN: 2041-1723. DOI: {10.1038/s41467-024-49910-w}. URL: {http://dx.doi.org/10.1038/s41467-024-49910-w}.

[34] Fabio Franchini. *An Introduction to Integrable Techniques for One-Dimensional Quantum Systems*. Springer International Publishing, 2017. ISBN: 9783319484877. DOI: 10.1007/978-3-319-48487-7. URL: http://dx.doi.org/10.1007/978-3-319-48487-7.

[35] Georgopoulos, Konstantinos and Emary, Clive and Zuliani, Paolo. "Modeling and simulating the noisy behavior of near-term quantum computers". In: *Phys. Rev. A* 104 (6 Dec. 2021), 062432. DOI: {10.1103/PhysRevA.104.062432}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.104.062432}.

[36] Tudor Giurgica-Tiron et al. "Digital zero noise extrapolation for quantum error mitigation". In: *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, Oct. 2020. DOI: 10.1109/qce49297.2020.00045. URL: http://dx.doi.org/10.1109/QCE49297.2020.00045.

[37] Shou-Shu Gong et al. "Plaquette Ordered Phase and Quantum Phase Diagram in the Spin-$\frac{1}{2}$ J_1–J_2 Square Heisenberg Model". In: *Phys. Rev. Lett.* 113 (2 July 2014), p. 027201. DOI: 10.1103/PhysRevLett.113.027201. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.027201.

[38] Gorini, Vittorio and Kossakowski, Andrzej and Sudarshan, Ennackal Chandy George. "Completely positive dynamical semigroups of N-level systems". In: *Journal of Mathematical Physics* 17.5 (1976), 821–825. DOI: {https://doi.org/10.1063/1.522979}.

[39] Hashim, Akel and Naik, Ravi K. and Morvan, Alexis and Ville, Jean-Loup and Mitchell, Bradley and Kreikebaum, John Mark and Davis, Marc and Smith, Ethan and Iancu, Costin and O'Brien, Kevin P. and Hincks, Ian and Wallman, Joel J. and Emerson, Joseph and Siddiqi, Irfan. "Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor". In: *Phys. Rev. X* 11 (4 Nov. 2021), 041039. DOI: {10.1103/PhysRevX.11.041039}. URL: {https://link.aps.org/doi/10.1103/PhysRevX.11.041039}.

[40] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2016, 770–778. DOI: {10.1109/CVPR.2016.90}.

[41] Hein, M. and Eisert, J. and Briegel, H. J. "Multiparty entanglement in graph states". In: *Phys. Rev. A* 69 (6 June 2004), 062311. DOI: {10.1103/PhysRevA.69.062311}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.69.062311}.

[42] Hein, Marc and Dür, Wolfgang and Eisert, Jens and Raussendorf, Robert and Van den Nest, Maarten and Briegel, H-J. "Entanglement in graph states and its applications". In: *Quantum computers, algorithms and chaos*. IOS Press, 2006, 115–218.

[43] Mohamed Hibat-Allah et al. "Recurrent neural network wave functions". In: *Phys. Rev. Res.* 2 (2 June 2020), p. 023358. DOI: 10.1103/PhysRevResearch.2.023358. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.2.023358.

[44] Wen Wei Ho and Timothy H. Hsieh. "Efficient variational simulation of non-trivial quantum states". In: *SciPost Physics* 6.3 (Mar. 2019). ISSN: 2542-4653. DOI: 10.21468/scipostphys.6.3.029. URL: http://dx.doi.org/10.21468/SciPostPhys.6.3.029.

[45] Holevo, A. S. and Werner, R. F. "Evaluating capacities of bosonic Gaussian channels". In: *Phys. Rev. A* 63 (3 Feb. 2001), 032312. DOI: {10.1103/PhysRevA.63.032312}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.63.032312}.

[46] Holmes, Zoë and Sharma, Kunal and Cerezo, M. and Coles, Patrick J. "Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus". In: *PRX Quantum* 3 (1 Jan. 2022), 010313. DOI: {10.1103/PRXQuantum.3.010313}. URL: {https://link.aps.org/doi/10.1103/PRXQuantum.3.010313}.

[47] He-Liang Huang et al. "Superconducting quantum computing: a review". In: *Science China Information Sciences* 63.8 (July 2020). ISSN: 1869-1919. DOI: 10.1007/s11432-020-2881-9. URL: http://dx.doi.org/10.1007/s11432-020-2881-9.

[48] Jamie Heredge and Maxwell West and Lloyd Hollenberg and Martin Sevior. *Non-Unitary Quantum Machine Learning*. 2024. arXiv: {2405.17388} (quant-ph). URL: {https://arxiv.org/abs/2405.17388}.

[49] J.V. Jos. *40 Years of Berezinskii-Kosterlitz-Thouless Theory*. EBL-Schweitzer. World Scientific, 2013. ISBN: 9789814417648. URL: https://books.google.it/books?id=roO6CgAAQBAJ.

[50] Kaining Zhang and Liu Liu and Min-Hsiu Hsieh and Dacheng Tao. "Escaping from the Barren Plateau via Gaussian Initializations in Deep Variational Quantum Circuits". In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh and Alekh Agarwal and Danielle Belgrave and Kyunghyun Cho. 2022. URL: {https://openreview.net/forum?id=jXgbJdQ2YIy}.

[51] Katabarwa, Amara and Gratsea, Katerina and Caesura, Athena and Johnson, Peter D. "Early Fault-Tolerant Quantum Computing". In: *PRX Quantum* 5.2 (June 2024). ISSN: 2691-3399. DOI: {10.1103/prxquantum.5.020101}. URL: {http://dx.doi.org/10.1103/PRXQuantum.5.020101}.

[52] Joris Kattemölle and Jasper van Wezel. "Variational quantum eigensolver for the Heisenberg antiferromagnet on the kagome lattice". In: *Phys. Rev. B* 106 (21 Dec. 2022), p. 214429. DOI: 10.1103/PhysRevB.106.214429. URL: https://link.aps.org/doi/10.1103/PhysRevB.106.214429.

[53] Kitaev, A Yu. "Fault-tolerant quantum computation by anyons". In: *Annals of physics* 303.1 (2003), 2–30. DOI: {https://doi.org/10.1016/S0003-4916(02)00018-0}.

[54] K Kraus. "General state changes in quantum theory". In: *Annals of Physics* 64.2 (1971), pp. 311–335. ISSN: 0003-4916. DOI: https://doi.org/10.1016/0003-4916(71)90108-4. URL: https://www.sciencedirect.com/science/article/pii/0003491671901084.

[55] Claudine Lacroix, Philippe Mendels, and Frederic Mila. *Introduction to Frustrated Magnetism: Materials, Experiments, Theory*. Springer, Jan. 2011. ISBN: 978-3-642-10588-3. DOI: 10.1007/978-3-642-10589-0.

[56] Cornelius Lanczos. "An iterative method for the solution of the eigenvalue problem of linear differential and integral operators". In: *J. Res. Nat. Bur. Standards* 45 (1950), pp. 225–280. DOI: 10.6028/jres.045.026.

[57] Larocca, Martin and Czarnik, Piotr and Sharma, Kunal and Muraleedharan, Gopikrishnan and Coles, Patrick J. and Cerezo, M. "Diagnosing Barren Plateaus with Tools from Quantum Optimal Control". In: *Quantum* 6 (Sept. 2022), 824. ISSN: 2521-327X. DOI: {10.22331/q-2022-09-29-824}. URL: {https://doi.org/10.22331/q-2022-09-29-824}.

[58] Ying Li and Simon C. Benjamin. "Efficient Variational Quantum Simulator Incorporating Active Error Minimization". In: *Phys. Rev. X* 7 (2 June 2017), p. 021050. DOI: 10.1103/PhysRevX.7.021050. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.021050.

[59] Lindblad, Goran. "On the generators of quantum dynamical semigroups". In: *Communications in Mathematical Physics* 48.2 (1976), 119–130. DOI: {https://doi.org/10.1007/BF01608499}.

[60] Chufan Lyu et al. "Symmetry enhanced variational quantum spin eigensolver". In: *Quantum* 7 (Jan. 2023), p. 899. ISSN: 2521-327X. DOI: 10.22331/q-2023-01-19-899. URL: http://dx.doi.org/10.22331/q-2023-01-19-899.

[61] M. Cerezo and Martin Larocca and Diego García-Martín and N. L. Diaz and Paolo Braccia and Enrico Fontana and Manuel S. Rudolph and Pablo Bermejo and Aroosa Ijaz and Supanut Thanasilp and Eric R. Anschuetz and Zoë Holmes. *Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing*. 2024. arXiv: {2312.09121} (quant-ph). URL: {https://arxiv.org/abs/2312.09121}.

[62] M Schumann and F K Wilhelm and A Ciani. "Emergence of noise-induced barren plateaus in arbitrary layered noise models". In: *Quantum Science and Technology* 9.4 (Aug. 2024), 045019. DOI: {10.1088/2058-9565/ad6285}. URL: {https://dx.doi.org/10.1088/2058-9565/ad6285}.

[63] Chanchal K Majumdar and Dipan K Ghosh. "On Next-Nearest-Neighbor Interaction in Linear Chain. II". In: *Journal of Mathematical Physics* 10.8 (1969), pp. 1399–1402. DOI: https://doi.org/10.1063/1.1664978.

[64] Daniel Manzano. "A short introduction to the Lindblad master equation". In: *Aip Advances* 10.2 (2020). DOI: https://doi.org/10.1063/1.5115323.

[65] Martin Larocca and Supanut Thanasilp and Samson Wang and Kunal Sharma and Jacob Biamonte and Patrick J. Coles and Lukasz Cincio and Jarrod R. McClean and Zoë Holmes and M. Cerezo. *A Review of Barren Plateaus in Variational Quantum Computing*. 2024. arXiv: {2405.00781} (quant-ph).

[66] Mary Beth Ruskai and Stanislaw Szarek and Elisabeth Werner. "An analysis of completely-positive trace-preserving maps on M2". In: *Linear Algebra and its Applications* 347.1 (2002), 159–187. ISSN: 0024-3795. DOI: {https://doi.org/10.1016/S0024-3795(01)00547-X}. URL: {https://www.sciencedirect.com/science/article/pii/S002437950100547X}.

[67] McClean, Jarrod R. and Boixo, Sergio and Smelyanskiy, Vadim N. and Babbush, Ryan and Neven, Hartmut. "Barren plateaus in quantum neural network training landscapes". In: *Nature Communications* 9.1 (Nov. 2018). ISSN: 2041-1723. DOI: {10.1038/s41467-018-07090-4}. URL: {http://dx.doi.org/10.1038/s41467-018-07090-4}.

[68] James McClung. "Constructions and applications of W-states". In: *Worcester Polytechnic Institute* (2020). DOI: https://digital.wpi.edu/downloads/mw22v793n?locale=en.

[69] Antonio A. Mele et al. "Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz". In: *Physical Review A* 106.6 (Dec. 2022). ISSN: 2469-9934. DOI: 10.1103/physreva.106.l060401. URL: http://dx.doi.org/10.1103/PhysRevA.106.L060401.

[70] Meyer, Carl D. *Matrix analysis and applied linear algebra*. USA: Society for Industrial and Applied Mathematics, 2000. ISBN: 0898714540.

[71] Michele Vischi and Giovanni Di Bartolomeo and Massimiliano Proietti and Seid Koudia and Filippo Cerocchi and Massimiliano Dispenza and Angelo Bassi. *Simulating photonic devices with noisy optical elements*. 2023. arXiv: {2311.10613} (quant-ph).

[72] Takahiro Mizusaki and Masatoshi Imada. "Quantum-number projection in the path-integral renormalization group method". In: *Phys. Rev. B* 69 (12 Mar. 2004), p. 125110. DOI: 10.1103/PhysRevB.69.125110. URL: https://link.aps.org/doi/10.1103/PhysRevB.69.125110.

[73] N. L. Diaz and Diego García-Martín and Sujay Kazi and Martin Larocca and M. Cerezo. *Showcasing a Barren Plateau Theory Beyond the Dynamical Lie Algebra*. 2023. arXiv: {2310.11505} (quant-ph). URL: {https://arxiv.org/abs/2310.11505}.

[74] Mahdi Naghiloo. *Introduction to Experimental Quantum Measurement with Superconducting Qubits*. 2019. arXiv: 1904.09291 [quant-ph].

[75] Masaaki Nakamura. "Mechanism of CDW-SDW Transition in One Dimension". In: *Journal of the Physical Society of Japan* 68.10 (Oct. 1999), 3123–3126. ISSN: 1347-4073. DOI: 10.1143/jpsj.68.3123. URL: http://dx.doi.org/10.1143/JPSJ.68.3123.

[76] Masaaki Nakamura. "Tricritical behavior in the extended Hubbard chains". In: *Phys. Rev. B* 61 (24 June 2000), pp. 16377–16392. DOI: 10.1103/PhysRevB.61.16377. URL: https://link.aps.org/doi/10.1103/PhysRevB.61.16377.

[77] Nielsen, Michael A and Chuang, Isaac L. *Quantum computing and quantum information*. Cambridge University Press, Cambridge, 2000. DOI: {https://doi.org/10.1017/CBO9780511976667}.

[78] K Nomura. "Correlation functions of the 2D sine-Gordon model". In: *Journal of Physics A: Mathematical and General* 28.19 (Oct. 1995), p. 5451. DOI: 10.1088/0305-4470/28/19/003. URL: https://dx.doi.org/10.1088/0305-4470/28/19/003.

[79] Yusuke Nomura. "Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry". In: *Journal of Physics: Condensed Matter* 33.17 (Apr. 2021), p. 174003. ISSN: 1361-648X. DOI: 10.1088/1361-648x/abe268. URL: http://dx.doi.org/10.1088/1361-648X/abe268.

[80] Yusuke Nomura and Masatoshi Imada. "Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy". In: *Phys. Rev. X* 11 (3 Aug. 2021), p. 031034. DOI: 10.1103/PhysRevX.11.031034. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.031034.

[81] Ortiz Marrero, Carlos and Kieferová, Mária and Wiebe, Nathan. "Entanglement-Induced Barren Plateaus". In: *PRX Quantum* 2 (4 Oct. 2021), 040316. DOI: {10.1103/PRXQuantum.2.040316}. URL: {https://link.aps.org/doi/10.1103/PRXQuantum.2.040316}.

[82] P. Singkanipa and D. A. Lidar. *Beyond unital noise in variational quantum algorithms: noise-induced barren plateaus and limit sets*. 2024. arXiv: {2402.08721} (quant-ph). URL: {https://arxiv.org/abs/2402.08721}.

[83] Park, Chae-Yeun and Killoran, Nathan. "Hamiltonian variational ansatz without barren plateaus". In: *Quantum* 8 (Feb. 2024), 1239. ISSN: 2521-327X. DOI: {10.22331/q-2024-02-01-1239}. URL: {https://doi.org/10.22331/q-2024-02-01-1239}.

[84] Pesah, Arthur and Cerezo, M. and Wang, Samson and Volkoff, Tyler and Sornborger, Andrew T. and Coles, Patrick J. "Absence of Barren Plateaus in Quantum Convolutional Neural Networks". In: *Physical Review X* 11.4 (Oct. 2021). ISSN: 2160-3308. DOI: {10.1103/physrevx.11.041011}. URL: {http://dx.doi.org/10.1103/PhysRevX.11.041011}.

[85] Preskill, John. "Quantum computing in the NISQ era and beyond". In: *Quantum* 2 (2018), 79.

[86] Pérez-García, David and Wolf, Michael M. and Petz, Denes and Ruskai, Mary Beth. "Contractivity of positive and trace-preserving maps under Lp norms". In: *Journal of Mathematical Physics* 47.8 (Aug. 2006). ISSN: 1089-7658. DOI: {10.1063/1.2218675}. URL: {http://dx.doi.org/10.1063/1.2218675}.

[87] Raginsky, Maxim. "Strictly contractive quantum channels and physically realizable quantum computers". In: *Physical Review A* 65.3 (Feb. 2002). ISSN: 1094-1622. DOI: {10.1103/physreva.65.032306}. URL: {http://dx.doi.org/10.1103/PhysRevA.65.032306}.

[88] Ragone, Michael and Bakalov, Bojko N. and Sauvage, Frédéric and Kemper, Alexander F. and Ortiz Marrero, Carlos and Larocca, Martín and Cerezo, M. "A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits". In: *Nature Communications* 15.1 (Aug. 2024). ISSN: 2041-1723. DOI: {10.1038/s41467-024-49909-3}. URL: {http://dx.doi.org/10.1038/s41467-024-49909-3}.

[89] Raussendorf, Robert and Browne, Daniel E and Briegel, Hans J. "Measurement-based quantum computation on cluster states". In: *Physical review A* 68.2 (2003), 022312.

[90] Ricard Puig and Marc Drudis and Supanut Thanasilp and Zoë Holmes. *Variational quantum simulation: a case study for understanding warm starts*. 2024. arXiv: {2404.10044} (quant-ph). URL: {https://arxiv.org/abs/2404.10044}.

[91] Richard V. Kadison. "A Generalized Schwarz Inequality and Algebraic Invariants for Operator Algebras". In: *Annals of Mathematics* 56.3 (1952), 494–503. ISSN: 0003486X, 19398980. URL: {http://www.jstor.org/stable/1969657}.

[92] Christopher Roth and Allan H. MacDonald. *Group Convolutional Neural Networks Improve Quantum State Accuracy*. 2021. arXiv: 2104.05085 [quant-ph].

[93] Christopher Roth, Attila Szabó, and Allan H. MacDonald. "High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks". In: *Phys. Rev. B* 108 (5 Aug. 2023), p. 054410. DOI: 10.1103/PhysRevB.108.054410. URL: https://link.aps.org/doi/10.1103/PhysRevB.108.054410.

[94] Subir Sachdev. "Quantum phase transitions". In: *Physics World* 12.4 (Apr. 1999), p. 33. DOI: 10.1088/2058-7058/12/4/23. URL: https://dx.doi.org/10.1088/2058-7058/12/4/23.

[95] Sack, Stefan H. and Medina, Raimel A. and Michailidis, Alexios A. and Kueng, Richard and Serbyn, Maksym. "Avoiding Barren Plateaus Using Classical Shadows". In: *PRX Quantum* 3.2 (June 2022). ISSN: 2691-3399. DOI: {10.1103/prxquantum.3.020365}. URL: {http://dx.doi.org/10.1103/PRXQuantum.3.020365}.

[96] Anders W. Sandvik. "Ground States of a Frustrated Quantum Spin Chain with Long-Range Interactions". In: *Phys. Rev. Lett.* 104 (13 Mar. 2010), p. 137204. DOI: 10.1103/PhysRevLett.104.137204. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.137204.

[97] Anders W. Sandvik, Leon Balents, and David K. Campbell. "Ground State Phases of the Half-Filled One-Dimensional Extended Hubbard Model". In: *Phys. Rev. Lett.* 92 (23 June 2004), p. 236401. DOI: 10.1103/PhysRevLett.92.236401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.236401.

[98] A.W. Sandvik. "Computational Studies of Quantum Spin Systems". In: *AIP Conference Proceedings* 1297.1 (2010), pp. 135–338. DOI: 10.1063/1.3518900. URL: https://aip.scitation.org/doi/abs/10.1063/1.3518900.

[99] Sarovar, Mohan and Proctor, Timothy and Rudinger, Kenneth and Young, Kevin and Nielsen, Erik and Blume-Kohout, Robin. "Detecting crosstalk errors in quantum information processors". In: *Quantum* 4 (2020), 321.

[100] Philipp Schmoll et al. "A programming guide for tensor networks with global SU(2) symmetry". In: *Annals of Physics* 419 (Aug. 2020), p. 168232. ISSN: 0003-4916. DOI: 10.1016/j.aop.2020.168232. URL: http://dx.doi.org/10.1016/j.aop.2020.168232.

[101] Kazuhiro Seki, Tomonori Shirakawa, and Seiji Yunoki. "Symmetry-adapted variational quantum eigensolver". In: *Phys. Rev. A* 101 (5 May 2020), p. 052340. DOI: 10.1103/PhysRevA.101.052340. URL: https://link.aps.org/doi/10.1103/PhysRevA.101.052340.

[102] Seneta, E. *Non-negative matrices and Markov chains*. New York: Springer, 2006. ISBN: 0387297650.

[103] Shettell, Nathan and Markham, Damian. "Graph states as a resource for quantum metrology". In: *Physical review letters* 124.11 (2020), 110502.

[104] Shor, Peter W. "Scheme for reducing decoherence in quantum computer memory". In: *Phys. Rev. A* 52 (4 Oct. 1995), R2493–R2496. DOI: {10.1103/PhysRevA.52.R2493}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.52.R2493}.

[105] Sukhwinder Singh and Guifre Vidal. "Tensor network states and algorithms in the presence of a global SU(2) symmetry". In: *Phys. Rev. B* 86 (19 Nov. 2012), p. 195114. DOI: 10.1103/PhysRevB.86.195114. URL: https://link.aps.org/doi/10.1103/PhysRevB.86.195114.

[106] Temme, Kristan and Bravyi, Sergey and Gambetta, Jay M. "Error Mitigation for Short-Depth Quantum Circuits". In: *Phys. Rev. Lett.* 119 (18 Nov. 2017), 180509. DOI: {10.1103/PhysRevLett.119.180509}. URL: {https://link.aps.org/doi/10.1103/PhysRevLett.119.180509}.

[107] Thanasilp, Supanut and Wang, Samson and Nghiem, Nhat Anh and Coles, Patrick and Cerezo, Marco. "Subtleties in the trainability of quantum machine learning models". In: *Quantum Machine Intelligence* 5.1 (May 2023). ISSN: 2524-4914. DOI: {10.1007/s42484-023-00103-6}. URL: {http://dx.doi.org/10.1007/s42484-023-00103-6}.

[108] Tüysüz, Cenk and Chang, Su Yeon and Demidik, Maria and Jansen, Karl and Vallecorsa, Sofia and Grossi, Michele. "Symmetry Breaking in Geometric Quantum Machine Learning in the Presence of Noise". In: *PRX Quantum* 5 (3 July 2024), 030314. DOI: {10.1103/PRXQuantum.5.030314}. URL: {https://link.aps.org/doi/10.1103/PRXQuantum.5.030314}.

[109] Ewout Van Den Berg, Zlatko K Minev, and Kristan Temme. "Model-free readout-error mitigation for quantum expectation values". In: *Physical Review A* 105.3 (2022), p. 032620. DOI: https://doi.org/10.1103/PhysRevA.105.032620.

[110] Van Den Berg, Ewout and Minev, Zlatko K and Kandala, Abhinav and Temme, Kristan. "Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors". In: *Nature physics* 19.8 (2023), 1116–1121.

[111] Tom Vieijra and Jannes Nys. "Many-body quantum states with exact conservation of non-Abelian and lattice symmetries through variational Monte Carlo". In: *Phys. Rev. B* 104 (4 July 2021), p. 045123. DOI: 10.1103/PhysRevB.104.045123. URL: https://link.aps.org/doi/10.1103/PhysRevB.104.045123.

[112] Tom Vieijra et al. "Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries". In: *Phys. Rev. Lett.* 124 (9 Mar. 2020), p. 097201. DOI: 10.1103/PhysRevLett.124.097201. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.097201.

[113] Ville Bergholm and Josh Izaac and Maria Schuld and Christian Gogolin and Shahnawaz Ahmed and Vishnu Ajith et al. *PennyLane: Automatic differentiation of hybrid quantum-classical computations*. 2022. arXiv: {1811.04968} (quant-ph).

[114] Luciano Loris Viteritti, Francesco Ferrari, and Federico Becca. "Accuracy of restricted Boltzmann machines for the one-dimensional $J_1 - J_2$ Heisenberg model". In: *SciPost Phys.* 12 (2022), p. 166. DOI: 10.21468/SciPostPhys.12.5.166. URL: https://scipost.org/10.21468/SciPostPhys.12.5.166.

[115] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. "Transformer Variational Wave Functions for Frustrated Quantum Spin Systems". In: *Physical Review Letters* 130.23 (June 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.236401. URL: http://dx.doi.org/10.1103/PhysRevLett.130.236401.

[116] Luciano Loris Viteritti et al. *Transformer Wave Function for the Shastry-Sutherland Model: emergence of a Spin-Liquid Phase*. 2024. arXiv: 2311.16889 [cond-mat.str-el].

[117] Wallman, Joel and Granade, Chris and Harper, Robin and Flammia, Steven T. "Estimating the coherence of noise". In: *New Journal of Physics* 17.11 (Nov. 2015), 113020. ISSN: 1367-2630. DOI: {10.1088/1367-2630/17/11/113020}. URL: {http://dx.doi.org/10.1088/1367-2630/17/11/113020}.

[118] Ling Wang and Anders W. Sandvik. "Critical Level Crossings and Gapless Spin Liquid in the Square-Lattice Spin-1/2 $J_1 - J_2$ Heisenberg Antiferromagnet". In: *Phys. Rev. Lett.* 121 (10 Sept. 2018), p. 107202. DOI: 10.1103/PhysRevLett.121.107202. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.107202.

[119] Wang, Samson and Fontana, Enrico and Cerezo, M. and Sharma, Kunal and Sone, Akira and Cincio, Lukasz and Coles, Patrick J. "Noise-induced barren plateaus in variational quantum algorithms". In: *Nature Communications* 12.1 (Nov. 2021). ISSN: 2041-1723. DOI: {10.1038/s41467-021-27045-6}. URL: {http://dx.doi.org/10.1038/s41467-021-27045-6}.

[120] Dave Wecker, Matthew B Hastings, and Matthias Troyer. "Progress towards practical quantum variational algorithms". In: *Physical Review A* 92.4 (2015), p. 042303. DOI: https://journals.aps.org/pra/pdf/10.1103/PhysRevA.92.042303.

[121] S.R. White and I. Affleck. "Dimerization and incommensurate spiral spin correlations in the zigzag spin chain: Analogies to the Kondo lattice". In: *Phys. Rev. B* 54 (14 Oct. 1996), pp. 9862–9869. DOI: 10.1103/PhysRevB.54.9862. URL: https://link.aps.org/doi/10.1103/PhysRevB.54.9862.

[122] David Wierichs, Christian Gogolin, and Michael Kastoryano. "Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer". In: *Physical Review Research* 2.4 (Nov. 2020). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.2.043246. URL: http://dx.doi.org/10.1103/PhysRevResearch.2.043246.

[123] David Wierichs et al. "General parameter-shift rules for quantum gradients". In: *Quantum* 6 (Mar. 2022), p. 677. DOI: 10.22331/q-2022-03-30-677. URL: https://doi.org/10.22331%2Fq-2022-03-30-677.

[124] Roeland Wiersema et al. "Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz". In: *PRX Quantum* 1.2 (Dec. 2020). ISSN: 2691-3399. DOI: 10.1103/prxquantum.1.020319. URL: http://dx.doi.org/10.1103/PRXQuantum.1.020319.

[125] Wiseman, H. M. and Milburn, G. J. "Quantum theory of field-quadrature measurements". In: *Phys. Rev. A* 47 (1 Jan. 1993), 642–662. DOI: {10.1103/PhysRevA.47.642}. URL: {https://link.aps.org/doi/10.1103/PhysRevA.47.642}.

[126] Yabo Wang and Bo Qi and Chris Ferrie and Daoyi Dong. *Trainability Enhancement of Parameterized Quantum Circuits via Reduced-Domain Parameter Initialization*. 2023. arXiv: {2302.06858} (quant-ph). URL: {https://arxiv.org/abs/2302.06858}.

[127] Jianwei Yang, Anders W. Sandvik, and Ling Wang. "Quantum criticality and spin liquid phase in the Shastry-Sutherland model". In: *Phys. Rev. B* 105 (6 Feb. 2022), p. L060409. DOI: 10.1103/PhysRevB.105.L060409. URL: https://link.aps.org/doi/10.1103/PhysRevB.105.L060409.

[128] Zhang, Hao-Kai and Liu, Shuo and Zhang, Shi-Xin. "Absence of Barren Plateaus in Finite Local-Depth Circuits with Long-Range Entanglement". In: *Phys. Rev. Lett.* 132 (15 Apr. 2024), 150603. DOI: {10.1103/PhysRevLett.132.150603}. URL: {https://link.aps.org/doi/10.1103/PhysRevLett.132.150603}.