Technische Universität München
TUM School of Natural Sciences

Tensor Network Methods in Many-Body Physics

Miguel Frías-Pérez

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Jonathan J. Finley, Ph.D.

Prüfende der Dissertation:
1. Hon-Prof. Dr. Ignacio Cirac
2. Prof. Dr. Michael Knap

Die Dissertation wurde am 20.02.2025 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 17.03.2025 angenommen.

Abstract

This thesis contributes to the development of algorithms for the simulation of different aspects of many-body physics with tensor networks.

In the context of classical statistical mechanics, we present a new algorithm that combines tensor networks with Metropolis-Hastings Markov chain methods. We demonstrate its performance in generating correlated updates for spin systems and show that the new technique offers a powerful alternative to traditional Monte Carlo methods, enabling more efficient sampling in complex systems.

A big challenge in tensor network simulations is their application for the study of out-of-equilibrium dynamics of quantum systems. This thesis presents two directions to overcome the limitations of current tensor network methods in dynamical scenarios. By carefully examining the entanglement structures that appear in the tensor networks that give rise to out-of-equilibrium expectation values, we show that the timescales that tensor networks can simulate can be extended.

Beyond many-body physics, we also explore what tensor networks can say rigorously about the impact of noise on the performance of near-term quantum devices. In particular, we develop a method to compute certifiable lower bounds on the minimum energy attainable by quantum circuits in the presence of different noise models. This approach provides insights into the limitations of quantum computers under realistic noise assumptions.

As a result of this work, we demonstrate how tensor network methods can be further developed and combined with other techniques, and we contribute to their role in solving complex computational problems in quantum and classical physics.

Zusammenfassung

Diese Arbeit trägt zur Entwicklung von Algorithmen für die Simulation verschiedener Aspekte der Vielkörperphysik mit Tensornetzen bei.

Im Kontext der klassischen statistischen Mechanik stellen wir einen neuen Algorithmus vor, der Tensornetzwerke mit Metropolis-Hastings-Markov-Ketten-Methoden kombiniert. Wir demonstrieren die Leistungsfähigkeit des Algorithmus bei der Erzeugung korrelierter Zustandsaktualisierungen für Spinsysteme und zeigen, dass der neue Ansatz eine leistungsstarke Alternative zu herkömmlichen Monte-Carlo-Methoden darstellt und effizienteres Monte-Carlo-sampling in komplexen Systemen ermöglicht.

Eine große Herausforderung bei der Simulation mit Tensornetzwerken ist ihre Anwendung für die Untersuchung der Dynamik von Quantensystemen außerhalb des Gleichgewichts. In dieser Arbeit werden zwei Wege aufgezeigt, die Grenzen der derzeitigen Tensornetzwerkmethoden in dynamischen Szenarien zu überwinden. Durch eine sorgfältige Untersuchung der Verschränkungsstrukturen, die in jenen Tensornetzwerken auftreten, welche zu Erwartungswerten außerhalb des Gleichgewichts führen, zeigen wir, dass die Zeitskalen, die Tensornetzwerke simulieren können, erweitert werden können.

Über die Vielkörperphysik hinaus untersuchen wir auch, was Tensornetzwerke rigoros über die Auswirkungen von Rauschen auf die Leistung von Quantengeräten in naher Zukunft sagen können. Insbesondere entwickeln wir eine Methode zur Berechnung zertifizierbarer unterer Schranken für die von Quantenschaltkreisen erreichbare Mindestenergie in Gegenwart verschiedener Rauschmodelle. Dieser Ansatz bietet Einblicke in die Grenzen von Quantenschaltungen unter realistischen Rauschannahmen.

Als Ergebnis dieser Arbeit zeigen wir, wie Tensornetzwerk-Methoden weiterentwickelt und mit anderen Techniken kombiniert werden können und wie sie zur Lösung komplexer Berechnungsprobleme in der Quanten- und klassischen Physik beitragen können.

List of publications

Articles included in this thesis:

1. **M. Frías-Pérez**, M. Mariën, D. Pérez-García, M. C. Bañuls, S. Iblisdir
   Collective Monte Carlo updates through tensor network renormalization
   SciPost Phys. 14, 123 (2023)

2. **M. Frías-Pérez**, M. C. Bañuls
   Light cone tensor network and time evolution
   Phys. Rev. B 106, 115117 (2022)

3. **M. Frías-Pérez**, L. Tagliacozzo, M. C. Bañuls
   Converting Long-Range Entanglement into Mixture: Tensor-Network Approach to Local Equilibration
   Phys. Rev. Lett. 132, 100402 (2024)

4. S. D. Mishra*, **M. Frías-Pérez***, R. Trivedi
   Classically Computing Performance Bounds on Depolarized Quantum Circuits
   PRX Quantum 5, 020317 (2024)
   * Equal contribution

Contents

Chapter 1

Introduction

The development of computing in the last century has completely changed the world. In a period of 70 years, we have passed from computers whose capabilities are comparable to those of a single person with a pen and paper to widespread devices with far larger computing capabilities than all humans combined. This development has profoundly reshaped how we study all sciences, from mathematics to economics or biology.

At the same time, physics has no shortage of hard computational problems. When we want to study systems composed of many particles, classical or quantum, textbooks tell us that we must solve well-posed mathematical problems in spaces whose dimension grows exponentially with the number of constituents in the system. For example, classical statistical mechanics deals with specific probability distributions that live in those spaces, while quantum many-body physics requires the study of linear algebra problems in them. This growth in dimensionality is particularly limiting: in many cases, in order to make contact with real systems, we are interested in the limit of a large or infinite number of particles.

This limitation calls for the development of numerical methods that can efficiently deal with the relevant physical problems. This call does not come only from being limited to studying only small systems exactly but also from the fact that nature itself is performing these computations efficiently around us. It is only our naive algorithms, classical or quantum, to simulate those problems that come with an exponential overhead. Many powerful techniques have emerged from this obstruction: Monte Carlo methods, dynamical mean field theory, series expansion techniques, or, more recently, tensor networks or quantum algorithms. These techniques come with an upside: as well as allowing precise studies of systems that are outside the reach of exact descriptions, where the exponentially large space is treated in full detail, they sometimes provide insight into the mathematical structure of the underlying physical problems.

The main theme of this thesis is the study and development of numerical

algorithms for one of the previous techniques: tensor networks. Tensor networks are a way to represent arrays with many indices in a compact fashion. In particular, they are extremely useful when representing the coefficients of a quantum-many body wavefunction. Initially, they were discovered in their one-dimensional version, matrix product states. Although there had been algorithms formulated, explicitly or implicitly, with them [1–4], it took some time until it was realized that they presented a variational class of quantum many-body states with many desirable properties for studying equilibrium states of one-dimensional systems [5–8]. Once the ansatz and algorithms were understood for ground states, there was an explosion of methods to tackle different aspects of quantum-many body systems: thermal states [9–12], low-lying excited states [13, 14], closed or open system dynamics [9, 10, 15], or extensions to target directly the thermodynamic limit [16, 17]. At the same time, extensions were soon developed to address critical, continuous, or higher-dimensional systems [18–20].

Thirty years after their inception, tensor networks are the go-to tool to simulate lattice models in low-dimensional systems. They are used to study challenging ground state or thermal phase diagrams [21–26], as well as dynamical phenomena [27–29] or even the simulation of quantum circuits in near-term quantum devices [30–32]. Nevertheless, they are an active area of research, with plenty of new developments in recent years.

In this thesis, we will present three new directions in which we have tried to expand the tensor network toolbox. Before diving into them, in Chapter 2, we will introduce the tensor network knowledge and notation used in the rest of the thesis. We will also present some basic ideas needed to understand the new applications of tensor networks developed in our work.

Then, in Chapter 3, we will study how tensor networks can be combined with Monte Carlo methods. In those methods, in order to study many-body systems, stochastic processes are constructed whose long-time properties reproduce a probability distribution of interest. However, the most extended methods to construct those processes use dynamics that are local. At a given step, only the state of a single or a few individual particles in the system is updated. This is fine for systems where particles interact weakly but suffers from severe limitations when the system in consideration is strongly correlated. In those cases, the times needed for the stochastic process to reach the long-time behavior will scale with the system size, even exponentially in the worst case. In our work, we will show how to construct new Monte Carlo dynamics that can greatly speed up the timescales to equilibration, and make possible the simulation of challenging many-body problems via Monte Carlo methods.

In Chapters 4 and 5, we will develop new algorithms to study dynamical phenomena with tensor networks. While the technical aspects of it are understood,

that is, there exist algorithms to integrate the differential equations governing the dynamics of a quantum many-body system using tensor networks, the techniques are severely limited in practice. In the general case, the numerical resources needed to keep an exact representation of the time-evolved state grow exponentially with time. In the chapters mentioned we will explore two different directions to go around this issue. In Chapter 4, we will study what happens if, instead of trying to approximate the exact time-evolved state, we target just a single expectation value. As we will see, we can write the dynamical expectation value as a tensor network and compare different strategies to evaluate it. We will find new strategies with different scalings or prefactors to the usual algorithms, signaling better ways to use the resources when simulating out-of-equilibrium physics with tensor networks. Chapter 5 follows a different strategy: out-of-equilibrium generic quantum-many body systems are expected to thermalize. That means that if we start with a state that is not an eigenstate of the Hamiltonian governing the dynamics or a superposition of a small number of them, after evolving by a time long enough, local expectation values will be captured by a thermal state compatible with the initial state. It can be shown that tensor networks capture those states. This poses a challenge: if the short and the long-time behavior is captured by tensor networks, is it possible to say something about the intermediate time scales? We will tackle this problem in Chapter 5 by studying the entanglement structures that appear in the tensor network simulation of the dynamics.

Afterward, in Chapter 6, we will explore what tensor networks can tell us about the impact of noise in the algorithms proposed for near-term quantum computers. In recent years, the progress towards constructing devices capable of quantum computation has been astonishing. Still, the present computers have sizes and error rates that make them incapable of being used to run quantum algorithms with a provable advantage. In the meantime, there have been plenty of proposals of heuristic algorithms to try to outperform classical techniques. Most of these proposals tackle physical problems; often, the task in consideration is finding the ground state of a Hamiltonian. Thus, it is desirable to study the impact of noise in the solutions found with the quantum device. In this cwhhapter, we will propose a way to compute rigorous bounds on the performance of these devices using tensor networks.

Finally, in Chapter 7, we present our conclusions and outlook, as well as new directions for exploration originating from the new techniques in this thesis.

Chapter 2

Preliminaries

2.1 Overview

In this introductory chapter, we will present the unifying theme of this thesis: tensor networks (TNs). First, we present them in their simplest setting: matrix product states (MPSs). From there, we generalize them to higher dimensions and discuss the physical ground for their success in many-body physics: area laws. In the latter two sections, we briefly introduce two connected fields where we find new applications of tensor networks in this thesis: Monte Carlo simulations in many-body physics and near-term quantum computation.

2.2 Tensor Networks

2.2.1 The many-body problem

Let us consider a quantum system made up of N individual systems. For simplicity, we assume that every single particle system has the same number of states, d. The postulates of quantum mechanics specify that an arbitrary state of the system is described by a vector living in the Hilbert space obtained by taking the tensor product of the individual Hilbert spaces. Thus, a generic pure state of the system has the form

$$|\Psi\rangle = \sum_{i_1,\dots i_N}^{d} c_{i_1,\dots i_N} |i_1, \dots i_N\rangle, \tag{2.2.1}$$

specified by d^N amplitudes, the entries of $c_{i_1,\dots i_N}$. This consideration, which is usually referred to as "the curse of dimensionality" in the context of quantum many-body physics, makes apparent a limitation in the numerical study of these

systems: for large N, describing the state by storing or specifying the entries of $c_{i_1,\dots i_N}$ is going to be inefficient. Even if one is interested in physical properties of the system, such as its local expectation values in equilibrium states, the storage cost of keeping exactly all the entries of the tensor $c_{i_1,\dots i_N}$ will limit the exact study to a small number of particles.

This limitation calls for a way to circumvent the problem. The one on which tensor networks are built on is the variational principle [33]. Its critical insight is to disregard the exponentially large Hilbert space and consider a subfamily of states in it, $|\Psi(x)\rangle$, parametrized by a manageable number k of parameters $x^i \in \mathbb{R}^k$. The subfamily has to be specified in such a way that it can be manipulated efficiently, i.e., it allows for the computation of expectation values. Then, if we are interested in finding out the energy and different physical properties of the ground state of a many-body system, we can tune the parameters to find the best approximation to the ground state inside the class by solving the following problem:

$$\min_x E_x = \min_x \langle \Psi(x)|H|\Psi(x)\rangle. \tag{2.2.2}$$

The resemblance of the obtained state to the actual ground state will be limited by the ability of the variational class to capture the relevant physics. More sophisticated schemes based on similar formulations exist for different tasks, such as simulating dynamics of quantum systems or studying thermal physics [34–37].

As we will see in the next sections, tensor network states offer a physically motivated class for the study of relevant many-body states in the Hilbert space. Furthermore, due to their simple nature, as they are based on linear algebra, they also offer a plethora of sophisticated techniques to manipulate and optimize them. However, before we delve into them, we will briefly introduce some graphical notation that will greatly help their formulation.

2.2.2 Tensor network diagrams

The main building block of tensor networks, as their name says, are tensors. Tensors are multidimensional arrays, such as the coefficient $c_{i_1,\dots i_N}$ in Eq. 2.2.1. In this subsection, we will introduce a graphical notation that will simplify the presentation of the algorithms and manipulations that we will present later.

Tensors are denoted by their rank, that is, the number of indices they contain: a rank-0 tensor is a scalar, a rank-1 is a vector, and a rank-2 is a matrix. The coefficients of the many-body wavefunction $c_{i_1,\dots i_N}$ are a rank-N tensor. Graphically, tensors can be denoted by geometrical shapes containing as many legs as indices they have.

We say that two indices are contracted along a common index j to denote the operation that takes two tensors, multiplies their entries, and sums along those

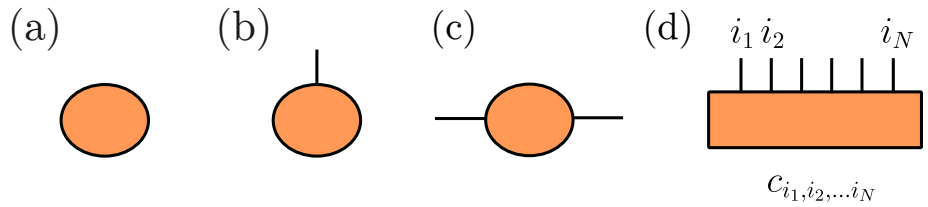


Figure 2.2.1: Tensor network diagrams representing (a) a scalar, (b) a vector, (c) a matrix, (d) the rank-N tensor representing the coefficients of the wavefunction 2.2.1.

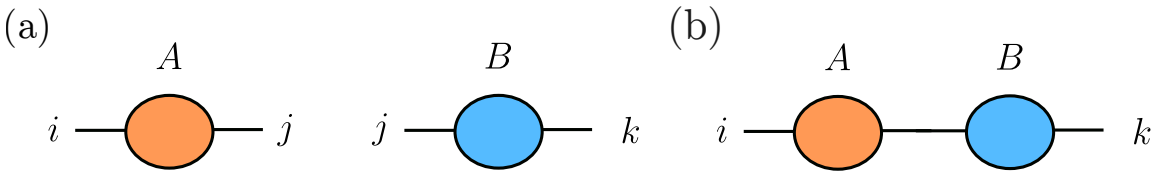


Figure 2.2.2: Tensor network diagram representing the matrix multiplication of A and B. In this diagram, we represent the indices explicitly but will omit them in general.

where the value of the index is repeated. For instance, the multiplication of matrices A and B, of sizes $m \times n$ and $n \times k$ respectively,

$$C_{ik} = \sum_j A_{ij} B_{jk}, \tag{2.2.3}$$

corresponds to the contraction of two rank-2 tensors along the column index of the first and the row index of the second matrix involved. The graphical notation for their multiplication can be seen in Figure 2.2.2.

2.2.3 Matrix product states

Let us return to the state defined in Eq. 2.2.1. A state is said to be a matrix product state of bond dimension D [38, 39], in the finite setting with open boundary conditions, if the wavefunction can be written as

$$|\Psi\rangle = \sum_{i_1,\ldots i_N}^{d} A_1^{i_1} A_2^{i_2} \ldots A_N^{i_N} |i_1, \ldots i_N\rangle, \tag{2.2.4}$$

where $A_n^{i_n}$ is a complex matrix for every possible value of i_n and n. Alternatively, we can view A_n as three-legged tensors, with i_n the so-called physical index and

the matrix indices referred to as virtual indices or bond indices. The sizes of the tensors are such that for a set of fixed indices $i_1, \dots i_N$, the wavefunction coefficient is a scalar. The maximum size of the virtual indices, D, is the bond dimension of the matrix product state. Using the graphical notation introduced in the previous section, we can represent the state with the diagram in Fig. 2.2.3.

Matrix product states exhibit a series of properties that make them very interesting from a quantum information and many-body physics point of view. Here below, we list some of the most relevant for this thesis:

- Efficient computation of local observables and overlaps

  Compared to the general state of Eq. 2.2.1, computing properties of an MPS does not involve a sum of exponentially many terms. As an illustrative example, consider the computation of the norm of one:

$$\langle \Psi | \Psi \rangle = \sum_{i'_1, \dots i'_N}^{d} \sum_{i_1, \dots i_N}^{d} \left(A_1^{i'_1} A_2^{i'_2} \dots A_N^{i'_N} \right)^* A_1^{i_1} A_2^{i_2} \dots A_N^{i_N} \langle i'_1, \dots i'_N | i_1, \dots i_N \rangle. \tag{2.2.5}$$

  If we regroup the items in the product and define the set of rank-4 tensors $E_n = \sum_{i_n}^{d} {A_n^{i_n}}^* \otimes A_n^{i_n}$, the norm's value can be computed numerically by contracting appropriately the tensors, as seen in Fig. 2.2.3 (b) and (c).

$$\langle \Psi | \Psi \rangle = E_1 E_2 \dots E_N. \tag{2.2.6}$$

  Thus, computing the product of the E_n matrices gives us the norm of an MPS. For the practical cost of the computation, it is convenient to use the low-rank structure of the tensors E_n explicitly [39]. An analogous scheme can be used to compute local expectation values or overlaps between two different MPSs.

- Translation invariance and extension to the infinite-system size limit

  In many fields of physics, we are often interested in systems that display translation invariance: the microscopic parameters of the Hamiltonian do not depend on the specific position in consideration. Similarly, another common theme in quantum many-body physics is the so-called thermodynamic limit, where we are interested in systems with an extremely large number of constituents in order to make contact with real materials. It is then useful, both from a theoretical or a practical point of view, to take the limit of the number of subsystems to infinity.

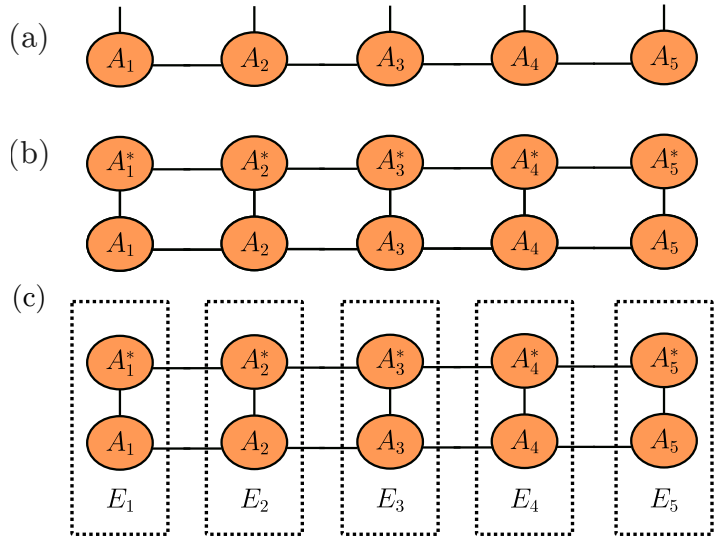


Figure 2.2.3: (a) Tensor network diagram representing an MPS with open boundary conditions. (b) Tensor network resulting from the computation of the norm of an MPS. (c) Transfer matrix picture of the contraction of the norm of an MPS.

It turns out that it is possible to impose both in MPSs [16, 40]. Indeed, matrix product states where the same tensor is repeated in all sites $A^{i_1}A^{i_2}A^{i_3}\ldots$, or a small number of tensors are repeated in a periodic fashion, $A^{i_1}B^{i_2}A^{i_3}B^{i_4}\ldots$ give rise to states with built-in translation invariance properties. Similarly, if we repeat the tensor an infinite number of times, they still allow for efficient manipulations with them [40]. In this manner, they constitute a very powerful ansatz to treat directly these scenarios.

- Bounded rank and entanglement

  Consider the connected subsystem comprising the particles in $R = \{1, \ldots n\}$ in a state described by an MPS of bond dimension D. If we compute its reduced density matrix ρ_R by tracing out the degrees of freedom outside of it, it can be seen that the resulting rank of the matrix is at most D, the bond dimension (see Fig. 2.2.4). This is the case regardless of the subsystem's size R. This result, consequently, bounds the entanglement entropy[1] between subsystem R and its complement R^c,

$$S(\rho_R) = -\operatorname{Tr}(\rho_R \log \rho_R) \leq \log D. \tag{2.2.7}$$

[1]Given a pure quantum state of a composite system, the entanglement entropy with respect to a partition into a region R and its complement R^c is defined as the von Neumann entropy of the reduced density matrix of R, ρ_R [41].

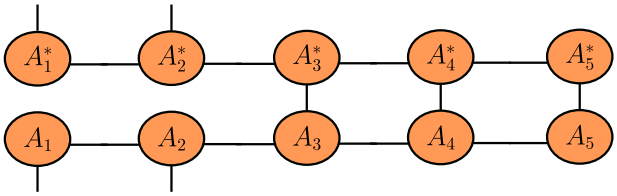


Figure 2.2.4: Tensor network diagram representing the computation of the reduced density matrix for the subsystem $R = \{1, 2\}$. Using the tensor network notation plus making use of the inequality for the rank of the product of two matrices, $\text{rank}(M_1 M_2) \leq \min(\text{rank}(M_1), \text{rank}(M_2))$, we can show that the rank of ρ_R is smaller or equal than D, the bond dimension.

Notice that this statement explicitly depends on the partition. If, instead, we were to compute the reduced density matrix of subsystem $R = \{1, 3, 5 \ldots 2n+1\}$, the rank of the resulting reduced density matrix could grow exponentially with the size of the region R, resulting in an entanglement entropy that potentially grows with the size of the region R.

- Matrix product operators

  The idea behind MPSs can be used to describe operators. The way to do so is to include an extra leg in the tensors in every site to describe both the bra and the ket indices of the operators. In this way, we can define matrix product operators (MPOs),

$$O = \sum_{i_1, \ldots i_N}^{d} W_1^{i_1, i_1'} W_2^{i_2, i_2'} \ldots W_N^{i_N, i_N'} |i_1, \ldots i_N\rangle\langle i_1', \ldots i_N'|. \tag{2.2.8}$$

  MPOs can be shown to describe relevant operators of one-dimensional many-body quantum systems: they capture exactly all short-range Hamiltonians [42], and can be shown to efficiently approximate thermal states [9, 43].

The previous list of properties that matrix product states satisfy explains why they are a very interesting class of states from the quantum information point of view, as well as hints as to why they are successful in capturing the ground states of gapped one-dimensional systems. In the next section 2.2.4, we will describe the so-called area law, which is deeply connected to their success in describing one-dimensional equilibrium states. Despite the incredible success of MPSs, early researchers working on them quickly realized that their immediate application to two-dimensional systems was not straightforward [44]. In section 2.2.5, we will discuss their generalizations to that case.

2.2.4 Area laws

Let us consider a thought experiment: pick a random pure state in a multipartite Hilbert space and compute its entanglement entropy across a bipartition into two subregions of equal size. It can be shown [45] that in the large size limit, the expected entanglement entropy will be close to maximal: the entanglement entropy will grow proportionally to the volume of the subregions. This is a reflection of something that we have seen in section 2.2.1: the Hilbert space contains exponentially many states. However, it seems unlikely that the states we can observe or prepare in nature, which are equilibrium states of simple Hamiltonians or states prepared after evolving a quantum system for a short time, can exploit and span this exponential complexity [46].

Indeed, it can be shown that many physical systems of interest exhibit what is usually referred to as an area law [47]: certain correlation measures computed for a region of the system do not depend on the size of the region, which would be an upper bound to the dependence they can exhibit, but instead depend on the size of the area that separates the region with the rest of the system. Examples of this behavior are the ground states of gapped one-dimensional Hamiltonians, which satisfy an area law for their entanglement entropy, or thermal states of local Hamiltonians on lattices for the mutual information.

This justifies using the MPS approximation for the ground states of gapped one-dimensional quantum systems, as it can also be shown that one-dimensional states with an area law, albeit for the Rényi entropy with $\alpha < 1$ [48], have an efficient MPS description. Furthermore, it also provides the understanding of how to generalize MPS to higher dimensions, namely, by constructing TN states that can mimic the patterns of entanglement in those geometries.

2.2.5 Higher-dimensional Tensor Network states

In higher-dimensional systems, the one-dimensional structure of MPSs leads to a mismatch between the locality of the state and the one of the Hamiltonian. The way to generalize tensor network states to higher dimensions consists in constructing tensor networks that mimic the locality of the Hamiltonian: for a lattice system, we place a tensor on every vertex of the lattice with an index that describes the degrees of freedom of the local system. At the same time, the tensor has virtual indices, or legs, that connect it to the other neighboring tensors (see Fig. 2.2.5). This generalization, which is called a Projected Entangled Pair State or PEPS [19], can be shown to satisfy an area law for the entanglement entropy in higher dimensions.

Similarly to MPSs, PEPSs are an extremely interesting class of states both from the many-body and the quantum information points of view. They represent many relevant two-dimensional states, such as the toric code, the resonating valence bond

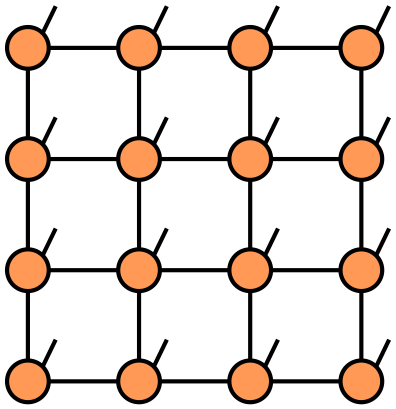


Figure 2.2.5: Tensor network diagram of a PEPS in a square lattice: a set of tensors are placed in the underlying lattice with virtual indices that connect them to their nearest neighbors and a physical leg that describes the state of the quantum system in each site.

state, or the two-dimensional cluster state [49, 50]. At the same time, they are an indispensable ansatz for numerical simulations of challenging strongly correlated systems [23, 24].

2.3 Monte Carlo methods

The many-body problem introduced in Subsection 2.2.1 calls for multiple approaches to simulate strongly correlated systems: distinct methods might be limited by different factors and provide results in complementary models. At the same time, when two methods are applicable to a given model, they can provide confirmation of the validity of the results if they agree.

In this section, we will introduce one of the most powerful approaches to deal with challenging classical and quantum strongly correlated systems, the Monte Carlo method [51–53]. For concreteness, let us focus on the study of classical many-body systems at finite temperature on a lattice. The central object of interest in those systems is the partition function

$$Z(\beta) = \sum_{\omega} e^{-\beta H(\omega)}, \tag{2.3.1}$$

where, $\beta = 1/T$ is the inverse temperature of the system, $H(\omega)$ represents the Hamiltonian of the system as a function of the microscopical configuration $\omega =$

$\{\sigma_1, \dots \sigma_N\}$ and the sum runs over all possible configurations of the system. A paradigmatic example of a model described in this framework is the Ising model, where classical spins $\sigma_i = \{-1, 1\}$ are placed in the vertices of a graph, with pairwise interactions between particles connected by an edge of the graph,

$$H(\omega) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j. \tag{2.3.2}$$

The partition function represents the normalization of the probability distribution of a given microscopic state at thermal equilibrium, the Boltzmann distribution. Consequently, computing the expectation value of an observable O in the system corresponds to taking the average

$$\langle O \rangle = \sum_{\omega} O(\omega) p(\omega) = \sum_{\omega} O(\omega) \frac{e^{-\beta H(\omega)}}{Z(\beta)}. \tag{2.3.3}$$

As in the previous section, computing this expectation value suffers from a dimensionality problem. The sum in the previous equation runs over a number of configurations that grows exponentially with the number of constituents of the system in consideration. This is extremely problematic, as we are interested in studying systems in the thermodynamic limit if we want to draw conclusions about the behavior of real materials. A possible way to overcome this limitation is based on sampling. If we are capable of getting samples from the probability distribution $p(\omega)$, we can obtain an estimate for the expectation value of O by taking the average over the obtained samples [54].

In the general case, sampling exactly from a given multi-dimensional distribution can be a hard task. In practice, approximate heuristic results can be obtained using Markov chains. In the rest of this section, we briefly introduce what a Markov chain is and show different ways to construct them so that, asymptotically, we can sample from the distribution in Eq. 2.3.3.

2.3.1 Markov chains

A Markov chain is a stochastic process in which the probability of each event depends only on the result of the immediate previous event. Given the sequence of stochastic events $X_0, X_1, X_2, \dots$, a Markov chain satisfies the Markov property,

$$P(X_n | X_0, X_1 \dots X_{n-1}) = P(X_n | X_{n-1}). \tag{2.3.4}$$

Let us consider a Markov chain over the configuration space of the Ising model: at every time step n, the random variable X_n takes values in the space of all possible spin configurations ω. As the configuration space is finite, we can write the probability distribution at a given step n as a matrix-vector product between

the transition matrix $P(X_n = \omega | X_{n-1} = \omega')$ and the probability distribution at the step before, $P(X_{n-1} = \omega')$,

$$P(X_n = \omega) = \sum_{\omega'} P(X_n = \omega | X_{n-1} = \omega') P(X_{n-1} = \omega'). \tag{2.3.5}$$

In this manner, the state of the chain at a time n can be understood as applying the n-th power of the transition matrix to the probability distribution of the initial configuration X_0. The spectral properties of the transition matrix will then govern the long-time behavior of the Markov chain. By construction, the transition probabilities $P(X_n = \omega | X_{n-1} = \omega')$ form a right stochastic matrix [2]. By the Gershgorin circle theorem, its eigenvalues are contained in the unit circle. In particular, it has at least an eigenvalue of value 1, whose left eigenvector is the flat probability vector. Additionally, if this eigenvalue is unique, the right eigenvector has nonnegative entries, which give the asymptotic probability of a configuration for the Markov chain at large times [55]. The asymptotic probability distribution is sometimes referred to as the "fixed point" of the chain. Markov chain Monte Carlo methods aim to construct processes whose long-time behavior is described by a probability distribution of interest.

It turns out that there is a simple way of enforcing the unicity of the leading eigenvector by imposing two conditions on the transition matrix: aperiodicity and irreducibility. Aperiodicity guarantees that we do not move across the configuration space in a periodic manner, while irreducibility ensures that it is possible to visit all the points in the configuration space. For the reader interested in the details, we refer them to [53] for a practical presentation and [55] for a more rigorous and complete introduction.

In the next section, we will describe a useful way of constructing Markov chains with transition matrices that have as unique fixed points the desired probability distribution in high-dimensional configuration spaces.

2.3.2 Metropolis-Hastings filter

As described in the previous section, we can construct stochastic processes that asymptotically sample a given probability distribution using Markov chains with adequate transition matrices. The last issue is how to construct explicitly those matrices. A possible way is to use the so-called detailed balance or reversibility condition. Using the notation of the previous section, this condition reads:

[2] A right stochastic matrix has nonnegative entries and columns whose entries sum up to 1. In the case of a transition matrix of a Markov chain, those conditions follow from the nonnegativity of the probabilities and conservation of the probabilities: for any $P(X_{n-1} = \omega')$, the corresponding probabilities in $P(X_n = \omega)$ have to be non-negative and normalized.

$$P(X_n = \omega | X_{n-1} = \omega')p(\omega') = P(X_n = \omega' | X_{n-1} = \omega)p(\omega), \tag{2.3.6}$$

where $p(\omega)$ is the probability distribution of interest, which will be the Boltzmann distribution in our case, and ω is short-hand notation for a specific spin configuration $\{\sigma_1, \dots \sigma_N\}$. Given a transition matrix of this form, it is straightforward to show that $p(\omega)$ is the leading right eigenvector of the matrix,

$$\sum_{\omega'} P(X_n = \omega | X_{n-1} = \omega')p(\omega') = \sum_{\omega'} P(X_n = \omega' | X_{n-1} = \omega)p(\omega) = p(\omega). \tag{2.3.7}$$

In the first equality, we have used the detailed balance condition, while in the second we have just used the normalization of the transition matrix.

A particular choice to fulfill the detailed balance condition is by using the so-called Metropolis or Metropolis-Hastings filter [56]. In order to satisfy Eq. 2.3.6, the transition probability is split into two parts: the probability of proposing a move $g(\omega|\omega')$ times the probability of accepting the move, $P_{\text{acc}}(\omega' \to \omega)$,

$$P(X_n = \omega | X_{n-1} = \omega') = g(\omega|\omega')P_{\text{acc}}(\omega' \to \omega). \tag{2.3.8}$$

If the acceptance probability is given by

$$P_{\text{acc}}(\omega' \to \omega) = \min\left\{1, \frac{g(\omega'|\omega)p(\omega)}{g(\omega|\omega')p(\omega')}\right\}, \tag{2.3.9}$$

the transition probability $P(X_n = \omega | X_{n-1} = \omega')$ satisfies detailed balance with respect to the probability distribution of interest $p(\omega)$, as can be seen by substituting the previous acceptance probability in Eq. 2.3.6. This particular choice of transition probability is usually referred to as the Metropolis-Hastings filter [56]. Furthermore, it can be shown that the transition matrix is a valid stochastic matrix [53] for any valid proposal probability $g(\omega'|\omega)$. This last fact is very relevant, as configuration spaces in many-body physics are exponentially large in the number of constituents of the system, and thus, only simple proposal probabilities can be constructed.

Despite the algorithm being valid for any choice of proposal probability $g(\omega|\omega')$, as convergence to the stationary distribution is guaranteed by the detailed balance condition, the algorithm's efficiency will completely depend on the choice of a clever proposal probability. In practice, its efficiency is reflected in two quantities: equilibration times and auto-correlation times. The former dictates the timescale in which Markov chains reach the equilibrium distribution. The latter, the auto-correlation time, is a heuristic quantity. From a practical point of view, Markov chains in many-body physics are run for very long times after they have equilibrated, in order to obtain many samples from the equilibrium distribution. This is less

costly than equilibrating a chain starting from an initial configuration every time we want to obtain a sample. However, successive configurations of a Markov chain are correlated; they may differ only locally if we use proposal probabilities based on local updates.

Knowing the two timescales is indispensable to estimate expectation values accurately [52]: if our simulations have not equilibrated, the expectation values are computed with respect to a different probability distribution. At the same time, if we disregard the auto-correlation times, the samples obtained are not independent, and potentially, the statistical errors might not decrease with the number of samples. Thus, it is necessary to develop algorithms where the two timescales are small, as they set the number of samples that we can obtain in a given simulation for a fixed amount of computer time.

In many-body simulations, there are usually two situations where the two timescales can potentially blow up: critical systems and frustrated systems. In both cases, these timescales will grow with the system size, possibly even in an exponential manner. Many techniques exist to alleviate these problems: cluster algorithms [57, 58], parallel tempering [59], replica exchanges [60, 61], etc. . However, these techniques have limited applicability: they have to be tailored to a specific scenario, and they cannot be generalized beyond a single model. Thus, there is a need for developing methods to simulate challenging models. In Chapter 3, we will see how tensor networks can be used to construct Markov chains that equilibrate and decorrelate much faster than the usual techniques.

2.4 Quantum Computation and errors in near term quantum computers

The complexity of classically simulating quantum mechanics we discussed in Section 2.2.1 poses at the same time an exciting question: if we do not know how to efficiently simulate a quantum mechanical system made up of many particles, is it possible to use those systems to encode and solve other problems which we also do not know how to treat efficiently with classical methods? It turns out that the answer to this question seems to be positive. There are problems that are believed to be intractable by classical computers, for which there exist efficient quantum algorithms [3]. As well as quantum algorithms to simulate quantum mechanical systems [62–64], there exist methods for problems outside the realm of quantum mechanics: the most remarkable ones are Shor's algorithm [65], Grover's algorithm [66], or the HHL

[3]We use the notion of efficiency in the computational complexity sense. An algorithm is efficient when it runs in a time and uses a memory that grows at most polynomially with the size of the problem's input.

algorithm [67].

In the rest of the section, we will present two basic ingredients to understand some of the results in this thesis: the circuit model of quantum computation and the impact of noise on quantum computation.

2.4.1 Circuit model of quantum computation

In the quantum circuit model of quantum computation [68], a computation is a sequence in which the qubits are initialized to known values, followed by performing unitary gates and measurements in the system. At the end of the circuit, a quantum state of interest is obtained as an output, or a measurement is performed, which are the outcome of the computation.

It can be shown that the circuit model is equivalent to other models of quantum computation, like adiabatic quantum computation [69, 70] or measurement-based quantum computation [71].

Two key theorems guarantee the soundness of the circuit model of quantum computation: the Solovay-Kitaev [72] and threshold theorems [73]. The first one guarantees that any unitary can be approximated with a discrete set of gates. The second states that under general models of local noise, a threshold noise rate exists below which any given circuit can be error-corrected. Both imply only a very mild overhead in the computation.

2.4.2 Noise in quantum computers

Despite the theoretical promises and the huge recent theoretical efforts [74–76], present-day quantum computers are still, mostly, in the so-called NISQ era[4]. The number of qubits in them is still small, of the order of 50 to 100, yet they are already too large to simulate classically. At the same time, the noise rate is still too large to perform error correction. Only very recently, error-correcting codes that decrease the probability of a logical error have been experimentally demonstrated [78, 79], with limited applicability, as the number of qubits can not grow too large yet. Nonetheless, many heuristic algorithms have been proposed to run in NISQ devices [80], so it is an interesting question to understand the impact of noise on those circuits.

A common way to model the noise in the circuit setting is to consider that with every unitary that is applied to a qubit, the qubit is also subjected to some noise due to imperfections in the implementation of the unitary and interaction with the environment. Several properties about the asymptotics of the circuit and the

[4]NISQ stands for Noisy Intermediate Scale Quantum (devices) [77].

timescales in which the noise will make the computation useless can be proved in this model.

Mathematically, the most general evolution of a quantum system is described with a quantum channel [68, 81]. This class of maps can account, among others, for the evolution of an isolated system, its interaction with an external environment, or measurements performed on it. Some examples of relevant quantum channels to model the effect of noise on a qubit are the depolarizing channel of strength p, $T(\rho) = (1-p)\rho + p\mathbb{I}/2$, or the bit-flip channel $T(\rho) = (1-p)\rho + pX\rho X$. Some quantum channels T satisfy so-called data-processing inequalities,

$$D\left(T(\rho)||T(\sigma)\right) \leq (1-\alpha)D\left(\rho||\sigma\right), \tag{2.4.1}$$

for the quantum relative entropy $D(\rho||\sigma) = \operatorname{Tr}\rho(\log\rho - \log\sigma)$, and for any two states ρ and σ. Here $0 < \alpha \leq 1$ is a constant. The quantum relative entropy is a measure of closeness between quantum states[5]. These inequalities characterize channels that drive states to a fixed point. Under repeated actions of the noise channel, the relative entropy will behave as

$$D\left(T^k(\rho)||T^k(\sigma)\right) \leq (1-\alpha)^k D\left(\rho||\sigma\right). \tag{2.4.2}$$

Thus, under the condition $0 < \alpha \leq 1$, any state that goes through the repeated application of the channel will converge to the same state.

In the circuit case, where unitary gates are intertwined with the applications of the noise in every qubit, our study has to be generalized. Still, as we will explain in Section 6.B of Chapter 6, data processing inequalities can be proved for the channels that describe the evolution of the noisy quantum circuit. Let us consider a specific result to illustrate the power of this reasoning. As we have seen, the action of depolarizing noise for a single qubit with noise strength p is described by the channel

$$T(\rho) = (1-p)\rho + p\frac{\mathbb{I}}{2}. \tag{2.4.3}$$

Its physical interpretation is simple: with probability p, we lose the qubit's state and it is replaced by the maximally mixed state. In the opposite case, the qubit remains intact. When the qubits in the computer are subjected to this type of noise, regardless of the unitary gates applied, the state after k unitaries are applied in every qubit will satisfy

$$D(\rho_k||\mathbb{I}/2^N) \leq (1-p)^{2k} D(\rho_0||\mathbb{I}/2^N), \tag{2.4.4}$$

[5]It is not a proper distance in the mathematical sense, as it does not satisfy the triangle inequality. However, its value can be used to value the trace norm between states through the Quantum Pinsker inequality [81].

where ρ_k is the state after k steps, ρ_0 is the initial state and N is the number of qubits [82]. The relative entropy between a state and the maximally mixed state can be directly translated into a statement about the entropy of the state,

$$D(\rho||\mathbb{I}/2^N) = N - S(\rho). \tag{2.4.5}$$

N is the number of qubits in the system. Thus, assuming the state of the qubits is initially pure, we get that after k applications of the channel

$$S(\rho_k) \geq N\left(1 - (1-p)^{2k}\right). \tag{2.4.6}$$

The implications of this inequality are strong: algorithms whose quantum circuit depth scales with the size of the input will be too sensitive to noise when run in a NISQ device [83], as the output state of the circuit will be too close to the maximally mixed state. Even algorithms for quantum mechanical tasks, for instance, ground state preparation, will not improve over classical algorithms for depths that scale faster than logarithmically with the system size [83–85].

At the same time, results based on entropic or convergence to the fixed point of the channel miss part of the story. Many NISQ algorithms prepare quantum states variationally: a circuit that contains gates dependent on some parameters is run with the goal of tweaking the parameters to obtain the desired output. Thus, the algorithm's results will not only be affected by the impact of the noise but also the optimal circuit will not be known. Furthermore, the entropic bounds presented before apply both to circuits that generate lots of entanglement as well as to circuits that do not generate entanglement at all. It is expected that highly entangled computations will be more sensible to noise than unentangled ones.

In light of these considerations, it is desirable to have good tools to study and assess the impact of noise in quantum circuits, even tools that do not need to be run on a quantum computer to give results, or tools that can work in a scalable way. As we will see in Chapter 6, tensor networks offer a platform to study the impact of noise in a quantum computer rigorously.

Chapter 3

Collective Monte Carlo updates with Tensor Networks

This chapter is based on the publication
Miguel Frías-Pérez, Michael Mariën, David Pérez-García, Mari Carmen Bañuls, Sofyan Iblisdir
Collective Monte Carlo updates through tensor network renormalization
SciPost Phys. 14, 123 (2023)
Structure, text, and figures have been adapted for this thesis

3.1 Overview

In this chapter, we present and explore a novel connection between tensor networks and Monte Carlo methods. Our primary concern here will be sampling configurations representative of the Boltzmann distribution of classical nearest-neighbor Hamiltonians at finite temperature. To achieve this goal, we introduce a Tensor Network Metropolis-Hastings (TNMH) Markov chain [54, 56], where the asymmetric prior, i.e. the distribution from which the new candidate configuration is drawn at each step, is an approximation to the target distribution, obtained via an inexpensive tensor network renormalization contraction. Our approach does not oppose but genuinely combines TN and Markov Chain Monte Carlo ideas. In this way, it features concrete advantages with respect to each strategy. The salient properties of the scheme introduced here are the following.

(i) It is *universal*. That is, it works identically for all instances of a given model. This is in contrast to other Monte Carlo algorithms where a powerful prior choice can only be built by relying on a deep insight about the target distribution and thus has limited applicability beyond the model for which

it has specifically been tailored. That is, for instance, the case of Wolff's algorithm, which performs extremely well for ferromagnetic Ising models but rather poorly for antiferromagnets or frustrated instances. In turn, we will show that our method fares consistently well for a variety of models that are all very different from one another.

(ii) The scheme produces collective updates. That is, the state of each degree freedom of the considered system is susceptible to change at each Monte Carlo step. We have found that the computational effort scales mildly with increasing acceptance rates in a broad variety of instances. Presumably, as a consequence, we have found that the number of Monte Carlo steps necessary to reach convergence is between $\sim 10^1$ and $\sim 10^3$ shorter than those of other well-established Monte Carlo algorithms for several instances of two and three-dimensional models of the Ising type.

(iii) As compared to algorithms that purely rely on a tensor network renormalization of the partition function, the shift to sampling results in the substitution of systematic errors with statistical errors, since our TNMH scheme satisfies the classical sufficient conditions for convergence. Thus, modest tensor network contraction schemes, too inaccurate for a direct evaluation of a chosen observable, can be successfully used in our method, as they still enable collective updates with sufficiently high acceptance rates.

We have tested our Markov chain systematically in a variety of instances of the Ising model defined on finite square lattices: ferro- and antiferromagnetic, frustrated, disordered, in two and three spatial dimensions. One may anticipate that for systems with large, or even diverging, correlation length, our scheme will perform increasingly poorly if the bond dimension is fixed. Our findings are consistent with this expectation, with drops in acceptance rates actually signaling phase transitions. But we have also observed that for ferro- and antiferromagnetic instances, acceptance rates remain fairly high for systems of considerable size across a phase transition, even with a bond dimension as low as $D = 2$. Equilibration and decorrelation times in our TNMH scheme have been found to be systematically lower than for Metropolis and Wolff's algorithms. As expected, frustrated and spin-glass instances have turned out to be challenging, not only for fundamental complexity-theoretic reasons but also because their study is complicated by ill-conditioning issues [86]. However, even in such cases, and without any optimization of our renormalization procedure, we have observed that acceptance rates stay high enough to be usable down to temperatures that can be considered low by nowadays state-of-the-art standards.

The rest of this chapter is organized as follows. The new algorithm is described in Section 3.2 in general terms. In Section 3.3, we explore its performance for

two-dimensional models. In particular, for a broad variety of instances of Ising models, we explore the role of the bond dimension in the acceptance rates, also in relation to the presence of critical temperatures. We further analyze equilibration and autocorrelation times, and demonstrate how the method can be used to obtain physical observables and chart phase diagrams. In Section 3.4, we demonstrate how the algorithm is also useful for three-dimensional systems, and illustrate it for ferromagnetic and antiferromagnetic instances of the Ising model in cubic lattices of up to 16^3 sites. Section 3.5 is a discussion of situations where our findings could find further applications. An outlook is provided in Section 3.6.

3.2 Markov chains and tensor network renormalisation

In this section, we present a collective Monte Carlo update where, given a current configuration, tensor network renormalization is used to propose a candidate and to decide whether it should be accepted or not. For the sake of concreteness, and with a view to the example computations that will be considered in the next two sections, we will focus on the Ising model on a square lattice. Generalization to other nearest-neighbor interactions, such as the Potts model, is immediate.

We start by fixing some notation. A lattice will be denoted as $\Lambda = (V, E)$, where V stands for its vertices and E for its edges. We will focus on systems made of two-state particles (classical spins) residing on the vertices. That is, the sample space will be $\Omega = \{-1, +1\}^{|V|}$. A spin at location $j \in V$ will be denoted by σ_j.

The Ising Hamiltonian associated with a spin configuration ω is defined as

$$H(\omega) = -\sum_{i \in V} h_i \sigma_i - \sum_{\langle i,j \rangle \in E} J_{ij} \sigma_i \sigma_j. \tag{3.2.1}$$

Our aim is to sample according to the Boltzmann distribution

$$\pi^{(\beta)}(\omega) = e^{-\beta H(\omega)}/Z(\beta) \; \forall \omega \in \Omega, \tag{3.2.2}$$

where β denotes the inverse temperature, and

$$Z(\beta) = \sum_{\phi \in \Omega} e^{-\beta H(\phi)} \tag{3.2.3}$$

is the partition function.

A Markov chain is a sequence of configurations $\omega(0), \omega(1), \ldots$ where the probability that the t-th element of this sequence is in some state ω only depends on the state of the previous element $\omega(t-1)$, and on some random numbers. That is, a

Markov chain is an evolution with short memory. If the process used to determine the state at each time step satisfies some general requirements reminded below, $\lim_{t\to\infty}\omega(t)$ is a state drawn according to the target probability distribution (3.2.2). A very simple Markov chain is the celebrated Metropolis algorithm, where, at most, one spin is flipped at each time step.

A powerful class of Markov chain is that introduced by Hastings in [56]. In the context of Statistical Physics, this class can be described as follows. Given a current spin configuration ω, a candidate configuration ω' is proposed according to some prior distribution $g^{(\beta)}(\omega'|\omega)$, from which we are able to draw. This candidate is next accepted as the new current state with probability

$$P_{\text{acc}}(\omega\to\omega') = \min\{1, \frac{g^{(\beta)}(\omega|\omega')}{g^{(\beta)}(\omega'|\omega)} \times \frac{\pi^{(\beta)}(\omega')}{\pi^{(\beta)}(\omega)}\}. \tag{3.2.4}$$

As will be shown shortly, this acceptance rule allows to satisfy reversibility (a.k.a. detailed balance), Eq.(3.2.8), one of the conditions which, when met, guarantees convergence to the target probability distribution. If the prior $g^{(\beta)}$ is symmetric in its arguments, the acceptance probability (3.2.4) reduces to the celebrated Metropolis algorithm formula. But in some situations, the generalization proposed by Hastings allows encoding some information about the target distribution in the possibly asymmetric prior $g^{(\beta)}(\omega'|\omega)$, in a beneficial way. It can, for example, result in a boosted exploration of the sample space along the iterations of the Markov chain. The Swendsen-Wang and the Wolff cluster algorithms are examples of Metropolis-Hastings construction [57, 58]. Actually, an ideal prior is one where $g^{(\beta)}(\omega'|\omega) = \pi^{(\beta)}(\omega')$, that is, the prior that consists in direct sampling according to the target probability distribution $\pi^{(\beta)}$. Of course, for generic instances of the Ising model, such an ideal prior is unavailable. But an approximation $\widetilde{\pi}^{(\beta)}$ to this ideal prior might be good enough for Monte Carlo. We will be concerned with such approximations that can be constructed through tensor network renormalization.

Let $n = |V|$ represent the system size, and let $\{1, 2, \ldots, n\}$ denote a certain sequential labeling of the vertices (see, e.g., figure 3.A.1 in App. 3.A). Using Bayes formula, the Boltzmann distribution can be expressed as

$$\pi^{(\beta)}(\omega) = \pi_1^{(\beta)}(\sigma_1) \prod_{k=2}^{n} \pi_k^{(\beta)}(\sigma_k|\sigma_1\ldots\sigma_{k-1}), \tag{3.2.5}$$

where $\pi_1^{(\beta)}$ stands for the marginal distribution of the first spin, and $\pi_k^{(\beta)}(\cdot|\sigma_1\ldots\sigma_{k-1})$ denotes the conditional distribution for the kth spin when the spins 1 through $k-1$ are fixed to values $\sigma_1,\ldots\sigma_{k-1}$. The marginal distribution for the first spin $\pi_1^{(\beta)}(\sigma_1)$ can be expressed as the ratio of two partition functions: $\pi_1^{(\beta)}(\sigma_1) = Z(\beta|\sigma_1)/Z(\beta)$, where $Z(\beta|\sigma_1)$ represents the partition function for a system with

the same nearest-neighbor Hamiltonian as in $Z(\beta)$ but where the first spin has been fixed to the value σ_1. As mentioned in the introduction, the partition function (3.2.3) of any nearest-neighbor Hamiltonian can be expressed exactly as a tensor network (TN), whose bond dimension is equal to the number of states accessible by each local degree of freedom. For the Ising model, this number is equal to two. In general, neither $Z(\beta|\sigma_1)$ nor $Z(\beta)$ can be evaluated exactly. But a TN renormalisation scheme yields approximations $\widetilde{Z}(\beta|\sigma_1)$, and $\widetilde{Z}(\beta)$ for each of these quantities (see Appendix 3.A). With them, one can construct an approximation $\widetilde{\pi}_1^{(\beta)}(\sigma_1) = \widetilde{Z}(\beta|\sigma_1)/\widetilde{Z}(\beta)$ to the true marginal distribution for the first spin. This Bernoulli distribution is next sampled. Let s_1 be the outcome obtained. With this fixed value for the first spin, one can compute an approximation $\widetilde{Z}(\beta|s_1, \sigma_2)$ of the partition function for each value σ_2 for the second spin. These approximations are then used to construct an approximation $\widetilde{\pi}_2^{(\beta)}(\sigma_2|s_1) = \widetilde{Z}(\beta|s_1, \sigma_2)/\widetilde{Z}(\beta|s_1)$ to the distribution for the second spin, conditioned on the value s_1 for the first spin. This second Bernoulli distribution is then sampled. And so on. For all other sites $k > 2$, the conditional probability distribution $\pi_k^{(\beta)}(\sigma_k|s_1 \dots s_{k-1})$ can be expressed as the ratio of two TN contractions $Z(\beta|s_1 \dots s_{k-1}\sigma_k)/Z(\beta|s_1 \dots s_{k-1})$, and a TN renormalisation scheme provides approximations $\widetilde{Z}(\beta|s_1 \dots s_{k-1})$ and $\widetilde{Z}(\beta|s_1 \dots s_{k-1}\sigma_k)$ to $Z(\beta|s_1 \dots s_{k-1})$ and $Z(\beta|s_1 \dots s_{k-1}\sigma_k)$ respectively. These approximations are in turn used to compute an approximation $\widetilde{\pi}_k^{(\beta)}(\sigma_k|s_1 \dots s_{k-1})$ to $\pi_k^{(\beta)}(\sigma_k|s_1 \dots s_{k-1})$, which is sampled and yields an outcome s_k. Fig. 3.2.1 illustrates the first two steps of this sequential sampling. The configuration $(s_1, \dots, s_n)$ obtained after the whole lattice is swept will have been drawn with probability

$$\widetilde{\pi}^{(\beta)}(s_1, \dots, s_n) \equiv \widetilde{\pi}_1^{(\beta)}(s_1) \prod_{k=2}^{n} \widetilde{\pi}_k^{(\beta)}(s_k|s_1 \dots s_{k-1}), \tag{3.2.6}$$

which the identity (3.2.5) shows to be an approximation to $\pi^{(\beta)}(s_1, \dots, s_n)$.

We will be interested in schemes where the Metropolis-Hastings probability to select a candidate ω' reads:

$$g^{(\beta)}(\omega'|\omega) \equiv \widetilde{\pi}^{(\beta)}(\omega'). \tag{3.2.7}$$

As explained in Appendix 3.A, the approximate probability $\widetilde{\pi}^{(\beta)}(\omega)$ can be evaluated for any configuration ω, and the update rule (3.2.4) can be implemented. Our construction is summarized in Algorithm 1.

Properties of the TNMH Markov chain

(i) The construction is *universal* in the sense that it is independent of the magnetic fields and couplings that define the Ising instance being considered.

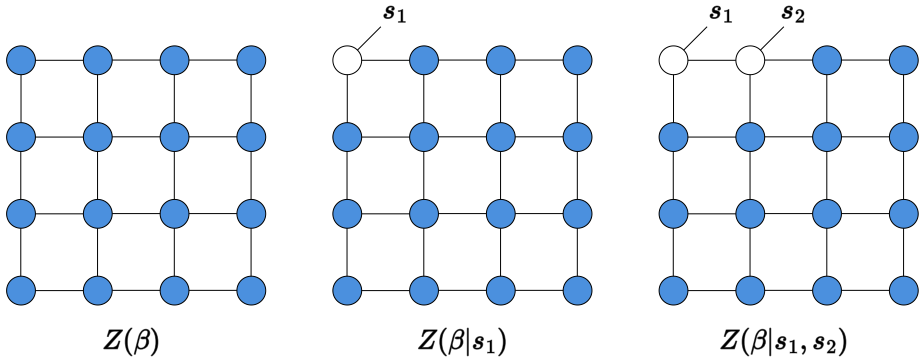


Figure 3.2.1: Pictorial illustration of the first two steps of the TNMH sequential sampling. White dots refer to sites where the spin value has been fixed.

Yet, it is *adaptive* in that the details of the Hamiltonian are taken into account when the tensors are constructed.

(ii) The constitution of the candidate is independent of the current configuration.

(iii) The update (3.2.7) is *collective and correlated*: in principle *all* spins of the system could be refreshed in a single Monte Carlo step, and the spin values proposed at different sites are conditioned by the correlations present in the tensor network. We believe this feature is the principal cause for the high acceptance rates and fast equilibration reported in the next section. Whereas a local update rule could have a hard time overcoming energy barriers, we expect our algorithm to be more capable of hopping between distant regions of the configuration space and escaping local minima in a single iteration of the Markov chain.

(iv) The transition matrix, i.e. the set of probabilities to transition from a configuration ω to a configuration ω', $\mathcal{T}(\omega \to \omega') = \widetilde{\pi}^{(\beta)}(\omega') \times P_{\text{acc}}(\omega \to \omega')$, is reversible (i.e. it satisfies detailed balance):

$$\pi^{(\beta)}(\omega)\, \mathcal{T}(\omega \to \omega') = \pi^{(\beta)}(\omega')\, \mathcal{T}(\omega' \to \omega). \tag{3.2.8}$$

That is, the expression

$$\pi^{(\beta)}(\omega)\widetilde{\pi}^{(\beta)}(\omega') \min\{1, \frac{\widetilde{\pi}^{(\beta)}(\omega)}{\widetilde{\pi}^{(\beta)}(\omega')} \times \frac{\pi^{(\beta)}(\omega')}{\pi^{(\beta)}(\omega)}\}$$

is manifestly symmetric in ω and ω'.

Furthermore, when numerical errors are small enough that all the conditioned partition functions $\widetilde{Z}(\beta|\sigma_1 \dots \sigma_k)$ are strictly positive (see Appendix 3.A), the Markov chain is also irreducible:

$$\mathcal{T}(\omega \to \omega') > 0 \quad \forall \omega, \omega' \in \Omega.$$

Thus, even if the distributions $\{\pi_k^{(\beta)} : k \in V\}$ turned out to be poorly approximated by the TN renormalization scheme used, it is still possible to guarantee that the Markov chain will eventually converge to the target probability distribution [87]. This last point will be illustrated with three-dimensional Ising models.

Algorithm 1 TNMH Markov chain

1: Compute the tensors associated with the distribution (3.2.2).
2: Set $t = 0$, and draw some initial configuration $\omega(0)$ according to any distribution over Ω.
3: If $t > t_{\max}$ go to 8.
4: Use the tensor network to draw a candidate configuration ω' according to Eq.(3.2.7).
5: Evaluate the probabilities $\widetilde{\pi}^{(\beta)}(\omega(t))$ and $\widetilde{\pi}^{(\beta)}(\omega')$.
6: Accept the change $\omega(t) \leftarrow \omega'$ with probability $\min\{1, \frac{\widetilde{\pi}^{(\beta)}(\omega|\omega')}{\widetilde{\pi}^{(\beta)}(\omega'|\omega)} \times \frac{\pi^{(\beta)}(\omega')}{\pi^{(\beta)}(\omega)}\}$.
7: $t \leftarrow t + 1$. Go to 3.
8: End.

Before turning to applications of Algorithm 1, there are two points we would like to stress. (i) Although the construction of a candidate at each iteration of the Markov chain is actually independent of its current configuration, the decision to accept this candidate *does* depend on the current state. That is, the transition probability $\mathcal{T}(\omega \to \omega')$ is *not* independent of ω. This can be seen explicitly:

$$\mathcal{T}(\omega \to \omega') = \widetilde{\pi}^{(\beta)}(\omega') \min\{1, \frac{\widetilde{\pi}^{(\beta)}(\omega)}{\widetilde{\pi}^{(\beta)}(\omega')} \times \frac{\pi^{(\beta)}(\omega')}{\pi^{(\beta)}(\omega)}\}. \tag{3.2.9}$$

Notice that a similar sampling scheme, based on Bayes' chain rule, has been proposed to directly sample an approximation to the Gibbs distribution [88]. In contrast, TNMH is not an approximate Gibbs sampler: the decision to accept or reject the candidate, depending on the Metropolis-Hastings ratio, marks an essential difference. (ii) Even though the tensor network contractions used in TNMH are (inevitably) approximate, the reversibility condition is *exactly* satisfied. This ensures the asymptotic convergence to the Gibbs distribution.

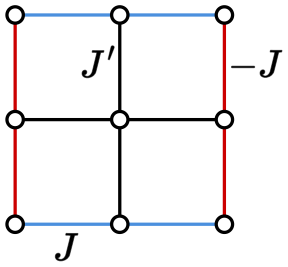


Figure 3.3.1: Distribution of the couplings on the $J' - J$ model. Black lines indicate bonds with a coupling of J', while red and blue bonds have couplings $-J$ and J, respectively.

3.3 Two-dimensional Ising models

In order to assess the potential of the construction presented in the previous section, we have run tests on instances of the two-dimensional Ising models chosen to cover a broad range of cases ($L \times L$ square lattice):

- Ferromagnetic : $J_{ij} = 1 \; \forall \; \langle i, j \rangle, \; h_i = 0 \; \forall i$.
- Antiferromagnetic : $J_{ij} = -1 \; \forall \; \langle i, j \rangle$, h_i constant across the whole lattice.
- $J' - J$ model: In this model $h_i = 0 \; \forall i$ and couplings alternate between even and odd rows or columns (see Fig.3.3.1):

$$\begin{aligned} &J_{2j-1,k} = J', \; J_{2j,k} = J, \\ &J_{k,2j-1} = J', \; J_{k,2j} = -J, \quad j = 1, \ldots, L/2 \end{aligned}$$

  The point $J = J'$, known as the fully frustrated square lattice Ising model (or Villain model), is characterized by extensive ground state degeneracy and maximal frustration.

- Edwards-Anderson spin glass: this disordered model is such that $h_i = 0 \; \forall i$, and J_{ij} are random couplings sampled from a Gaussian distribution with zero mean and unit variance [89].

We will be interested in the following observables: the energy per spin,

$$\varepsilon = \frac{1}{|V| Z(\beta)} \sum_{\omega \in \Omega} H(\omega) \; e^{-\beta H(\omega)},$$

the magnetization density,

$$m = \frac{1}{|V| Z(\beta)} \sum_{\omega \in \Omega} \left| \sum_{i \in V} \sigma_i \right| e^{-\beta H(\omega)},$$

the staggered magnetization density [1],

$$m_s = \frac{1}{|V|Z(\beta)} \sum_{\omega\in\Omega} \left| \sum_{i\in V} \mathrm{sign}(i)\ \sigma_i \right| e^{-\beta H(\omega)},$$

where $\mathrm{sign}(i)$ is equal to ± 1 in a checkerboard manner. Finally, we will consider also the magnetic susceptibility, defined for a system with a uniform magnetic field as

$$\chi = \partial m / \partial h.$$

Unless stated otherwise, we will be considering open boundary conditions.

Role of the bond dimension

A crucial ingredient of the algorithm described in the previous section is the substitution of exact partition functions $Z(\beta|\sigma_1 \dots \sigma_{k-1})$ with approximations $\widetilde{Z}(\beta|\sigma_1 \dots \sigma_{k-1})$ obtained by tensor network renormalization. Amongst all available methods for this renormalization, we have used the matrix product state (MPS) renormalization scheme described in [90] (see also Appendix 3.A). It is a choice of simplicity, which turned out to be sufficient for our purposes. We, however, would like to stress that the analogous TNMH Markov chain can be defined using any other contraction scheme, and some might yield better results than those presented here. In MPS renormalization, both the accuracy of the approximation and the computational effort increase with an integer parameter, the bond dimension, commonly denoted D. We thus expect the total variation distance between the target and the prior distribution,

$$\|\pi^{(\beta)} - \widetilde{\pi}^{(\beta)}\|_{\mathrm{TV}} = \frac{1}{2} \sum_{\omega\in\Omega} |\pi^{(\beta)}(\omega) - \widetilde{\pi}^{(\beta)}(\omega)|, \tag{3.3.1}$$

to decrease with increasing values of D. As a result, Monte Carlo rejection rates should decrease as the bond dimension grows large.

To characterize the behavior of our method, we have explored the interplay between the bond dimension, the temperature, and the rejection rate for the four different models mentioned above (Fig. 3.3.2). In all cases, we have verified that the rejection rate decreases with increasing bond dimension, and even modest values of the bond dimension may yield virtually rejection-free updates. At the same time, for a fixed D, rejection rates increase in the vicinity of critical points.

[1] When the external magnetic field is uniformly naught, averaging over Monte Carlo samples would result in zero (staggered) magnetization even at temperatures where the system is known to exhibit a finite spontaneous magnetization. The absolute values appearing in our definition of the (staggered) magnetization are meant to counter this artifact.

This can be understood considering that, presumably, the distance $\|\pi^{(\beta)} - \widetilde{\pi}^{(\beta)}\|_{\mathrm{TV}}$ for fixed D will increase with the true correlation length.

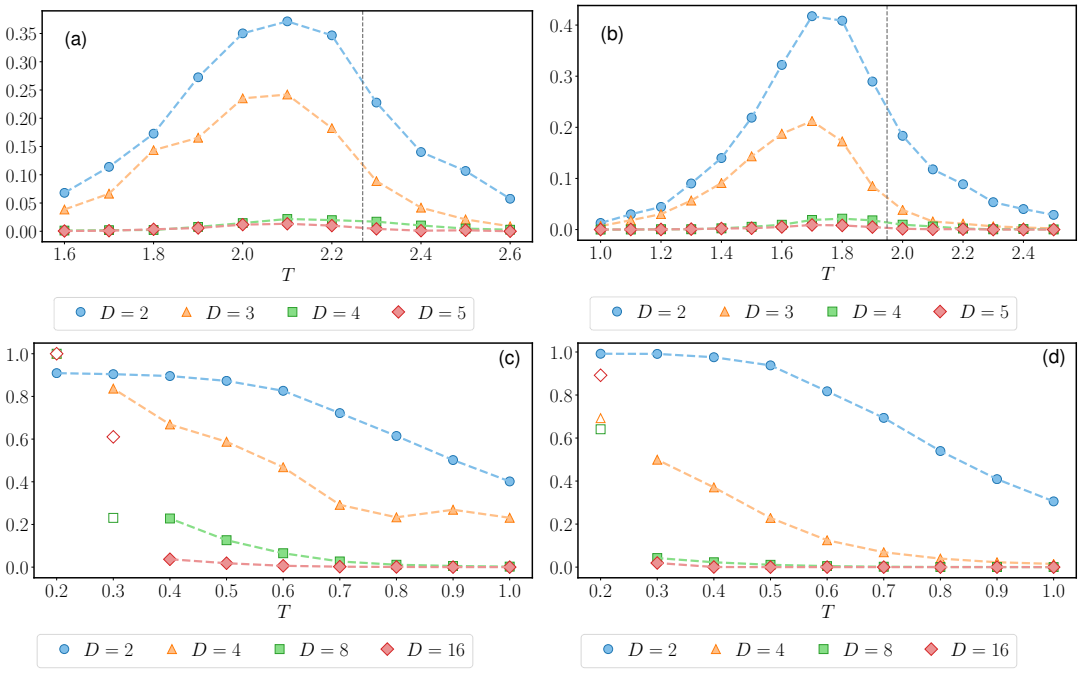


Figure 3.3.2: TNMH rejection rates as a function of the temperature for four different instances of the Ising model at a fixed system size: (a) Ferromagnetic, (b) Antiferromagnetic with a constant magnetic field $h = 2$, (c) Fully frustrated, (d) Gaussian spin glass. The geometry is that of a 32×32 square lattice with open boundary conditions in all four cases. For each model, the rejection rate was obtained by averaging over 40 independent chains, each run for a model-dependent number of steps. For the fully frustrated and spin glass cases, the points at temperatures where we believe our method starts to suffer from ill-conditioning issues are indicated by empty markers. For the ferro- and antiferromagnetic instances, the vertical lines indicate criticality in the thermodynamic limit.

Fig. 3.3.2a demonstrates these features for the ferromagnetic case. This model exhibits, in the limit of large system sizes, a second-order phase transition at $T_c = 2/\log(1+\sqrt{2}) \approx 2.269$ from a magnetically ordered to a paramagnetic phase [91]. Still, for the system size considered, 32×32, rejection rates remain remarkably low (below 0.4) across the whole temperature range that surrounds the critical temperature. Actually, even a bond dimension as low as $D = 2$ appears to already be sufficient to achieve our goal of producing collective updates frequently. More details regarding the vicinity of the critical point are provided in Fig. 3.3.3a, where

we have plotted the rejection rate as a function of the system size for different bond dimensions. Even for systems as large as 256×256, acceptance rates of about 0.4 can be obtained using only a bond dimension $D = 4$. As can be appreciated from the inset of this figure, our data suggest that the bond dimension only needs to grow *logarithmically* with the system size in order to maintain the acceptance rate above a threshold value. Our observations for the antiferromagnetic case (Fig. 3.3.2b) are similar.

For the fully frustrated case of the $J' - J$ model (Fig. 3.3.2c), we obtain lower acceptance rates, as compared to the two previous cases, but still high enough that the Markov chain is usable down to $T = O(10^{-1})$. As expected, the rejection rate increases when approaching the $T = 0$ critical point. Still, the minimal cost curve $D = 2$ is sufficient to obtain decent acceptance rates down to at least $T = 0.2$, and increasing the bond dimension again suppresses rejection events. At very low temperatures, acceptance rates drop dramatically and numerical instabilities typical of frustrated systems pointed out in [86] set in. Some strategies exist to mitigate these effects, but their discussion is beyond the scope of the present work. Again, we have looked at the rejection rate as a function of the system size for different bond dimensions (see Fig. 3.3.3b). Even though this instance is more challenging, the example in the figure demonstrates that at $T = 0.4$ perfectly usable acceptance rates of about 0.2 or higher can be obtained for systems of size up to 128×128 using a bond dimension, $D = 6$, for which computations are not too demanding. Fig. 3.3.4 provides additional data regarding the $J' - J$ model, beyond the fully frustrated point $J = J'$. It is remarkable that the observed maxima of rejection rates are consistent with the predicted critical lines of this model.

Our findings for the Edwards-Anderson spin glass, Fig. 3.3.2d, are qualitatively similar to those for the fully frustrated case, presumably because this spin glass is also critical at $T = 0$ [89].

Improved approximations of the contraction will generally result in a higher acceptance rate. But actually, as far as this acceptance rate does not vanish and scales well with the system size, the TNMH scheme should be applicable.

Equilibration and decorrelation

Equilibration and auto-correlation times are the two crucial time scales in Monte Carlo simulations. The former controls the number of steps needed by the Markov chain to decouple from the initial distribution (that is, the distribution from which the first configuration of the chain is sampled) and reach the desired equilibrium distribution. The latter determines the minimal time between two consecutive sample extractions in order to guarantee statistical independence. Since these times are typically difficult to bound, let alone calculate, rigorously, heuristic diagnostics are commonly used to estimate them. We have used two such heuristics to provide

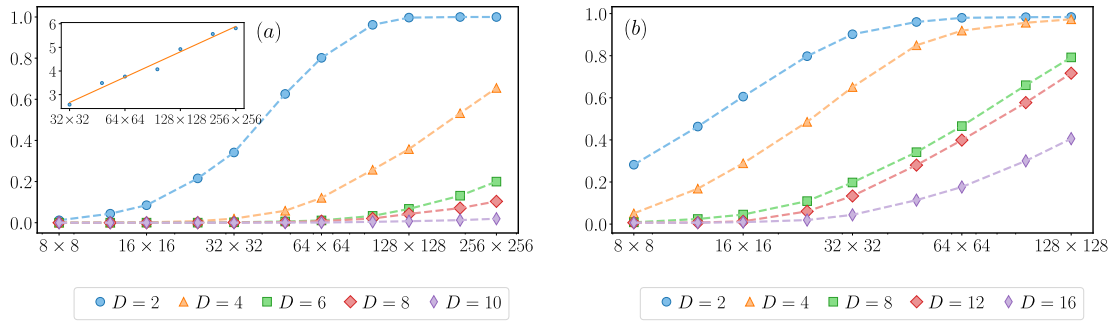


Figure 3.3.3: (a) TNMH rejection rates near the ferromagnetic Ising phase transition ($T \simeq 2.27$) as a function of the system size for different bond dimensions. Dashed lines are simply a guide to the eye. Inset: Bond dimension needed to maintain a fixed rejection rate (in this case 0.25, although the behaviour seems to be independent of the value chosen) as a function of the system size. The fit shows that the increase in the bond dimension seems to be only logarithmic. (b) TNMH rejection rates for the fully frustrated Ising model at $T = 0.4$ as a function of the system size for different bond dimensions.

evidence that these two time scales are relatively short for the TNMH scheme of Section 3.2.

Sophisticated equilibration diagnostics can be devised for specific problems. In particular, for the Edwards-Anderson spin glass, we have run the following test, discussed in [61, 93]. Let $\langle X \rangle$ stand for the thermal average of an observable X, and $[x]_{\text{av}}$ denote the disorder average of a quantity x that might depend on the coupling constants $\{J_{ij}\}$. We have considered the disorder averaged energy.

$$[\langle H \rangle]_{\text{av}} = -\int \prod_{\langle ab \rangle} \frac{dJ_{ab}}{\sqrt{2\pi}} e^{-J_{ab}^2/2} \sum_{\langle ij \rangle} J_{ij} \langle \sigma_i \sigma_j \rangle. \tag{3.3.2}$$

At equilibrium, integrating by parts allows us to prove that

$$\Delta \equiv \frac{1}{|V|}[\langle H \rangle]_{\text{av}} + \frac{1}{|V|}\beta\Big(|E| - \sum_{\langle ij \rangle}[\langle \sigma_i \sigma_j \rangle^2]_{\text{av}}\Big) = 0, \tag{3.3.3}$$

where $\sum_{\langle ij \rangle}[\langle \sigma_i \sigma_j \rangle^2]_{\text{av}}$ is a quantity known as the link overlap. Starting configurations of Markov chains, drawn according to an easy distribution, typically have high energy and small link overlap. As a result, Δ typically has a non-zero value when the Markov chain is started. As Monte Carlo steps are taken, this value decreases in magnitude. It is a common heuristic to decide equilibration has occurred when Δ is below a given threshold value.

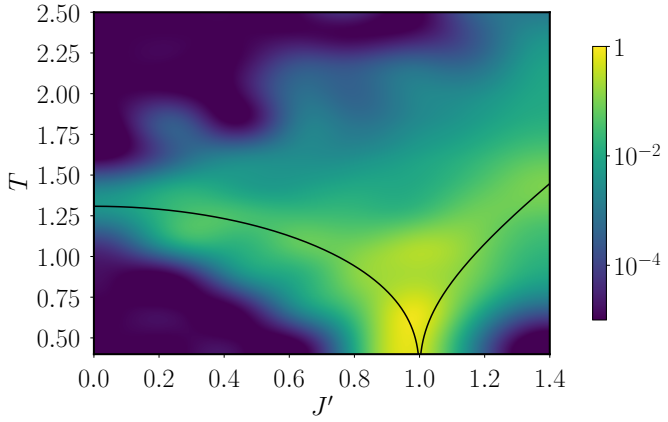


Figure 3.3.4: TNMH rejection rate for the $J' - J$ model as a function of the temperature and the value of one of the couplings (the other has been set to unity). Computations made with a bond dimension $D = 4$. Rejection rates were obtained as averages over 40 independent Markov chains, each run for 200 steps. Lattice dimensions: 32×32. The phase separatrix predicted in [92] is shown in black.

We have implemented this heuristic test of convergence with 10^4 disorder realizations using the TNMH Markov chain. We have observed that after a few time steps, Δ drops drastically from its initial value before stagnating at a small but finite value. For the vast majority of disorder realizations, the Markov chain behaved well. That is, it shows frequent jumps between different configurations from one time step to the next. However, for some disorder realizations, a few Markov chains (typically less than a sixth of the chains we simulate for a given disorder realization) remained stuck in their original configurations. This inertia is what causes Δ to stagnate. Increasing the bond dimension did not help. Actually, we believe the issue is rather related to an ill-conditioning of the tensor network contraction, induced by frustration, an effect previously reported in Ref. [86]. As a result, the approximate probability weights can be off their true value by orders of magnitude [2]. As can be seen from Eq.(3.2.4), this mismatch affects the acceptance rates. That is when the Markov chain hits a configuration ω such that $\widetilde{\pi}^{(\beta)}(\omega)/\pi^{(\beta)}(\omega) \ll 1$, it may remain stuck for a long time, as we have observed.

Various strategies are possible to mitigate the effect of ill-conditioning. One is to work with greater machine precision, using, for example, the techniques described in Ref. [94]. Another is to consider a variation of the TNMH Algorithm 1. A very simple variation consists of interspersing spin flip Metropolis sweeps

[2]Interestingly, it is not clear that such configurations actually correspond to local minima or maxima of the energy.

Δ	0.25	0.15	0.05	0.025
Metropolis	$> 10^5$	$> 10^5$	$> 10^5$	$> 10^5$
PT	$> 10^5$	$> 10^5$	$> 10^5$	$> 10^5$
PT + ICM	$1.1 \cdot 10^3$	$1.6 \cdot 10^3$	$3.2 \cdot 10^3$	$4.3 \cdot 10^3$
TNMH	$> 10^2$	$> 10^2$	$> 10^2$	$> 10^2$
TNMH + Metropolis	3	3	5	5

Table 3.3.1: First row: target value of Δ, as defined by Eq.(3.3.3). Next rows: each entry represents the number of lattice sweeps necessary to decrease Δ below the value indicated in the same column for the given algorithm. An entry of the form '$> a$' indicates that more than a iterations are needed to reach the desired value of Δ. The system considered is the same as in Fig. 3.3.5.

in between TNMH moves. We have tested this possibility. Our findings are reported in Fig. 3.3.5b and in Table 3.3.1, where we compare our times with state-of-the-art methods: parallel tempering (PT) and parallel tempering combined with isoenergetic cluster moves (PT + ICM) [61, 93]. This comparison clearly shows that the combination of TNMH with single flip MC sweeps allows us to outperform these methods by orders of magnitude. This result is interesting: whereas both TNMH and the Metropolis algorithm show poor performance individually (for different reasons; ill-conditioning in the case of TNMH, locality in the case of single flip updates), their alternating use is drastically more efficient than either of them.

So far, we have counted equilibration times in steps of the Markov chain, which has conceptual relevance. From a practical point of view though, it is also interesting to know how the TNMH compares to other methods when looking at program execution times. To make such a comparison fairly is a delicate issue because we have not sought to optimize our code at all: a detailed comparison with, e.g., Metropolis sweeps, which is simpler to optimize, is a project in itself. We can, however, provide *indicative* times related to the data presented in Table 3.3.1. With our setup, the times to get to $\Delta < 0.025$ are $> 3.93 \times 10^7$ sec, 1.76×10^6 sec, and 6.75×10^4 sec for the parallel tempering method, the parallel tempering method supplemented with isoenergetic cluster moves, and TNMH respectively. We believe that these estimates credibly signal the *practical* potential of the TNMH Markov chain introduced here.

We next move to autocorrelation times, that is, after equilibration is reached, the time needed between two sample extractions to guarantee (sufficient) independence. Given an observable X, we study the time correlation function

$$C_X(t) = \frac{\langle X(t_0)X(t_0+t)\rangle - \langle X(t_0)\rangle\langle X(t_0+t)\rangle}{\langle X^2(t_0)\rangle - \langle X(t_0)\rangle^2}, \tag{3.3.4}$$

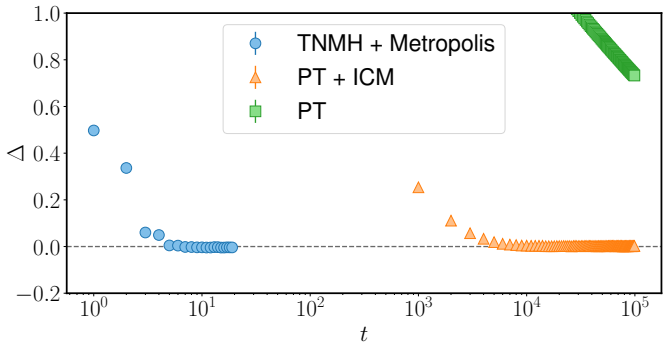


Figure 3.3.5: Difference between the (disordered averaged) energy per spin computed from the Hamiltonian and computed from the link overlap, Eq. (3.3.3), as a function of the number of iterations for three algorithms: TNMH + Metropolis (blue), (PT + ICM) (orange) and PT (green). Bond dimension for the TNMH moves: $D = 16$. Lattice dimensions: 32×32 (open boundary conditions). The symbol t represents the number of iterations for TNMH and the number of lattice sweeps for PT and PT+ICM. The error bars, smaller than the symbols, have been computed by estimating the variance of the disorder. The number of Markov chains used for the thermal average is 30 in all cases, the number of different instances used for the disordered average is 10^4, and the temperature has been set to $T = 0.212$.

where t_0 is assumed larger than the equilibration time. As discussed in [52], at large t, we expect $C_X(t)$ to decay exponentially, with a time scale set by the decorrelation time. As the exponential tail has a fixed amount of noise, controlled by the number of samples, a useful measure to determine that time scale is the integrated correlation time

$$\tau_X^{\text{int}}(t) \equiv 1 + 2\sum_{t'=1}^{t} C_X(t'). \tag{3.3.5}$$

It can also be shown that it is approximately the factor that enhances the variance when averaging over samples that are not sufficiently decorrelated [52]. The two quantities are plotted in Fig. 3.3.8a for the magnetization of a Fully Frustrated Ising model on a 32×32 lattice at $T = 1$. The motivation for choosing this observable is that often the energy is a poor choice to measure the decorrelation of samples in a Markov chain. At low temperatures, it is impossible to distinguish global changes in a configuration of a Markov chain from local motion around a local minima just by tracking the energy. In this particular case, choosing an observable that breaks the global spin flip symmetry of the model in consideration allows us to assess the ergodicity of the scheme, since for any configuration with energy E and

magnetization m, there exists another with magnetization $-m$ and same energy. The data shown in Fig. 3.3.8a shows that TNMH outperforms a local algorithm by almost two orders of magnitude. Furthermore, we expect that the difference in performance can only increase as the temperatures are lowered or the system size is increased.

To further illustrate sample-to-sample decorrelation in our algorithm, we have considered the fully frustrated Ising model and represented in Fig. 3.3.8b snapshots at different times for TNMH and for Metropolis sweeps starting from the same configuration. The difference between both techniques is striking: while the configurations appearing in our technique seem to bear no resemblance to one another from one acceptance to the next, memory of the initial configuration can still be appreciated visually after ten Metropolis sweeps.

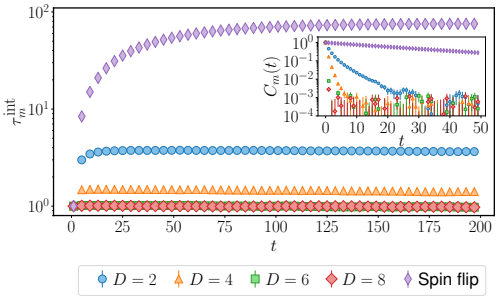


Figure 3.3.6: (a)

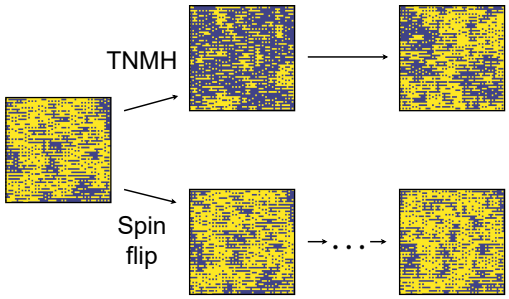


Figure 3.3.7: (b)

Figure 3.3.8: (a) Integrated correlation time (3.3.5) and (inset) decay of the autocorrelation function (3.3.4) of the magnetization as a function of time, for different bond dimensions on a Fully Frustrated Ising model on a 32×32 lattice with open boundary conditions at $T = 1$. The error bars have been computed by estimating the variance of the observables using the same samples. (b) Snapshots of the evolution of two different Markov chains starting from the same configuration, for the fully frustrated Ising model on a 64×64 lattice at $T = 0.5$. The top plots display the configuration obtained after three consecutive steps of the TNMH method (with bond dimension $D = 8$), while those below show configurations after Metropolis sweeps at times $t = 0, 1, 10$.

Observables

We now turn to the estimation of observables from the samples output by the TNMH Markov chain. We have focused on the ferromagnetic case and the antiferromagnetic one with an external field. The absolute value of the magnetisation for the ferromagnetic case is plotted on Fig. 3.3.11a (inset) and is in good agreement

with the theory [91]. Fig. 3.3.11a also shows estimates for the fourth order Binder cumulant [51], $g = (3 - \langle m^4 \rangle / \langle m^2 \rangle^2)/2$. One can appreciate that the phase transition point is correctly signaled by the locus where all data sets meet, as expected. In Fig. 3.3.11b, we have represented the staggered susceptibility as a function of the temperature and the external magnetic field for an antiferromagnet. Our findings seem to be in good agreement with previous studies of this model [95, 96]. When the external field is naught, the ferromagnetic phase transition around $T_c = 2/\ln(1+\sqrt{2})$ is recovered, as expected, since for a square lattice, a local change of variables allows mapping between antiferro and ferromagnetic instances of the Ising model. As the field increases, the temperature at which the phase transition takes place decreases. The intuition for this fact is as follows: at $h = 0$ and below the critical temperature one has antiferromagnetic order. In the large h limit, at the same temperature, all spins would align with the external field and one would have ferromagnetic order. Thus, some phase boundary must be encountered when going from one to the other.

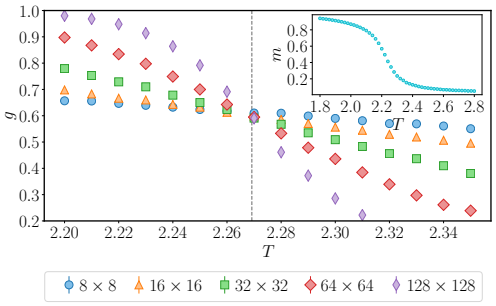


Figure 3.3.9: (a)

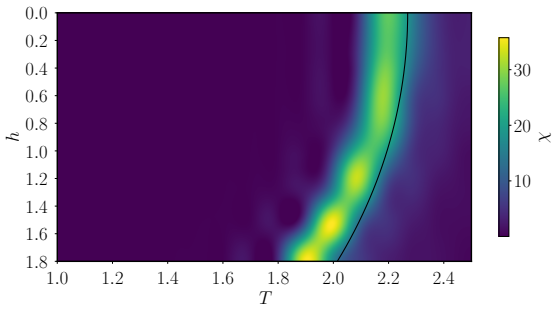


Figure 3.3.10: (b)

Figure 3.3.11: (a) Binder ratio as a function of temperature for the two-dimensional ferromagnetic Ising model obtained through TNMH with a bond dimension $D = 6$. The data approximately cross at one point, signaling a phase transition, in great concordance with the theoretical result $T \approx 2.269$. Inset: Magnetisation of the ferromagnetic Ising model on a square 64×64 lattice with open boundary conditions. The error bars have been computed via a jackknife analysis for the Binder cumulant and by estimating the variance for the magnetization, and are smaller than the symbols. (b) Susceptibility χ for an antiferromagnetic Ising model with an external field, on a 64×64 square lattice obtained through TNMH with a bond dimension $D = 6$. Black: theoretical prediction of the critical line in the thermodynamical limit.

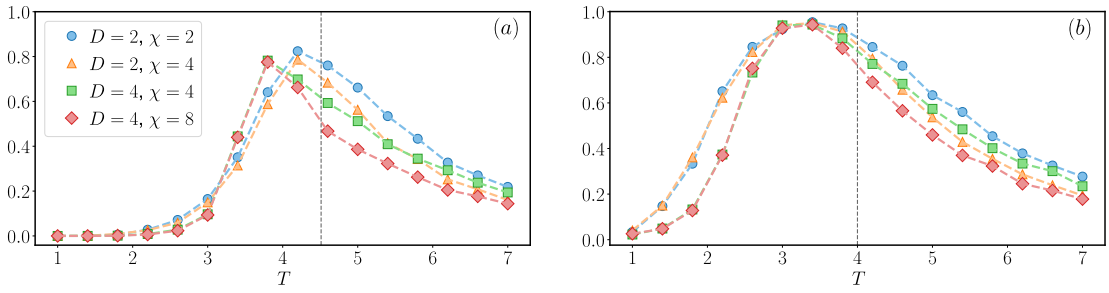


Figure 3.4.1: TNMH rejection rates for the three-dimensional Ising model as a function of the temperature. Plot (a) corresponds to a uniform ferromagnet and (b) to a uniform antiferromagnet in a field $h = 3$. D denotes the PEPS bond dimension, while χ stands for the boundary bond dimension used when compressing the PEPS associated with a plane of the lattice. A lattice of size $16 \times 16 \times 16$ was used, with open boundary conditions, and for each bond dimension and temperature, 50 chains were run for 150 steps each. The critical temperature is $T_c \approx 4.512$ [100] for the ferromagnetic Ising model, and $T_c \approx 4$ [101] for the antiferromagnetic with this field (We attribute the offset with respect to this value to finite size effects.).

3.4 Three-dimensional Ising models

Just as for planar systems, the partition function of a three-dimensional Ising model can be expressed as a tensor network. As a consequence, our TNMH algorithm immediately extends to three dimensions. The approximate contraction of a three-dimensional TN is, however, a more demanding problem than its two-dimensional analog. Still, TN renormalization schemes can be applied to find an approximation to the contraction [97–99]. We have chosen an unsophisticated renormalization scheme involving projected entangled pair states (PEPS). Two cutoff parameters now govern the effort put in the TNMH for this implementation: a boundary PEPS bond dimension D, and a boundary MPS bond dimension χ (see Appendix 3.A for details). We have considered two instances of the Ising model: ferromagnetic, and antiferromagnetic with an external magnetic field. The upshot is that our TNMH performs very well, even with rather low values for D and χ.

Fig. 3.4.1 shows the interplay between the two parameters D and χ, the temperature, and the rejection rate. Again, the peak in the rejection rate signals the presence of a critical point (displaced due to finite-size effects). As this critical point is approached, rejection rates increase much faster than in two dimensions, and putting in more computational effort by increasing D and χ now produces milder drops in rejection rates. We attribute this situation to an increase in correlations in the system due to a higher coordination number for each spin. Still, these

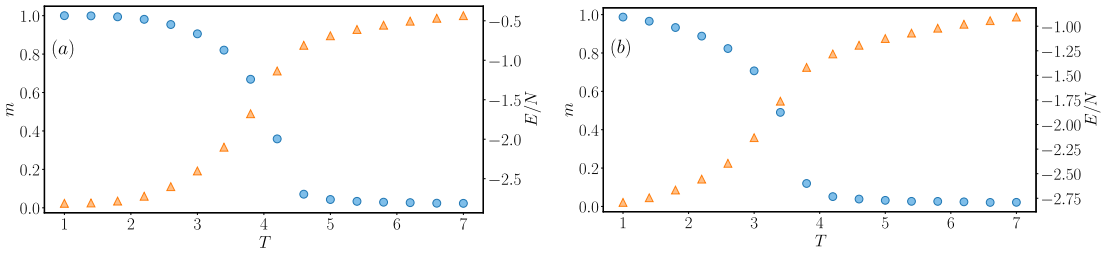


Figure 3.4.2: Magnetisation (blue) and energy (orange) per spin of the ferromagnetic Ising model (a) and staggered magnetization and energy per spin of the antiferromagnetic Ising model in an external field (b) on a cubic $16 \times 16 \times 16$ lattice with open boundary conditions. The error bars are smaller than the symbols. The largest bond dimensions used to obtain the curves were $D = 4, \chi = 8$.

preliminary results are very encouraging, since using a non-optimized contraction scheme, and modest values for the parameters D and χ, usable acceptance rates (> 0.12 and > 0.05 for the ferro- and antiferromagnetic case respectively) have been found across the whole temperature range considered, for systems as large as $16^3 = 4096$ spins.

Since observables can also be expressed as a TN, it is possible to estimate them using a direct contraction [90], and one might then wonder if the sampling procedure, which in itself requires a TN contraction, provides an advantage with respect to such a direct calculation. But, while the TNMH algorithm can succeed with a very undemanding approximate TN contraction (i.e. using very low bond dimensions), achieving a result of comparable quality by direct contraction generally requires more computational effort. To make this point more concrete, we have compared the value of the average energy in the three-dimensional ferromagnetic case, as obtained with the TNMH scheme and with direct TN contractions with different values of the bond dimensions (Fig. 3.4.3). We observe that, at temperatures where the direct contraction with up to $(D, \chi) = (8, 16)$ was not sufficient to obtain an accurate estimate of the energy, the TNMH with $(D, \chi) = (2, 2)$ was successful, since it produced decent acceptance rates, and eventually provided good samples thanks to irreducibility and reversibility.

3.5 The XY model

In this section, we describe how to extend our scheme to models that have degrees of freedom that are not discrete. For the sake of concreteness, we focus on a simple instance, the XY model in the square lattice, and explicitly describe how to write

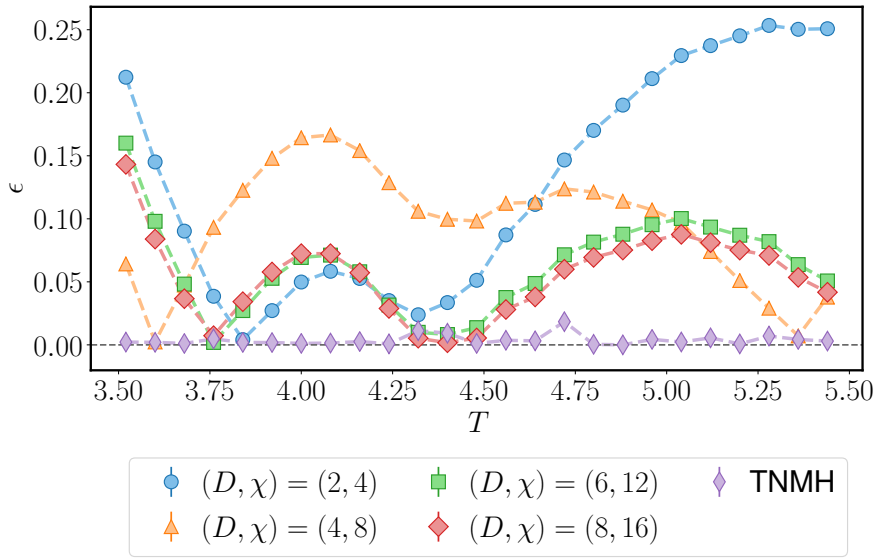


Figure 3.4.3: Relative error ϵ in the average energy per spin computed via different techniques for different temperatures (near the phase transition) for the three-dimensional ferromagnetic Ising model. The reference values are obtained using Wolff's algorithm, purple rhombi are obtained using samples from our algorithm with $(D, \chi) = (2, 2)$, and the other results are obtained taking derivatives of the logarithm of the approximate contraction of the TN representing the partition function.

its partition function as a tensor network with finite bond dimension. The reader interested in how this construction works for other models can consult the reference [102].

In the absence of a vector potential, the XY model describes a lattice of planar spins, interacting through the Hamiltonian

$$H_{XY} = -\sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j), \tag{3.5.1}$$

where the local variables are the angles $\{0 \leq \theta_i < 2\pi : i \in V\}$. Although these variables are continuous, this model can be mapped into a system that allows the use of a variation of the sampling method used for the Ising model. First, a duality transformation establishes an equivalence between (3.5.1) and a system of integer variables residing on the (oriented) links of the lattice involved in four-body

interactions [103–106]. That is, the partition function takes the form

$$Z(\beta) = \lim_{N\to\infty} \prod_{l\in E} \left(\sum_{n_l=-N}^{N} I_{n_l}(\beta) \right) \prod_{i\in V} F^{n_3^{(i)},n_4^{(i)}}_{n_1^{(i)},n_2^{(i)}}, \tag{3.5.2}$$

where $n_1^{(i)}, n_2^{(i)}, n_3^{(i)}, n_4^{(i)}$ are the values for the links meeting at site i. $I_{n_l}(\beta)$ are the modified Bessel functions of the first kind, and

$$F^{n_3,n_4}_{n_1,n_2} = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{i\theta(n_1+n_2-n_3-n_4)} = \delta_K(n_1+n_2-n_3-n_4), \tag{3.5.3}$$

where δ_{K} denotes the Kronecker delta function.

At fixed β, $I_{n_l}(\beta)$ decays fast and truncating the sum in Eq. (3.5.2) is a sensible approximation. The partition function of the XY model can thus be approximated by a tensor network where the degree of freedom at each bond takes value in a finite set. In the language of Appendix 3.A, the tensor at each site i would now be

$$A^{(i)}_{n_2n_4n_1n_3} = \left(\prod_{k=1}^{4} I_{n_k}(\beta) \right)^{1/2} F^{n_3,n_4}_{n_1,n_2}, \tag{3.5.4}$$

and the contraction of the TN made up of these tensors would give an approximation $\widetilde{Z}(\beta)$ to the partition function $Z(\beta)$. Similarly, the marginal probability density of the spin at a site i, $\widetilde{\pi}^{(\beta)}(\theta_i)$, can be approximated by replacing the tensor at site i with

$$A^{(i)}_{n_2n_4n_1n_3}(\theta_i) = \left(\prod_{k=1}^{4} I_{n_k}(\beta) \right)^{1/2} \frac{e^{i\theta(n_1+n_2-n_3-n_4)}}{2\pi}, \tag{3.5.5}$$

and normalizing the contraction to the approximate partition function previously obtained. Using renormalization to approximately contract tensor networks, and the inverse sampling method, a candidate configuration $\omega' = \{\theta'_i : i \in V\}$ can be drawn and accepted or rejected, as we did for Ising models with Algorithm 1. A vector potential could be included [104, 105], and other continuous variable systems admit a similar construction [107].

On top of the bond dimension used for the renormalization, the number of terms kept in the series expansion of the transfer matrix in Eq. (3.5.2) is another parameter that governs the accuracy of the contraction. As for the 3D Ising model discussed above, a tensor network with a low value for this parameter may be accurate enough to sample from and propose moves for a Markov chain, but not precise enough to compute the observables with a single contraction.

A detailed study of the XY model is beyond the scope of this work. However, we have made preliminary computations that show acceptance rates comparable to

those of the ferromagnetic Ising model. In order to see how correlated the proposed collective moves are, we have computed the mutual information between the updates at different sites of the TNMH algorithm and compared it to that obtained from a local algorithm, Figure 3.5.1. The instance chosen for the comparison is the zero-field uniform XY model on a 16×16 lattice at a temperature $T = 0.5$. The difference in the results is noteworthy and demonstrates that the TNMH algorithm is indeed capable of producing global correlated updates.

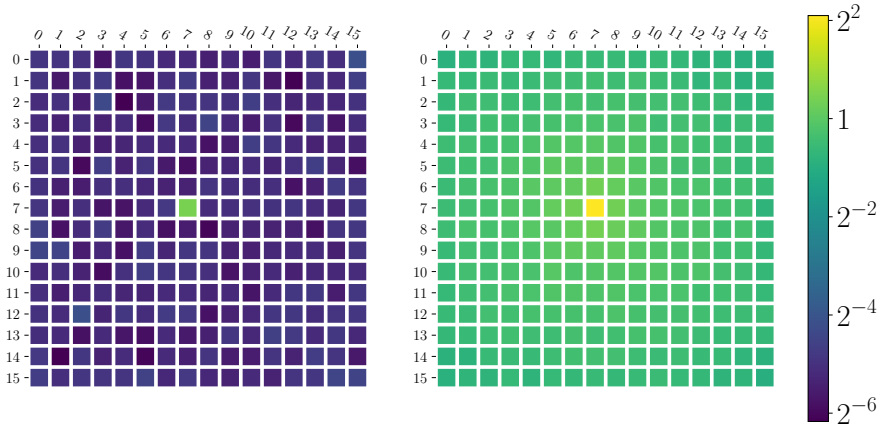


Figure 3.5.1: Mutual information in bits between the updates at different sites, that is, in the changes of the angles after a sweep of the Markov chain, $I(\theta_i(t+1) - \theta_i(t) : \theta_j(t+1) - \theta_j(t))$ for two schemes. Left: a local algorithm (a Metropolis single spin flip where the proposed local updates were chosen from $U(0, 2\pi)$). Right: TNMH. The numerical experiment was conducted on a homogeneous XY model with no external field on a 16×16 lattice at a temperature of $T = 0.5$. The bond dimension used in TNMH was $D = 20$, and the number of terms kept in the series of Eq. (3.5.2) was $N = 4$, which gave a TN with a local dimension $d = 9$.

3.6 Discussion

The interplay between Monte Carlo and tensor network methods is a rich and vastly unexplored subject. While various previous works have reported on using Monte Carlo sampling for tensor network contractions, we have here presented an analysis of the converse: a new class of Markov chain Monte Carlo algorithms for many-body classical systems based on tensor network renormalization. This class belongs to the family of Metropolis-Hastings schemes. Our construction produces collective updates. It is also irreducible and reversible; as such, asymptotic convergence towards the target probability distribution is guaranteed. We emphasize its universal

nature: it works the same for any nearest-neighbor Hamiltonian with finite local degrees of freedom.

We have benchmarked our scheme for a variety of instances of the two-dimensional Ising model defined on a square lattice. For ferromagnets and antiferromagnets, very high acceptance rates have been observed for larger systems, even with modest values of the bond dimension. Besides, drops in acceptance rates have been shown to signal criticality. Looking at equilibration and decorrelation times, the scheme compares extremely well with single spin flip updates and Wolff algorithm. As expected, the scheme's performance is lower for frustrated and disordered instances than for the ferro- and antiferromagnets. Still, our results are very encouraging. In particular, for disordered instances, equilibration appears to be occurring orders of magnitude faster than for state-of-the-art techniques such as parallel tempering supplemented with isoenergetic cluster moves, both when time is counted in Monte Carlo steps and in seconds.

We have also demonstrated the potential of the method for three-dimensional systems, by testing it on ferromagnetic and antiferromagnetic instances. In this case, we have observed faster equilibration as compared to Wolff algorithm and, remarkably, even with a simple contraction strategy and small bond dimension, the scheme can be shown to remain usable at near-critical temperatures, whereas a much more costly direct TN contraction results in considerable errors.

We have used simple procedures to implement tensor network renormalization, and we have made no particular effort to write an efficient code. For these reasons, we believe the results presented here could be substantially improved. It would also be very interesting to study what can be gained by using other renormalization schemes for approximate contractions of tensor networks [108]. For example, schemes involving disentanglers would be a natural option in this regard [109]. Also for future work is the study of how TNMH Markov chains combine with parallel tempering [59].

A major advantage of our construction is its versatility. Our scheme can be used to construct a collective update Markov chain in other settings, such as the XY model, different lattices, or interaction terms. In principle, lattice systems with long-range interactions could also be considered. For instance, given an Ising Hamiltonian H where the interactions decay with the distance as a power law, one can associate an auxiliary Hamiltonian H_ϱ where all interactions within some range ϱ are identical to H, and all interactions beyond ϱ have been truncated. One can next construct a tensor network prior from this Hamiltonian H_ϱ. Two parameters would now govern the Markov chain: the bond dimension and the range ϱ. We have also restricted ourselves to scalar degrees of freedom in this work. But similar constructions to those presented in Section 3.5 show that TNMH sampling should also work for matrix models, in particular lattice gauge theories.

A natural variation of our work would be to depart from tensor network representations and use a quantum device to prepare Gibbs states and estimate the probability of drawing a given configuration [110, 111]. Such a device would be called as an external subroutine in (classical) Metropolis-Hastings iterations. Just as our 3D computations have revealed that inaccurate contraction schemes could still be useful for sampling, it would be very interesting to investigate how much computational power such quantum devices retain when imperfect. These ideas will be studied elsewhere.

Finally, it would be instructive to develop a mathematical perspective on the schemes presented here. In particular, we believe it would be meaningful to identify a non-trivial model for which the mixing time associated with our TNMH scheme could be upper bounded, *e.g.* using a log-Sobolev inequality [112]. It would be insightful to establish the dependence of the log Sobolev constant with the bond dimension.

3.A MPS renormalisation

We here review the relation between tensor networks and partition function [90, 97, 108, 113, 114]. The setup is a slight generalization of that of Section 3.2. That is, we consider a nearest-neighbor classical Hamiltonian

$$H(\omega) = \sum_{\langle i,j \rangle} \varphi_{ij}(\sigma_i, \sigma_j), \tag{3.A.1}$$

on a lattice $\Lambda = (V, E)$, where the local variables σ_i now take value in any finite set, which size we are going to denote d. For the sake of simplicity, we will consider square lattices, and first limit ourselves to two-dimensional systems for now. At fixed inverse temperature β, the partition function can be expressed as

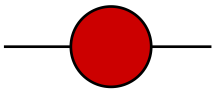
$$Z(\beta) = \sum_{\omega \in \Omega} \prod_{\langle i,j \rangle \in E} W_{ij}(\sigma_i, \sigma_j), \tag{3.A.2}$$

where W_{ij} is a $d \times d$ matrix, whose entries represent all possible contributions of the bond $\langle i, j \rangle$ to the Boltzmann weight of the model, i.e. $W_{ij}(\sigma, \sigma') = e^{-\beta \varphi_{ij}(\sigma, \sigma')}$. As an example, for the Ising model without an external magnetic field, the energy associated with a given bond $\langle i, j \rangle$ reads $\varphi_{ij}(\sigma, \sigma') = -J_{ij}\sigma\sigma'$, and the 2×2 matrix W_{ij} is

$$W_{ij} = \begin{pmatrix} e^{\beta J_{ij}} & e^{-\beta J_{ij}} \\ e^{-\beta J_{ij}} & e^{\beta J_{ij}} \end{pmatrix}. \tag{3.A.3}$$

We will use the diagrammatical notation in which a tensor is represented by a vertex or a small shape, with as many legs sticking out as there are indices; and

where joining two lines represents a contraction of the corresponding indices. For example, a matrix W_{ij} is represented as follows,

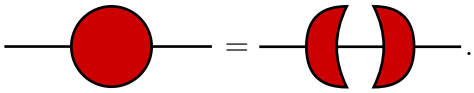


(3.A.4)

$Z(\beta)$ can be expressed as a tensor network if we shift from a description in terms of matrices associated with the bonds of the lattice (3.A.2) to a description in terms of tensors associated with its vertices. Let us consider some vertex i with four neighbors and let $e(i)$ denote the vertex to its right. We decompose $W_{i,e(i)}$ as:

$$W_{i,e(i)}(\sigma,\sigma') = \sum_{\mu=1}^{d} L_i(\sigma,\mu) R_{e(i)}(\mu,\sigma'). \tag{3.A.5}$$

Graphically,



(3.A.6)

This can be achieved e.g. through a singular value decomposition (SVD) $W_{i,e(i)} = U_{i,e(i)}\Sigma_{i,e(i)}V^{\dagger}_{i,e(i)}$, and by setting $L_i = U_{i,e(i)}\sqrt{\Sigma}_{i,e(i)}$, $R_{e(i)} = \sqrt{\Sigma}_{i,e(i)}V^{\dagger}_{i,e(i)}$. Similarly, if $n(i), w(i), s(i)$ denote vertices located above, to the left, and below i respectively, three additional SVD provide the decompositions

$$W_{w(i),i}(\sigma,\sigma') = \sum_{\nu=1}^{d} L_{w(i)}(\sigma,\nu) R_i(\nu,\sigma'), \tag{3.A.7}$$

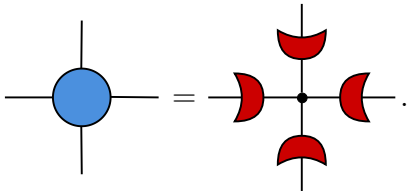
$$W_{i,n(i)}(\sigma,\sigma') = \sum_{\rho=1}^{d} B_i(\sigma,\rho) T_{n(i)}(\rho,\sigma'), \tag{3.A.8}$$

$$W_{s(i),i}(\sigma,\sigma') = \sum_{\tau=1}^{d} B_{s(i)}(\sigma,\tau) T_i(\tau,\sigma'). \tag{3.A.9}$$

We associate a 4-index tensor $A^{(i)}(\sigma)$ with each site i having four neighbours and each spin value σ, whose components are

$$A^{(i)}_{\mu\nu\rho\tau} = \sum_{\sigma=1}^{d} L_i(\sigma,\mu) R_i(\nu,\sigma) B_i(\sigma,\rho) T_i(\tau,\sigma). \tag{3.A.10}$$

In diagrammatic notation, Eq. 3.A.10 reads



(3.A.11)

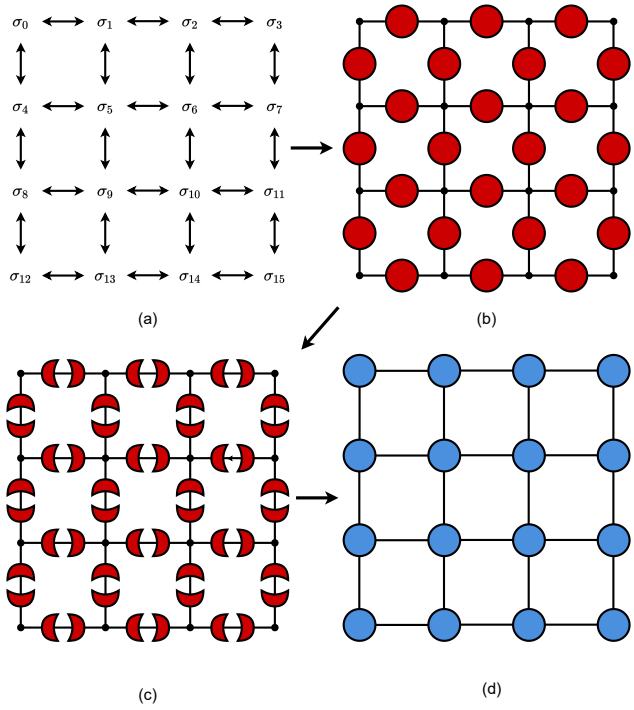


Figure 3.A.1: Graphical depiction of the construction of the TN associated with the partition function of a nearest-neighbor classical Hamiltonian (4×4 lattice in this illustration). (a) We start with a labeling of the vertices of the lattice in consideration. (b) Diagrammatic representation of the Boltzmann weights (red circles) associated with each edge; their contraction yields the partition function. (c) and (d) Singular value decomposition of each W matrix, and regrouping into tensors associated with each vertex of the lattice.

For a system with open boundary conditions, vertices with only three or two neighbors are dealt with likewise. With these tensors, the partition function can be expressed as

$$Z(\beta) = \mathcal{C}(\{A^{(i)}\}), \tag{3.A.12}$$

where $\mathcal{C}(\{A^{(i)}\})$ denotes the contraction of all the tensors associated with all sites. The entire process from (3.A.2) to (3.A.12) is illustrated on Figure 3.A.1 for a 4×4 lattice.

Similarly, one can construct a TN representation of the partition function with some fixed value s for the degree of freedom at site i, $Z(\beta|\sigma_i = s)$. It is for instance sufficient that for each neighbour of i, j, we replace $W_{i,j}(\sigma, \sigma')$ with $W^{(s)}_{ij}(\sigma, \sigma') = \delta_{s,\sigma} W_{i,j}(\sigma, \sigma')$. The ratio of the two quantities, $Z(\beta|\sigma_i = s)/Z(\beta)$, would exactly be the marginal probability $\pi^{(\beta)}_i(s)$ of that spin being in state s, that is, a ratio of two TN contractions. Similarly, one can express any conditional

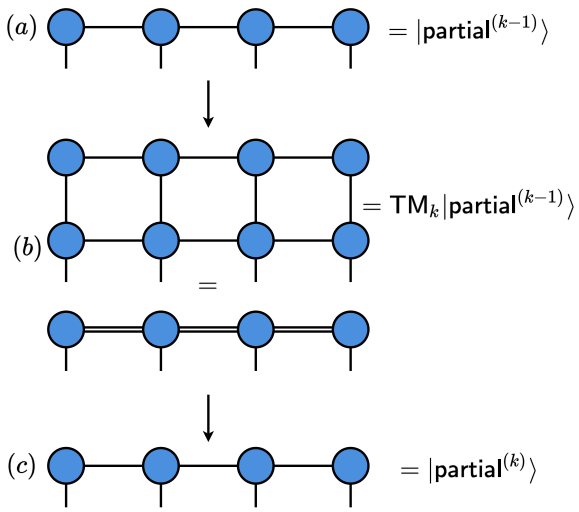


Figure 3.A.2: Graphical depiction of the process of contracting a two-dimensional lattice. (a) Approximation to the contraction of the $k-1$ top most rows, $|\mathsf{partial}^{(k-1)}\rangle$. (b) The approximation to $|\mathsf{partial}^{(k)}\rangle$ is constructed by applying the MPO associated with row k, TM_k, on the MPS obtained by approximate contraction of the $k-1$ first rows, $|\mathsf{partial}^{(k-1)}\rangle$. The result of that MPO-MPS multiplication is a MPS with a larger bond dimension. (c) Using standard MPS techniques, the product can be approximated by a MPS with lower bond dimension.

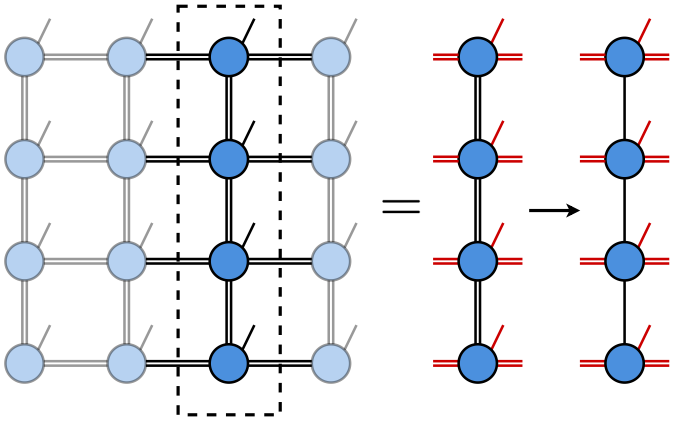


Figure 3.A.3: Renormalisation of a black column of a PEPS. For the truncation of the bond dimension of a given column, its tensors are singled out. The physical bond dimension and the bond dimension that connects these tensors to other columns, drawn in red in this diagram, are treated as the physical dimension of an auxiliary MPS. The bond dimension of that MPS is reduced using a standard truncation algorithm. The resulting tensors of the obtained MPS with lower bond dimension are inserted back into the PEPS. While this procedure has no guarantee of optimality, it is computationally cheap and works well in practice.

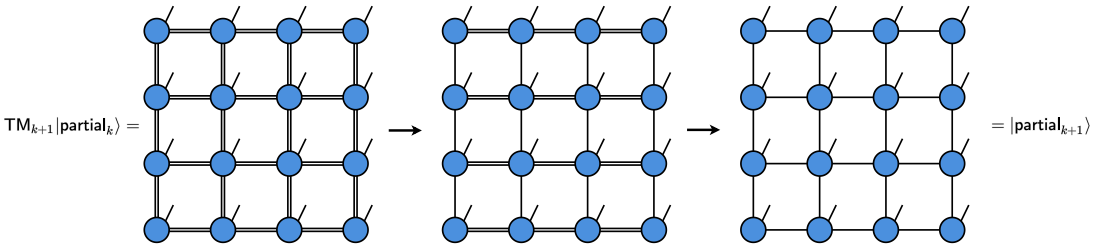


Figure 3.A.4: Renormalisation of a PEPS used to apply the TNMH algorithm to three-dimensional systems. The bond dimension is first reduced along horizontal bonds and next along vertical bonds.

probability $\pi_k^{(\beta)}(\sigma_k|s_1 \dots s_{k-1})$ as a ratio of two TN contractions:

$$\pi_k^{(\beta)}(\sigma_k|s_1 \dots s_{k-1}) = \frac{Z(\beta|s_1 \dots s_{k-1}\sigma_k)}{Z(\beta|s_1 \dots s_{k-1})}. \tag{3.A.13}$$

As explained in Section 3.2, if one were able to evaluate TN contractions exactly, one would have the means to sample according to the Boltzmann distribution exactly. In general, it is only possible to compute approximations to the contractions appearing in the ratio (3.A.13) and as a result, get an approximation $\widetilde{\pi}_k^{(\beta)}$ to $\pi_k^{(\beta)}$. Instead of using these approximations for direct sampling with systematic errors, one can use them as a prior for a reversible Metropolis-Hastings Markov chain. The impossiblity to carry out exact TN contraction then translates into more controllable statistical errors.

For the approximate contraction of an $L \times L$ lattice, we have used one of the simplest schemes available [38, 90]. We define $|\mathsf{top}\rangle$ to be the tensor resulting from contracting all the top row tensors along horizontal edges; the remaining free indices after this contraction are legs pointing downward. Similarly, we will call transfer matrix the tensor resulting from a contraction of the tensors along a horizontal bulk row; the transfer matrix resulting from contracting the tensors of row k will be denoted TM_k, $2 < k < L$. Finally, in analogy to $|\mathsf{top}\rangle$, we will denote $|\mathsf{bot}\rangle$ the contraction of bottommost tensors. With these notations, the partition function can be expressed as

$$Z(\beta) = \langle \mathsf{bot}|\mathsf{TM}_{L-1} \dots \mathsf{TM}_2|\mathsf{top}\rangle. \tag{3.A.14}$$

Both $|\mathsf{top}\rangle$ and $\langle \mathsf{bot}|$ are matrix product states (MPS), whereas the transfer matrices TM_k are matrix product operators (MPO), all with a bond dimension and a 'physical' dimension equal to d; their length is equal to L. Our approximation of $Z(\beta)$ is obtained by estimating the right hand side of (3.A.14) sequentially. We initialise $|\mathsf{partial}^{(1)}\rangle \equiv |\mathsf{top}\rangle$, and for $k \in \{2 \dots L-1\}$, we define $|\mathsf{partial}^{(k)}\rangle$ to be an MPS approximation to $\mathsf{TM}_k|\mathsf{partial}^{(k-1)}\rangle$ obtained by matrix product state renormalisation, see Fig. 3.A.2. $Z(\beta)$ is finally approximated with $\langle \mathsf{bot}|\mathsf{partial}^{(L-1)}\rangle$. The cutoff parameter D sets the accuracy of the approximation. There are many methods available for the renormalization. Throughout this work, we have mostly used the scheme based on successive SVD [39]. Two-site variational compression has been used to explore the equilibration of the two-dimensional Ising model with Gaussian disorder [39].

The same method allows to approximate the partition function of a system where some spins have been set to definite values, $\widetilde{Z}(\beta|\sigma_1 \dots \sigma_k)$. The only difference is that for such a site i with spin value σ_i, the tensor (3.A.10) is replaced with

$$L_i(\sigma_i, \mu)R_i(\nu, \sigma_i)B_i(\sigma_i, \rho)T_i(\tau, \sigma_i). \tag{3.A.15}$$

As claimed in section 3.2, an approximate Boltzmann weight $\widetilde{\pi}(\omega)$ can be evaluated since, using Bayes theorem, this probability can be expressed as

$$\frac{\widetilde{Z}(\beta|\sigma_1)}{\widetilde{Z}(\beta)} \times \ldots \times \frac{\widetilde{Z}(\beta|\sigma_1 \ldots \sigma_n)}{\widetilde{Z}(\beta|\sigma_1 \ldots \sigma_{n-1})}. \tag{3.A.16}$$

Two remarks are in order. First, if the tensors $A^{(i)}$ are well conditioned and if D is high enough, the approximations to partition functions we construct will be strictly positive. So will then be the approximated probabilities (3.2.7), and the TNMH is irreducible. Second, if the tensors $\{A^{(i)}\}$, the MPS $|\mathsf{top}\rangle$, $|\mathsf{bot}\rangle$ and the transfer matrices $\mathsf{TM_k}$ are stored, a TNMH update of the whole lattice can be performed at a computational cost that scales linearly with the lattice size.

Plaquette interactions can be dealt with similarly. Using singular value decomposition for the Boltzmann weights and regrouping all the matrices relating to a given site, one obtains a ($\pi/4$ rotated) square lattice for the partition function. Bayes formula can thus again be used for sampling.

We have dealt with three-dimensional models in a similar fashion. Assuming an $L \times L \times L$ lattice, the identity (3.A.12) can again be obtained after a sequence of SVD; $A^{(i)}$ is now a six-leg tensor (for a bulk spin). (3.A.14) is also still valid, but $|\mathsf{top}\rangle, |\mathsf{bot}\rangle$ are now projected entangled pair states (PEPS) instead of MPS, and the transfer matrices TM_k projected entangled pair operators (PEPOs) instead of MPOs. Just as in two dimensions, without any cutoff, the bond dimension of $\mathsf{TM}_{L-1} \ldots \mathsf{TM}_2|\mathsf{top}\rangle$ would generically grow exponentially with L, and renormalisation is in order. There exists a plethora of methods to contract three-dimensional tensor networks [97, 115]. We have not aimed at optimality and have opted for simplicity. Again, denoting $|\mathsf{partial}_k\rangle$ the approximate contraction of the first k layers of the TN, the core of the renormalization consists in constructing a PEPS approximation $|\mathsf{partial}_{k+1}\rangle$ for the contraction $\mathsf{TM}_{k+1}|\mathsf{partial}_k\rangle$. When a PEPO is superimposed on a PEPS, the resulting state is a PEPS with a larger bond dimension.

The bond dimension of $\mathsf{TM}_{k+1}|\mathsf{partial}_k\rangle$ has been reduced by recycling the technique used to compress the bond dimension of MPS. First, an index reshuffling allows us to regard each column of the PEPS $\mathsf{TM}_{k+1}|\mathsf{partial}_k\rangle$ as an MPS, with an effective physical index at each site given by lumping the original physical index of the PEPS with the horizontal virtual indices at that site. The virtual bonds of that MPS are the vertical virtual bonds of the corresponding column of the PEPO. These 'thick' bonds are compressed (or renormalized) as before. This compression along columns is illustrated in Fig. 3.A.3. The resulting tensors from the compression are then inserted back into the PEPS, and the same compression is next performed on the horizontal bonds of the PEPS. See Fig. 3.A.4 for a depiction of this PEPS renormalization. Two parameters now govern the accuracy of the

approximation: the bond dimension of the PEPS $|\mathsf{partial}_k\rangle$, D, and the cutoff for the approximate contraction of two rows of a PEPS, χ [115].

Chapter 4

Light cone tensor network and time evolution

This chapter is based on the publication
Miguel Frías-Pérez, Mari Carmen Bañuls
Light cone tensor network and time evolution
Phys. Rev. B 106, 115117 (2022)
Structure, text and figures have been adapted for this thesis

4.1 Overview

Simulating the dynamics of out-of-equilibrium quantum systems is one of the most challenging open problems for tensor networks and other techniques. In one-dimensional systems, limitations of TN methods for dynamics are well understood: in global quenches, the entanglement may grow fast [48, 116, 117], and the true state can escape the descriptive power of the TN ansatz. This so-called entanglement barrier limits the applicability of the MPS [5, 8, 118, 119] description, and makes it difficult to predict the asymptotic long-time behavior, even when local observables in this limit are expected to be well-described by a thermodynamic ensemble, itself well approximated by an MPO [9, 10, 42, 43, 120, 121]. A number of methods have been suggested to try to overcome this issue and extract information about the long-time behavior of local properties [122–132]. While there is no universal solution, understanding the entanglement structures in the evolution TN can be crucial to identify the most adequate one for practical computations.

In particular, the transverse folding strategy [123, 133, 134] avoids the explicit representation of the evolved state as a MPS and instead focuses on contracting a TN that represents exactly (up to Trotter errors) the time-dependent observables. Instead of the standard evolution in time direction, the folding algorithm

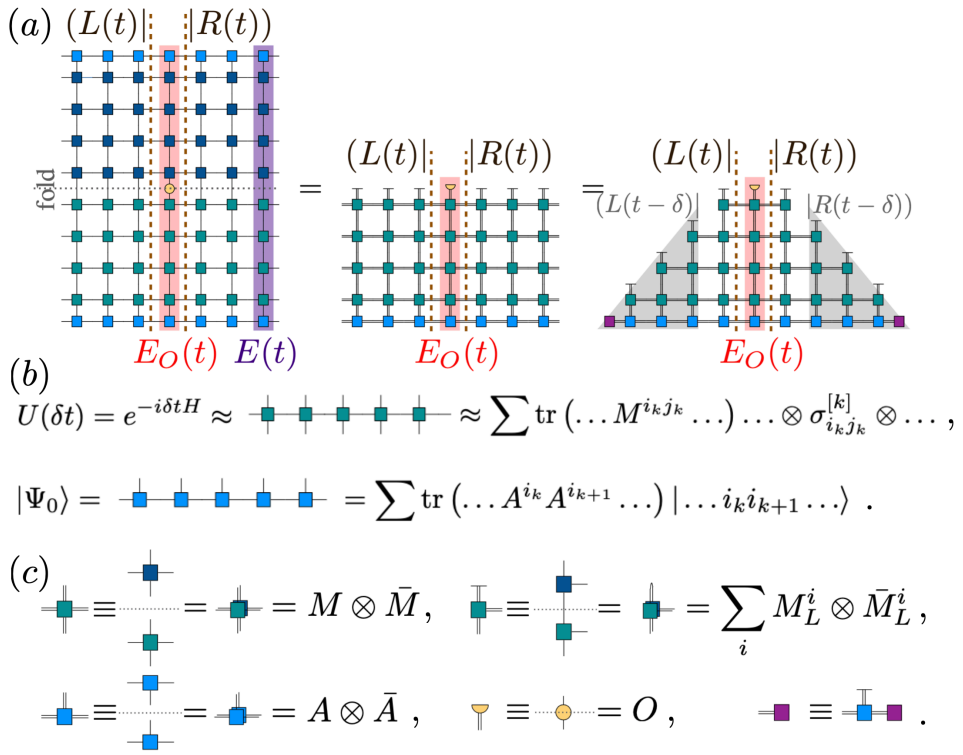


Figure 4.1.1: (a, b) Schematic construction of the minimal TN for the expectation value of a local operator O after a global quench in a translationally invariant setting. At time $t = M\delta$ the expectation value $\langle\Psi(t)|O|\Psi(t)\rangle$ corresponds to a two dimensional TN. After folding, the exact light cone is obtained after removing the mutually canceling gates. (c) Graphical notation for folded TN diagrams through the paper.

contracts the TN along space. In some scenarios, this allows local observables to be computed to longer times than other approaches [135], and it is an exact strategy for certain models [136]. Recently, there has been a rekindled interest in this approach, triggered by the interpretation of the network in terms of an influence functional [137–139].

In local lattice models, the velocity of propagation of information is upper-bounded [140–142] and the exact TN for observables has a light cone structure. While there have been proposals that exploit this fact to reduce the cost of the numerical simulation of the evolved state with TN [143–148], and with quantum simulation [149], until now, the potential of combining it with the transverse strategy has not been explored.

Here we propose a strategy to exploit this property, a *transverse light cone contraction* of the TN (TLCC). As in the original transverse folding, the TLCC does

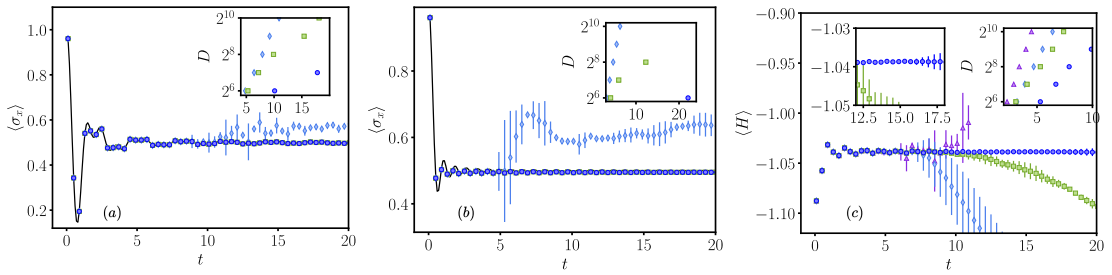


Figure 4.1.2: Evolution after a global quench from the initial state $|X+\rangle$, for the integrable [(a) $g = 0.5$, (b) $g = 1$] and non-integrable [(c) $g = -1.05$, $h = 0.5$] Ising model. The main plots show the transverse magnetization $\langle\sigma_x(t)\rangle$ (a, b) and the energy density (c) computed with different algorithms using respectively bond dimension $D = 128$ (a,b) and 512 (c). Error bars represent the difference with respect to the results obtained with $D' = D/2$. The TLCC contraction has been obtained both with the standard MPS truncation (green squares) and the hybrid truncation of [134] (dark blue circles). For comparison, we also show the results of standard iTEBD (blue diamonds) and Heisenberg picture DMRG (purple triangles).[1] For the integrable case (a,b) we also show the analytic result (black line). The insets show the scaling of the bond dimension required to keep constant precision in each algorithm. In the integrable case, this is compatible (at the later times) with a polynomial increase $D \sim t^{\alpha}$, consistent with observations in [133, 150]. In the non-integrable case, the increase is compatible with an exponential growth for both truncation methods, but using the hybrid truncation exhibits a slower rate than standard ones, such that longer times can be reached with the same bond dimension. The left inset in (c) shows a zoom of the main plot to better appreciate the differences.

not directly suffer from the entanglement growth in the state, and will be more efficient than standard algorithms when entanglement in the time direction grows slower than in the spatial one. But the TLCC improves the efficiency with respect to the transverse folding in all cases, by reducing the computational effort to that of approximating the minimal network describing the time-dependent observables in a Trotterized evolution. We demonstrate explicitly its performance for global quenches and different-time thermal correlators at infinite temperature, and investigate how the strategy can make use of the (more efficient) physical light cone determined by the Lieb-Robinson velocity [141]. We also discuss possible extensions to other interesting quantities.

4.2 Light cone tensor network for global quenches

The one-dimensional global quench is a natural test bench for time-evolution TN algorithms. At time $t = 0$, the system is prepared in a state that can be written as a MPS (e.g., a product state), and then it is let to evolve under a fixed Hamiltonian. For simplicity, we restrict the discussion to a nearest-neighbor model and a translationally invariant case, but the construction generalizes straightforwardly to any model with local (finite-range) interactions and some non-translationally invariant scenarios.

The transverse folding proposal of [123] starts from a two-dimensional TN whose contraction represents some time-dependent observable, such as a local expectation value. This TN can be constructed from a Suzuki-Trotter approximation of the evolution operator, where the evolution for a discrete step of time δ can be approximated as an MPO [9, 10] with a small bond dimension, constructed from a product of two-body gates [42]. The TN for the observable at time $t = M\delta$ is obtained by applying M copies of this MPO with the initial state, which yields the evolved state, and contracting the operator of interest between this and its adjoint.

While standard TN algorithms as TEBD or tMPS [9, 16, 151, 152] compute the observable by contracting the network in the time direction, the transverse folding strategy performs the contraction in the spatial direction, after folding the TN in half, such that tensors for the same site and time step in the ket and the bra are grouped together (see figure 4.1.1a). After folding, the growth of entanglement in the time direction can be slower than in the spatial one, with the most dramatic difference observed for integrable systems [133, 150], but occurring also in generic cases, as the ones shown here. When this difference in growth is present, the transverse strategy allows reaching longer times than standard algorithms.

For a translationally invariant system in the thermodynamic limit, the transverse contraction reduces to an expectation value of the form $(L(t)|E_O(t)|R(t))$, where $(L(t)|$ and $|R(t))$ are the dominant left and right eigenvectors of the transfer operator $E(t) = \sum_i \bar{A(t)}^i \otimes A^i(t)$, and $E_O(t) = \sum_i \bar{A(t)}^i \otimes A^j(t)\langle i|O|j\rangle$ [119]. Here, $A^i(t)$ represents the *concatenated* [153] local tensor of the time-dependent state, itself a MPO. In the transverse folding strategy, the boundary vectors $(L(t)|$ and $|R(t))$ are approximated by MPS. This approximation can be found, for instance, via a power iteration or a Lanczos algorithm, using repeated MPO-MPS contractions.

Such strategies do not take into account that the TN has a light cone structure. Because the individual gates are local, outside the causal cone of the operator, each gate cancels with its adjoint. This ensures that each of the required boundary vectors (dominant eigenvectors of the transfer operator) corresponds precisely to the contraction of a triangular network as depicted in fig. 4.1.1. We can approximate directly the contraction of such triangle in the space direction by a MPS. This

strategy, which we call *transverse light cone contraction* (TLCC), allows us to obtain $(L(t)|$ and $|R(t))$ in a fixed number of steps (proportional to M). Furthermore, once we have found the vectors for M time steps, we can directly obtain them for $M+1$ by applying a single MPO (as illustrated in the figure), which increases the length by one, and approximating the result via a single truncation step. This step can be performed using standard MPS truncation algorithms, which reduce the bond dimension by minimizing the distance between the truncated vector and the original one. However, for this particular problem, the *hybrid* truncation algorithm proposed in [134], which effectively *evolves* the bond of the boundary vector according to the real-time dynamics, yields a much more efficient use of the available bond dimension (see also insets of fig. 4.1.2).

The TLCC strategy results in a more efficient algorithm than the originally proposed folding, which required iterative MPO-MPS contractions until convergence of the dominant eigenvectors, run independently for each different time step (in particular, for the cases analyzed in this work, we find the power iteration required several tens of MPO-MPS contractions per time step). Notice, nevertheless, that if the bond dimension used is large enough, both the original folding algorithm and the TLCC should result in the same boundary vector. What ultimately determines the applicability of transverse strategies is thus the amount of *entanglement* present in the transverse network.

To probe the performance of the method, we consider a quantum Ising chain, initialized in a product state $|X+\rangle = \lim_{N\to\infty}[(|0\rangle + |1\rangle)/\sqrt{2}]^{\otimes N}$. We then apply the Hamiltonian,

$$H_{\mathrm{I}} = \sum_i \left(J\sigma_i^z\sigma_{i+1}^z + g\sigma_i^x + h\sigma_i^z \right), \tag{4.2.1}$$

and compute local expectation values after time evolution. In all the following we fix $J=1$, and a Trotter step $\delta = 0.1$, and vary the parameters of the model to study integrable ($g = \{0.5,\, 1\}$, $h=0$) and non-integrable ($g=-1.05$, $h=0.5$) regimes. Figure 4.1.2 shows the results and demonstrates that the TLCC can efficiently simulate the integrable quenches. In the non-integrable regime, the required bond dimension grows much faster with time, but the method is still advantageous as compared to standard evolution, much more so when the truncation is performed as in [134] (see right inset of fig. 4.1.2c).

4.3 Light cone tensor network for transport coefficients

The same idea can be adapted to the computation of other dynamical quantities. It is the case of thermal correlators, of the form $C_{1,2}(t,\ell,\beta) = \mathrm{Tr}(\rho_\beta O_2^{[\ell]}(t) O_1^{[0]}(0))$,

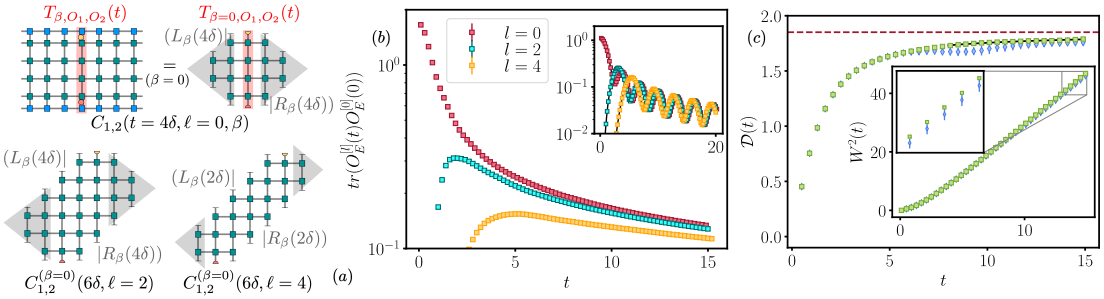


Figure 4.3.1: (a) Schematic construction of the minimal TN for two-point correlators at infinite temperature for different times and distances. For more details, see Appendix 4.A. (b) Energy autocorrelations $C_{EE}(t,\ell,\beta=0)$ obtained from the TLCC method at several distances as a function of time in the integrable ($g=0.5$, inset) and non-integrable ($g=-1.05$, $h=0.5$, main plot) Ising chain at $\beta=0$. The error bars (smaller than the size of the marker) show the difference between results with two different bond dimensions (D,D') [for the inset (200, 100), for the main plot (500, 200)]. In the inset, the black curves represent the results coming from the analytical solution of the model. (c) Spatial variance 4.3.2 of the normalized autocorrelations (4.3.2) (inset) and corresponding diffusion constant (main plot) in the non-integrable case obtained from the TLCC (green squares) and TEBD (blue diamonds) with $D=1024$, with error bars showing the difference with respect to $D'=512$. The solid black line in the main plot shows a fit of the form $\mathcal{D}_E\exp(b/t)$, which predicts the asymptotic value $\mathcal{D}_E\approx 1.9$ (red dotted line).

where $\rho_\beta = e^{-\beta H}/Z$ is the thermal equilibrium state at inverse temperature β, $Z=\mathrm{Tr}(e^{-\beta H})$ is the partition function, $O_k^{[\ell]}(t)$ is a (local) operator acting on-site ℓ at time t, and $O_k(t)=U(t)^\dagger O_k U(t)$ is the time-evolved operator in Heisenberg picture. Since $[\rho_\beta,H]=0$, the thermal state is invariant under the evolution, and using $\rho_\beta\propto\rho_{\beta/2}\rho_{\beta/2}^\dagger$ we can write (up to normalization), $C_{1,2}(t,\ell,\beta)\propto\mathrm{Tr}(U(t)^\dagger\rho_{\beta/2}^\dagger O_2^{[\ell]}U(t)O_1^{[0]}\rho_{\beta/2})$. Using a MPO approximation to $\rho_{\beta/2}$ (obtained with standard TN methods [9–11, 154]), and the Trotterized real-time evolution as in the previous section, this quantity can be expressed as a two dimensional folded TN, which can be contracted in the temporal [155–157] or spatial (transverse) [133] direction.

Due to the invariance of the thermal state, each local observable generates also a light cone structure that can be exploited in the TLCC approach. Now the cancellation of gates outside the causal cone of the operators occurs both at the upper and the lower parts of the network (see figure 4.3.1a), and the minimal TN has a rectangular form, resembling a pillow, a structure which was used in [158] to evaluate correlators in random quantum circuits. The TLCC strategy again

requires contracting a triangular TN corresponding to the lateral corners of the figure to obtain boundary vectors ($L_\beta(t)|$ and $|R_\beta(t)$). [2] If both operators act on the same site ($\ell = 0$), the time-dependent correlators can be expressed as a contraction $(L_\beta(t)|T_{\beta,O_1,O_2}(t)|R_\beta(t))$, with a single MPO $T_{\beta,O_1,O_2}(t)$ constructed from concatenating the local tensors for the unitaries, the operators and the states (see fig. 4.3.1a). For correlators at non-zero distance ℓ the minimal TN becomes elongated (fig. 4.3.1a, lower diagrams). To approximate its contraction, the boundary vectors ($L_\beta(t)|$ and $|R_\beta(t)$) for a certain time t are first grown to incorporate, respectively, O_1 at the bottom of the TN, and O_2 at the top. These extended vectors contain the evolution steps up to time $t+2\delta$, and can be contracted together to obtain the correlators at $\ell = 1$ for times $t + 3\delta$ and $t + 4\delta$. The vectors can be then evolved again, following the TN structure, which does not increase their length, but allows access to correlators at any later time $t + (2 + k)\delta$ and distances $\ell = k$, $k+1$. Applying this systematically we can obtain all non-vanishing correlators. This generalizes trivially to operators on more than one site, or with MPO structure.

Here we illustrate the simplest case, infinite temperature, where $\rho_{\beta=0} \propto \mathbb{I}$ and the contour of the TN becomes uncorrelated. We consider the energy density operator

$$O_E^{[i]} := J\sigma_i^z\sigma_{i+1}^z + \frac{g}{2}(\sigma_i^x + \sigma_{i+1}^x) + \frac{h}{2}(\sigma_i^z + \sigma_{i+1}^z), \tag{4.3.1}$$

which can be written as a MPO of range 2. Figure 4.3.1b shows our results for the correlators $C_{EE}(t, \ell, \beta = 0)$ as a function of time for several distances in the non-integrable ($g = -1.05$, $h = 0.5$, main plot) and integrable ($g = 0.5$, $h = 0$, inset) cases.

Especially interesting is the possibility of ab initio calculations of transport properties [159] in non-integrable models. In particular, diffusion constants can be related to the spatial spreading in time of autocorrelations of a density [127, 160, 161]. Normalizing the correlators as $\tilde{C}_{EE}(0, \ell) := C_{EE}(t, \ell)/\sum_\ell C_{EE}(0, \ell)$, a diffusion constant $\mathcal{D}(t)$ may be obtained from their spatial variance [160]

$$W^2(t) := \sum_\ell \tilde{C}_{EE}(t, \ell)\ell^2 - \left(\sum_\ell \tilde{C}_{EE}(t, \ell)\ell\right)^2, \tag{4.3.2}$$

as $\frac{\partial W^2}{\partial t} = 2\mathcal{D}(t)$. Figure 4.3.1c shows the (linearly growing) variance $W^2(t)$ (main plot), and the corresponding diffusion constant (inset) obtained from the correlators for the non-integrable case. The diffusion constant is well fitted by a function $\mathcal{D}(t) = \mathcal{D}_E \exp(b/t)$, compatible with saturation to a constant $\mathcal{D}_E \approx 1.9$ in the

[2] Different from the global quench above, in this case, each iteration of the algorithm grows the boundary vectors in two time steps.

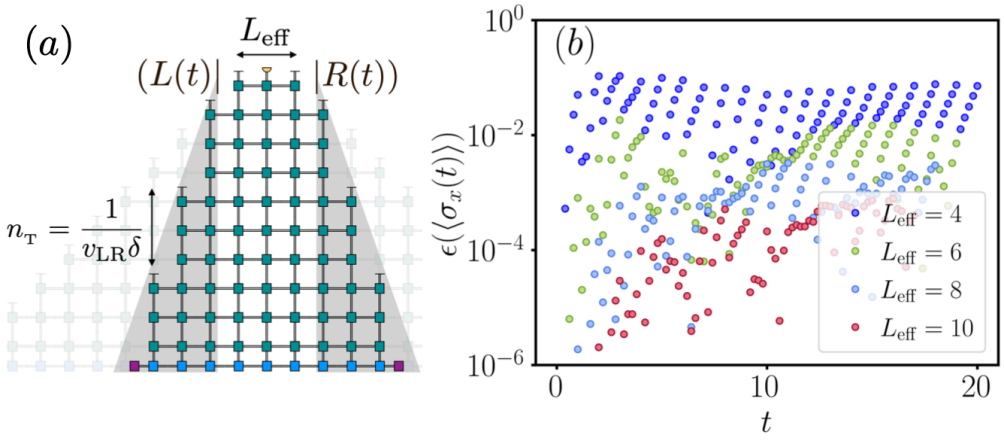


Figure 4.4.1: (a) The physical velocity defines a much narrower light cone than the Trotterization (background). (b) Relative difference between $\langle\sigma_x\rangle$ computed with the LR and Trotter light cones for the integrable global quench of fig. 4.1.2a, for which $n_{\text{T}} = 10$ and different sizes of the subsystem L_{eff}, with $D = 200$ in all cases.

asymptotic regime [3]. While TEBD (blue diamonds) produces close values for the same quantities, the error is appreciable in the diffusion constant already at short times.

4.4 The physical light cone

In general, we expect that the physical light cone is much narrower than the trivial one from the Trotterization, used in the previous sections. We could thus approximate the TN by a light cone one in which the slope corresponds to the maximal physical velocity v_{LR}. This can be achieved by implementing a more efficient TLCC growing iteration, in which $n_{\text{T}} = 1/(v_{\text{LR}}\delta)$ time steps are applied at once every time a space site is contracted (fig. 4.4.1a). Notice that this light cone is not exact, but has (exponential) corrections. Thus it is convenient to consider the light cone for a subsystem of size L_{eff} that includes the support of the operator.[4]

To probe this reduced light cone we choose an integrable instance, ($g = 0.5$, $h = 0$), for which the Lieb-Robinson velocity is known ($v_{\text{LR}} = 1$, corresponding to $n_{\text{T}} = 10$ with our Trotter step), and simulate the global quench of fig. 4.1.2a. Compared to TLCC for the full light cone with the same bond dimension, we

[3]The fit $D(t) = \mathcal{D}_E \exp(b/t)$ is a heuristic choice that describes our data well over a range of fitting windows and allows us to extrapolate to the limit of infinite time. We have also tried successfully fits with polynomials of $1/t$, and found compatible results.

[4]Equivalently, we can *insert* in the middle of the column, corresponding to an earlier time.

observe (fig. 4.4.1) that the physical one, determined by v_{LR}, captures indeed the correct evolution: while the narrower light cone deviates from full results, the errors are reduced exponentially (until the level of original truncation error) by considering a small window L_{eff}.

4.5 Discussion

We have presented a strategy that builds on the transverse folding [123] to approximate time-dependent observables in a one-dimensional quantum system. Noticing the exact light cone structure of the TN and implementing its transverse contraction, it is possible to compute long-time properties in a more efficient manner. Combined with the hybrid truncation [134], this allows us to reach longer times with a smaller bond dimension whenever the temporal entanglement grows slower than the physical one, which, as we have seen, happens not only for integrable systems. It is possible to use the physical upper bound of the Lieb-Robinson velocity to further restrict the width of the relevant TN and define a more efficient iteration.

We have evaluated the performance of the TLCC strategy for integrable and non-integrable global quenches and for transport properties at infinite temperature. With minimal changes, the method extends to other scenarios, such as finite temperature or non-translationally invariant setups including impurities or a contact between two chains. It is furthermore possible to adapt the strategy to other more complex dynamical quantities.

The basic TLCC does not require additional hypothesis to truncate observables or states. Its convergence can be systematically explored as the bond dimension is increased. What ultimately limits the validity of the strategy is the entanglement in the time direction, which strongly depends on the setup and the model [133, 150, 162]. The behavior of the TLCC can thus provide useful information to determine optimal strategies for different problems. Another parameter in the approximation is the Trotter step, which is known to affect the entanglement growth in standard algorithms [152]. Since simulations with different δ may be necessary to extrapolate the exact results, it is also interesting to study how varying δ affects our observations. Further interesting avenues for future investigation are exploring the TN cut according to different velocities, to explore the propagation of correlations in the TN and effectively measure v_{LR}.

4.A Tensor networks for thermal response functions

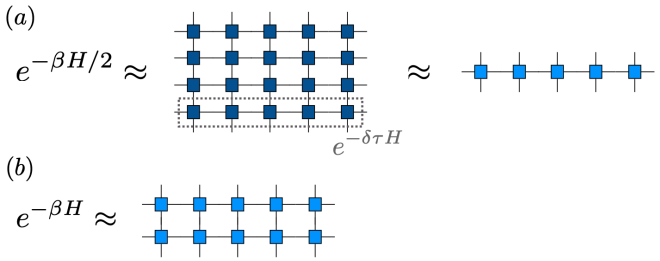


Figure 4.A.1: Schematic description of a TN algorithm to find the MPO approximation of the Gibbs ensemble. (a) The MPS approximation of the thermofield state found by iteratively applying small imaginary time steps onto the maximally entangled state is equivalent to an MPO approximation of the exponential operator. (b) Tracing out the ancillary degrees of freedom is equivalent to considering $(e^{-\beta H/2})^\dagger e^{-\beta H/2}$, which is guaranteed to be positive.

This section shows explicitly how to construct the TN of Fig. 3(a) in the main text, for the case of arbitrary inverse temperature β. We are interested in correlators of the form $C_{1,2}(t,\ell,\beta) = \mathrm{Tr}(\rho_\beta U(t)^\dagger O_2^{[\ell]} U(t) O_1^{[0]})$, where $\rho_\beta = e^{-\beta H}/Z$ is the thermal equilibrium state at inverse temperature β, $Z = \mathrm{Tr}(e^{-\beta H})$ is the partition function, $O_k^{[\ell]}$ is a (local) operator acting on-site ℓ, and $U(t) = e^{-iHt}$ is the time-evolution operator for time t. In order to write this quantity as the contraction of the TN shown in the text, we start by finding an MPO approximation of the thermal state. This can be achieved with several algorithms [9–11, 154]. Here we illustrate (see Fig. 4.A.1) the (possibly) most common algorithm, based on a purification and a Trotter expansion of ρ_β.

The purification approach is equivalent to considering a thermofield double state, i.e. a pure state of the form

$$|\Psi_\beta\rangle \propto e^{-\beta H/2}|\Phi\rangle, \tag{4.A.1}$$

where $|\Phi\rangle$ is a maximally entangled state of the system and an ancillary copy of it. Tracing out the ancillary system results (up to normalization) in the Gibbs ensemble ρ_β. For any basis $\{|n\rangle\}$ of the system, we can write

$$|\Psi_\beta\rangle \propto e^{-\beta H/2} \sum_n |n,n\rangle. \tag{4.A.2}$$

The most frequently used TNS algorithm for thermal equilibrium states proceeds by approximating $|\Psi_\beta\rangle$ by an MPS in a basis in which each system site is grouped

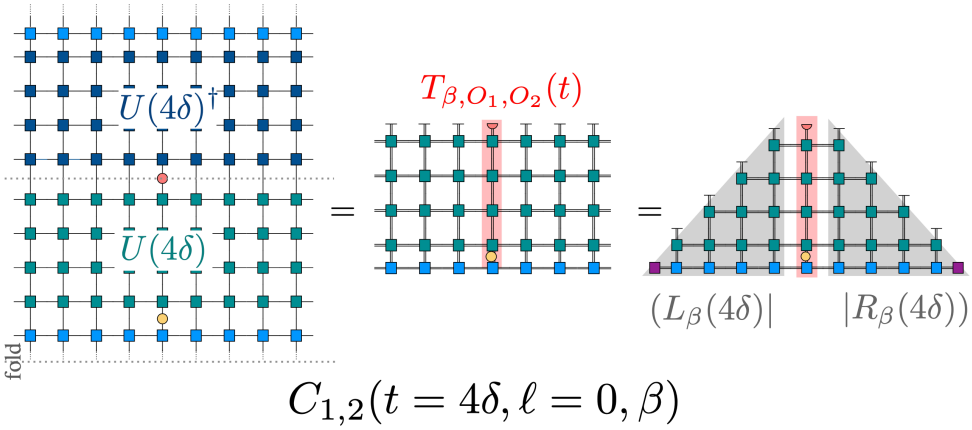


Figure 4.A.2: TN for the thermal correlators. The left diagram shows the network for the observable (4.A.4) (the open dotted lines above are connected to the corresponding ones on the lower edge). A folded version across the dotted horizontal line, shown in the middle, can be written in terms of the same double tensors defined in Fig. 1 of the main text with the thermal MPO tensors defining the boundary. For local O_2, local unitaries cancel out, resulting in the light cone structure on the right, analogous to the one for pure state quenches.

with an ancillary one (forming effective sites of dimension d^2). To do so, the state is initialized to the maximally mixed one between system and ancilla (equivalent to the vectorized identity operator), i.e. a MPS with bond dimension one. The exponential operator $e^{-\beta H/2}$ can be discretized as the product of a finite number $M = \beta/(2\delta\tau)$ of imaginary time steps of length $\delta\tau$. If the Hamiltonian is local, each of them, can be approximated by a MPO (for instance, for nearest-neighbor models, one can use an even-odd Trotter approximation) and successively applied on the MPS, acting on the system degrees of freedom. After each step, a standard truncation can be performed (using any of the algorithms for Trotterized time evolution), such that after a fixed number of steps $M = \frac{\beta}{2\delta\tau}$, an MPS approximation to (4.A.2) is obtained. As sketched in Fig. 4.A.1(a), this procedure is equivalent to approximating the exponential by an MPO (with the Frobenius norm characterizing the quality of the approximation in the operator level). Whereas this procedure could be run for the full inverse temperature, to obtain an MPO approximation of $e^{-\beta H}$, the truncation in MPO-MPS products does not preserve positivity. Instead, an MPS approximation of the thermofield state results necessarily in a positive density matrix (with purification structure) when tracing the ancillae, as shown schematically in Fig. 4.A.1(b).

The TN for the thermal response functions can be constructed by applying the operator O_1 followed by the Trotterized time evolution on this MPO, and finally

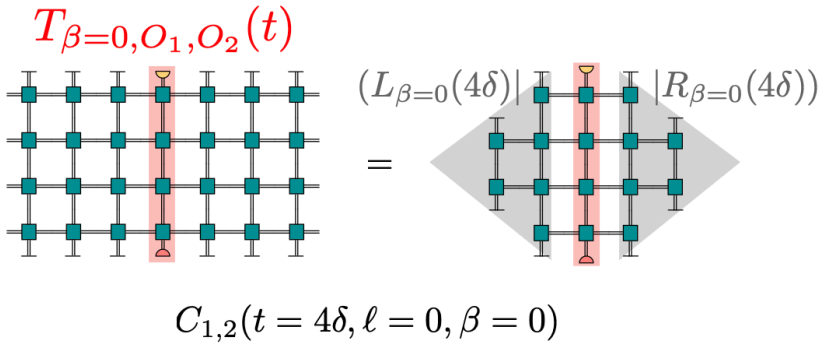


Figure 4.A.3: TN for the thermal correlators at infinite temperature. In the particular case $\beta = 0$, the thermal MPO is exact and proportional to identity, and local unitaries cancel around local O_1, resulting in a diamond shape.

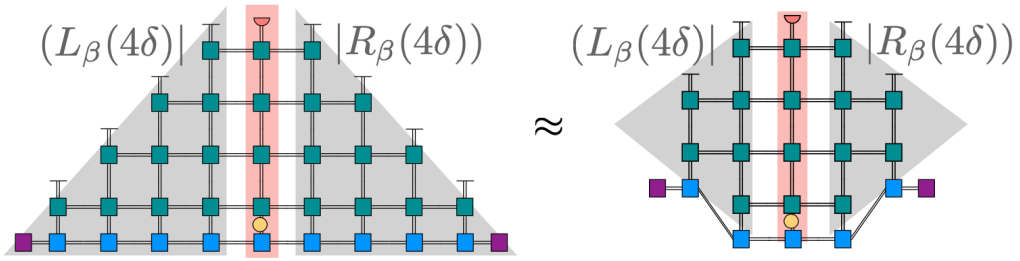


Figure 4.A.4: TN for the thermal correlators at finite temperature. In the general case, the thermal MPO is not exact, and the diamond-shaped TN will approximate the full one.

applying O_2 (possibly on a different site) and taking the trace. This results in a TN that is periodic in the time direction. Folding (or flattening) it results in a doubled TN, similar to the one obtained for pure state evolution, as illustrated in Fig. 4.A.2.

The simplest case is that of infinite temperature ($\beta = 0$), when the thermal state has an exact MPO representation with bond dimension one, since $\rho_{\beta=0} \propto \mathbb{I}$. Then the local unitary matrices that represent the real-time evolution cancel also around O_1 exactly, and the TN to be contracted has a diamond shape (see fig. 4.A.3).

At arbitrary temperature, the cancellation around O_1 is no longer exact, since the MPO representation of the thermal state is only approximate, and local unitaries do not commute exactly with it. Thus we can consider the diamond-shaped TN in this case to be an approximation of the infinite one (see Fig. 4.A.5). We expect this to introduce a small error. Notice that in more standard evolution algorithms (i.e. those that evolve the MPS in real-time), exploiting this light cone structure has also been shown to be useful in infinite systems, for instance by considering an

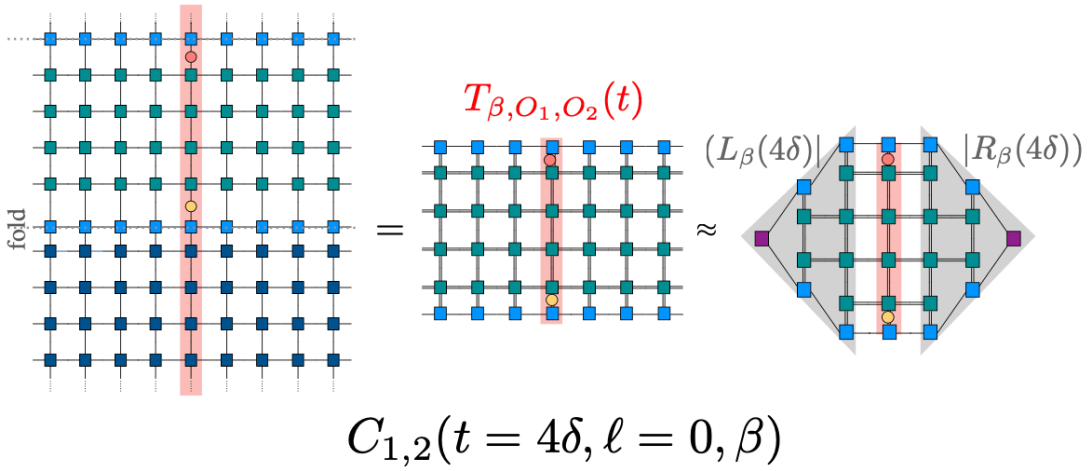


Figure 4.A.5: Alternative construction of the TN for the thermal correlators using the purification structure.

expanding window embedded in an infinite MPS [146–148].

Finally, notice that the TN construction described above is not unique. If we make use of the property

$$\begin{aligned} UO_1e^{-\beta H}U^\dagger &= UO_1e^{-\beta H/2}U^\dagger Ue^{-\beta H/2}U^\dagger \\ &= UO_1e^{-\beta H/2}U^\dagger e^{-\beta H/2}, \end{aligned} \tag{4.A.3}$$

which has been previously exploited for the simulation of real-time evolution of thermal states [155–157], and apply also the cyclic property of the trace, we can write the quantity of interest as (up to a normalization factor)

$$C_{1,2}(t,\ell,\beta) \propto \mathrm{Tr}\left(e^{-\beta H/2}O_2U(t)O_1e^{-\beta H/2}U(t)^\dagger\right). \tag{4.A.4}$$

This results in a different TN, illustrated in Fig. 4.A.5, in which the upper and lower boundaries are given by the purification tensors. For infinite temperature, both constructions are equivalent, but for the general case, they will give rise to different approximation errors. A systematic analysis of these alternatives, the approximation error, and its dependence on temperature will be carried out elsewhere.

4.B Error estimates for the different approaches

In the insets of Figure 2 of the main text, we show the scaling of the bond dimension needed with time to maintain fixed precision in the global quench scenario. Here in

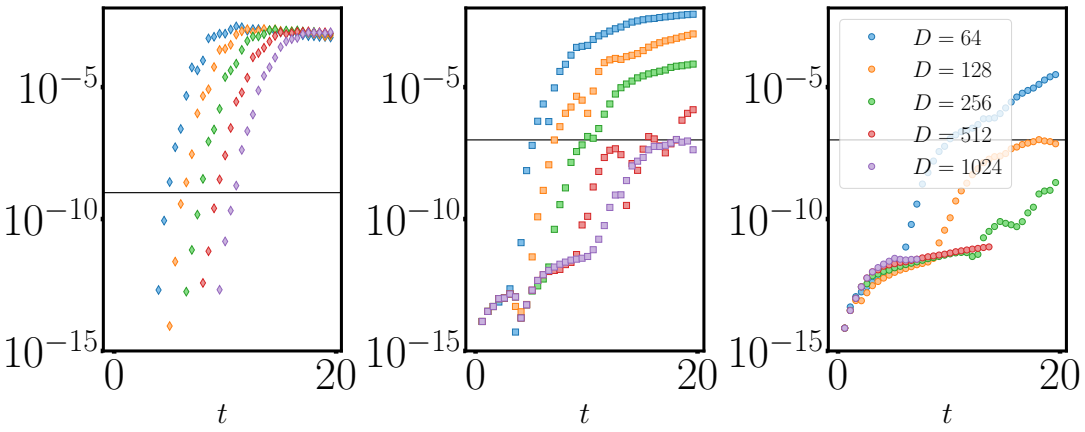


Figure 4.B.1: Scaling of the error measures for the different algorithms in the non-critical integrable case ($J = 1, g = 0.5, h = 0$). The left plot shows the growth of the sum of the squares of the discarded Schmidt weights for the different bond dimensions simulated with TEBD, and the center and right plots show the deviation of the expectation value of the identity from one as a function of time for the TLCC algorithm with the standard MPS truncation (center) and hybrid truncation from [134] (right). The black lines in each plot show the constant level of error we use to make the scaling in the insets of Fig. 4.1.2 of the main text.

this appendix, we provide some extra information on how the scaling is computed and what we mean by fixed precision.

A standard way to bound the error in TN simulations is to sum the squares of the discarded Schmidt weights in every truncation [152]. For the results from TEBD and Heisenberg-picture DMRG, we use that as a measure of the growth in the errors during the simulation. Notice that in the case of TEBD, as we are simulating an infinite system that measure is only an heuristic, as it does not provide an upper bound in the error that one can incur when evaluating some observable. For the TLCC, both when we use the standard and the hybrid truncation from [134] to truncate the boundary vectors, we keep track of the deviation of the expectation value of the identity. As explained in the main text, knowledge of the boundary vectors $(L(t)|$ and $|R(t))$ gives access to out-of-equilibrium expectation values by computing the expectation value $(L(t)|E_O(t)|R(t))$. That includes the possibility of computing the expectation value of the identity, which should be one for any properly normalized state. The deviation with respect to this value, when using MPS approximations for $(L(t)|$ and $|R(t))$, gives a good measure of the error of the TLCC.

In order to perform the scaling analysis, we compute as a function of time the quantities mentioned above for the different TN algorithms with simulations with

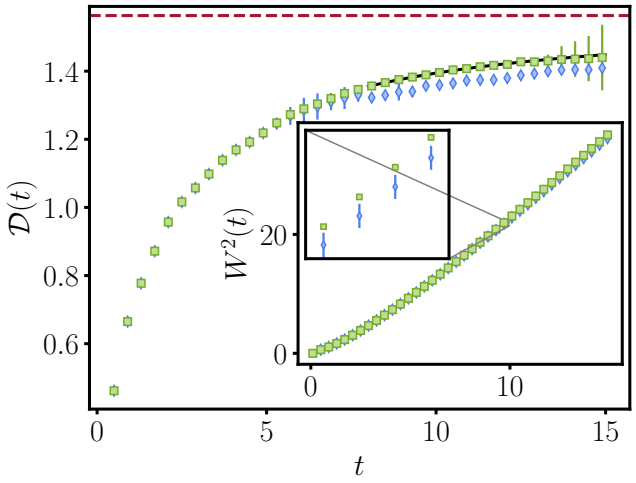


Figure 4.C.1: Spatial variance (Eq. 4.3.2) of the normalized autocorrelations (inset) and corresponding diffusion constant (main plot) in the non-integrable Ising case with $(J = 1, g = 1.4, h = 0.9)$ obtained from the TLCC (green squares) and TEBD (blue diamonds). The solid black line in the main plot shows an heuristic fit of the form $\mathcal{D}_E \exp(b/t)$, which predicts the asymptotic value $\mathcal{D}_E \approx 1.6$ (red dotted line). The bond dimensions used for both methods are 1024, and the error bars show the difference with the results of bond dimension 512.

different bond dimensions. Setting a constant value of the precision, that is, of the truncation errors in the TEBD and Heisenberg-picture DMRG cases and of the deviation from the identity for the TLCC, we can keep track of the times when the different bond dimensions exceed the desired precision threshold.

4.C Alternative values of the parameters

In the main text, for simplicity, we focused on a particular point in parameter space in the non-integrable case. Here, we show results obtained in a different case, described by the couplings $(J = 1, g = 1.4, h = 0.9)$ and studied in [127, 161]. The spatial variance and diffusion constant obtained with TLCC for this case are shown in Fig. 4.C.1, along with results obtained with TEBD. As seen in the main text, the introduction of TLCC allows in this case as well to extend the range where it is possible to reliably simulate out-of-equilibrium dynamics by a factor of around two.

Chapter 5

Long-time evolution with tensor networks

This chapter is based on the publication
Miguel Frías-Pérez, Luca Tagliacozzo, Mari Carmen Bañuls
Converting Long-Range Entanglement into Mixture: Tensor-Network Approach to Local Equilibration
Phys. Rev. Lett. 132, 100402 (2024)
Structure, text and figures have been adapted for this thesis

5.1 Overview

During the out-of-equilibrium evolution of quantum many-body systems, initially localized correlations can propagate over arbitrarily large distances. As a result, the entanglement in the system increases rapidly [116, 117, 163–166] and simple TN ansatzes have limited applicability. By focusing on the local description of the state, nevertheless, one can circumvent the exponential complexity originating from such non-local correlations. Significant simplifications using this approach have already been observed [123, 126–130, 132, 167–170]. In this chapter, we develop this idea and propose a new algorithm that explicitly identifies long-range entanglement in the system and trades it for mixture (see also [129]). In particular, our method focuses on separating fast and slow propagating degrees of freedom. Already from an early time we observe that the fast degrees of freedom mediate non-local correlations whose effect on local observables is that of a statistical mixture. Using that knowledge, we devise a TN algorithm that provides an effective description of the time-evolved state as a density matrix, represented by a matrix product operator (MPO) with bounded bond dimension, and accurately

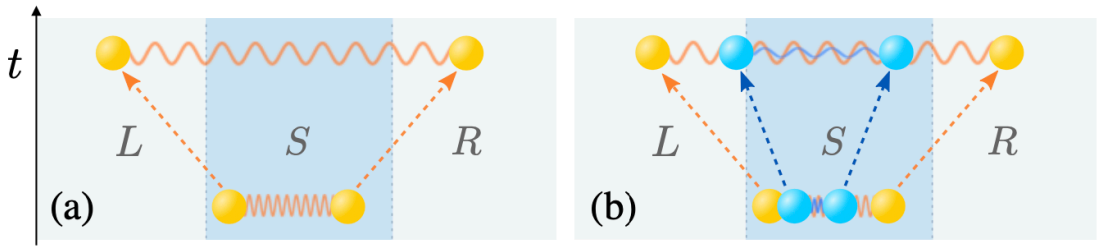


Figure 5.1.1: (a) A correlated QP pair initially located inside region S creates long-distance (pure LR) entanglement when both QPs propagate outside of S. (b) At that time, slower QPs (blue) can still contribute to entanglement between S and its surroundings.

captures the long-time behavior of local operators.

To benchmark the algorithm, we simulate quenches in the transverse field Ising model and its non-integrable generalization. We find that our results agree with those dictated by, respectively, the analytical solutions and the prediction of the diagonal ensemble [171–173].

5.1.1 Quenches and quasiparticles

We focus on quantum quenches [116, 174], in which a one-dimensional system in the thermodynamic limit is prepared at $t = 0$ in a product state $|\psi_0\rangle$, and later evolved with an entangling Hamiltonian H. Our aim is to compute the time-dependent expectation value of a local observable O.

The initial state has a finite energy density above the ground state, and thus a large occupation of excited states, often described as quasiparticles states (QP). The subsequent dynamics can be described as the radiation of entangled pairs of QPs [116]. In translational invariant systems indeed QPs possess a well-defined dispersion relation $\epsilon(k)$, and a QP wave-packet centered at k_0 propagates with group velocity $v_{k_0} = \partial_k \epsilon(k)|_{k_0}$, while momentum conservation enforces equal occupation of states that propagate with opposite momenta.

5.1.2 Identifying long-distance entanglement

In order to identify the long-distance contributions to the entanglement we need to focus on a subsystem S of ℓ neighboring spins, and label L and R the remaining left and right regions of the system (see Fig. 5.1.1). Entanglement across a bipartition (e.g. L vs SR) is created by correlated pairs of QPs with support on both sides of the cut. We define as *long-distance entanglement* the contribution to the entanglement generated by any QP pair supported on L and R, but not on S.

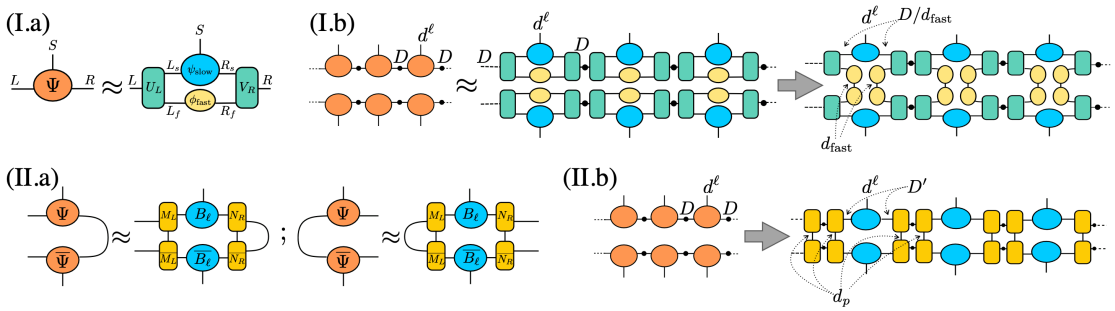


Figure 5.1.2: First iteration in the transformation of long-range entanglement into mixture in TN. (I.a) Decomposition (5.1.3) for a block S, represented by a single tensor; (I.b) The density matrix of the full state (left) can be written in terms of the new tensors (middle) when we apply the decomposition to each block [black circles represent the inverse Schmidt values matrix, inserted to compensate the central gauge in (I.a)]. We then substitute each long-range component with the product of its marginals, giving rise to a mixed description (right). (II.a) The heuristic algorithm directly searches for tensors that (approximately) preserve the reduced density matrices for LS and SR. (II.b) The density matrix for the full state is replaced by a purification defined by the solution. The structure is analogous to that in (I.b) (in each case we have indicated the relevant dimensions).

We can visualize this in the cartoon of Fig. 5.1.1(a), where we consider a quench that excites a single pair of entangled QPs, initially located inside S, that later travel with opposite velocities $\pm v$. [1] In this simple example, S becomes entangled with the rest when one member of the pair leaves the region, but after $t \sim \ell/v$ both QPs have left, and the total state factorizes into the product of an entangled state of LR and a state of S.

In a more general scenario such complete factorization is not possible, since S and LR might contain other sources of entanglement. The simplest generalization depicted in Fig. 5.1.1(b) involves two pairs of QPs, one *fast* (orange) and one *slow* (blue). When the former has left S the latter is still partially in S. At that stage, it is still possible to disentangle a part of L and R from S. We can indeed identify a factorization of the Hilbert spaces, $L \equiv L_s \otimes L_f$ and $R \equiv R_s \otimes R_f$, such that the subsystem $L_f \otimes R_f$ is not entangled with S. In the simple scenario presented above, $L_f \otimes R_f$ is defined by the degrees of freedom that describe the fast pair, and we can factorize the state as $|\psi^{(\text{slow})}_{L_s S R_s}\rangle \otimes |\phi^{(\text{fast})}_{L_f R_f}\rangle$, where the second factor captures the pair of fast modes.

[1] This kind of simple dynamics is actually realized in a toy model of local unitaries [123, 133] and in some concrete physical realizations such as the split quenches studied in [175, 176].

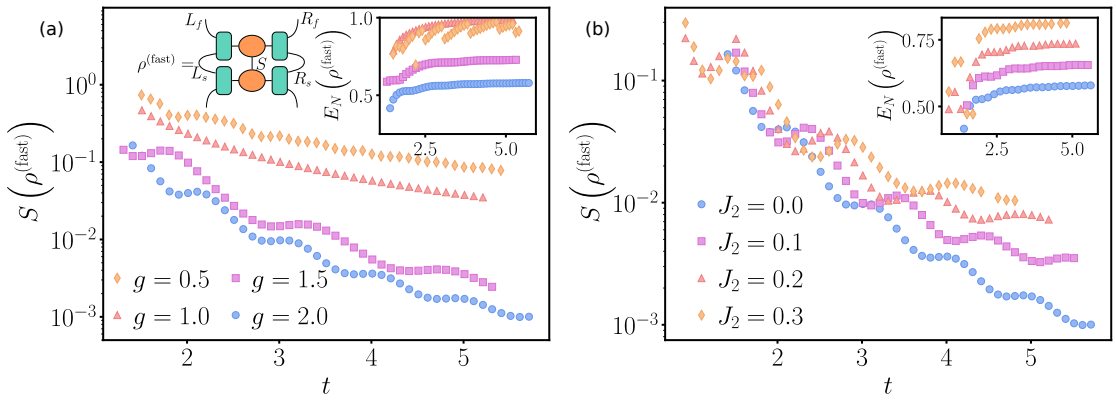


Figure 5.1.3: Entropy (main plots) and left-right logarithmic negativity (insets) of the reduced density matrix for the *fast* degrees of freedom [see sketch in panel (a)], computed from the decomposition in fig. 5.1.2(I.a) at different times (a) for several integrable quenches ($g \neq 0$, $J_2 = 0$) and (b) non-integrable ones ($g = 2$, $J_2 \neq 0$).

5.1.3 Trading long-distance entanglement for mixture

In a translationally invariant system, we can repeat the cartoon picture above for each block of ℓ sites. When computing the expectation value of a local observable, all the degrees of freedom are traced out, except those supporting the observable. For fast entangled pairs separated by at least ℓ sites, one partner will necessarily be traced out, leaving the other in a mixed state.

Thus, when focusing on local observables, we can describe the system as a mixed state where, for each fast pair, we have substituted the entangled state $\rho^{(\text{fast})}_{L_f R_f}$ by $\tilde{\rho}^{(\text{fast})}_{L_f R_f}$, the product of the mixed states obtained by tracing out each partner in the pair,

$$\rho^{(\text{fast})}_{L_f R_f} = |\phi^{(\text{fast})}_{L_f R_f}\rangle\langle\phi^{(\text{fast})}_{L_f R_f}| \to \tilde{\rho}^{(\text{fast})}_{L_f R_f} = \rho^{(\text{fast})}_{L_f} \otimes \rho^{(\text{fast})}_{R_f}, \tag{5.1.1}$$

where $\rho^{(\text{fast})}_{L(R)} \equiv \text{Tr}_{R(L)}\left(\rho^{(\text{fast})}_{LR}\right)$ are the reduced density matrices of each QP. This provides a more efficient local description in terms of entanglement, as the long-distance components have been removed.

5.1.4 From the QP intuition to a TN algorithm

Standard MPS algorithms for this setup attempt to represent the time-evolved state as a MPS, parametrized by one or few tensors A_m of size $d \times D \times D$, where d is the physical dimension of the chain sites and D the bond dimension [16, 40, 177]. Since the half-chain entropy of the MPS is upper-bounded by $\log D$ [38,

119, 178], the typical linear growth in time of entropy after a quench [116, 163] requires an exponential increase of the dimensions of the tensors to maintains a constant precision. As a result, given finite computational resources, standard TN algorithms [152] only give reliable predictions for relatively short times (typically of the order $Jt \simeq 10$ with J the relevant energy scale).

Translating the QP intuition above to the TN setting we can however obtain a more efficient TN description for the local observables. Let us, for simplicity, assume that each MPS tensor represents precisely ℓ sites [2].

Singling out one subsystem, the whole state can thus be written as

$$|\Psi\rangle = \sum_{\alpha s_\ell \beta} C^{s_\ell}_{\alpha\beta} |\Phi^L_\alpha\rangle |s_\ell\rangle |\Phi^R_\beta\rangle, \tag{5.1.2}$$

where the sum is over orthonormal bases of the block and its left and right environments $\{|\Phi^{L/R}\rangle\}$, meaning that we use the gauge in which the tensor $C^{s_\ell}_{\alpha\beta}$ for the subsystem $\{|s\rangle\}$ is the orthogonality center of the MPS [16, 40, 119, 179].

If there are long-range entangled degrees of freedom that decouple from S, the state will have a product structure, with one component completely disentangled from the physical degrees of freedom. There will thus exist basis transformations (disentanglers [18, 180]) on L and R identifying the decomposition $L = L_f \otimes L_s$ and $R = R_f \otimes R_s$, such that

$$U_L^\dagger \otimes \mathbb{I}_S \otimes V_R^\dagger |\Psi\rangle = |\psi_{L_s S R_s}\rangle \otimes |\phi_{L_f R_f}\rangle, \tag{5.1.3}$$

as depicted in Fig. 5.1.2(I.a). If U_L and V_R exist, they can be determined by minimizing the Euclidean distance between the left and right-hand side of (5.1.3) [181]. Namely, given the evolved state in the form (5.1.2), and for fixed dimension d_{fast} of the long-range component, we iteratively optimize each of the disentanglers U_L and V_R and the vectors $|\psi_{L_s S R_s}\rangle$ and $|\phi_{L_f R_f}\rangle$ until the distance converges. Since the dimension d_{fast} of the fast factors L_f and R_f is unknown, the procedure needs to be repeated for different trial values.

Assuming the above procedure succeeds, we can now transform the identified long-range entanglement into mixture by applying the substitution (5.1.1) to each consecutive subsystem in the chain. As schematically shown in Fig. 5.1.2(I.b), this exchanges the pure MPS description of the system by a mixed (purified) MPO with smaller bond dimension D/d_{fast}, at the expense of additional (purification) indices.

When the decomposition (5.1.3) is exact, this mixed state has the property that the reduced density matrices for $S_\ell R$ (resp. LS_ℓ) are unchanged. As a consequence, also the reduced density matrix for two adjacent blocks, and thus

[2] This can be always obtained by blocking together ℓ tensors of the original systems to yield a single tensor.

all expectation values and correlation functions with support on up to ℓ sites are preserved. Larger supports can be achieved working with a larger ℓ, at the cost of longer disentangling times. We can then continue the evolution with the original Hamiltonian acting on the physical indices. Fast degrees of freedom will continue propagating entanglement through the subsystem S. Thus, the step of finding and mixing the long-range entangled components is repeated periodically.

5.1.5 An improved TN algorithm

In practice, the best decomposition we identify using this strategy retains some residual entropy between fast and slow degrees of freedom $S(\rho_{\text{fast}})$. We apply the truncation when this entropy falls below a given threshold η_S. This results into small errors when building the mixed state that reflect in the evolution of local observables [see Fig. 5.1.4(a)]. Therefore we introduce an improved heuristic TN algorithm in which, after solving (5.1.3), if $S(\rho_{\text{fast}}) < \eta_S$ we propose an effective purification ansatz, described by three rank-3 tensors M_L, B_ℓ, N_R with fixed dimensions $d_p D D'$, $d^\ell D'^2$ and $d_p D' D$, where d_p is a chosen dimension for the purification index and $D' = D/d_p$ [see Fig. 5.1.2(c)].

The purification tensors are found variationally by minimizing the distance between the reduced density matrices ρ_{SR} (and ρ_{LS}) obtained from the original state (5.1.2) and from the purification. As an initial guess, we use the tensors obtained from solving (5.1.3), and we use gradient descent for the minimization problem.

The dimension of the purification indices doubles after each mixing step. In practice, after a few steps we keep it fixed by discarding the smallest eigenvalues of its reduced density matrix since we observe that the error induced by this truncation is negligible compared to other sources of errors in the simulation [3]

5.2 Numerical results

We benchmark the algorithm using the transverse field Ising model with an additional integrability-breaking next-to-nearest-neighbor interaction,

$$H = -\sum_i \left(\sigma_i^z \sigma_{i+1}^z + g\sigma_i^x + J_2 \sigma_i^z \sigma_{i+2}^z\right). \tag{5.2.1}$$

Initially in a product state $|X+\rangle \equiv \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, corresponding to the $g \to \infty$ ground state, the system is quenched to finite values of g, and potentially J_2.

[3] Other schemes would be possible to try and optimize the purification sites, for instance, in the spirit of the optimal purification of [182].

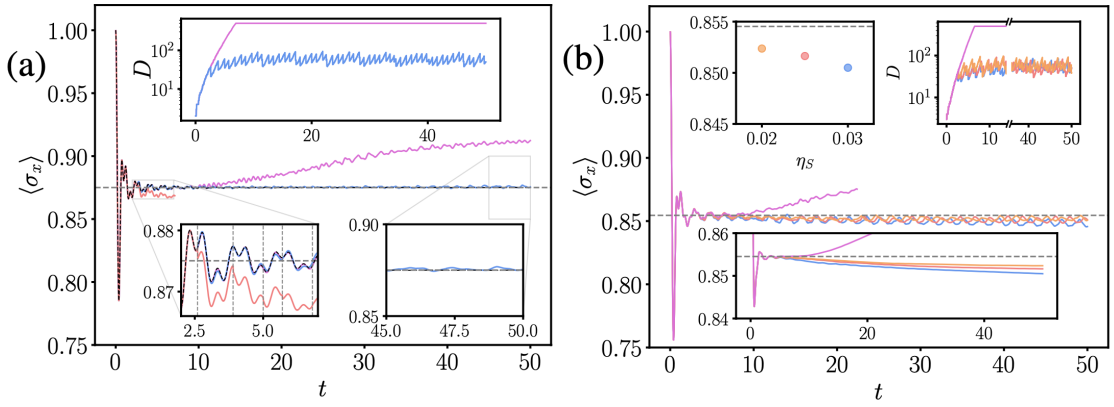


Figure 5.1.4: (a) Evolution of the transverse magnetization for the integrable quench ($g = 2$, $J_2 = 0$) as obtained with the direct decomposition of fig. 5.1.2(I) (orange) and the heuristic algorithm of fig. 5.1.2(II) (blue). For comparison, we show the exact result (dashed black line), the equilibration value (dashed grey) and the iTEBD result with $D = 500$ (pink). The lower insets show close-ups of the short- and long-time regions. In the left one, the vertical dashed lines indicate mixing steps, inducing jumps in the simple but not the heuristic algorithm. (b) Same observable for the non-integrable quench ($g = 2$, $J_2 = 0.1$) using the heuristic algorithm. Different colors indicate different residual entropy thresholds η_S. Our simulations converge towards the diagonal ensemble value $\langle\sigma_x\rangle_{\rm DE}$ (dashed gray line). This can be more clearly seen in the time average of the observable (lower inset) or the value of the time average at time $t = 50$ as a function of η_S (upper-left inset). For both quenches, the upper (right) inset shows the bond dimension as a function of time.

To explore systematically the structure of the time-evolved MPS, we simulate accurately the short-time evolution using the infinite time-evolving block decimation (iTEBD) algorithm [16] with a large enough bond dimension. We then probe the disentangling of fast and slow degrees of freedom at different times solving (5.1.3) for a subsystem of $\ell = 2$ sites for several quenches.

As shown in Fig. 5.1.3, after a certain time we can identify subsystems of L and R that practically disentangle from the local region and carry long-range entanglement between the environments, measured by the logarithmic negativity [183, 184] of the reduced density matrix for the fast degrees of freedom $\mathrm{Tr}_{L_s S R_s} |\Psi\rangle\langle\Psi|$. For all quenches, the residual entropy $S(\rho_{\rm fast})$ decays fast with time, with slower rate when quenching into the ordered phase [4]. As expected, the time at which fast degrees

[4]This is consistent with the distribution of momentum excitations in the initial state, computed in the exact free-fermionic formulation, which is peaked at finite velocities when the quench is within the same phase, and at zero velocity otherwise (see the Appendix 5.D).

of freedom start decoupling from the local system depends linearly on ℓ [inset of Fig. 5.1.3(a)].

We now use the TN algorithm described above and check its performance simulating the long-time evolution of local observables. Fig. 5.1.4(b) shows the results for the integrable quench $g = 2$, $J_2 = 0$ and the only non-vanishing (due to the Z_2 symmetry) single-site observable $\langle\sigma_x\rangle$. Solving (5.1.3) and trading entanglement by mixture captures the qualitatively correct dynamics, but introduces discontinuous jumps (orange line and lower-left inset) that we attribute to the decomposition not being exact. The improved heuristic algorithm achieves much better results (blue line and lower-right inset). While the approximation induces some residual oscillations, these are very small and the results are close to the exact equilibration even after considerably long times. In contrast, iTEBD results with a maximal bond dimension $D = 500$ (pink line) start to severely deviate from the exact solution at around $t = 10$.

Our algorithm uses instead a much smaller bond dimension: during the time evolution steps, the required bond dimension of the purified MPS grows exponentially, but each time we trade some of the entanglement for mixture, the bond dimension gets halved [5]. As we iterate the procedure, the maximum required bond dimension tends to a constant $D \sim 100$ (upper inset).

In Fig. 5.1.4(b) we repeat a similar analysis for the non-integrable quench $g = 2$, $J_2 = 0.1$. The different colors show the results for different thresholds for the residual entropy η_S. Smaller values require longer evolution between mixing steps, and thus larger bond dimension, but improve the results systematically.

Our results systematically approach the long-time limit predicted by the diagonal ensemble $\langle\sigma_x\rangle_{\mathrm{DE}} = 0.852$ [6].

Even for the largest threshold, the relative error in the time-averaged value after $t \sim 50$ is below 1% (0.848). Also the time-averaged magnetization exhibits a similar precision, as shown in the lower inset. Also in this case the largest bond dimension saturates with t (upper right inset).

In the appendices, we also repeat the calculations in the integrable case directly using the free-fermionic formalism [185] where we can track the coherence we discard in our simulation and confirm it does not play a relevant role in the long-time dynamics of local observables.

5.3 Discussion

By identifying the long-range entanglement produced in the out-of-equilibrium dynamics after a quantum quench and converting it into mixture, we have pro-

[5] We use $d_p = 2$ for our ansatz.

[6] Extrapolated from exact finite size results.

posed an explicit approach to avoid the entanglement barrier and simulate the out-of-equilibrium dynamics of an infinite quantum chain with MPS using finite computational resources.

Our approach is inspired by the intuitive understanding of entanglement dynamics in terms of the radiation of QP. It relies on the hypothesis that fast degrees of freedom propagate correlations to steadily growing distances, contributing to the linear growth of entanglement across the system, but only as statistical mixture to sufficiently local observables. Our numerical results for the Ising chain show that the intuition is accurate in the free-fermionic case, which best fits the QP picture.

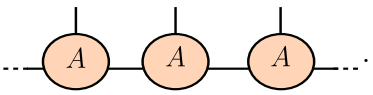
We have generalized our intuition to a heuristic algorithm that goes beyond the QP picture and also performs well in non-integrable regimes of the model. It is thus important to pursue further characterization of the algorithm in order to chart its potential and limitations.

5.A Description of the algorithms

In the main text, we discussed a way to transform the long-range entanglement generated in a quench into mixture using tensor networks. The transformation is such that it reduces the bond dimension of the TN, while aiming to preserve the local observables and their evolution. In this appendix, we provide a complete description of the algorithms we use for that purpose. First, in order to set the notation and introduce some useful concepts, we briefly describe the uniform MPS formalism. For a more complete description, we refer the reader to the excellent review [40], whose notation we follow here.

5.A.1 Uniform MPS

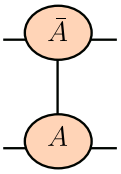
Uniform MPS (uMPS) are TN states that represent one-dimensional quantum many-body states in the thermodynamic limit. These states are constructed by repeating infinitely many times a tensor along a line (or repeating in a cyclic manner a finite set of them, if we work with a unit cell larger than one). In graphical TN notation,

$|\Psi(A)\rangle =$ 

. (5.A.1)

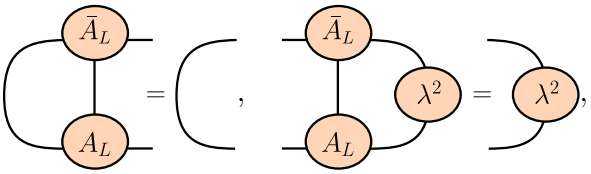
Where A is a $d \times D \times D$ tensor, d being the physical dimension of the sites of the chain and D the bond dimension. A central object in uMPS simulations is the

transfer operator, defined as the following four-legged tensor:

$E(A) =$  . (5.A.2)

We can interpret this object as a $D^2 \times D^2$ matrix (left vs. right indices). Many of the important properties of the state are captured by the dominant left and right eigenvectors of this matrix. For instance, they determine the reduced density matrix of any (connected) subsystem and thus appear in the calculation of any local expectation value. In the generic case, these dominant eigenvectors are non-degenerate [119]. If the state is normalized, the largest eigenvalue is 1, and the leading eigenvectors correspond to the fixed points of the transfer operator seen as a map from the left to the right indices and vice-versa.

It is possible to exploit the gauge freedom in the TN to transform these eigenvectors into a canonical form, which is particularly simple. In particular, in the so-called *left-canonical* form, the MPS tensor for a normalized state satisfies the conditions

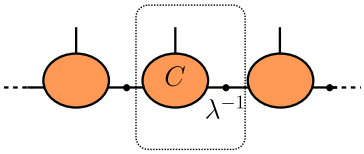
 (5.A.3)

where λ^2 is a diagonal matrix with positive entries. Cutting (5.A.1) in half across one cut defines D states for the left half-chain $|\Phi^L_\alpha\rangle$, and the same number for the right half-chain, indexed by the (now open) virtual leg. The first condition in (5.A.3) ensures that the D states on the left half-chain are orthonormal. Alternatively, one can impose a *right-canonical* form A_R in which the conditions are the symmetrical ones and orthonormality is satisfied by the states to the right of the cut.

In practice, it is instead common to work in a *mixed* gauge, in which one can define [40] a single central tensor $C^s_{\alpha\beta} = \sum_\gamma (A_L)^s_{\alpha\gamma}\lambda_{\gamma\beta}$, where α (γ) is the left (right) virtual index of the tensor A_L, and s denotes its physical index. The state can then be written as a half-chain of left-canonical tensors to the left of C and a half-chain of right-canonical ones to the right, such that both left and right virtual bonds of C are in orthonormal bases. Taking the tensor product of these two bases with the single-site physical basis, the state can be written as,

$$|\Psi\rangle = \sum_{\alpha s \beta} C^s_{\alpha\beta} |\Phi^L_\alpha\rangle |s\rangle |\Phi^R_\beta\rangle. \tag{5.A.4}$$

We can rewrite the full uMPS in terms of the single tensor C, as



(5.A.5)

In this work, we are interested in the long-range entanglement between distant regions of the system L and R, separated by a block S of finite size ℓ. Thus, it is convenient to use the central canonical form with respect to the whole block S, such that we can study the long-range entanglement at the level of the virtual indices of the MPS.

After blocking the ℓ sites of the subsystem S, the wave function of the system can be written as

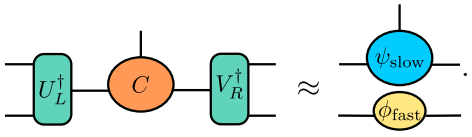
$$|\Psi\rangle = \sum_{\alpha s_\ell \beta} C^{s_\ell}_{\alpha\beta} |\Phi^L_\alpha\rangle |s_\ell\rangle |\Phi^R_\beta\rangle. \tag{5.A.6}$$

As we will see in the next three subsections, our study of the long-range entanglement will allow us to construct a low-rank decomposition of the tensor C when partitioned between L and SR (or LS versus R), at the cost of introducing indices that couple the tensors in the bra and the ket. This decomposition will be such that the local expectation values and their evolution will be preserved, while the long-range coherences will be discarded.

5.A.2 Detection of long-range entanglement

In this subsection we describe the algorithm that we use to identify the long-range entanglement in our MPS.

As discussed in the main text, if there are long-range entangled degrees of freedom that decouple from S, the state will factorize in a tensor product with one component completely disentangled from the physical degrees of freedom in S. There will thus exist basis transformations on L and R identifying the decomposition $L = L_f \otimes L_s$ and $R = R_f \otimes R_s$, such that the degrees of freedom in the Hilbert spaces L_f and R_f are in a tensor product structure with the rest of the system. Graphically,



(5.A.7)

To look for such a decomposition we numerically minimize the square of the Euclidean distance between the left and the right-hand side of the previous diagram (5.A.7). The variables in the optimization problem are the unitary matrices U_L and V_R, which respectively determine the left and right basis transformations,

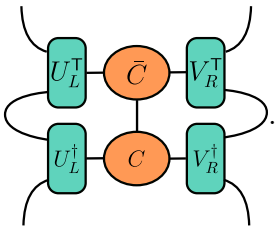
and the two resulting vectors, $|\psi_{L_sSR_s}\rangle$ and $|\phi_{L_fR_f}\rangle$. The full optimization is a complicated non-linear problem. However, as a function of each of the individual pieces, the cost function is quadratic and can be solved using basic linear algebra. Hence, to find the decomposition we use an alternating scheme, first used in [181]: we fix all the variational tensors but one, and substitute the free piece by the one that minimizes the value of the cost function. When the unitaries are fixed, the optimal for the two vectors are just the leading left and right singular vectors across the partition L_sSR_s versus L_fR_f. For the case of the unitaries, after imposing the unitarity constraint, the cost function is linear in each of them. That kind of problem can also be easily solved with a singular value decomposition of the environment [180].

The algorithm is efficient, with cost scaling as $O(D^3)$ with the bond dimension, and we find that it shows a good performance. In particular, we find components of the wave function that mediate long-range entanglement and decouple from the slower degrees of freedom, as expected according to the intuition drawn from the quasiparticle picture.

5.A.3 Simple truncation algorithm

In the previous section we have discussed the TN algorithm we use to identify the degrees of freedom that contribute to the long-range entanglement in our MPS representation of the time-evolved state. In practice, we combine the previous algorithm with an iTEBD simulation that provides us with a faithful representation of the state at short times and later allows us to evolve the local tensors in the ansatz.

Starting from a pure state (uMPS) description of the initial state, we evolve unitarily with iTEBD. At periodic time intervals, we look for the above decomposition of the simulated wave function. The extent to which our time-evolved state has the sought structure can be characterized in several ways. We choose to use the entanglement entropy between the fast and slow degrees of freedom, which can be identified after we have transformed the basis of the bond dimension indices according to the found unitaries. We thus need to compute the entropy of the reduced density matrix $\rho^{(\text{fast})}_{L_fR_f} = \text{Tr}_{L_sSR_s} |\Psi\rangle\langle\Psi|$, which can be graphically represented as

$\rho^{(\text{fast})} =$ 

. (5.A.8)

A vanishing entanglement entropy means that our state has, in the found virtual

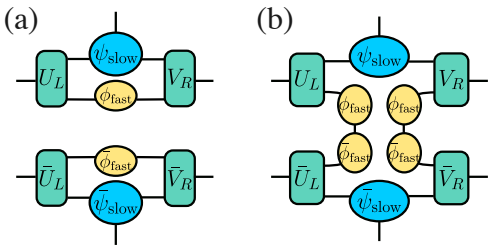


Figure 5.A.1: Graphical representation of the density matrices of the entire state before (a) and after the truncation of the fast degrees of freedom (b).

bases, exactly the desired tensor product structure, $|\Psi\rangle = |\psi^{(\text{slow})}_{L_s S R_s}\rangle \otimes |\phi^{(\text{fast})}_{L_f R_f}\rangle$. In this case, it is possible to modify the state in such a way that local expectation values are preserved while discarding some long-range correlations, as described in the main text. More concretely, we substitute the density operator of the full system $|\Psi\rangle\langle\Psi|$, represented graphically in Fig. 5.A.1 (left), by a mixed state (right) that, in the basis of the virtual degrees of freedom in which the state factorizes, corresponds to

$$|\psi^{(\text{slow})}_{L_s S R_s}\rangle\langle\psi^{(\text{slow})}_{L_s S R_s}| \otimes \rho^{(\text{fast})}_{L_f} \otimes \rho^{(\text{fast})}_{R_f}, \tag{5.A.9}$$

with $\rho^{(\text{fast})}_{L(R)} \equiv \text{Tr}_{R(L)}\left[\rho^{(\text{fast})}_{LR}\right]$ as defined in the main text.

The state on Fig. 5.A.1(a) is an exact representation of the tensor C in the case in which the time-evolved state presents the exact sought structure. The state on the right (Fig. 5.A.1(b)) has a different structure: it represents a mixed state. Despite the two states being globally different, the two have the same reduced density matrices on the subsystems LS, as graphically shown in Fig. 5.A.2. An analogous computation shows that the reduced density matrix for SR is also preserved by this substitution. Furthermore, as can be seen from Fig. 5.A.1, the rank of the state (or bond dimension) across the cut L versus SR is reduced by a factor which is given by the dimension of the fast degrees of freedom.

Because the state of the chain can also be explicitly written as a sequence of blocks, using the form (5.A.5), we can perform the same transformation in all blocks of the chain at the same time [i.e., we substitute all the tensors C in (5.A.5) as shown in Fig. 5.A.4]. After this substitution, it can be shown that the leading left and right eigenvectors of the transfer operator for the block are unchanged (see Fig. 5.A.2). This can be used to show graphically that the reduced density matrices of up to two blocks (2ℓ consecutive sites) are preserved (Fig. 5.A.3).

In practice, as can be seen from Fig. 3 of the main text, the decomposition we obtain variationally never reaches full separation between fast and slow degrees of freedom and as a result the substitution is never exact. However, the fast-slow entropy seems to decay fast with time. Our simple truncation consists

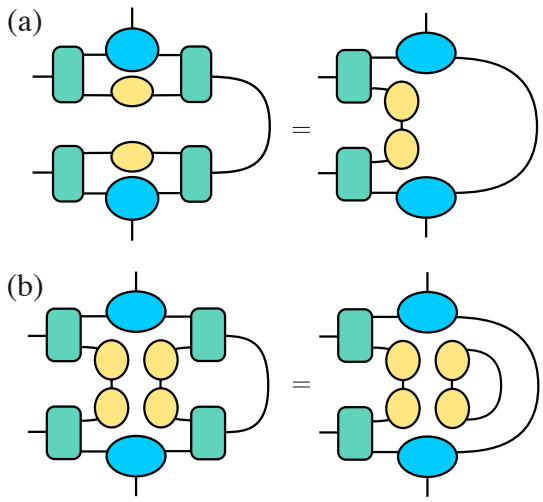


Figure 5.A.2: Reduced density matrix on the subsystem LS before (a) and after (b) the truncation. Using the unitarity of V_R, one can check that the reduced density matrices are identical (right-hand side of the equalities). The state $|\phi^{(\text{fast})}_{L_f R_f}\rangle$ is normalized, and thus, the tensor contraction in the right-hand side of the equation in (b) gives 1. This same diagram also shows that the fixed points are unchanged by the mixing step: notice that by tracing subsystem R, we get a condition equivalent to the second equality in 5.A.3 (namely, that $\sum_{s_\ell \beta} C^{s_\ell}_{\alpha\beta} \bar{C}^{s_\ell}_{\alpha'\beta} = (\lambda^2)_{\alpha\alpha'}$). Both the state before and after the truncation give the same resulting tensor. A similar argument can be made for the left fixed point.

just on performing an iTEBD simulation, while monitoring the "disentangling" entanglement entropy. When it falls below a certain threshold, we substitute all the C tensors on the present state by its tensor product approximation, and mix the fast degrees of freedom according to the diagrams above. After that, because the last step is not exact, we normalize again the state. The results from this evolution are the orange line in Fig. 4(a) of the main text. In them, it can be appreciated that at the times where the truncation is performed, we incur in an instantaneous error in the local observable. However, as we evolve the state again the dynamics seems to be correctly captured modulo the instantaneous shift introduced by the truncation. The same phenomenon repeats itself as we perform more and more truncations in the same simulation. A closer-up figure of the results for this truncation in the main text can be seen in Fig. 5.A.5. The qualitative agreement with the true dynamics of this simple algorithm corroborates the intuitive picture discussed in the main text, and justifies using these tensors as starting point for an improved method that corrects for those discontinuous jumps in a heuristic manner. We discuss the improved technique in the next section.

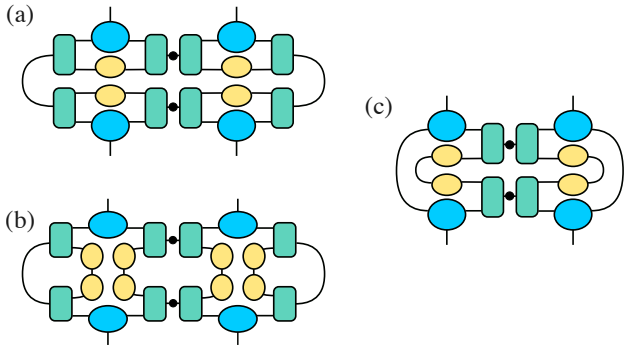


Figure 5.A.3: Reduced density matrices on two consecutive blocks of the system, before (a) and after (b) the truncation of the fast degrees of freedom. As explained in the text, the fixed points are also unchanged by the transformation. Using the unitarity of the two basis transformations in the virtual degrees of freedom, one can show that the expectation value for both is given by the same contraction (c).

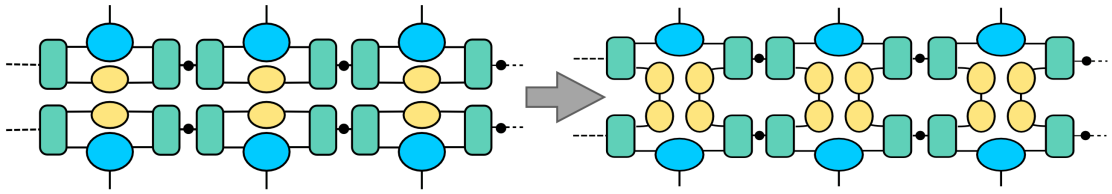


Figure 5.A.4: After identifying the long range entanglement, we truncate the fast degrees of freedom in all block simultaneously.

5.A.4 Heuristic truncation

Inspired by the results of our initial truncation, we present in this section a small modification of the previous algorithm that attempts to correct the shifts induced by the fact that we do not find an exact factorization between fast and slow degrees of freedom.

Our improved heuristic algorithm takes advantage of this fact and tries to find an alternative purification form that generalizes the one of the simple algorithm and guarantees the correct marginals for LS and SR. The ansatz we use is composed by three tensors M_L, B_ℓ and N_R. We initialize their values from the pieces of the decomposition recovered by the simple algorithm and modify them to ensure the desired property (see Fig. 5.A.6 and Fig. 5.A.7).

In order to improve the purification, we minimize the sum of the Euclidean distances between the new and previous reduced density matrices, namely we are interested in

$$\text{argmin}_{M_L,B_\ell,N_R} \left(\|\rho_{LS} - \tilde{\rho}_{LS}\|^2 + \|\rho_{SR} - \tilde{\rho}_{SR}\|^2 \right),$$

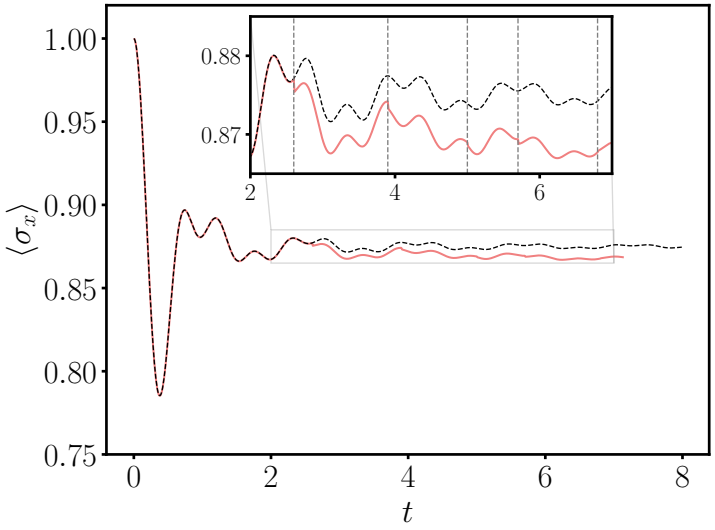


Figure 5.A.5: Evolution of the transverse magnetization for the integrable quench ($g = 2, J_2 = 0$) as obtained with the algorithm discussed in the section above, in orange. The exact result is shown by the black dashed line. The inset shows a close-up of the evolution. In the inset, the vertical gray dashed lines show the times in which the truncation is performed.

where $\tilde{\rho}$ denotes the new reduced density matrices obtained from the mixed state defined by the tensors M_L, B_ℓ and N_R, which are the variables in the optimization. To minimize the cost function we use a gradient-descent scheme, as the optimization contains terms which are of fourth order in the entries of the tensors. The dimensions of the tensors, and the fact that they have purification legs guarantee that the original pure state will be approximated by a mixed state with a lower bond dimension.

The same as in the simple algorithm, we combine this decomposition with an iTEBD simulation. As the evolution unfolds, we monitor the behavior of the entanglement entropy between fast and slow degrees of freedom. Once it falls below a pre-established threshold η_S, we perform the truncation described by the direct decomposition and group the pieces obtained in it into the tensors M_L, B_ℓ and N_R, as can be seen in Fig. 5.A.6. Lastly, we use these tensors as a starting guess for the heuristic optimization described in the paragraph above, which modifies them slightly to correct for the small deviation in the reduced density matrices. Once the optimization is stopped, we substitute the pieces for all the tensors C in our chain, normalize again the state and continue the evolution. After the tensors C are substituted by their lower-rank counterparts for the first time, the structure of the uMPS describing the purification contains a sequence of tensors B_ℓ that carry

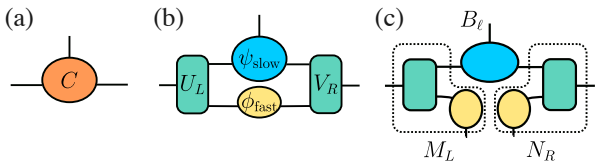


Figure 5.A.6: (a) Tensor C obtained from the uMPS dynamical simulation in the mixed gauge. (b) Using the entanglement decomposition, we find an approximation to the tensor C in terms of the different pieces that mediate the fast and slow degrees of freedom. (c) We group the pieces according to the dotted lines to construct the new tensors M_L, B_ℓ and N_R, which will be the starting point for the heuristic optimization variables.

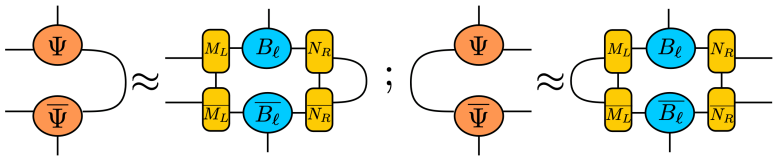


Figure 5.A.7: In order to improve our purification, we minimize the squares of the euclidean distances between the left and the right hand side of the two approximate equations above. We minimize the sum of the two distances, as we need a resulting state that satisfies both.

the physical indices, separated by two *purification* tensors N_R and M_L, themselves connected by a diagonal of inverse Schmidt values. Notice that the subsequent time evolution is applied only to the physical indices of each B_ℓ tensor, but not to the purification ones. Nevertheless, these tensors mediate the entanglement between physical sites, which in general will continue building up in time. We thus need to continue monitoring the fast-slow entanglement entropy of the B_ℓ tensors in order to apply further truncations as required.

Two extra aspects of the algorithm deserve a comment. The first one is that with each truncation/mixing step, the number of purification sites between a pair of blocks increases by two. In principle, since there is a large freedom in the representation of the purification indices, it would be possible to look for a more efficient description of this set of tensors (e.g. minimally entangled, in the spirit of [182]). In this work, for simplicity, we group all the purification legs together and work with a unit cell that has two tensors, one with the physical dimension and another one with the purification dimension. The dimension of the purification index would thus in principle grow exponentially with the number of truncations. In order to efficiently deal with it, we perform a SVD of the purification tensor, splitting virtual vs purification legs, and implement a hard cutoff on this dimension (with a value of a $d_{\text{purification}} = 1000$, for our simulations). Then we continue the

evolution while keeping the purification dimension fixed. The second aspect that deserves a mention refers to the performance of the heuristic truncation without using the input from the entanglement decomposition. We observe that whenever we do not make use of it, the heuristic optimization results (i.e. the distance between the correct marginals and the density matrices of the varitional mixed state) are not good, and thus, we incur in errors that accumulate during the simulation.

5.B Measure of long-range entanglement

In the main text we use the logarithmic negativity [183, 184] to quantify the entanglement between the left part of the system L and the right part of the system R that is mediated by the long-range degrees of freedom. In this section, for the sake of completeness of our discussion, we briefly introduce its definition and mention our motivation for using it to characterize the long-range entanglement.

Given a bipartite mixed state ρ_{LR}, the logarithmic negativity is defined as

$$E_N(\rho_{LR}) = \log_2 \|\rho_{LR}^{\Gamma_L}\|_1, \tag{5.B.1}$$

where $\|A\|_1 = \mathrm{Tr}\sqrt{A^\dagger A}$ is the trace norm of a matrix, i.e. the sum of its singular values and $\rho_{LR}^{\Gamma_L}$ denotes the partial transposition with respect to the subsystem L. If we write ρ_{LR} in the tensor product basis

$$\rho_{LR} = \sum_{ijkl} \rho_{ik,jl} \left(|i\rangle\langle j|\right)_L \otimes \left(|k\rangle\langle l|\right)_R , \tag{5.B.2}$$

then the partial transpose reads

$$\rho_{LR}^{\Gamma_L} = \sum_{ijkl} \rho_{ik,jl} \left(|i\rangle\langle j|\right)_L \otimes \left(|l\rangle\langle k|\right)_R . \tag{5.B.3}$$

A non-zero logarithmic negativity indicates that the state under consideration is entangled, and it can be shown that it is a proper entanglement monotone [184]. In the main text we use it to characterize the entanglement mediated by the fast degrees of freedom, as it is useful when the state under consideration is mixed. This is the case in our study, as at finite times the fast degrees of freedom have not decoupled completely from the rest of excitations present in the system and they still carry some residual entanglement with the former.

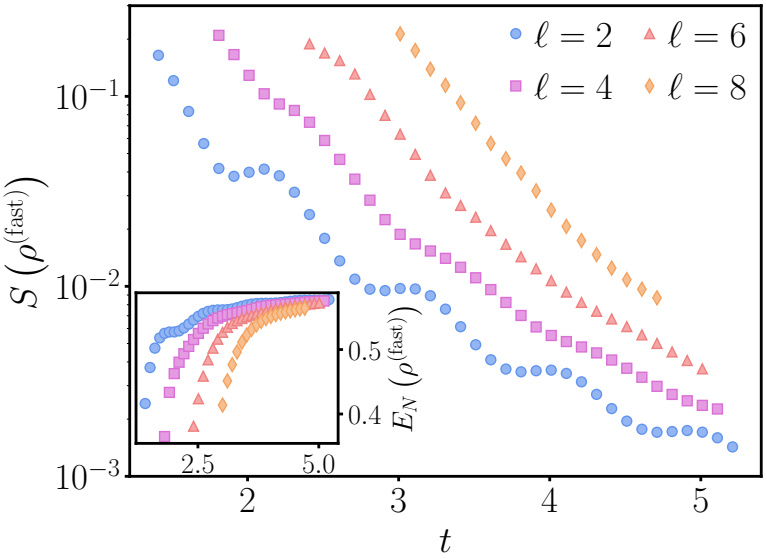


Figure 5.C.1: Decay of the entropies (main plot) and behavior of the logarithmic negativity of the fast degrees of freedom (inset) as a function of time for different sizes of the subsystem S, ℓ. The parameters for the quench are $g = 2, J_2 = 0$, the same phase diagram point that we use in the main text.

5.C Dependence on the block size

In this section, we show how the results presented in Fig. 3 of the main text depend on the size ℓ of the subsystem S. As can be seen from Fig. 5.C.1, the larger the size of S, the longer it takes for the fast degrees of freedom to disentangle. Increasing the size of the subsystem S linearly leads to an apparently linear increase in the timescale in which excitations disentangle from S. Still, the decay of the entropy seems to be exponentially fast with time. In addition, if we wait for long enough, the long-range entanglement found by our algorithm seems to saturate to the same value for all ℓ. This seems to indicate that the same excitations that have decoupled from a subsystem S of size $\ell = 2$ propagate and eventually decouple from larger subsystems, provided we wait long enough.

5.D Free fermion calculations

In the integrable case ($J_2 = 0$), the model can be mapped to an exactly solvable free fermionic system. We can use this solution to perform the same calculations as with the TN algorithm, and thus corroborate and gain a deeper understanding of the observed numerical results.

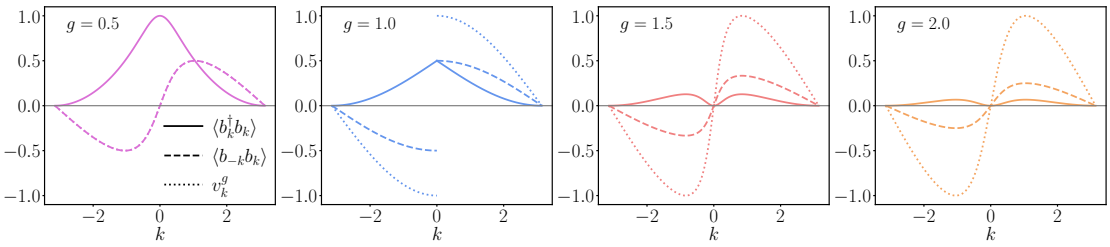


Figure 5.D.1: Occupation of the eigenmodes of the quench Hamiltonian (solid lines) and their group velocity (dotted lines) as a function of their momentum in the integrable quenches shown in figure 3(a) in the main text. Dashed lines show the initial coherence between opposite momenta. In the leftmost plot ($g = 0.5$), the coherence between opposite momenta and the group velocity are on top of each other.

5.D.1 Disentangling the long-range components

Fig. 3(a) of the main text shows that, for quenches from the ground state at $g \to \infty$, the residual entropy between fast and slow degrees of freedom decays slower for smaller values of g, even though for all the cases studied the decay is compatible with an exponentially fast decrease in time. At the same time, the saturation value of the logarithmic negativity carried by the fast degrees of freedom seems to grow as we scan the final value of the transverse field from $g = 2$ to $g = 0.5$. In fact, at $g = 1$, the logarithmic negativity seems to saturate to 1, the maximal possible value for a bipartite state of two qubits [183].

We can relate this behavior to the distribution of quasi-particles in the initial state, illustrated in figure 5.D.1 for the studied quenches. As we scan from $g = 2$ to $g = 0.5$ in the phase diagram, the occupation number of the freely propagating modes of the quench Hamiltonian grows. Something similar happens to the coherences between opposite momenta. This is reminiscent of the behavior of the logarithmic negativities in the TN computations and of the fact that as we quench to smaller values of g, the energy of the initial state according to the final Hamiltonian gets closer to the middle of the spectrum. Furthermore, by analysing the group velocities one can get some intuition of the rates of decay of the residual entropy. For a quench inside the paramagnetic phase, the maximum of the occupation coincides with the maximal group velocity. In contrast, when we quench to the ferromagnetic phase, the maximally occupied mode has vanishing group velocity. This could be an explanation of the slower decay rate observed when quenching to the ferromagnetic phase.

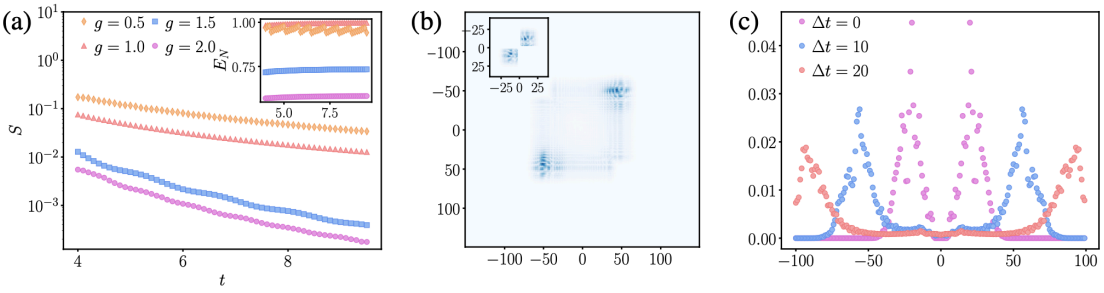


Figure 5.D.2: We reproduce the TN calculations for an integrable quench using the free fermionic formalism. For a system of size $N = 800$ we consider a block S of 4 sites and identify a pair of modes, respectively in L and R, that disentangle from S, as explained in the text. The residual fast-slow entropy (a) shows the same behavior as in the TN calculation, as does the logarithmic negativity between the L and R fast components (inset). (b) shows a heatmap of the contribution to the spatial correlation matrix $\langle a_i^\dagger a_j \rangle$ of the coherences of this long-range pair when it is identified at $t = 5$ (inset) and after an additional evolution for $\Delta t = 20$. Lastly, (c) shows the sum of these contribution as a function of the distance between the insertions of the operators in the correlation matrix.

5.D.2 Eliminating the long-range coherences

The free fermionic chain can also provide insight on the transformation performed by the TN algorithm. A Jordan-Wigner transformation maps the spin chain to a system of N fermionic modes, created by operators $a_i^\dagger$, $i = 1, \dots N$, that fulfill standard commutation relations $\left\{a_i^\dagger, a_j\right\} = \delta_{i,j}$, with δ the Kronecker delta. Because the resulting Hamiltonian involves only quadratic terms in a_i, $a_j^\dagger$, the fermionic system is free, and ground states and their evolution are described by Gaussian states. By virtue of Wick's theorem, these are completely characterised by two-point correlation functions, which can be organized in the $2N \times 2N$ correlation matrix [185] $\Gamma_{i,j} = \langle \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j^\dagger \rangle$, where the operator-valued vector

$$\boldsymbol{\alpha}^\dagger = \left(a_1, a_2, \dots, a_N, a_1^\dagger, a_2^\dagger, \dots, a_N^\dagger\right)$$

holds all creation and annihilation operators of the chain. Other sets of fermions $\boldsymbol{c} = U\boldsymbol{\alpha}$ can be defined through symplectic transformations, which conserve the anticommutation relations.

We can now divide the system in three parts, namely a central region S containing few fermions and playing the role of the local block in the above description, and two regions L and R, containing respectively all the fermions to the left or right of S. Because the state is Gaussian, all correlations of operators supported in

one connected region are determined by the restriction of the correlation matrix to the modes in that region. The symplectic transformation representing a unitary that acts on the left (right) virtual leg of the MPS would in this setting act only on the modes defined in L (R).

The free-fermionic equivalent of the TN algorithm should try to single out, after some evolution time, at least one fermionic mode localized in L and entangled only with modes in R. Namely we want to identify modes $l = \sum_j U^L_{1,j}\alpha^L_j$ and $r = \sum_j U^R_{1,j}\alpha^R_j$ such that $\langle \boldsymbol{\lambda} \cdot \boldsymbol{\alpha}^{S\dagger}\rangle = \langle \boldsymbol{\rho} \cdot \boldsymbol{\alpha}^{S\dagger}\rangle = 0$, where $\boldsymbol{\rho}^\dagger = (r, r^\dagger)$, $\boldsymbol{\lambda}^\dagger = (l, l^\dagger)$ and $\boldsymbol{\alpha}^{S(L,R)}$ is the restriction of $\boldsymbol{\alpha}$ to the modes in S (L, R). To this end, after discarding modes that are pure in either L or R, we optimize numerically the linear combinations above that produce the lowest overlap with modes in S. Once these two modes are isolated, we can compute their joint reduced density matrix, study its entanglement properties and compare them to those found in the TN case. This reduced density matrix would be analogous to $\rho^{(\text{fast})}_{L_f R_f}$. In the context of free fermions, converting the long-range entanglement to mixture is achieved by setting the coherences of the modes, that is, the matrix elements $\langle \boldsymbol{\rho} \cdot \boldsymbol{\lambda}^\dagger\rangle$ to zero. As in the TN case, after the truncation we can continue the evolution and check the effect in the local observables.

In order to compare to the TN results, we have applied this procedure to different quenches for a system of $N = 800$ fermionic modes, in which we singled out a central region of four sites. Fig. 5.D.2(a) shows the residual entanglement between the l, r pair and the modes in S (main panel) for several values of the quench parameter, together with the negativity between the two modes l and r (inset), to be directly compared to the results in Fig. 3 in the main text. Both quantities show good agreement with the TN results and confirm that in all scenarios, as time passes, the identification of l, r becomes more and more accurate, and that both modes keep being robustly entangled with each other. Notice that the agreement is not only qualitative, but the magnitude of the entanglement quantities is very similar to those found with the TN algorithm, which supports our interpretation of the TN transformation as correctly identifying the quasiparticles that carry away long-range entanglement.

For the quench $g = 2$, after time $t = 5$, we track the effect of the long-range coherences identified at that time in the observables. Fig. 5.D.2(b) shows (in absolute value) the contribution of the coherences of this pair to the ($a^\dagger a$ part of the) correlation matrix on the spatial basis (inset) as well as how this has evolved after a time $\Delta t = 20$ has elapsed. The empty region around $(0, 0)$ corresponds to our block S. As can be seen from the figure, the coherences only contribute to matrix elements very far away from the diagonal, which correspond to very non-local correlation functions. In contrast, the local expectation values, which are represented in this matrix by the diagonals close to the main one, are practically

zero. This fact indicates that the evolution of these coherences does not have an effect on local expectation values, and thus they can be safely discarded, and the state of these modes approximated by a mixture.

As a way to visualize how the contributions of the $l - r$ coherences drift to more and more non-local observables, we show in panel (c), for several instants of time, the sum of the n-diagonal of the coherence matrix in Fig. 5.D.2(b) as a function of n. This corresponds to the contribution of the $l - r$ coherences to the quantity $\sum_n |\langle a^\dagger_{n+x} a_n \rangle|$, and shows how, as the evolution unfolds, the coherences propagate towards more and more non-local expectation values, never affecting the local ones, as can be seen from the location of the peaks in the figure and the almost vanishing values around the origin.

These findings agree with the ones found by TN, and show that also in the Gaussian formalism it is possible to decohere some of the long-range entanglement generated during a quench while preserving the evolution of the local expectation values.

Chapter 6

Classically Computing Performance Bounds on Noisy Circuits

This chapter is based on the publication
Sattwik Deb Mishra*, **Miguel Frías-Pérez***, Rahul Trivedi
Classically Computing Performance Bounds on Depolarized Quantum Circuits
PRX Quantum 5, 020317 (2024)
Structure, text and figures have been adapted for this thesis
* Equal contribution

6.1 Overview

Fault-tolerant quantum computers hold promise for outperforming classical computers at several computational tasks. One of the most explored computational tasks is the problem of finding the ground state of a given many-body Hamiltonian — a problem that naturally arises in studying equilibrium properties of condensed matter systems [186]. Moreover, classical optimization problems can also be framed as finding ground states of commuting Hamiltonians [187]. Unsurprisingly, quantum algorithms for finding Hamiltonian ground states have been extensively studied [70, 111, 188] in search of a possible quantum advantage [189–191] — algorithms based on phase estimation and adiabatic evolution have been proposed for solving this problem, and have even been shown to be efficient for specific classes of Hamiltonians [192, 193]. Furthermore, due to the constraints on available quantum hardware, there has been intense activity in exploring hardware-efficient heuristics for solving this problem such as quantum adiabatic algorithms or variational quantum algorithms [80, 194–196].

Current noisy-intermediate scale quantum devices, however, do not perform quantum error correction and consequently noise places a severe constraint on the performance of these quantum algorithms. From a theoretical standpoint, it has thus become of interest to develop no-go results by providing theoretical bounds on the minimum energy that a noisy quantum circuit can achieve for a given Hamiltonian — if a classical algorithm [197–199] could obtain an energy better than this lower bound, then we can conclude that a reduction in noise rate is necessarily needed for a possible quantum advantage. An approach to assessing the impact of noise on quantum algorithms is to directly simulate the circuit more so since the presence of noise in quantum circuits is expected to make them easier to classically simulate [200, 201]. In fact, there have been several recent demonstrations of noisy quantum circuit simulations using tensor network methods [30, 32, 127]. However, most of the tensor network methods lack rigorous accuracy guarantees and cannot certify an accurate simulation of the quantum circuit. In particular, they are expected to deviate significantly from the circuit output as the noise rate continues to decrease and thus fall short of rigorously providing a no-go result for quantum advantage.

Alternatively, this problem can be approached analytically using tools from quantum information theory. For instance, Refs. [83, 85, 202] analyzed the increase in entropy of the quantum state due to noise, and showed that it can allow for an analytical lower bound on the attainable minimum energy. However, while providing rigorous no-go results, these analyses were circuit-architecture independent and were thus expected to underestimate the impact of noise. Certain circuit architectures are expected to significantly worsen the impact of noise, and this phenomena has been theoretically demonstrated in random quantum circuits models [84, 203]. However, it remains unclear if it is possible to provide an architecture-dependent lower bound for a specific engineered quantum circuit.

In this chapter, we propose a method for efficiently computing rigorous bounds on the performance of any *specified* quantum circuit in the presence of a constant rate of depolarizing noise. The key insight behind our proposed method is the formulation of a Lagrangian dual corresponding to the circuit dynamics, which allows us to account for the circuit architecture in addition to the increase in the entropy, or equivalently, the decrease in the purity of the quantum state. We show that the Lagrangian dual yields a hierarchy of classically computable lower bounds on energy, with respect to a specified Hamiltonian, obtained at the output of a noisy quantum circuit. We provide numerical and analytical evidence that this formulation can capture the circuit-architecture dependent propagation of errors through the noisy quantum circuit and thus provide more stringent lower bounds than currently available. Our work is, in part, motivated by the application of Lagrangian duality to provide performance bounds on classical physical systems

[204–209] and quantum optical devices [210, 211].

6.2 Notation

Given a finite-dimensional Hilbert space $\mathcal{H}$, we use $\mathcal{D}_1(\mathcal{H})$ to denote the set of all density matrices on $\mathcal{H}$, and $\mathcal{M}(\mathcal{H})$ to denote the set of Hermitian linear operators on $\mathcal{H}$. Unless otherwise mentioned, for any linear operator A on $\mathcal{H}$, $\|A\|$ will denote its operator norm i.e. the maximum singular value of A and $\|A\|_F = \sqrt{\mathrm{Tr}(A^\dagger A)}$ will denote its Frobenius norm.

We use the computer-science big-O notation for function asymptotics. In particular, given two functions $f, g : [0, \infty) \to [0, \infty)$, $f = O(g)$ if, for some $c > 0$, $f(x) \leq cg(x)$ as $x \to \infty$ and $f = \Omega(g)$ if, for some $c > 0$, $f(x) \geq cg(x)$ as $x \to \infty$.

6.3 Duality based bounds

6.3.1 Single-qubit example

As a simple illustrative example of the Lagrangian dual formulation, we first consider a single-qubit circuit [Fig. 6.3.1(a)] — consider a qubit initially in $|0\rangle$, with a gate $U = e^{-i\theta Y}$ being applied on it followed by depolarizing noise with probability p. We would like to find the parameter θ to minimize the energy corresponding to the Hamiltonian $H = \Delta Z$ — in the absence of noise ($p = 0$), it is straighforward to verify that this would be accomplished by setting $\theta = \pi/2$ to obtain an energy $-\Delta$.

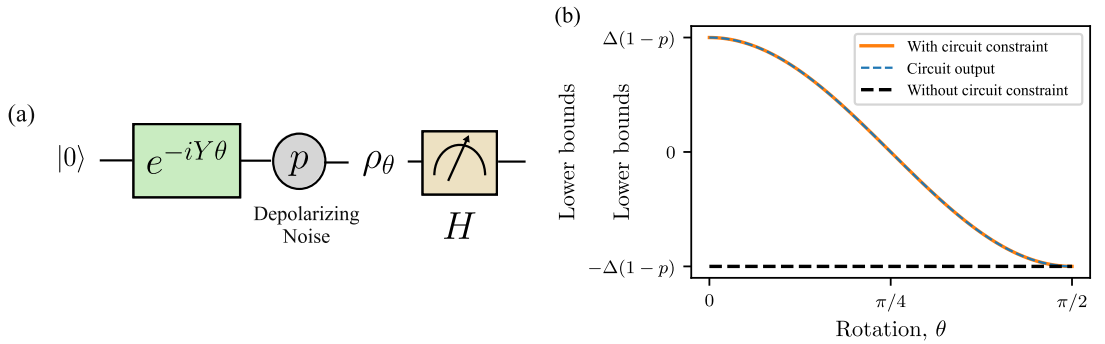


Figure 6.3.1: Comparison of bounds, with and without accounting for circuit constraints, on the minimum energy corresponding to the Hamiltonian $H = \Delta Z$ attainable by the output of the single qubit circuit shown in the schematic. The circuit consists of a Y-axis rotation followed by depolarizing noise acting with probability p.

However, in the presence of depolarizing noise, the qubit will necessarily be in a mixed state. The extent to which the state is mixed can be quantified with a purity measure, for instance the von Neumann entropy of the qubit state, its trace purity or even higher order Renyi Entropies [212]. For concreteness, we use the trace purity measure of a state ρ: $P(\rho) = \mathrm{Tr}(\rho^2)$ — $P(\rho) = 1$ if and only if ρ is a pure state, else $P(\rho) < 1$. Now, since the state ρ_θ at the output of the single-qubit circuit in Fig. 6.3.1(a) is obtained by applying the depolarizing noise channel to a single qubit pure state, $P(\rho_\theta) = P_0 := p^2/4 + (1-p/2)^2 < 1$. Since ρ_θ is necessarily mixed, it cannot produce the pure ground state of the Hamiltonian H perfectly irrespective of the choice of θ – in fact, this simple observation can be used to lower bound the energy that can possibly be obtained at the output of the circuit by minimizing it with respect to states with purity at-most P_0 i.e. solving the following optimization problem

$$\begin{aligned} \underset{\rho \in \mathcal{D}_1(\mathbb{C}^2)}{\text{minimize}} \quad & \mathrm{Tr}(H\rho) \\ \text{subject to} \quad & P(\rho) \leq P_0, \end{aligned} \tag{6.3.1}$$

where $\mathcal{D}_1(\mathbb{C}^d)$ is the space of density matrices on the Hilbert space $\mathbb{C}^d$. The optimization problem in Eq. 6.3.1 is solved by $\rho = (1-p/2)|1\rangle\langle 1| + (p/2)|0\rangle\langle 0|$ with energy $-\Delta(1-p)$. This bound clearly exhibits the intuitively expected dependence on the noise rate p — if $p = 0$, then the energy attained coincides with the ground state energy of $-\Delta$, and if $p = 1$, it is simply the energy obtained by the maximally mixed state.

However, this bound does not account for the unitary being applied on the qubit, and a better bound can be obtained by explicitly accounting for the circuit. To do so, we use the method of Lagrange duality [204, 213]. For this, we extend the problem in Eq. 6.3.1 by adding an additional constraint due to the circuit:

$$\begin{aligned} \underset{\rho \in \mathcal{D}_1(\mathbb{C}^2)}{\text{minimize}} \quad & \mathrm{Tr}(H\rho) \\ \text{subject to} \quad & \rho = \mathcal{E}_\theta(\rho_0), \\ & P(\rho) \leq P_0, \end{aligned} \tag{6.3.2}$$

where $\rho_0 = |0\rangle\langle 0|$ and $\mathcal{E}_\theta$ is the channel corresponding to the unitary $e^{-iY\theta}$ followed by the single-qubit depolarizing noise. Note that the purity constraint $P(\rho) \leq P_0$ is redundant and is already implied by the circuit constraint $\rho = \mathcal{E}_\theta(\rho_0)$. However, as we will see in the following discussion, while redundant constraints do not impact the solution of an optimization problem, depending on the specific technique used to obtain a lower bound on the problem, they can have a considerable impact.

To provide a lower bound on this optimization, we construct its Lagrangian

$\mathcal{L}(\sigma, \lambda)$ by introducing Lagrange multipliers $\sigma \in \mathcal{M}(\mathbb{C}^2)$, and $\lambda \geq 0$,

$$\begin{aligned}\mathcal{L}(\rho, \sigma, \lambda) = \mathrm{Tr}[H\rho] + \mathrm{Tr}[\sigma(\rho - \mathcal{E}_\theta(|0\rangle\langle 0|))] \\ + \lambda(P(\rho) - P_0).\end{aligned} \tag{6.3.3}$$

$\mathcal{L}(\rho, \sigma, \lambda)$ can be considered to be a modified energy function which, in addition to the energy $\mathrm{Tr}[H\rho]$, also penalizes violation of the two constraints: $\rho = \mathcal{E}_\theta(\rho_0)$ imposed by the circuit *and* $P(\rho) \leq P_0$ on the purity of the state ρ. Minimizing the Lagrangian with respect to ρ, we obtain the *dual function*,

$$g(\sigma, \lambda) = \min_\rho \mathcal{L}(\rho, \sigma, \lambda), \tag{6.3.4}$$

which is a function of σ, λ, the *dual variables*. It follows from the principle of Lagrange duality that for any σ and $\lambda \geq 0$, $g(\sigma, \lambda)$ is a lower bound on the energy attained by the circuit. This can easily be seen from Eq. 6.3.3 by noting that when $\mathcal{L}$ is evaluated at the circuit output $\rho_\theta = \mathcal{E}_\theta(|0\rangle\langle 0|)$, we obtain

$$\mathcal{L}(\rho_\theta, \sigma, \lambda) = \mathrm{Tr}[H\rho_\theta] + \lambda(P(\rho_\theta) - P_0) \leq \mathrm{Tr}[H\rho_\theta],$$

since $P(\rho_\theta) \leq P_0$ and $\lambda \geq 0$. Since from Eq. 6.3.4 $g(\sigma, \lambda)$ is the smallest attainable value of $\mathcal{L}(\rho, \sigma, \lambda)$ on varying ρ, we obtain $g(\sigma, \lambda) \leq \mathrm{Tr}[H\rho_\theta]$. We emphasize that the dual function $g(\sigma, \lambda)$, evaluated at any $\sigma, \lambda \geq 0$, is a lower bound on the energy $\mathrm{Tr}[H\rho_\theta]$ attained by the circuit, and the best lower bound can be obtained by maximizing $g(\sigma, \lambda)$ with respect to σ, λ. Furthermore, since the construction of the dual function explicitly accounts for the circuit constraint, it gives a better bound than obtained from the problem in Eq. 6.3.1 i.e. by just accounting for the final purity of the state. This is exhibited in Fig. 6.3.1(b), where $\max_{\sigma,\lambda\geq 0} g(\sigma, \lambda)$ compared with $-\Delta(1-p)$ and it can be seen that the dual function provides a better lower bound for most values of θ. For the simple example of a single qubit, the duality-based bound that we can compute coincides exactly with the circuit output and thus models it exactly. As we will see in the next sections, this will not be the case for circuits over a large number of qubits.

6.3.2 General formulation

We can now extend the duality lower bound to more general quantum circuits [Fig. 6.3.2] — consider a quantum circuit of depth d, consisting of unitaries $U_1, U_2 \ldots U_d$ that has been designed to approximate the ground state of a target Hamiltonian H of N qubits. In the presence of noise, the state becomes increasingly mixed as the unitaries are applied to it — while it is typically hard to compute exactly how mixed the state is, an analytical upper bound on several purity measures can be obtained. In particular, Refs. [83, 202] establish explicit upper bounds for two

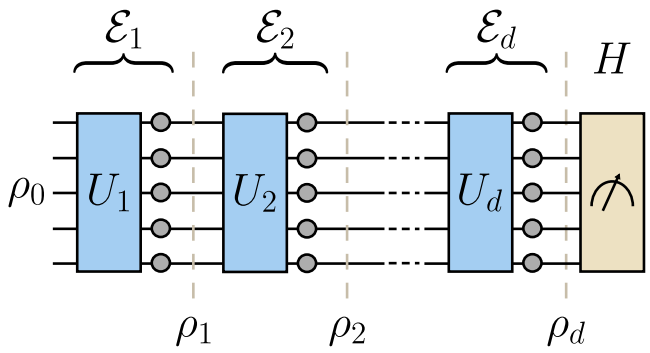


Figure 6.3.2: Schematic depiction of the problem setting considered in this paper. The unitaries (colored boxes) implement a quantum algorithm to prepare an approximation of the ground state of a target Hamiltonian H in the absence of noise. Each layer of unitary is followed by single-qubit depolarizing noise (gray circles) on the qubits applied with a probability p.

purity measures after t time-steps — the information content of the quantum state, as well as its trace purity.

Lemma 1. *[Refs. [83], [202]] Suppose ρ_t is the quantum state on N qubits obtained from an initial pure state after applying t unitaries and single qubit depolarizing channels, then*

$$I(\rho_t) := N + \mathrm{Tr}[\rho_t \log_2(\rho_t)] \leq N(1-p)^t,$$
$$P_{tr}(\rho_t) := \mathrm{Tr}(\rho_t^2) \leq 2^{-N(1-(1-p)^t)},$$

where p is the probability of applying the depolarizing noise at each time step independently on each qubit.

Both the information content and trace purity can be viewed as measures of how mixed the given state is. Both are largest for a pure state ($I(|\psi\rangle\langle\psi|) = N$ and $P_{\mathrm{tr}}(|\psi\rangle\langle\psi|) = 1$), and are lowest for the maximally mixed state ($I(\mathbb{I}/2^N) = 0$ and $P_{\mathrm{tr}}(\mathbb{I}/2^N) = 2^{-N}$).

In the remainder of this subsection, we denote by P_t an upper bound on the purity of the state at t^{th} time-step — we will formulate the results of this subsection for general convex purity measures, and specialize them to concrete purity measures (such as information content or trace purity) in the following sections. Now, as with the single-qubit case, the energy attained at the output of the circuit can now

be written as,

$$\begin{aligned} \underset{\rho_1,\rho_2\ldots\rho_d\in\mathcal{S}}{\text{minimize}} \quad & \mathrm{Tr}(H\rho_d) \\ \text{subject to} \quad & \rho_t = \mathcal{E}_t(\rho_{t-1}),\ t\in\{1,\ldots,d\}, \\ & P(\rho_t)\leq P_t,\ t\in\{1,\ldots,d\}, \end{aligned} \tag{6.3.5}$$

where $\mathcal{E}_t(\cdot)$ is the quantum channel that applies the unitary U_t for the t^{th} layer of the circuit followed by depolarizing noise acting individually on the qubits, and ρ_0 is a fixed and known initial state. Importantly, in Eq. 6.3.5, $\mathcal{S}$ is the set of N qubit operators over which we allow the states $\rho_1,\rho_2\ldots\rho_d$ to vary — this set can be chosen to be *any* set containing density matrices over N qubits $\mathcal{D}_1((\mathbb{C}^2)^{\otimes N})$ since the circuit constraints ($\rho_t=\mathcal{E}_t(\rho_{t-1})$) enforce $\rho_1,\rho_2\ldots\rho_d$ to be valid density matrices. For instance, $\mathcal{S}$ can be chosen to be just the set of N-qubit Hermitian operators, or the set of N-qubit Hermitian operators with unity trace. As we will see below, the choice of this set together with the purity measure determines the form of the dual function.

To construct the dual function corresponding to Eq. 6.3.5 — we introduce the dual variables $\vec{\sigma}=\{\sigma_1,\sigma_2\ldots\sigma_d\}$, which are $N-$qubit Hermitian operators, corresponding to the circuit constraints and $\vec{\lambda}=\{\lambda_1,\lambda_2\ldots\lambda_d\geq 0\}$ corresponding to the purity constraints. The Lagrangian is now constructed by adding penalties corresponding to the circuit constraints and purity constraints at each time step to the output energy:

$$\begin{aligned} \mathcal{L}(\vec{\rho},\vec{\sigma},\vec{\lambda}) &= \mathrm{Tr}[H\rho_d] + \sum_{t=1}^{d}\mathrm{Tr}\Big[\sigma_t\big(\rho_t-\mathcal{E}_t(\rho_{t-1})\big)\Big] + \\ &\qquad \sum_{t=1}^{d}\lambda_t\Big[P(\rho_t)-P_t\Big], \\ &= \sum_{t=1}^{d}\mathrm{Tr}(\rho_t H_t) + \lambda_t\Big[P(\rho_t)-P_t\Big], \end{aligned} \tag{6.3.6}$$

where $H_d = H+\sigma_d$, $H_t=\sigma_t-\mathcal{E}_{t+1}^{\dagger}(\sigma_{t+1})$ for $t\in\{1,2\ldots d-1\}$ The dual function is obtained by minimizing the Lagrangian with respect to $\rho_1,\rho_2\ldots\rho_d\in\mathcal{S}$ i.e.

$$\begin{aligned} g(\vec{\sigma},\vec{\lambda}) &= \min_{\rho_1,\rho_2\ldots\rho_d\in\mathcal{S}}\mathcal{L}(\vec{\rho},\vec{\sigma},\vec{\lambda}), \\ &= \sum_{t=1}^{d}\min_{\rho_t\in\mathcal{S}}\Big[\big(\mathrm{Tr}(\rho_t H_t)+\lambda_t P(\rho_t)\big)-\lambda_t P_t\Big], \\ &= \sum_{t=1}^{d}\Big(\mathcal{F}_{\mathcal{S},P}(H_t,\lambda_t)-\lambda_t P_t\Big) \end{aligned} \tag{6.3.7a}$$

where

$$\mathcal{F}_{\mathcal{S},P}(H,\lambda) = \min_{\rho\in\mathcal{S}} \Big(\mathrm{Tr}[H\rho] + \lambda P(\rho) \Big). \tag{6.3.7b}$$

As with the single-qubit example, the dual function is a lower bound on the energy produced at the circuit output for any $\vec{\sigma}, \vec{\lambda} \geq 0$ i.e.

$$g(\vec{\sigma},\vec{\lambda}) \leq \mathrm{Tr}[\rho_d H] \;\text{ for all } \sigma_t \in \mathcal{M}((\mathbb{C}^2)^{\otimes N}), \lambda_t \geq 0.$$

The function $\mathcal{F}_{\mathcal{S},P}(H,\lambda)$ can be interpreted as a generalized free energy corresponding to the Hamiltonian H at temperature λ which depends on both the domain $\mathcal{S}$ and the purity measure P. For instance, if the purity measure is taken to be the information content $I(\rho) = N - \mathrm{Tr}[\rho \log_2(\rho)]$, then it reduces to the Gibbs free energy with an offset of $N\lambda$. However, by choosing different purity measures P as well as different domains $\mathcal{S}$, the dual function allows us to obtain a family of bounds on the noisy quantum circuit. As we will see in the next section, certain choices of P and $\mathcal{S}$ provide lower bounds that can be classically computed.

Consider first the best lower bound that can be obtained from the dual function. In the following proposition, we show that the best lower bound attained by the dual function is exactly equal to the energy attained by the quantum circuit, and the choice of dual variables $\sigma_1, \sigma_2 \ldots \sigma_d$ that yields the largest value dual function corresponds to the Heisenberg picture evolution of the Hamiltonian H.

Proposition 1. *For the dual function defined in Eq. 6.3.7, it follows that its maximum over the dual variables is equal to the output energy of the noisy circuit i.e.*

$$\underset{\substack{\sigma_1,\sigma_2\ldots\sigma_d\in\mathcal{M}((\mathbb{C}^2)^{\otimes N}) \\ \lambda_1,\lambda_2\ldots\lambda_d\geq 0}}{\text{maximum}} g(\vec{\sigma},\vec{\lambda}) = \mathrm{Tr}[H\mathcal{E}_d\mathcal{E}_{d-1}\ldots\mathcal{E}_1(\rho_0)],$$

and the maximum is attained at

$$\sigma_d = -H, \sigma_t = -\mathcal{E}_{t+1}^\dagger\mathcal{E}_{t+2}^\dagger\ldots\mathcal{E}_d^\dagger(H),$$

and $\lambda_1 = \lambda_2 = \ldots \lambda_d = 0$.

Proof: The proof of this proposition follows simply by noting that, from the definition,

$$\mathcal{F}_{\mathcal{S},P}(0,0) = 0.$$

Now, if $\sigma_d = -H$, and $\sigma_t = -\mathcal{E}_{t+1}^\dagger\mathcal{E}_{t+2}^\dagger\ldots\mathcal{E}_d^\dagger(H)$, then $H_t = 0$. Hence, we obtain that at this value of $\vec{\sigma}$ and at $\vec{\lambda} = 0$, $g(\vec{\sigma},\vec{\lambda}) = \mathrm{Tr}[\rho_0\mathcal{E}_1^\dagger\mathcal{E}_2^\dagger\ldots\mathcal{E}_d^\dagger(H)] = \mathrm{Tr}[H\mathcal{E}_d\mathcal{E}_{d-1}\ldots\mathcal{E}_1(\rho_0)]$. Since $\mathrm{Tr}[H\mathcal{E}_d\mathcal{E}_{d-1}\ldots\mathcal{E}_1(\rho_0)]$ is also an upper bound of $g(\vec{\sigma},\vec{\lambda})$, the proposition follows. □

This proposition establishes that finding the best dual bound is equivalent to exactly simulating the circuit, which we expect to be hard to do on classical computers. This hardness fundamentally stems from the fact that the dual variables $\sigma_1, \sigma_2 \dots \sigma_d$ are operators in an exponentially large space. However, since the dual function $g(\vec{\sigma}, \vec{\lambda})$ is a lower bound on the output energy for any $\vec{\sigma}, \vec{\lambda}$, a natural approach to evaluate a lower bound would be restricted σ_i to subsets of $\mathcal{M}((\mathbb{C}^2)^{\otimes N})$ where the dual function could be evaluated efficiently — the specific subset would depend on the choice of the purity measure. In Section 6.4, we will see that the dual function obtained on choosing the purity measure to be trace purity and the domain $\mathcal{S} = \mathcal{M}((\mathbb{C}^2)^{\otimes N})$ of N-qubit Hermitian operators can be evaluated efficiently if $\sigma_1, \sigma_2 \dots \sigma_d$ are chosen to be matrix product operators (MPOs) of bond-dimension poly(N). In Section 6.5, we will consider the dual function obtained on choosing the purity measure to be the information content of the state, in which case $\sigma_1, \sigma_2 \dots \sigma_d$ can be restricted to the space of geometrically local Hamiltonians, allowing for an exact evaluation of the dual function.

Restricting the dual variables $\sigma_1, \sigma_2 \dots \sigma_d$ to a subset of $\mathcal{M}((\mathbb{C}^2)^{\otimes N})$ raises the question of whether the maximum value that the dual function can attain within this restricted set of dual variables gives a better lower bound on the energy compared to neglecting the circuit constraints and just accounting for the purity of the final state i.e. does the duality based bound still account for the circuit architecture. Our next proposition answers this question affirmatively, and shows that a better lower bound can be obtained as long as the restricted set of dual variables contains 0.

Proposition 2. *Suppose $\mathcal{S}_\sigma \subset \mathcal{M}((\mathbb{C}^2)^{\otimes N})$, such that $0 \in \mathcal{S}_\sigma$, then*

$$\underset{\substack{\sigma_1,\sigma_2\dots\sigma_d\in\mathcal{S}_\sigma \\ \lambda_1,\lambda_2\dots\lambda_d\geq 0}}{\text{maximum}}\ g(\vec{\sigma},\vec{\lambda}) \geq \underset{\rho\in\mathcal{S}, P(\rho)\leq P_d}{\text{minimize}}\ \mathrm{Tr}(H\rho)$$

Proof: Since $0 \in \mathcal{S}_\sigma$,

$$\underset{\lambda\geq 0}{\text{maximum}}\ g(\{0\dots 0\},\{0\dots\lambda\}) \leq \underset{\substack{\sigma_1,\sigma_2\dots\sigma_d\in\mathcal{S}_\sigma \\ \lambda_1,\lambda_2\dots\lambda_d\geq 0}}{\text{maximum}}\ g(\vec{\sigma},\vec{\lambda}).$$

Now, we can note that

$$g(\{0\dots 0\},\{0\dots\lambda\}) = \mathcal{F}_{S,P}(H,\lambda) - \lambda P_d.$$

It can be noted that $g(\{0\dots 0\},\{0\dots\lambda\})$ is simply the dual function of the convex problem

$$\begin{aligned} \underset{\rho\in\mathcal{S}}{\text{minimize}} \quad & \mathrm{Tr}(H\rho) \\ \text{subject to} \quad & P(\rho) \leq P_d. \end{aligned}$$

Furthermore, this convex problem trivially satisfies the Slater's conditions [213, 214]. This can be checked by noting that the Slater's conditions are satisfied if there is a $\rho \in \mathcal{S}$ such that $P(\rho) < P_d$ — this follows by noting that $P(I/2^N) < P_d$ and $I/2^N \in \mathcal{D}_1((\mathbb{C}^2)^{\otimes N}) \subseteq \mathcal{S}$. Since Slater's conditions are satisfied, this problem is strongly dual and consequently the optimal duality-bound is equal to the solution of the optimization problem i.e.

$$\underset{\lambda \geq 0}{\text{maximum}}\ g(\{0\ldots0\},\{0\ldots\lambda\}) = \underset{\rho\in\mathcal{S}, P(\rho)\leq P_d}{\text{minimize}}\ \mathrm{Tr}(H\rho),$$

which proves the proposition. □

While this proposition indicates that accounting for the circuit constraints while constructing the lower bound results in an improvement over only accounting for the final purity even with restricted space of dual variables, it says nothing about the extent to which the bound improves. We expect the improvement to be strongly dependent on the purity function P, the domain set $\mathcal{S}$, and the dual set $\mathcal{S}_\sigma$ used in formulating and evaluating the bound. In the next section, we consider a specific formulation of the dual function that uses the trace purity measure, and show that the lower bound obtained on accounting for the circuit constraints can be exponentially better than if the circuit constraints were not accounted for.

6.4 Trace purity-based lower bound

6.4.1 Formulation

In this section, we consider now a specific choice of the purity function and the domain $\mathcal{S}$ that results in a dual function that can be computed exactly when the dual variables are parametrized as matrix product operators with poly(N) bond dimension. We choose the purity measure to be trace purity $P(\rho) = P_{\mathrm{tr}}(\rho) = \mathrm{Tr}(\rho^2)$, and the domain $\mathcal{S}$ in Eq. 6.3.5 to be the space of Hermitian N-qubit operators $\mathcal{M}((\mathbb{C}^2)^{\otimes N})$. It then follows that $\mathcal{F}_{\mathcal{S},P}(H,\lambda)$ defined in Eq. 6.3.7b evaluates to

$$\mathcal{F}_{\mathcal{S},P}(H,\lambda) = -\frac{\mathrm{Tr}(H^2)}{4\lambda},$$

and therefore, we obtain that

$$g(\vec{\sigma},\vec{\lambda}) = -\mathrm{Tr}[\rho_0\mathcal{E}_1^\dagger(\sigma_1)] - \sum_{t=1}^{d}\left(\frac{\mathrm{Tr}(H_t^2)}{4\lambda_t} + \lambda_t P_t\right),$$

where $H_d = H + \sigma_d$ and $H_t = \sigma_t - \mathcal{E}_{t+1}^\dagger(\sigma_{t+1})$. Furthermore, for this dual function, it is possible to perform the maximization over $\vec{\lambda}$ analytically to obtain

$$\begin{aligned} h(\vec{\sigma}) &= \underset{\vec{\lambda}\geq 0}{\text{maximum}} \quad g(\vec{\sigma}, \vec{\lambda}), \\ &= -\operatorname{Tr}[\rho_0 \mathcal{E}_1^\dagger(\sigma_1)] - \sum_{i=1}^{d} \sqrt{P_t \operatorname{Tr}(H_t^2)}. \end{aligned} \tag{6.4.1}$$

From the expression for $h(\vec{\sigma})$, we immediately notice that if $\sigma_1, \sigma_2 \dots \sigma_d$ are restricted to be matrix product operators with bond dimension D, then $h(\vec{\sigma})$ can be evaluated classically in time NdD^4. However, as we established in proposition 1, the best lower bound is obtained $h(\vec{\sigma})$ when evaluating it at $\vec{\sigma}$ corresponding to a Heisenberg picture evolution of Hamiltonian H. While for most problems of interest (e.g. where H is a local or spatially-local Hamiltonian), H can be represented as a matrix product operator of a modest bond dimension, the unitaries involved in the circuit can, in general, grow its bond dimension exponentially. A natural choice of $\sigma_1, \sigma_2 \dots \sigma_d$ would then be to perform time-evolving block decimation (TEBD) [9, 118, 151, 215, 216] on the Heisenberg evolution and compress the operators in each step into bond-dimension D i.e. at $\vec{\sigma}^{h,D} = \{\sigma_1^{h,D}, \sigma_2^{h,D}, \dots, \sigma_d^{h,D}\}$

$$\begin{aligned} &\sigma_d^{h,D} = -H \text{ and,} \\ &\sigma_t^{h,D} = \Pi_D \mathcal{E}_{t+1}^\dagger(\sigma_{t+1}^{h,D}) \text{ for } t \in \{1, 2 \dots d-1\}, \end{aligned} \tag{6.4.2}$$

where Π_D compresses an N-qubit operator to an operator with a bond-dimension D [38].

Duality-bound and TEBD truncation errors. If a Heisenberg picture TEBD simulation, for some bond dimension D, of the noisy quantum circuit is exact, then by Proposition 1, the duality-based bound $h(\vec{\sigma}^{h,D})$ is exactly equal to the expected energy at the output of the circuit. In practice, for small bond dimensions D, the TEBD algorithm is not exactly correct but incurs an error. However, as shown below, an upper bound on this error can also be efficiently computed for the TEBD algorithm. Consequently, tracking the error incurred in the TEBD algorithm allows us to calculate another lower bound on the output of the quantum circuit i.e. if the TEBD algorithm produces an estimate E_{TEBD} the output energy E of the circuit within an additive error δ, then $E_{\text{TEBD}} - \delta$ also lower bounds the energy E. A natural question to ask is if the duality-based bounds are more informative than the bound obtained from just a TEBD simulation.

Consider now the problem of estimating the TEBD error following Ref. [9]. The TEBD estimate of the energy at the circuit output, E_{TEBD}, can be expressed as

$$E_{\text{TEBD}} = \operatorname{Tr}\left(\mathcal{E}_1(\rho_0)\left(\prod_{t=2}^{d} \Pi_D \mathcal{E}_t^\dagger\right)(H)\right),$$

while the true energy at the circuit output can be expressed as

$$E = \mathrm{Tr}\left(\mathcal{E}_1(\rho_0)\left(\prod_{t=2}^{d}\mathcal{E}_t^\dagger\right)(H)\right).$$

Denoting by ρ_t the state of the qubits in the quantum circuit at time-step t, $\rho_t = \mathcal{E}_t\mathcal{E}_{t-1}\ldots\mathcal{E}_1(\rho_0)$, we note that

$$\begin{aligned}
&E - E_{\mathrm{TEBD}} \\
&= \mathrm{Tr}\left(\rho_1\left(\prod_{t=2}^{d}\mathcal{E}_t^\dagger - \prod_{t=2}^{d}\Pi_D\mathcal{E}_t^\dagger\right)(H)\right), \\
&= \sum_{t=2}^{d}\mathrm{Tr}\left(\rho_1\left(\prod_{s=2}^{t-1}\mathcal{E}_s^\dagger\right)\left(\mathcal{E}_t^\dagger - \Pi_D\mathcal{E}_t^\dagger\right)\left(\prod_{s=t+1}^{d}\Pi_D\mathcal{E}_s^\dagger\right)(H)\right), \\
&= \sum_{t=2}^{d}\mathrm{Tr}\left(\rho_{t-1}\left(\sigma_{t-1}^{h,D} - \mathcal{E}_t^\dagger(\sigma_t^{h,D})\right)\right) = \sum_{t=1}^{d-1}\mathrm{Tr}\big(\rho_t H_t\big),
\end{aligned}$$

where, in the last step, we have used the fact that, by definition, $H_t = \sigma_{t-1}^{h,D} - \mathcal{E}_t^\dagger(\sigma_t^{h,D}))$. Now, an upper bound on the error $|E - E_{\mathrm{TEBD}}|$ can be obtained via

$$|E - E_{\mathrm{TEBD}}| \leq \sum_{t=1}^{d-1}|\mathrm{Tr}(\rho_t H_t)| \leq \sum_{t=1}^{d-1}\|H_t\|_F\,, \tag{6.4.3}$$

where $\|A\|_F = \sqrt{\mathrm{Tr}(A^\dagger A)}$ and we have used the fact that, by the Holder's inequality, $|\mathrm{Tr}(\rho_t H_t)| \leq \|\rho_t\|_1\|H_t\| \leq \|H_t\|_F$ since $\|\rho_t\|_1 = 1$ and $\|H_t\| \leq \|H_t\|_F$. We point out that an important reason why we express the error bound in terms of the Frobenius norm of H_t, instead of its operator norm, is because the Frobenius norm can be efficiently computed if H_t is a matrix product operator of a small bond dimension (which is the case while performing the TEBD simulation). The deviation bound in Eq. 6.4.3 implies a lower bound

$$E \geq E_{\mathrm{TEBD}} - \delta = -\mathrm{Tr}[\rho_0\mathcal{E}_1^\dagger(\sigma_1)] - \sum_{i=1}^{d}\sqrt{\mathrm{Tr}(H_t^2)}. \tag{6.4.4}$$

This bound is significantly worse than the duality-based bound in Eq. 6.4.1 as $P_t \ll 1$ for all time steps t. The key reason why just accounting for a worst-case accumulation of TEBD errors yields a loose lower bound is that the upper bound in $|E - E_{\mathrm{TEBD}}|$ *does not* account for the decrease in the trace purity of the quantum state in the presence of noise, which is explicitly factored into the formulation of the dual.

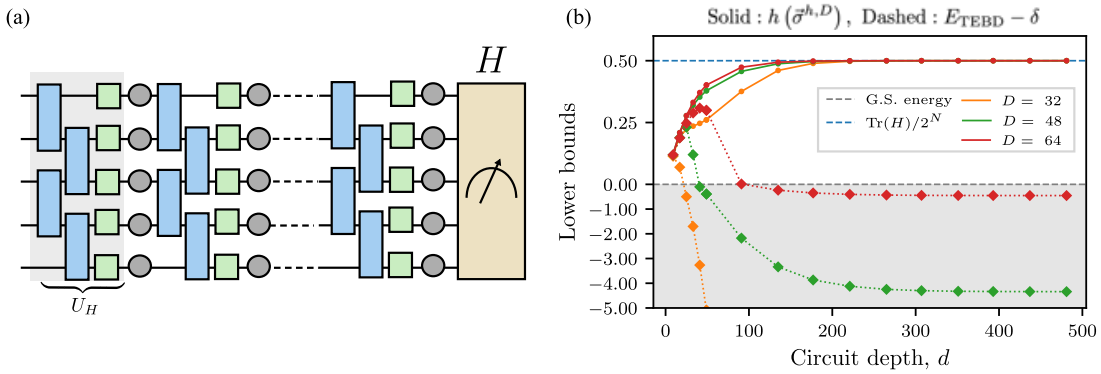


Figure 6.4.1: (a) Schematic of benchmark circuits considered for 1D spin systems: colored boxes indicate unitaries and grey circles depolarizing noise. Two-qubit unitaries are chosen to be $\exp(-i\theta X \otimes X)$ and single-qubit unitaries are independently Haar random. The Hamiltonian is chosen to be $H = -U_H\big(\sum_i Z_i\big)U_H^\dagger$, where U_H is the first layer of unitaries, making H a 4-local commuting Hamiltonian. The first layer of unitaries U_H thus transforms the initial state $|0\rangle^{\otimes N}$ into the ground state of H. The last $(d-1)/2$ layers are chosen to be the inverse of the previous $(d-1)/2$ layers — in the absence of noise, the output of the circuit is the ground state of H. (b) Plot shows trace purity-based dual bound ($h(\vec{\sigma}^{h,D})$ in Eq. 6.4.1) (solid lines, circular markers) and bound obtained by only considering the TEBD errors ($E_{\mathrm{TEBD}} - \delta$ in Eq. 6.4.4) (dotted lines, diamond markers) for the ground state (G. S.) energy of the target Hamiltonian, as a function of circuit depth d for a system of $N = 40$ spins, with two-qubit gate parameter $\theta = 0.05$, depolarizing noise rate of $p = 3\%$ and varying MPO ansatz bond dimensions D. Grey dashed line indicates G.S. energy, grey shaded area indicates region of trivial bounds (less than G.S. energy), blue dashed line indicates energy of the completely mixed state $\mathbb{I}/2^N$. The y-axis is scaled by a constant multiplicative factor in the trivial region for visual clarity. The Hamiltonian is shifted and scaled such that its spectrum is in $[0, 1]$.

In Fig. 6.4.1, we numerically exhibit the difference between the bound in Eq. 6.4.4 and $h(\vec{\sigma}^{h,D})$ for a 1D circuit on $N = 40$ qubits [Fig. 6.4.1(a)], which is chosen to find the ground state of a commuting 1D Hamiltonian (see the figure caption for the exact circuit and Hamiltonian). As can be seen from Fig. 6.4.1(b), the lower bound computed from the trace purity-based dual is significantly larger, and thus more representative of the impact of noise on the output energy, than the lower bound provided by Eq. 6.4.4. We point out that the dual variables $\vec{\sigma}^{h,D}$ obtained by TEBD in the Heisenberg picture are not necessarily the globally optimal choice in the space of all MPOs with bond dimension D to evaluate the dual function $h(\vec{\sigma})$. The function $h(\vec{\sigma})$ can potentially be optimized beyond the

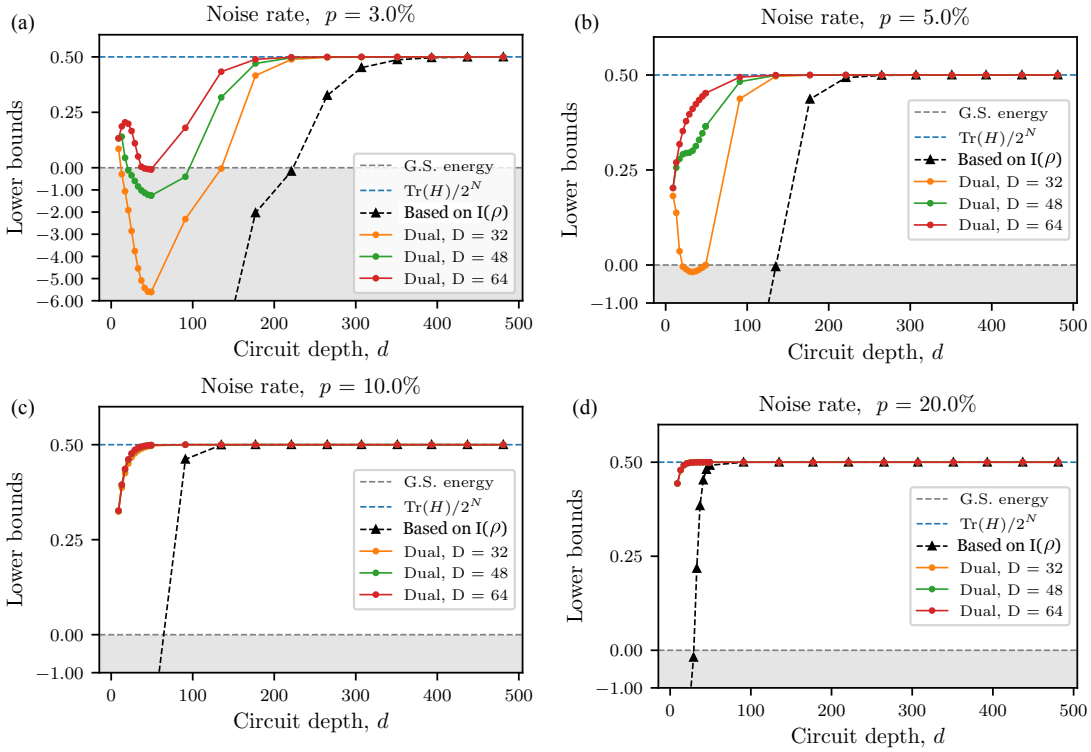


Figure 6.4.2: Comparison of the trace purity-based dual bound ($h(\vec{\sigma}^{h,D})$ in Eq. 6.4.1) and the bound based on just the information content of the output state ($\ell^I_{\lambda_c}$ in Eq. 6.4.6) for 1D many-body spin systems. Both bounds are lower bounds on the ground state (G. S.) energy of the same target Hamiltonian as considered for the results in Fig. 6.4.1 (see description of Hamiltonian in caption of Fig. 6.4.1). Plots show the bounds as a function of brick-wall quantum circuit depth d [Fig. 6.4.1(a)] for a system of $N = 40$ spins, varying MPO ansatz bond dimensions D, with depolarizing noise rates of (a) $p = 3\%$, (b) $p = 5\%$, (c) $p = 10\%$, and (d) $p = 20\%$. Note that at $p = 10\%$ and 20%, approximately the same duality-based lower bound is obtained for different bond dimensions. Two-qubit unitaries in the brick-wall circuit are chosen to be $\exp(-i\theta X \otimes X)$ with $\theta = 0.1$ and single-qubit unitaries are independently Haar random. Grey dashed line indicates G.S. energy, grey shaded area indicates region of trivial bounds (less than G.S. energy), blue dashed line indicates energy of the completely mixed state $\mathbb{I}/2^N$. The y-axis is scaled by a constant multiplicative factor in the trivial region for visual clarity. The Hamiltonian is shifted and scaled such that its spectrum is in $[0, 1]$.

TEBD-based value to obtain better lower bounds. In practice, we observe that local optimization of $h(\vec{\sigma})$ with a gradient-based method starting from the initial point of $\vec{\sigma} = \vec{\sigma}^{h,D}$ yields only a modest improvement over $h\left(\vec{\sigma}^{h,D}\right)$.

Next, we study the improvement that the duality-based bounds that account for the circuit constraint provide over bounds in the existing literature that just account for the information content at the circuit output. In particular, we numerically compare the best lower bound ℓ_D^{dual} that we can obtain by evaluating $h(\vec{\sigma})$ at $\sigma_1, \sigma_2 \ldots \sigma_d \in \text{MPO}_D$ (the space of all N-qubit MPOs of bond dimension D),

$$\ell_D^{\text{dual}} = \underset{\sigma_1,\sigma_2\ldots\sigma_d\in\text{MPO}_D}{\text{maximize}} h(\vec{\sigma}),$$

to the lower bound ℓ^I analyzed in Ref. [85],

$$\ell^I = \underset{\rho:I(\rho)\leq N(1-p)^d}{\text{minimize}} \text{Tr}(H\rho),$$

i.e. where they accounted only for the decreased information content $I(\rho) = N - \text{Tr}[\rho \log_2 \rho]$ of the final state as per Lemma 1. First, we show that there exists a Hamiltonian and a 1D circuit where ℓ_D^{dual}, with $D = O(N)$, scales super-exponentially with the depth of the circuit and thus captures the propagation of errors through the circuit, while ℓ^I scales at-most exponentially with the circuit depth.

Proposition 3. *There exists a 1D circuit and a N-qubit Hamiltonian H with $\text{Tr}(H) = 0$ and $\|H\| = N$, such that $\ell^I = -N(1-p)^{O(d)}$, while $\ell_D^{\text{dual}} = -N(1-p)^{\Omega(d^2)} + O(\sqrt{N})$ for a choice of $D \leq O(N)$.*

Proof sketch (see appendix 6.A for details): Consider a Clifford circuit chosen at random from the ensemble of entangle-unentangle circuits analyzed in Ref. [84] — it was shown for this ensemble that, for a 1D circuit, on average, the energy of the output state with respect to a 2-local Hamiltonian converges to the energy of the maximally mixed state as $\sim \text{poly}(N) \times (1-p)^{\Omega(d^2)}$. Consider now the Hamiltonian $H = -\sum_{i=1}^N Z_i$ and initial state $|0\rangle^{\otimes N}$. In the Heisenberg picture, each Z_i will be mapped to exactly one Pauli string under the action of Clifford gates [217, 218]. Since a Pauli string is representable as an MPO of bond dimension 1, σ_t obtained from Heisenberg picture evolution will be a sum of N MPOs of bond-dimension 1 and will thus be a MPO of bond-dimension at most N. Thus, from Proposition 1, the purity-based dual exactly matches the energy of the output of the quantum circuit, which will scale as $-N(1-p)^{\Omega(d^2)}$ [84, 219]. As the bound without circuit constraints is agnostic to the unitaries in a circuit, it is also a lower bound on the circuit where all the unitaries are just the identity operation. For this trivial circuit, the energy of the state after d layers of just depolarizing noise scales as $-N(1-p)^d$. Hence, the lower bound without circuit constraints $\ell^I = -N(1-p)^{O(d)}$.

6.4.2 Numerical studies

For non-Clifford circuits, MPO parametrization with bond dimension poly(N) is no longer expected to fully capture Heisenberg picture evolution. Even so, we numerically demonstrate for a 1D spin system that the circuit dual function Eq. 6.4.1 can be used to compute circuit-specific bounds that perform better than bounds that only take into account the information content of the output state. We compute the bounds by evaluating the dual function $h(\vec{\sigma})$ at the dual variables obtained from TEBD on the Heisenberg picture as in Eq. 6.4.2. Fig. 6.4.2 shows numerical studies of the bounds $h(\vec{\sigma}^{h,D})$ computed in this manner — we consider a 1D spin system of size $N = 40$ and circuits designed to prepare the ground state of a commuting local Hamiltonian (see figure caption for details). The plots in Fig. 6.4.2 show the bounds for MPO ansatzes with different bond dimensions D plotted against the circuit depth d for circuits with noise rates $p = 3\%$ [Fig. 6.4.2(a)], $p = 5\%$ [Fig. 6.4.2(b)], $p = 10\%$ [Fig. 6.4.2(c)], and $p = 20\%$ [Fig. 6.4.2(d)]. However, for the lowest noise rate $p = 3\%$, the circuit dual bounds at intermediate depths are trivial i.e. lower than the ground state energy of H — this can be attributed to the fact that the intermediate depth regime is the regime where the MPO ansatz is least representative. For shorter depths, the bond dimension of the Heisenberg picture operator would not have grown very much, while for very long depths, the action of the depolarizing noise reduces the bond dimension of the Heisenberg picture operator.

Figure 6.4.2 also compares the trace purity-based dual bound to the information content-based bound. However, since the duality-based bound is exactly computable on a classical computer, to make a fair comparison we need to use a certifiable method for computing the information-content based bound. In particular, using Lagrangian duality, the information content-based bound can be reframed in terms of the Gibbs free energy of the problem Hamiltonian i.e.

$$\begin{aligned} \ell^I &= \underset{\rho: I(\rho) \leq N(1-p)^d}{\text{minimize}} \ \text{Tr}(H\rho) \\ &= \underset{\lambda \geq 0}{\text{maximize}} \ \lambda S_d + G(H, \lambda), \end{aligned} \tag{6.4.5}$$

where $S_d = N - N(1-p)^d$ and $G(H, \lambda) = -\lambda \log \text{Tr} \exp(-H/\lambda)$ is the Gibbs free energy of H at temperature λ. However, since H is generally a many-body Hamiltonian, an accurate evaluation of $G(H, \lambda)$ can only be guaranteed at sufficiently high temperatures [220]. Thus, instead of evaluating the bound ℓ^I in Eq. 6.4.5, we introduce a lower bound λ_c on the temperature λ and evaluate

$$\ell^I_{\lambda_c} = \underset{\lambda \geq \lambda_c}{\text{maximize}} \ \lambda S_d + G(H, \lambda), \tag{6.4.6}$$

For spatially local Hamiltonians, λ_c can be chosen depending on the norms of the local terms in the Hamiltonian, the dimensionality of the lattice, and the interaction

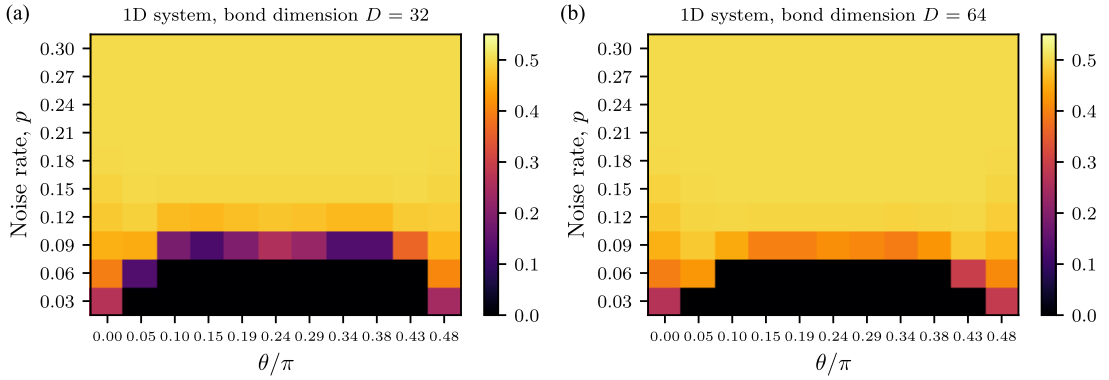


Figure 6.4.3: Trace purity-based dual bounds ($h(\vec{\sigma}^{h,D})$ in Eq. 6.4.1) for a 1D system of $N = 32$ spins, the same target Hamiltonian as considered for the results in Fig. 6.4.1 (see description of Hamiltonian in caption of Fig. 6.4.1), and brick-wall quantum circuits [Fig. 6.4.1(a)] where two-qubit unitaries are chosen to be $\exp(-i\theta X \otimes X)$ and single-qubit unitaries are independently Haar random. Plots show dual bounds as a function of noise rate p and circuit parameter θ for circuit depth $d = 25$ and bond dimensions (a) $D = 32$ and (b) $D = 64$. The Hamiltonian is shifted and scaled such that its spectrum is in $[0, 1]$.

range. In our calculations, we make the choice of $\lambda_c = 8e^3$ — this is based on Ref. [220] which, to the best of our knowledge, provides the only rigorous algorithm that works for evaluating $G(H, \lambda)$ at temperatures above λ_c.

We see from Fig. 6.4.2 that the dual provides a tighter lower bound on the output than the bound based on just the information content of the output state. The information content-based bounds shown in Fig. 6.4.2 are also trivial (i.e. lower than the ground-state energy) for intermediate *and* short depths — this is due to the temperature lower bound that needs to be introduced to ensure computability of the Gibbs free energy. We also observe that the separation between the information content-based and circuit dual bounds increases with the bond dimension D as the MPO ansatz becomes more expressive with increasing bond dimension. In the limit of large circuit depth at non-zero depolarizing noise rates, the state of the circuit approaches the completely mixed state, and we observe that both bounds also approach the energy corresponding to the completely mixed state.

Finally, we demonstrate that the dual bounds are able to capture the extent of entanglement being generated in a circuit. In the brick-wall quantum circuits we consider, the two-qubit gates in the circuit $\exp(-i\theta X \otimes X)$ are parametrized by an angle θ which controls the entanglement being produced — for example, at $\theta = 0, \pi/2$, there is no entanglement at all. Fig. 6.4.3 shows the bounds as a function of the angular parameter θ and the noise rate p, for constant bond dimensions D and circuit depths d, for a 1D system of $N = 32$ spins and Fig. 6.4.4(d,e) show the

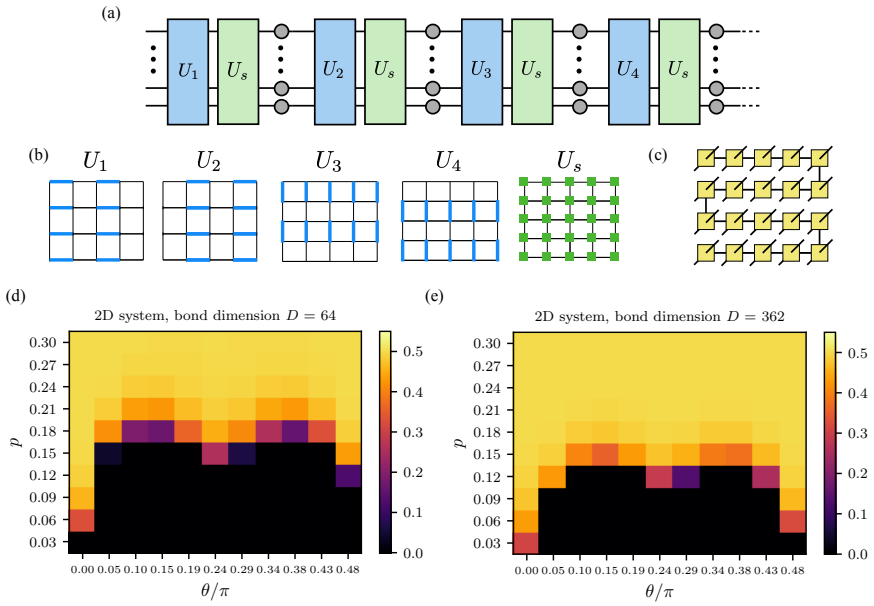


Figure 6.4.4: (a) Schematic of quantum circuits considered for 2D spin systems on a square lattice: colored boxes indicate unitaries and grey circles depolarizing noise. Each unitary layer consists of two-qubit unitaries $\exp(-i\theta X \otimes X)$ (blue boxes) followed by independently Haar random single-qubit unitaries (green boxes). The first $d/2$ layers serve to increase the entanglement in the state. The remaining layers invert the action of the previous $d/2$ such that, in the absence of noise, the output of the circuit is the ground state of $H = -\sum_{\langle i,j \rangle} Z_i Z_j$ where Z_i is the Pauli-Z operator for the i^{th} spin and $\langle i, j \rangle$ indicates nearest-neighbors. (b) Circuits considered have a brick-wall structure: two-qubit unitary layers cycle between gates on odd horizontal edges (U_1), even horizontal edges (U_2), odd vertical edges (U_3), and even vertical edges (U_4). Single-qubit gates (U_s) are applied on every qubit after every two-qubit gate layer. (c) Structure of the MPO considered for 2D spin systems: yellow squares indicate tensors at each site in the 2D lattice and lines emerging from them indicate tensor indices. Horizontal lines indicate bond indices with dimension D and diagonal lines indicate physical indices. (d,e) Trace purity-based dual bounds ($h(\vec{\sigma}^{h,D})$ in Eq. 6.4.1) for the ground state energy of the target Hamiltonian H as a function of noise rate p and circuit parameter θ for a system of $N = 36$ spins arranged in a 6×6 lattice, circuit depth $d = 32$ and MPO bond dimensions (d) $D = 64$ and (e) $D = 362$. The Hamiltonian is shifted and scaled such that its spectrum is in $[0, 1]$.

same for a 2D system of $N = 36$ spins in a 6×6 lattice. For the 2D system, we consider the MPO ansatz to have a 'snake-like' bond structure on the 2D lattice

[Fig. 6.4.4(c)] — such a snake-like structure is a numerically convenient approach for performing TEBD for 2D systems. This ansatz is useful for moderate system sizes but due to gates along the vertical edges of the lattice the bond dimension required grows rapidly. For larger system sizes, we expect that a tensor network ansatz that matches the architecture of the circuit [221–223] would give better bounds. For both the 1D and 2D systems, the target Hamiltonians are shifted and scaled such that the ground state energies are zero and any bounds lower than zero are considered trivial and represented as zero in the plots — the black regions in the plots thus correspond to trivial bounds. We observe that, near $\theta = 0, \pi/2$, where the entanglement is small, the MPO ansatz of constant bond dimension used for the bounds is able to capture it and we obtain non-trivial bounds for small noise rates $p \approx 6\%$. For values of θ away from these limits, the region of triviality is larger but non-trivial dual bounds can still be obtained for higher noise rates.

6.4.3 Non-depolarizing noise models

Up until now, we have modeled the noise present in the circuit as depolarizing. However, noise in several experimental systems might have a more complex structure. In this subsection, we consider these other noise models and show that the duality-based bounding procedure can be reformulated slightly to provide informative bounds even without the assumption of the depolarizing noise. The only assumption that we make that the noise channel under consideration has a full Kraus rank (i.e. the Kraus operators describing the noise channel span the entire space of singe qubit operators). This assumption could be seen as a reasonable physical assumption for sufficiently generic noise models — if the Kraus operators are interpreted effectively as operators that randomly act on the qubit when it is experiencing noise, the Kraus operators not spanning the full space of linear operators would mean that the noise is special and does *not* apply an entire subspace of operators on the qubit. Nevertheless, for channels that do not have full Kraus rank, the methods presented in this section do not apply and we leave it as an open problem for future work.

First, consider noise channels that are unital and primitive (i.e. have identity as a fixed point) — in this case, the noise channel $\mathcal{N}$ with noise rate p can always be expressed as

$$\mathcal{N}(\rho) = (1-p)\rho + V\Bigg(\sum_{P\in\{X,Y,Z\}} p_P PU\rho U^\dagger P\Bigg)V^\dagger, \tag{6.4.7}$$

for a single-qubit unitaries U, V and $p_X, p_Y, p_Z \in (0,1)$ with $p = p_X + p_Y + p_Z < 1$. In this case, as detailed in appendix, it follows from an application of a corollary of Ref. [224] that Lemma 1 can be extended to this class of channels with the noise rate being chosen as $\min(p_X, p_Y, p_Z)$.

Lemma 2 (Follows from Corollary 5.6 of Ref. [224]). *Suppose ρ_t is the quantum state on N qubits obtained from an initial pure state after applying t unitaries followed by single qubit noise channels of the form of Eq. 6.4.7, then*

$$P_{tr}(\rho_t) := \mathrm{Tr}(\rho_t^2) \leq 2^{-N(1-(1-\min(p_x,p_y,p_z))^t)}. \tag{6.4.8}$$

The case of non-unital noise channels is more complex, and it is not possible to get architecture independent bounds on the entropy or trace-purity of the time-dependent state of the quantum circuit. We restrict ourselves to the case where the non-unital noise channel under consideration has a unique fixed point τ — this noise channel, then, tends to drive the output of a quantum circuit on N qubits to the state $\tau^{\otimes N}$. Therefore, instead of using a trace-purity constraint while formulating the dual, as we have for unital noise channels, we instead use a constraint on $\left\|\rho_t - \tau^{\otimes N}\right\|_F^2$, where $\|X\|_F^2 = \mathrm{Tr}(X^\dagger X)$ is the Frobenius norm of X. To obtain a simple analytical upper bound on $\left\|\rho_t - \tau^{\otimes N}\right\|_F^2$, we assume that the two-qubit gates in the quantum circuit are diagonal in the basis of eigenvectors of τ. This is the case, for e.g., if the non-unital noise channel is amplitude damping and the two-qubit gates used in the quantum circuit are all control phase gates. We point out that if prior information about the fixed point of the noise channel is known (for e.g. from a previous noise tomography), then a universal gate set can always be chosen such that the two-qubit gates satisfy this requirement.

More concretely, suppose that the unitary in time step t has single-qubit gates $V_{t,\alpha}^{(1)}$. We also assume that the noise channel, $\mathcal{N}$, is non-unital and has a full Kraus rank, which we expect to be true for noisy systems if the noise is sufficiently generic. Denoting the fixed point of $\mathcal{N}$ by τ, we show in appendix that if $\mathcal{N}$ has a full Kraus rank, then

$$\mathcal{N} = \Lambda_{\tau,q}\mathcal{N}', \tag{6.4.9}$$

for some channel $\mathcal{N}'$ which also has τ as a fixed point and $\Lambda_{\tau,q}(X) = (1-q)X + q\mathrm{Tr}(X)\tau$ for $0 < q < 1$. The parameter q can be interpreted as the probability with which the noise channel traces out the qubit and replaces it with τ, and it can be computed by solving the following semi-definite program

$$\begin{aligned} \max_{q} \quad & q \\ \text{s.t.} \quad & \Phi_{\mathcal{N}} - qI \otimes \tau \succeq 0, \\ & q > 0, q < 1. \end{aligned} \tag{6.4.10}$$

where $\Phi_{\mathcal{N}}$ is the Choi state corresponding to $\mathcal{N}$. For such noise channels and unitary circuits,

Lemma 3. *[Follows from Lemma 1 of Ref. [85]] Suppose ρ_t is the quantum state on N qubits obtained from an initial state ρ_0 after applying t unitaries followed by single qubit noise channel $\mathcal{N}$ of the form of Eq. 6.4.9, then it follows that*

$$D(\rho_t\|\tau^{\otimes N}) \leq D(\rho_0\|\tau^{\otimes N})(1-q)^t + 2\sum_{s=0}^{t-1}\sum_{\alpha}(1-q)^{t-s}\log_2(\|\tau^{-1/2}V_{\alpha,s}\tau V_{\alpha,s}^{\dagger}\tau^{-1/2}\|) \tag{6.4.11}$$

where $D(\rho_1||\rho_2) = \mathrm{Tr}[\rho_1\log_2\rho_1 - \rho_1\log_2\rho_2]$ is the quantum relative entropy between ρ_1 and ρ_2.

We outline a full proof of this lemma in appendix B. To translate the upper bound on the quantum relative entropy to an upper bound on $\left\|\rho_t - \tau^{\otimes N}\right\|_F$, we note that

$$\left\|\rho_t - \tau^{\otimes N}\right\|_F^2 \leq 2\left\|\rho_t - \tau^{\otimes N}\right\|_1 \leq \left[2D(\rho_t\|\tau^{\otimes N})\right]^{1/2},$$

where we have used that $\|O\|_2^2 \leq \|O\|_1\|O\|$ and the Pinsker's inequality $\|\rho_1 - \rho_2\|_1 \leq \sqrt{\frac{1}{2}D(\rho_1\|\rho_2)}$. We remark that our bounds become trivial (i.e.$\to\infty$) when τ is not full rank which would also imply, by the quantum Perron-Frobenius theorem, that the noise channel doesn't have a full Kraus rank.

Following the same procedure as in Section 6.3, we can now formulate the following optimization problem for the energy at the output quantum circuit.

$$\begin{aligned} \underset{\rho_1,\rho_2\cdots\rho_d\in\mathcal{S}}{\text{minimize}} \quad & \mathrm{Tr}(H\rho_d) \\ \text{subject to} \quad & \rho_t = \mathcal{E}_t(\rho_{t-1}),\ t\in\{1,\ldots,d\}, \\ & \left\|\rho_t - \tau^{\otimes N}\right\|_F^2 \leq d_t^2,\ t\in\{1,\ldots,d\}, \end{aligned} \tag{6.4.12}$$

where, instead of a trace purity bound as in Eq. 6.3.5, we use an upper bound on the Frobenius norm distance between ρ_t and $\tau^{\otimes N}$ with d_t given by lemma 3. Again, introducing the dual variables $\sigma_1,\sigma_2\ldots\sigma_d \in \mathcal{M}((\mathbb{C}^2)^{\otimes N})$ and $\lambda_1,\lambda_2\ldots\lambda_t \geq 0$, we can construct the Lagrangian

$$\mathcal{L}(\vec{\rho},\vec{\sigma},\vec{\lambda}) = \mathrm{Tr}[H\rho_d] + \sum_{t=1}^{d}\mathrm{Tr}\Big[\sigma_t\big(\rho_t - \mathcal{E}_t(\rho_{t-1})\big)\Big] + \sum_{t=1}^{d}\lambda_t\Big[\left\|\rho_t - \tau^{\otimes N}\right\|_F^2 - P_t\Big]. \tag{6.4.13}$$

Minimizing the Lagrangian over $\vec{\rho}$ and then maximizing it over $\vec{\lambda}$ yields the dual function $h(\vec{\sigma})$

$$h(\vec{\sigma}) = -\operatorname{Tr}(\sigma_1 \mathcal{E}_1(\rho_0)) + \sum_{t=1}^{d} \left[\operatorname{Tr}(H_t \tau^{\otimes N}) - d_t \sqrt{\operatorname{Tr}(H_t^2)} \right],$$

where, again, $H_d = H + \sigma_d$ and $H_t = \sigma_t - \mathcal{E}_{t+1}^{\dagger}(\sigma_{t+1})$ for $t \in \{1, 2 \dots d-1\}$. Similar to the case in the previous sub-sections, due to Lagrangian duality, $g(\vec{\sigma})$ evaluated at any $\vec{\sigma}$ provides a lower bound on the energy at the output of the circuit. Following the strategy in the previous sections, we again evaluate $g(\vec{\sigma})$ at $\vec{\sigma}^h$ given by a TEBD algorithm in the Heisenberg picture (Eq. 6.4.2). As an example, the bounds obtained in a 1D circuit on $N = 30$ qubits, at a noise rate of 3%, with noise modelled by the non-unital replacement channel $\mathcal{N}(X) = (1-q)\rho + q \operatorname{Tr}(X)\tau$, where we assume $\tau = \mathbb{I}/2 + \epsilon Z$. Here ϵ controls how "non-unital" the noise channel is — $\epsilon = 0$ corresponds to the previously studied case of a depolarizing noise channel and in the limit of $\epsilon \to \pm 1/2$, we obtain an amplitude damping channel (which does not have a full Kraus rank). We also compute the bounds obtained by considering only the quantum relative entropy of the circuit output (ρ_d) with respect to the noise channel fixed point $\tau^{\otimes N}$ as given by Eq. 6.4.11 — to obtain this, we note that given an upper bound on $D(\rho_d \| \tau^{\otimes N})$, we can translate it to an upper bound on $\left\| \rho_d - \tau^{\otimes N} \right\|_1$ via the Pinkser's inequality i.e.

$$\left\| \rho_d - \tau^{\otimes N} \right\|_1 \leq \sqrt{\frac{1}{2} D(\rho_d \| \tau^{\otimes N})},$$

and therefore the expected energy $\operatorname{Tr}(H\rho_d)$ at the circuit observable can deviate from $\tau^{\otimes N}$ by at-most $\|H\| \, \|\rho_d - \tau^{\otimes N}\|_1$. Thus we obtain the lower bound

$$\begin{aligned} \operatorname{Tr}(H\rho_d) &\geq \operatorname{Tr}(H\tau^{\otimes N}) - \|H\| \, \|\rho_d - \tau^{\otimes N}\|_1, \\ &\geq \operatorname{Tr}(H\tau^{\otimes N}) - \|H\| \sqrt{\frac{1}{2} D(\rho_d \| \tau^{\otimes N})}. \end{aligned} \tag{6.4.14}$$

We find that using the duality-based bound continues to give informative bounds which are significantly better compared to the bounds attained by just accounting for the distance between the noisy output state and the noise-channel fixed point $\tau^{\otimes N}$. Specifically, as the noise channel becomes increasingly non-unital, then the dual formulation continues to provide non-trivial bounds since it accounts for the circuit architecture, and the bounds attained without accounting for the circuit architecture become trivial for even slightly non-unital channels.

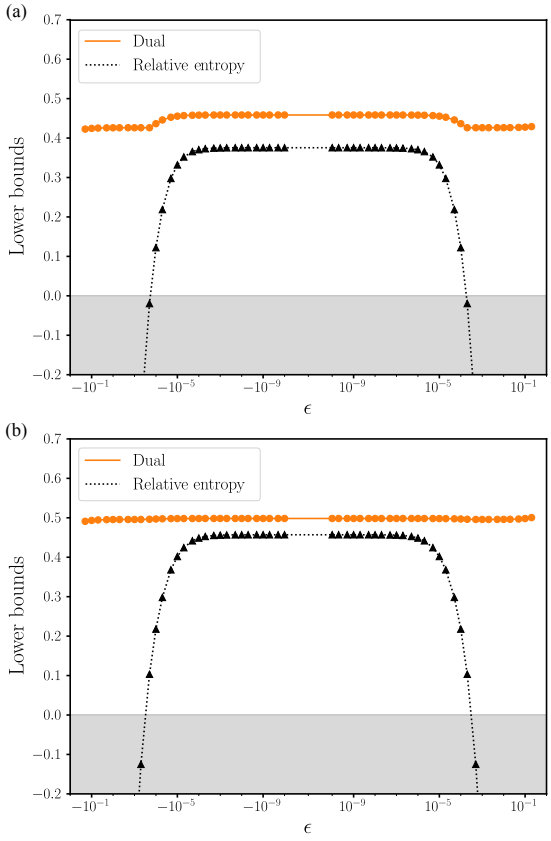


Figure 6.4.5: Bounds on the output energy for the case of a non-unital noise channel. The structure of the circuits are identical to those considered in Fig. 6.4.1(a) with parameters $N = 32$, $d = 102$ and $\theta = 0.1$. The local noise after each layer of unitaries is given by $\mathcal{N}(X) = (1-q)X + q\,\mathrm{Tr}(X)\tau_\epsilon$, with $\tau_\epsilon = (1/2+\epsilon)|0\rangle\langle 0| + (1/2-\epsilon)|1\rangle\langle 1|$, with $q = 0.03$ in (a) and $q = 0.05$ in (b). In orange, we show the bounds derived on the output energy from our dual problem with an MPO ansatz with $D = 128$ for the dual variables $\sigma_1, \sigma_2 \ldots \sigma_d$. The black lines show the lower bound $\mathrm{Tr}(H\tau^{\otimes N}) - ||H||||\rho_d - \tau^{\otimes N}||_1$ which is obtained by disregarding the circuit architecture.

6.5 Using duality with Information content

In the previous sections, we have investigated the impact of noise using the trace purity as a measure of the mixedness in the noisy circuit. The trace purity-based dual function Eq. 6.4.1 contains terms with Frobenius norms $\sqrt{\mathrm{Tr}(H_t^2)}$ which, in the worst case, could grow exponentially with the system size N. Hence, the trace purity-based dual tends to become trivial in the limit of large system size and intermediate circuit depths. An alternative better conditioned purity measure is the information

content based on the Von-Neumann entropy, $I(\rho) = N - S(\rho) = N + \mathrm{Tr}(\rho \log_2(\rho))$. In this section, we formulate a duality-based bound using the information content as a purity measure. However, as we illustrate below, the duality bound here is harder to compute than the one based on trace-purity for general spin model — to still gauge the efficacy of this bound, we numerically study in the simpler but physically relevant case of Gaussian fermions. While our results are suggestive that using the information-content based bounds could be useful for spin models, the associated dual function is harder to compute classically — we leave it as an open problem to develop classical algorithms to compute informative bounds using this strategy for spin models.

Considering the information content, the free energy defined in Eq. 6.3.7b becomes the Gibbs free energy with an offset,

$$\begin{aligned}\mathcal{F}_{\mathcal{S},P}(H,\lambda) &= \inf_{\rho\succeq 0,\mathrm{Tr}(\rho)=1}\Big(\mathrm{Tr}[H\rho] + \lambda I(\rho)\Big),\\ &= N\lambda + \inf_{\rho\succeq 0,\mathrm{Tr}(\rho)=1}\big(\mathrm{Tr}(H\rho) - \lambda S(\rho)\big),\\ &= N\lambda - \lambda \log \mathrm{Tr}\exp\left(-H/\lambda\right), \end{aligned} \tag{6.5.1}$$

which together with Eq. 6.3.7 yields,

$$\begin{aligned}\tilde{g}(\vec{\sigma},\vec{\lambda}) = &-\mathrm{Tr}[\rho_0 \mathcal{E}_1^\dagger(\sigma_1)]\\ &+\sum_{t=1}^{d}\Big(-\lambda_t \log \mathrm{Tr}\exp\left(-H_t/\lambda_t\right) + \lambda_t(N - I_t)\Big),\end{aligned} \tag{6.5.2}$$

where $I_t = N(1-p)^t$ is the analytical bound on the information content under depolarizing noise defined in Lemma 1.

To benchmark the performance of the information content-based dual, we consider Gaussian fermionic systems where the dual function Eq. 6.5.2 can be computed exactly. We study N fermions arranged on a lattice and choose H to be a quadratic Hamiltonian,

$$H = i\sum_{\alpha,\alpha',x,x'} h_{x,x'}^{\alpha,\alpha'} c_x^\alpha c_{x'}^{\alpha'},$$

where c_x^1, c_x^2 are the Majorana operators for the fermion at point x on the lattice, and $h_{x,x'}^{\alpha,\alpha'}$ are real numbers specifying H. We additionally assume the unitaries in the circuit that prepares the ground state of H from an initial vacuum state to be Gaussian unitaries.

Since both Gaussian unitaries and the depolarizing channel map a quadratic Hermitian operator to another quadratic Hermitian operator, Proposition 1 indicates

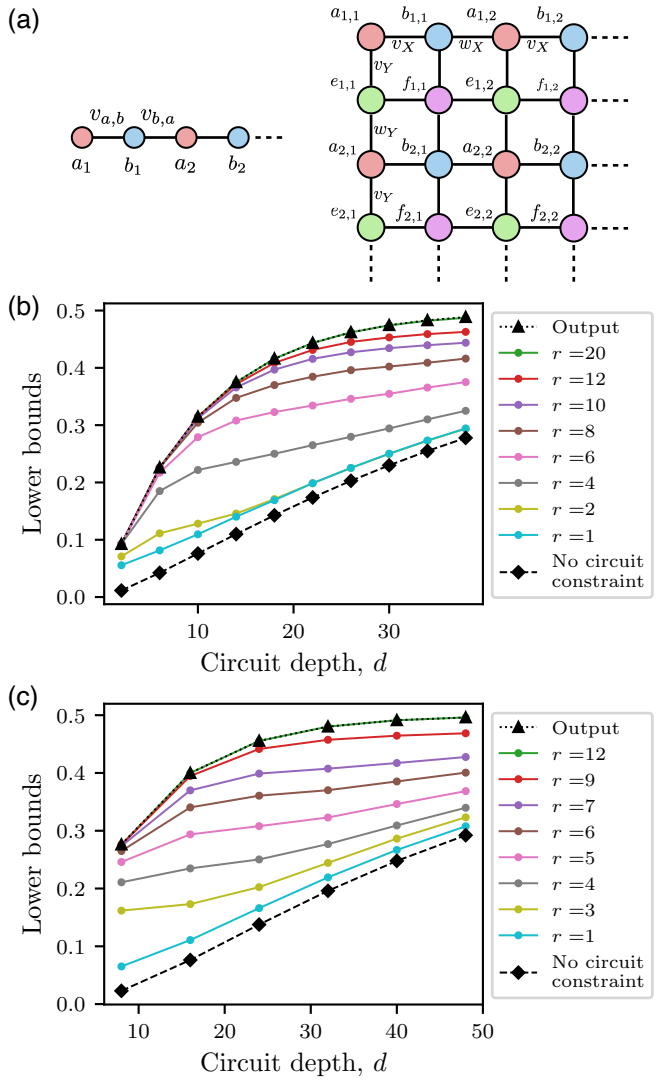


Figure 6.5.1: (a) Schematic of SSH model quadratic fermionic Hamiltonians with alternating hopping strengths in 1D and 2D. (b,c) Comparison of information content-based dual bounds with and without circuit constraints, and the output energies of noisy Gaussian circuits for systems consisting of (b) $N = 48$ fermions arranged in a 1D lattice, (c) $N = 49$ fermions arranged in a 7×7 2D lattice. Dual bounds are shown for ansatzes with varying interaction range r. The horizontal axis represents the depth d of a Gaussian brick-wall circuit that outputs the ground state of the SSH model. Fermions are independently subject to depolarizing noise with probability $p = 5\%$ after every unitary layer. The Hamiltonians are shifted and scaled such that their spectrum is in $[0, 1]$.

that the dual function is maximized for σ_t which themselves are quadratic Hermitian operators. This motivates the following ansatz for σ_t

$$\sigma_t = i \sum_{\substack{\alpha,\alpha',x,x' \\ d(x,x')\leq r}} s^{\alpha,\alpha'}_{x,x';t} c^{\alpha}_x c^{\alpha'}_{x'},$$

for real $s^{\alpha,\alpha'}_{x,x';t}$. In our study, we restrict σ_t to be local operators with interaction range r while maximizing $g(\vec{\sigma},\vec{\lambda})$ to obtain the lower bound — when $r \sim$ lattice size, we expect to obtain the best possible lower bound but since the ansatz always includes the point $\vec{\sigma}=0$, we expect from Proposition 2 to obtain a bound better than that predicted by only considering the information content of the output state, even for small r. Choosing σ_t to be quadratic Hermitian operators allows for exact, classically efficient computation of the Gibbs free energy terms in the dual function and, furthermore, even the circuit output can be computed exactly by considering the covariance matrix describing the state — see appendix B for details. We obtain bounds by maximizing $g(\vec{\sigma},\vec{\lambda})$ through a gradient-based local optimization algorithm (L-BFGS-B), starting from the initial point where σ_t are chosen to be the Heisenberg picture evolution of $-H$, but projected on to the space of quadratic fermionic Hamiltonians with interaction range r after each time step — much like the compression into MPOs of bond dimension D in Eq. 6.4.2.

Figure 6.5.1 shows a numerical study of the bounds that we obtain — we consider systems with ~ 50 fermions arranged both on 1D [Fig. 6.5.1(b)] and 2D lattices [Fig. 6.5.1(c)] and experiencing depolarizing noise at a rate of 5%. H is chosen to be a SSH model, nearest neighbor Hamiltonian with alternating hopping strengths [Fig. 6.5.1(a)]. For the 1D benchmarks, we choose

$$H = \sum_x \left(v_{a,b} a^\dagger_x b_x + v_{b,a} b^\dagger_x a_{x+1} + \text{h.c.} \right), \tag{6.5.3a}$$

and for 2D benchmarks, we choose

$$\begin{aligned} H = \sum_{x,y} \Bigg(& \sum_{\substack{p,q\in \\ \{(a,b),(e,f)\}}} \left(v_{p,q} p^\dagger_{x,y} q_{x,y} + v_{q,p} q^\dagger_{x,y} p_{x+1,y} + \text{h.c.} \right) \\ & + \sum_{\substack{p,q\in \\ \{(a,e),(b,f)\}}} \left(v_{p,q} p^\dagger_{x,y} q_{x,y} + v_{q,p} q^\dagger_{x,y} p_{x,y+1} + \text{h.c.} \right) \Bigg), \end{aligned} \tag{6.5.3b}$$

where we choose $v_{a,b} = v_{ef} = v_x, v_{b,a} = v_{f,e} = w_x, v_{a,e} = v_{b,f} = v_y$ and $v_{e,a} = v_{f,b} = w_y$. For the numerical studies shown in Fig. 6.5.1, H is shifted and scaled such that the ground and highest excited state energies are zero and one, respectively. We

consider circuits of depth d consisting of two-mode Gaussian unitaries arranged in a brick-wall layout, where the first $d/2$ layers are composed of randomly generated two-mode Gaussian unitaries that serve to increase the entanglement in the state. The remaining layers invert the action of the previous $d/2$ such that, in the absence of noise, the output state is the initial state, which is chosen to be the ground state of H. In Fig. 6.5.1, for comparison, we also include the exact output of the noisy Gaussian circuit, as well as the bound obtained by neglecting the circuit constraints and only considering the information content of the output state. As expected, we find that on accounting for the circuit constraint, we obtain bounds that are more representative of the output. We also observe that the dual bounds get closer to the output as the dual ansatz's interaction range r increases, since the ansatz becomes more expressive.

6.6 Discussion

In conclusion, we demonstrate a method to rigorously lower bound the performance of any given quantum circuit subject to a constant rate of depolarizing noise. We achieve this by constructing a Lagrangian dual specific to the circuit, which takes into consideration not only the decreasing purity of the state through the circuit due to noise, but also the details of the gates in the circuit, allowing the study of the effect of entanglement generation in the circuit that can worsen the detrimental effects of noise. We presented numerical studies in spin systems and showed that it is possible to efficiently calculate circuit-specific lower bounds that are tighter than bounds obtained by just considering the information content of the output state. We provided an interpretation of the trace purity-based dual evaluated at dual variables obtained from TEBD in the Heisenberg picture in terms of the compression errors. We also showed that the dual can be formulated in terms of the information content of the state instead of trace purity — we computed information content-based circuit dual bounds for Gaussian fermionic systems where the Gibbs free energy can be computed exactly.

Our method opens the door to promising avenues of future research. Larger-scale tensor-network numerics can allow us to also study higher dimensional circuits, with a number of qubits reaching state-of-the-art experiments. Numerical optimization algorithms can be explored for obtaining better lower bounds than those from evaluating the dual function at a specific point. Moreover, extensions of the methods of this paper to continuous-time would better capture the experimental system, and even allow us to apply this method for understanding quantum adiabatic algorithms [70]. Finally, the Lagrangian dual formulation of lower bounds, apart from being a numerical tool, could also shed rigorous theoretical insights in understanding resilience of quantum circuit architectures to noise.

6.A Proof of proposition 4

We first recall the main result of Ref. [84] which analyzed error propagation in a family of random quantum circuits. Specifically, they considered brick-wall quantum circuits of depth d (assumed to be even) with unitaries $U_1, U_2 \dots U_d$, where

$$U_{d/2+1} = U_{d/2}^\dagger, U_{d/2+2} = U_{d/2-1}^\dagger, U_{d/2+3} = U_{d/2-2}^\dagger \dots U_d = U_1^\dagger.$$

The unitaries $U_1, U_2 \dots U_{d/2}$ are chosen randomly depending on the circuit architecture. We will specifically consider the 1D case, where $U_1, U_3, U_5 \dots$ are formed by applying random 2-qubit gates between qubits $(1,2), (3,4)(5,6) \dots$ and $U_2, U_4, U_6 \dots$ are formed by applying random 2-qubit gates between qubits $(2,3), (4,5), (6,7) \dots$. All the two qubit gates are chosen independently at random from an ensemble that forms a 2-design. Furthermore, we consider the noisy setting where depolarizing noise with probability p is applied to each qubit after every unitary layer. Ref. [84] establishes the following result characterizing the average energy of the output state for a 2-local Hamiltonian. While the result of Ref. [84] holds for arbitrary two-local Hamiltonians, we will only consider the Hamiltonian $H = -\sum_{i=1}^N Z_i$.

Lemma 4 (Ref. [84])**.** *The expectation value E of the Hamiltonian $H = -\sum_{i=1}^N Z_i$ with respect to the the output state of a circuit chosen randomly from the ensemble described above satisfies,*

$$\text{Prob}\left(\left|E + N(1-p)^{\Omega(d^2)}\right| \leq \alpha_0 \sqrt{N}\right) \geq 1 - 2e^{-2\alpha_0^2/2d^2}.$$

Proof of proposition 4: Consider the Hamiltonian $H = -\sum_{i=1}^N Z_i$, and choose the two qubit gates to be a Haar-random Clifford gate — since Haar-random Clifford gates form a 2-design [225], we can use lemma A.1. H satisfies $\text{Tr}(H) = 0$ and has operator norm $\|H\| = N$. We evaluate the dual function at the dual variables obtained by Heisenberg picture evolution of $-H$,

$$\sigma_d = -H, \sigma_t = -\mathcal{E}_{t+1}^\dagger \mathcal{E}_{t+2}^\dagger \dots \mathcal{E}_d^\dagger(H).$$

Note that for Clifford circuits, Z_i will be mapped to a single Pauli string [218], which is expressible as an MPO with bond dimension 1, and consequently σ_t will have a bond-dimension of at-most N. Since the Heisenberg picture evolution can be captured exactly with a MPO ansatz of bond dimension $O(N)$, the dual bound ℓ_D^{dual} with $D \leq O(N)$ is exactly equal to the energy E of the output of the noisy quantum circuit. Now, from lemma A.1, it follows that there must exist at least one 1D circuit such that

$$\ell_D^{\text{dual}} = -N(1-p)^{\Omega(d^2)} + O(\sqrt{N}),$$

with $D \leq O(N)$.

We now consider the lower bound ℓ^I obtained by only considering the information content of the state at the output of the circuit and neglecting the circuit constraints:

$$\ell^I = \underset{\rho: I(\rho) \leq N(1-p)^d}{\text{minimize}} \ \text{Tr}(H\rho).$$

A negative upper bound on ℓ^I which converges to $\text{Tr}(H/2^N) = 0$ exponentially with d can be obtained by computing $\text{Tr}(H\rho)$ at

$$\rho = \left(p_d|0\rangle\langle 0| + (1-p_d)\frac{\mathbb{I}}{2}\right)^{\otimes N} \text{ where } p_d = (1-p)^d.$$

Note that this ρ satisfies $I(\rho) = N(1-p)^d$, and consequently $\text{Tr}(H\rho)$ is an upper bound on ℓ^I. For $H = \sum_{i=1}^N Z_i$, we then obtain $\text{Tr}(H\rho) = -N(1-p)^d$ which implies $\ell^I \leq -N(1-p)^{O(d)}$. □

6.B Non-depolarizing noise channels

In this subsection, we obtain bounds on trace purity in the presence of non-depolarizing noise channels. The bounds presented here are already contained in or can be straightforwardly obtained from existing results (for e.g. in Refs. [85, 202, 224]). We include this appendix for a self-contained derivation of the results used in the main text.

Unital noise channels. We first consider unital noise channels that are also primitive. It is a standard result from the characterization of qubit channels that such a noise channel can be expressed as

$$\mathcal{N}(X) = (1-p)X + \sum_{a \in \{x,y,z\}} p_a \mathcal{U}(A\mathcal{V}(X)A), \tag{6.B.1}$$

where $p = p_x + p_y + p_z$ can be considered to be the noise rate, and $\mathcal{V}(\cdot) = V \cdot V^\dagger$, $\mathcal{U}(\cdot) = U \cdot U^\dagger$ are unitary channels with V, U being unitaries. Furthermore, if $\mathcal{N}$ to be primitive (i.e. have $\mathbb{I}/2$ as the unique fixed point), then $p_x, p_y, p_z > 0$. To obtain a bound on the trace purity of the qubit's state in a quantum circuit impacted by such a noise, we need the following lemma from Ref. [224]. Here, $\|\cdot\|_{2\to 2}$ of a super-operator refers to the Schatten-2 norm i.e.

$$\|\mathcal{E}\|_{2\to 2}^2 = \sup_{X \neq 0} \frac{\text{Tr}(\mathcal{E}(X)^\dagger \mathcal{E}(X))}{\text{Tr}(X^\dagger X)}.$$

Furthermore, for $\sigma \succ 0$, $D_2(\rho||\sigma)$ is the sandwiched 2-Renyi divergence that is given by

$$D_2(\rho||\sigma) = \log_2 \text{Tr}\big(\sigma^{-1/2}\rho\sigma^{-1/2}\rho\big).$$

In particular, note that if ρ is a $N-$qubit density matrix and $\sigma = (\mathbb{I}/2)^{\otimes N}$, we obtain that

$$D_2(\rho||\sigma) = N + \log \mathrm{Tr}(\rho^2).$$

Lemma 5 (Corollary 5.6 from Ref. [224]). *Suppose $\Lambda_{\sigma,p}$ is the channel described by $\Lambda_{\sigma,q}(X) = (1-q)X + q\,\mathrm{Tr}(X)\sigma$. For $\sigma \succ 0$, define Γ_σ to be the superoperator $\Gamma_\sigma(X) = \sigma X \sigma^{-1}$. Suppose $\mathcal{N}$ be a channel with σ as its fixed point, then If $\|\Gamma_\sigma^{-1/2}\mathcal{N}\Lambda_{\sigma,q}^{-1}\Gamma_\sigma^{1/2}\|_{2\to 2} \leq 1$, then for any $N > 0$ and $N-$qubit density matrix ρ*

$$D_2(\mathcal{N}^{\otimes N}(\rho)||\sigma^{\otimes N}) \leq \alpha D_2(\rho||\sigma^{\otimes N}),$$

where $\alpha = 2^{2(1-\|\sigma^{-1}\|^{-1})\log_2(1-p)/\log_2(\|\sigma^{-1}\|)}$.

Proof of lemma 2: We pick $\mathcal{N}$ to be the unital noise channel in Eq. 6.B.1, and $\sigma = \mathbb{I}/2$. With this choice, we have that $\Lambda_{\mathbb{I}/2,q}(X) = (1-q)X + q\mathrm{Tr}(X)\mathbb{I}/2$ and $\Gamma_\sigma(X) = 4X$. Note that $\|\Gamma_{\mathbb{I}/2}^{-1/2}\mathcal{N}\Lambda_{\sigma,p}^{-1}\Gamma_{\mathbb{I}/2}^{1/2}\|_{2\to 2} = \|\mathcal{N}\Lambda_{\mathbb{I}/2,p}^{-1}\|_{2\to 2}$ and that $\Lambda_{\mathbb{I}/2,p}^{-1}(X) = (1-p)^{-1}\big(X - p\mathrm{Tr}(X)\mathbb{I}/2\big)$. Further analysis is simplified in the Pauli basis — since both $\mathcal{N}$ and $\Lambda_{\mathbb{I}/2,q}^{-1}$ are unital superoperators, it follows that, if written as 4×4 matrices in the Pauli basis, they have the form

$$\Lambda_{\mathbb{I}/2,q}^{-1} \cong \begin{bmatrix} 1 & 0 \\ 0 & \tilde{\Lambda}_{\mathbb{I}/2,q}^{-1} \end{bmatrix} \text{ and } \mathcal{N} \cong \begin{bmatrix} 1 & 0 \\ 0 & \tilde{\mathcal{N}} \end{bmatrix}.$$

Furthermore we can explicitly calculate $\tilde{\Lambda}_{\mathbb{I}/2,q}^{-1}$ and $\tilde{\mathcal{N}}$ to obtain

$$\tilde{\Lambda}_{\mathbb{I}/2,q}^{-1} = \frac{1}{1-q}\mathbb{I} \text{ and } \tilde{\mathcal{N}} = (1-p)\mathbb{I} + \tilde{U}\begin{bmatrix} p_x - p_y - p_z & 0 & 0 \\ 0 & p_y - p_x - p_z & 0 \\ 0 & 0 & p_z - p_x - p_y \end{bmatrix}\tilde{V},$$

where $\tilde{U}$ is a 3×3 unitary matrix with matrix elements given by $\tilde{U}_{a,a'} = \mathrm{Tr}(AUA'U^\dagger)$ for $A, A' \in \{X, Y, Z\}$, and $\tilde{V}$ is defined similarly. Now, $\|\mathcal{N}\Lambda_{\mathbb{I}/2,q}^{-1}\|_{2\to 2} = \max(1, \|\tilde{\Lambda}_{\mathbb{I}/2,q}^{-1}\tilde{\mathcal{N}}\|)$ and

$$\|\tilde{\Lambda}_{\mathbb{I}/2,q}^{-1}\tilde{\mathcal{N}}\| = \frac{1}{1-q}\|\tilde{N}\| \leq \frac{1}{1-q}\Big(1-p+\max(p_x-p_y-p_z, p_y-p_x-p_z, p_z-p_x-p_y)\Big).$$

Now, we clearly have that $\max(p_x - p_y - p_z, p_y - p_x - p_z, p_z - p_x - p_y) \leq p - \min(p_x, p_y, p_z)$ and thus $\|\tilde{\Lambda}_{\mathbb{I}/2,q}^{-1}\tilde{\mathcal{N}}\| \leq (1 - \min(p_x, p_y, p_z))/(1-q)$. Thus, choosing $q = \min(p_x, p_y, p_z)$ yields that $\|\Gamma_{\mathbb{I}/2}^{-1/2}\mathcal{N}\Lambda_{\sigma,p}^{-1}\Gamma_{\mathbb{I}/2}^{1/2}\|_{2\to 2} = \|\mathcal{N}\Lambda_{\mathbb{I}/2,q}^{-1}\|_{2\to 2} \leq 1$ — thus, we can now apply lemma 5 with $\sigma = \mathbb{I}/2$, $q = \min(p_x, p_y, p_z)$ which yields $\alpha =$

$1 - \min(p_x, p_y, p_z)$. In particular, if ρ is a $N-$qubit density matrix, we obtain from lemma 5 that

$$D_2\left(\mathcal{N}^{\otimes N}(\rho)\middle\|\frac{\mathbb{I}^{\otimes N}}{2^N}\right) \leq \big(1 - \min(p_x, p_y, p_z)\big) D_2\left(\rho\middle\|\frac{\mathbb{I}^{\otimes N}}{2^N}\right)$$

Now consider the setting where starting from a pure state ρ_0, a sequence of N-qubit unitaries $U_1, U_2 \dots U_d$ is applied interspersed with the noise channel $\mathcal{N}$ acting on each qubit. Since $D_2(U_i \rho U_i^\dagger \| (\mathbb{I}/2)^{\otimes N}) = D_2(\rho \| (\mathbb{I}/2)^{\otimes N})$, the final state $\rho_d = \mathcal{N}^{\otimes N}\mathcal{U}_d \mathcal{N}^{\otimes N}\mathcal{U}_{d-1} \dots \mathcal{N}^{\otimes N}\mathcal{U}_1(\rho_0)$ satisfies

$$D_2\left(\rho_d\middle\|\frac{\mathbb{I}^{\otimes N}}{2^N}\right) \leq \big(1 - \min(p_x, p_y, p_z)\big)^d D_2\left(\rho_0\middle\|\frac{\mathbb{I}^{\otimes N}}{2^N}\right) = N(1 - \min(p_x, p_y, p_z))^d.$$

This completes the proof. □

Non-unital noise channels. Next, we consider non-unital noise channels on $\mathbb{C}^d$ (where we are typically interested in $d = 2$) — we will restrict ourselves to noise channels which have a full Kraus rank i.e. the Kraus operators $K_1, K_2 \dots K_{d^2}$ span the entire space of operators on $\mathbb{C}^d$. Given a noise rate p, we will assume that the noise channel $\mathcal{N}$ is given by

$$\mathcal{N}(X) = (1 - p)X + p \sum_{i=1}^{d^2} K_i X K_i^\dagger, \tag{6.B.2}$$

where $\sum_{i=1}^{d^2} K_i^\dagger K_i = \mathbb{I}$. A common example of such a channel would be $\mathcal{N}(X) = (1 - p)X + p\mathrm{Tr}(X)\tau$, where $\tau \succ 0$ i.e. a channel that traces the qudit and replaces it with a, generally non-identity, full rank state. Physically, this would be a good model for an environment that disentangles the qubits in the quantum circuit, and brings them to a finite temperature state.

For channels of the form of Eq. 6.B.2, the following lemma straightforwardly follows.

Lemma 6. *Suppose $\mathcal{N}$ is a channel of the form given in Eq. 6.B.2 with unique fixed point $\tau \succ 0$, $\exists$ another channel $\mathcal{N}'$ with fixed point τ such that $\mathcal{N}(X) = \Lambda_{\tau,q}\mathcal{N}'(X)$, where $\Lambda_{\tau,q}(X) = (1-q)X + q\,\mathrm{Tr}(X)\tau$, for some $q \in (0, 1)$. Furthermore, the largest such q can be computed by solving the semi-definite program*

$$\begin{aligned} \max_{q} \quad & q \\ \text{s.t.} \quad & \Phi_{\mathcal{N}} - q\mathbb{I} \otimes \tau \succeq 0, \\ & q > 0, q < 1. \end{aligned} \tag{6.B.3}$$

where $\Phi_{\mathcal{N}}$ is the Choi state of $\mathcal{N}$.

Proof: Since $\mathcal{N}$ has a full Kraus rank, we can find linearly independent operators $L_1, L_2 \dots L_{d^2}$ such that $\sum_{i=1}^{d^2} L_i^\dagger L_i = \mathbb{I}$ and $\mathcal{N}(X) = \sum_{i=1}^{d^2} L_i X L_i^\dagger$. Furthermore, since $L_1, L_2 \dots L_{d^2}$ are linearly independent, it follows that for any $M \in \mathbb{C}^{d\times d}$, $\mathrm{Tr}(L_i^\dagger M) = 0 \; \forall \; i \in \{1, 2 \dots d^2\}$ implies that $M = 0$. Consequently, $\exists \lambda_0 > 0$ such that

$$\forall M \in \mathbb{C}^{d\times d} : \sum_{i=1}^{d^2} \Big|\mathrm{Tr}(L_i^\dagger M)\Big|^2 \geq \lambda_0 \mathrm{Tr}(M^\dagger M). \tag{6.B.4}$$

Now, suppose $q \in (0,1)$, and $\mathcal{N}' = (\mathcal{N} - q\,\mathrm{Tr}(\cdot)\tau)/(1-q)$. It is clear that $\mathcal{N}'$ is trace-preserving. We need to establish that $\mathcal{N}'$ is also completely positive. For that, consider the Choi State corresponding to $\mathcal{N}'$, $\Phi_{\mathcal{N}'}$:

$$\begin{aligned}
\Phi_{\mathcal{N}'} &= (\mathcal{N}' \otimes \mathrm{id})(|\Phi\rangle\langle\Phi|), \\
&= \frac{1}{1-q}\Bigg(\sum_{i=1}^{d^2} \sum_{j,k=1}^{d} L_i |j\rangle\langle k| L_i^\dagger \otimes |j\rangle\langle k| - q\mathbb{I}\otimes\tau \Bigg), \\
&= \frac{1}{1-q}\Bigg(\sum_{i=1}^{d^2} \sum_{j,k,j',k'=1}^{d} (L_i)_{j',j} (L_i)^*_{k',k} |j'\rangle\langle k'| \otimes |j\rangle\langle k| - qI\otimes\tau \Bigg).
\end{aligned}$$

For $\mathcal{N}'$ to be completely positive, it is necessary and sufficient for $\Phi_{\mathcal{N}'} \succeq 0$. To impose this condition, consider a state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$, then

$$\begin{aligned}
\langle\psi|\Phi_{\mathcal{N}'}|\psi\rangle &= \frac{1}{1-q}\Bigg(\sum_{i=1}^{d^2} \sum_{j,k,j',k'=1}^{d} (L_i)_{j',j} (L_i)^*_{k',k} \psi^*_{j',j}\psi_{k',k} - q\langle\psi|\mathbb{I}\otimes\tau|\psi\rangle \Bigg), \\
&= \frac{1}{1-q}\Bigg(\sum_{i=1}^{d^2} \Big|\mathrm{Tr}(L_i^\dagger \Psi)\Big|^2 - q\langle\psi|\mathbb{I}\otimes\tau|\psi\rangle \Bigg),
\end{aligned}$$

where $\Psi \in \mathbb{C}^{d\times d}$ is the state $|\psi\rangle$ reshaped as a matrix. Using Eq. 6.B.4, we obtain that

$$\langle\psi|\Phi_{\mathcal{N}'}|\psi\rangle \geq \frac{1}{1-q}\Bigg(\mathrm{Tr}(\Psi^\dagger\Psi)\lambda_0 - q\,\|\tau\|\,\||\psi\rangle\|^2 \Bigg) = \frac{\lambda_0 - q\,\|\tau\|}{1-q}\,\||\psi\rangle\|^2 .$$

Thus, for $q = \lambda_0/\,\|\tau\|$, we obtain that $\Phi_{\mathcal{N}'} \succeq 0$ and hence $\mathcal{N}'$ is completely positive. This establishes that there exists $q > 0$ such that $\mathcal{N} = (1-q)\mathcal{N}' + q\,\mathrm{Tr}(\cdot)\tau$. Furthermore, it trivially follows that $\mathcal{N}'(\tau) = \tau$ from this definition of $\mathcal{N}'$. It therefore also follows that $\mathcal{N} = \Lambda_{\tau,q}\mathcal{N}'$. The optimization problem written for the calculation of q in Eq. 6.4.10 is simply a reformulation of the condition that the Choi state of $\mathcal{N}'$ needs to be positive semi-definite. □

Finally, we now consider the setting of $N-$qubits which have unitaries $U_1, U_2 \dots U_d$ applied to them, along with noise $\mathcal{N}$ acting on each qubit at every time-step. As described in the main text, we will assume that each unitary U_t is composed of two-qubit and single-qubit gates, with the gate set being chosen such that the two-qubit gates leave $\sigma^{\otimes 2}$ invariant (i.e. the two-qubit gate is diagonal on the basis of eigenvectors of σ).

Proof of lemma 3: Suppose ρ is a $N-$qubit density matrix, and consider the state $\mathcal{N}^{\otimes N}(\rho)$. We will first establish that

$$D(\mathcal{N}^{\otimes N}(\rho)||\tau^{\otimes N}) \leq (1-q)D(\rho||\tau^{\otimes N}) \tag{6.B.5}$$

Expressing $\mathcal{N} = \Lambda_{\tau,q}\mathcal{N}'$, $\mathcal{N}^{\otimes N} = \Lambda_{\tau,q}^{\otimes N}\mathcal{N}'^{\otimes N}$. Suppose $\rho' = \mathcal{N}'^{\otimes N}(\rho)$ and then $D(\mathcal{N}^{\otimes N}(\rho)||\tau^{\otimes N}) = D(\Lambda_{\tau,q}^{\otimes N}(\rho')||\tau^{\otimes N})$. We point out that $\Lambda_{\tau,q}$ can be viewed as generated from the Lindbladian $\mathcal{L}_\tau = \mathrm{id} - \mathrm{Tr}(\cdot)\tau$ with evolution time $t = -\log(1-q)$. It now follows from theorem 19 of Ref. [226] that if $\tilde{\mathcal{L}}_\tau := \sum_{i=1}^N \mathcal{L}_\tau^{(i)}$, where $\mathcal{L}_\tau^{(i)}$ is the Lindbladian $\mathcal{L}_\tau$ acting on the i^{th} qubit, satisfies $D(e^{\tilde{\mathcal{L}}_\tau t}(\rho')||\tau^{\otimes N}) \leq e^{-t}D(\rho'||\tau^{\otimes N})$. Setting $t = -\log(1-q)$ in this inequality, we obtain that $D(\Lambda_{\tau,q}^{\otimes N}(\rho')||\tau^{\otimes N}) \leq (1-q)D(\rho'||\tau^{\otimes N})$. Now, this yields that

$$\begin{aligned} D(\mathcal{N}^{\otimes N}(\rho)||\tau^{\otimes N}) &\leq (1-q)D(\mathcal{N}'^{\otimes N}(\rho)||\tau^{\otimes N}), \\ = (1-q)D(\mathcal{N}'^{\otimes N}(\rho)||\mathcal{N}'(\tau^{\otimes N})) &\leq (1-q)D(\rho||\tau^{\otimes N}). \end{aligned}$$

Next, we consider one of the unitaries applied in the circuit, U_t. The two-qubit gates in this unitary, by assumption, leave $\tau^{\otimes 2}$ unchanged. Thus, if $V^{(2)}$ is a two-qubit unitary that satisfies this condition, we have that

$$D(V^{(2)}\rho V^{(2)\dagger}||\tau^{\otimes N}) = D(\rho||V^{(2)\dagger}\tau^{\otimes N}V^{(2)}) = D(\rho||\tau^{\otimes N}).$$

Furthermore, if $V^{(1)}$ is a single qubit gate, we obtain from the data processed triangle inequality that

$$D(V^{(1)}\rho V^{(1)\dagger}||\tau^{\otimes N}) \leq D(\rho||\tau^{\otimes N}) + D_\infty(V^{(1)}\tau V^{(1)\dagger}||\tau).$$

Suppose that the single qubits applied in U_t are $V_{t,\alpha}^{(1)}$ for $\alpha \in \{1, 2, 3 \dots\}$, then from the above two inequalities we obtain that

$$D(U_t\rho U_t^\dagger||\tau^{\otimes N}) \leq D(\rho||\tau^{\otimes N}) + \sum_\alpha D_\infty(V_{t,\alpha}^{(1)}\tau V_{t,\alpha}^{(1)\dagger}||\tau). \tag{6.B.6}$$

Recursing Eqs. 6.B.5 and 6.4.7, we obtain that $\rho_t = U_tU_{t-1}\dots U_1\rho_0U_1^\dagger\dots U_{t-1}^\dagger U_t^\dagger$ satisfies

$$D(\rho_t||\tau^{\otimes N}) \leq (1-q)^t D(\rho_0||\tau^{\otimes N}) + \sum_{s=0}^{t-1}\sum_\alpha (1-q)^{t-s} D_\infty(V_{s+1,\alpha}^{(1)}\tau V_{s+1,\alpha}^{(1)\dagger}||\tau)$$

Finally, this bound can be translated to a bound on $\left\|\rho_t - \sigma^{\otimes N}\right\|_F$ by a standard application of Pinsker's inequality as shown

$$\left\|\rho_t - \tau^{\otimes N}\right\|_F^2 \leq 2\left\|\rho_t - \tau^{\otimes N}\right\|_1 \leq \sqrt{2D(\rho_t||\tau^{\otimes N})},$$

which completes the proof of the lemma. □

6.C Free fermion formulation

In this section, we describe our approach to numerical studies of Gaussian fermionic systems. We describe the fermionic system model and how the energy of the state at the output of the noisy circuit and the information content-based dual function (6.5.2) can be efficiently calculated by utilizing covariance matrices and quadratic Hamiltonian transformations.

6.C.1 Model

We consider N fermions arranged in a lattice and define for each lattice site x the Majorana operators c_x^1, c_x^2 as satisfying the following anti-commutation relations,

$$\begin{aligned} &\{c_x^1, c_{x'}^1\} = \{c_x^2, c_{x'}^2\} = \delta_{xx'} \\ &\{c_x^1, c_{x'}^2\} = 0, \ \forall x, x' \in \text{lattice}. \end{aligned} \tag{6.C.1}$$

We choose the target Hamiltonian H to be quadratic,

$$H = i \sum_{\alpha,\alpha',x,x'} h_{x,x'}^{\alpha,\alpha'} c_x^\alpha c_{x'}^{\alpha'}.$$

Defining the operator valued vector,

$$\vec{c} = \left(c_1^1, \ c_2^1, \ \ldots, \ c_N^1, \ c_1^2, \ c_2^2, \ \ldots, \ c_N^2\right)^T, \tag{6.C.2}$$

the target Hamiltonian can be written as $H = i\vec{c}^{\,T} h \vec{c}$ where the matrix h specifying the Hamiltonian H is real and anti-symmetric. The ground state energy of the target Hamiltonian can be computed by diagonalization of its compact representation h [185].

6.C.2 Calculation of the noisy circuit output

The output energy, with respect to the quadratic target Hamiltonian H, can be computed from the covariance matrix of the state at the end of the circuit. We define the covariance matrix of a state ρ as,

$$\gamma_{x,x'}^{\alpha,\alpha'} = i\text{Tr}(\rho[c_x^\alpha, c_{x'}^{\alpha'}]), \tag{6.C.3}$$

where $[.,.]$ denotes the commutator. The energy of a state ρ with respect to the target Hamiltonian H is then [185],

$$\mathrm{Tr}(\rho H) = -\frac{\mathrm{Tr}(h\gamma)}{2}. \tag{6.C.4}$$

The covariance matrix of the state after each instance of Gaussian unitary or noise in the circuit can be calculated by evolving the covariance matrix γ_0 of the vacuum (γ_0 can be analytically calculated [185]) using the following transformations,

1. *Unitary:* Any Gaussian unitary can be expressed as $U = \exp(-iH_U)$, where $H_U = i\vec{c}^{\,T} h_U \vec{c}$ is the generating quadratic Hamiltonian, and it transforms the covariance matrix of the state as [185],

$$\gamma \to e^{2h_U} \gamma e^{-2h_U}. \tag{6.C.5}$$

2. *Noise:* Depolarizing noise with probability p acting on the fermion at lattice site x transforms the covariance matrix in the following manner [185, 227],

$$\gamma \to p\tilde{\gamma} + (1-p)\gamma, \tag{6.C.6}$$

where the matrix $\tilde{\gamma}$ is obtained from the input covariance matrix γ by zeroing out the rows and columns corresponding to the fermion at site x, i.e.,

$$\tilde{\gamma}^{\alpha,\alpha'}_{x',x''} = \begin{cases} 0, & \text{if } x' = x \text{ or } x'' = x \\ \gamma^{\alpha,\alpha'}_{x',x''}, & \text{else.} \end{cases} \tag{6.C.7}$$

6.C.3 Calculation of the dual function

We choose the dual variables $\sigma_1, \sigma_2, \ldots, \sigma_d$ to be quadratic Hamiltonians, i.e. we consider the following ansatz,

$$\sigma_t = i \sum_{\substack{\alpha,\alpha',x,x' \\ d(x,x')\leq r}} s^{\alpha,\alpha'}_{x,x';t} c^{\alpha}_x c^{\alpha'}_{x'} = i\vec{c}^{\,T} s_t \vec{c},\ t \in \{1,\ldots,d\},$$

where $d(x,x')$ is a distance measure between two lattice sites and the coefficients $s^{\alpha,\alpha'}_{x,x';t}$ are real and can be arranged into an anti-symmetric matrix s_t. σ_t are thus local Hamiltonians with interaction range r.

In order to evaluate the circuit dual function (6.5.2), we calculate the effect of the circuit channels on the dual variables first, i.e. $\mathcal{E}^{\dagger}_t(\sigma_t)$. Each such channel is the composition of a Gaussian unitary channel and the depolarizing channel and both the unitary and noise channels map quadratic operators to quadratic operators. Hence, the channel's action can be expressed as a transformation of the compact Majorana representation s_t of the dual variable σ_t using the following transformation rules,

1. *Unitary:* Any Gaussian unitary can be expressed as $U = \exp(-iH_U)$, where $H_U = i\vec{c}^{\,T} h_U \vec{c}$ is the generating quadratic Hamiltonian. Conjugation of the dual variable $\sigma_t = i\vec{c}^{\,T} s_t \vec{c}$ by U transforms it into another quadratic operator $\tilde{\sigma}_t = U\sigma_t U^\dagger = i\vec{c}^{\,T}\tilde{s}_t\vec{c}$, with the transformed Majorana representation $\tilde{s}_t$,

$$\tilde{s}_t = e^{2h_U} s_t e^{-2h_U}. \tag{6.C.8}$$

This transformation can be derived by expanding the conjugation using the Baker-Campbell-Hausdorff formula $e^B A e^{-B} = A + [B,A] + \frac{1}{2!}[B,[B,A]] + \frac{1}{3!}[B,[B,[B,A]]] + \ldots$, and using the commutation relation,

$$\begin{aligned}[c_{x_1}^{\alpha_1} c_{x_2}^{\alpha_2}, c_{x_3}^{\alpha_3} c_{x_4}^{\alpha_4}] = &- c_{x_1}^{\alpha_1} c_{x_3}^{\alpha_3} \delta_{\alpha_2,\alpha_4}\delta_{x_2,x_4} + c_{x_1}^{\alpha_1} c_{x_4}^{\alpha_4}\delta_{\alpha_2,\alpha_3}\delta_{x_2,x_3} \\ &- c_{x_3}^{\alpha_3} c_{x_2}^{\alpha_2}\delta_{\alpha_1,\alpha_4}\delta_{x_1,x_4} + c_{x_4}^{\alpha_4} c_{x_2}^{\alpha_2}\delta_{\alpha_1,\alpha_3}\delta_{x_1,x_3}.\end{aligned} \tag{6.C.9}$$

2. *Noise:* The noise channel $\mathcal{N}_x$ corresponding to depolarizing noise on the fermion at site x with probability p transforms the dual variable σ_t as,

$$\mathcal{N}_x(\sigma_t) = (1-p)\sigma_t + p\mathrm{Tr}_x(\sigma_t)\otimes\frac{\mathbb{I}}{2}, \tag{6.C.10}$$

where $\mathrm{Tr}_x(.)$ denotes partial trace over fermion at site x. Expanding out the partial trace term using linearity,

$$\begin{aligned}\operatorname*{Tr}_x(\sigma_t)\otimes\frac{\mathbb{I}}{2} &= \sum_{\alpha',\alpha'',x',x''} i s_{x',x'';t}^{\alpha',\alpha''} \operatorname*{Tr}_x(c_{x'}^{\alpha'} c_{x''}^{\alpha''})\otimes\frac{\mathbb{I}}{2}, \\ &= \sum_{\alpha',\alpha'',x',x''} i s_{x',x'';t}^{\alpha',\alpha''}(1-\delta_{x,x'})(1-\delta_{x,x''}) c_{x'}^{\alpha'} c_{x''}^{\alpha''},\end{aligned} \tag{6.C.11}$$

where we obtain the second line by using $\mathrm{Tr}(c_{x'}^{\alpha'}) = 0$ and $s_{x',x';t}^{\alpha',\alpha'} = 0\ \forall(x',\alpha')$. Thus, the noise channel transforms the dual variable into another quadratic operator $\tilde{\sigma}_t = i\vec{c}^{T}\tilde{s}_t\vec{c}$ with the transformed Majorana representation,

$$\tilde{s}_t = p s_t + (1-p) z_t, \tag{6.C.12}$$

where, using (6.C.11), the matrix z_t is obtained from s_t by zeroing out the rows and columns corresponding to the fermion at site x, i.e.,

$$z_{x',x'';t}^{\alpha',\alpha''} = \begin{cases} 0, & \text{if } x' = x \text{ or } x'' = x \\ s_{x',x'';t}^{\alpha',\alpha''}, & \text{else.}\end{cases} \tag{6.C.13}$$

To calculate the dual function (6.5.2), we need to calculate the Gibbs free energy of $H_t = \sigma_t - \mathcal{E}^\dagger_{t+1}(\sigma_{t+1})$ where H_t is a quadratic operator itself because, as mentioned earlier, the channel $\mathcal{E}^\dagger_{t+1}$ maps quadratic operators to quadratic operators. Therefore, $H_t = i\vec{c}^{\,T} h_t \vec{c}$ where the Majorana representation h_t can be obtained using the transformations described above. We calculate the free energy by diagonalizing H_t. This can be achieved by diagonalizing h_t — as h_t is anti-symmetric, it is possible [185] to find an orthogonal transformation O such that,

$$O^T h_t O = \bigoplus_{\alpha,x} \begin{pmatrix} 0 & \epsilon_{x;t} \\ -\epsilon_{x;t} & 0. \end{pmatrix}$$

We define the operator valued vector $\vec{\tilde{c}} = O^T \vec{c} = (\tilde{c}^1_1, \tilde{c}^1_2, \ldots, \tilde{c}^1_N, \tilde{c}^2_1, \tilde{c}^2_2, \ldots, \tilde{c}^2_N)^T$ and, due to the orthogonality of O, the operators $\tilde{c}^\alpha_x$ satisfy the Majorana anti-commutation relations (6.C.1). Using the newly defined Majorana operators, H_t can be written in a diagonal form as,

$$H_t = i \sum_x \epsilon_{x;t} \left(\tilde{c}^1_x \tilde{c}^2_x - \tilde{c}^2_x \tilde{c}^1_x \right).$$

Using the diagonal form, the Gibbs free energy for H_t is,

$$\begin{aligned} \mathcal{F}(H_t, \lambda_t) &= \log \operatorname{Tr} \exp \left(-\frac{i}{\lambda_t} \sum_x \epsilon_{x;t} \left(\tilde{c}^1_x \tilde{c}^2_x - \tilde{c}^2_x \tilde{c}^1_x \right) \right) \\ &= \log \prod_x \operatorname{Tr} \exp \left(-\frac{i\epsilon_{x;t}}{\lambda_t} \left(\tilde{c}^1_x \tilde{c}^2_x - \tilde{c}^2_x \tilde{c}^1_x \right) \right). \end{aligned}$$

The operator exponentials above can be represented in the Fock space basis of the fermion at site x,

$$\exp \left(-\frac{i\epsilon_{x;t}}{\lambda_t} \left(\tilde{c}^1_x \tilde{c}^2_x - \tilde{c}^2_x \tilde{c}^1_x \right) \right) = \begin{pmatrix} e^{\epsilon_{x;t}/\lambda_t} & 0 \\ 0 & e^{-\epsilon_{x;t}/\lambda_t}. \end{pmatrix}$$

The free energy then simplifies to,

$$\mathcal{F}(H_t, \lambda_t) = \sum_x \log(e^{\epsilon_{x;t}/\lambda_t} + e^{-\epsilon_{x;t}/\lambda_t}),$$

and therefore it can be calculated just from the diagonalization of the $2N \times 2N$ dimensional matrix h_t.

Chapter 7

Outlook

In this thesis, we have contributed to developing tensor network techniques. Despite being a mature subject, they still prove to be a common theme when studying the patterns of entanglement of many quantum particles interacting together, as well as high-dimensional data in a more general setting.

In the first chapter of the thesis, Chapter 1, we have introduced the computational challenge posed by many-body systems. Then, in Chapter 2, we have presented tensor networks in detail, outlining some of the techniques to manipulate them, as well as commenting on their present use and their limitations. At the same time, we have introduced the basics of two related fields where we find new applications of tensor networks: Monte Carlo methods in statistical physics and quantum computation.

In the next chapters, we have presented the new applications of tensor network methods found in this thesis: in Chapter 3, we have shown how tensor networks can generate collective updates for many-body Monte Carlo methods. Using cheap tensor network contractions, we can sample probability distributions that approximate the multi-dimensional probability distribution of interest. By inserting this procedure in a Markov chain, we can correct the bias of the imperfect contraction and obtain unbiased estimates of the observables of the many-body system. Compared to the usual cluster methods, the use of tensor network techniques presents some advantages. First, the tensor network prior is "universal", i.e. it includes all the details of the Hamiltonian of interest. Usual methods rely on clever collective updates that must be tailored for specific Hamiltonians. The advantages that they offer might disappear when the interaction terms are slightly modified or the dimensionality of the problem is changed. However, all this information goes into the tensor network that we use to propose the samples, so it is naturally included in our scheme. Secondly, our scheme proposes global updates: in a single step of our method, all the "spins" in the system are potentially updated. This is a great advantage compared to the single-spin methods, which are still the workhorse

behind many Monte Carlo studies. Lastly, our scheme presents an advantage compared to schemes where expectation values are computed only via tensor network contractions: using a Markov chain corrects the bias any finite bond dimension contraction will have, and the results we obtain show only statistical error, which decreases with the number of samples. Furthermore, our algorithm opens up new possibilities for future research: it would be interesting to explore the models where our algorithm can be applied. We believe that our technique is very competitive in models with disorder, as in those models the usual TN and Monte Carlo techniques fail. TN methods for statistical mechanics are best formulated with translation invariance, which disordered models lack. At the same time, cluster methods are hard to apply to those models, and algorithms with local updates might struggle with very long equilibration and decorrelation times. In comparison, our method naturally takes into account the disorder in the Hamiltonian and can still be used without translation invariance.

In Chapters 4 and 5, we have tackled one of the bigger questions open in the field of tensor networks. Their use for studying dynamical and out-of-equilibrium phenomena. As explained in Chapter 2, standard out-of-equilibrium methods suffer from a phenomenon that impedes the simulation of dynamics to long times, the entanglement barrier. In this thesis, we have explored two different ways to overcome this problem. In Chapter 4, we have abandoned the ambition of having a tensor network representation of the time-evolved state or operator of interest. Instead, we have written the simplest tensor network that gives rise to the out-of-equilibrium expectation value when contracted, and we have examined different ways to contract it. Remarkably, we have found that contracting those tensor networks in the space direction, as opposed to the usual contractions carried out in the time direction, gives rise to different scaling of the entanglement, from which a different scaling of the bond dimension follows. In non-interacting and integrable models, the growth of the temporal entanglement seems to be sub-linear with time, allowing us to continue the simulations to much larger times. For the non-integrable models we have tried, the entanglement growth still seems linear. However, the prefactors are much milder than in the standard direction, allowing us to simulate longer times before the exponential growth of the bond dimension required kicks in. This increase in the accessible time scales allows us to study better the out-of-equilibrium evolution of a quantum quench or the phenomenon of diffusion in non-integrable models. In Chapter 5, we have explored a different direction: in the scenario of a global quantum quench, "generic" many-body systems are expected to thermalize. At long times, local expectation values are expected to be captured by thermal states compatible with the state's initial conserved charges. In many cases, it can be shown that these ensembles have tensor network descriptions. This sets a challenge: at short and long times, there exists a tensor

network description for the behavior of local expectation values. Is it possible to develop methods that predict the behavior at intermediate times? In our work, we have tackled this question, and shown, that in certain models, it is possible to capture the thermalization process using tensor networks by correctly identifying the patterns of entanglement present in the time-evolved state.

Lastly, in Chapter 6, we have focused on a more practical matter. The recent spectacular advances towards constructing quantum computers have fueled the expectations on the field of quantum computation. At the moment, there exist devices that can carry out computational tasks that would take huge amounts of effort to carry out on classical computers. Many people hope that this early-stage quantum computer will be able to have practical uses, despite the noise present in them and the heuristic nature of the algorithms presented for them. In particular, there is hope that variational quantum circuits would outperform classical algorithms, either for preparing ground states of complicated quantum many-body systems or for some optimization tasks. In this chapter, we show how to assess correctly these expectations with tensor networks. We construct a cost function that can be evaluated classically with tensor networks and gives lower bounds on the performance of the quantum circuit for the tasks mentioned before. The main novelty of our work is that we can go beyond the bounds based on the growth of the entropy predicted by the noise. We include information about the specific gates and geometry of the circuit, which helps us outperform previous bounds. Similarly, our new technique performs better for noise models that go beyond the simple assumption of local independent depolarizing noise, which can be relevant for some platforms where amplitude damping or qubit decay are the most common sources of error.

Acknowledgments

I want to express my gratitude to those who supported me during my PhD. I landed at MPQ during hard times, and I am thankful to the people who have made these years into a time I will look back fondly.

First of all, I would like to thank my advisor, Mari Carmen Bañuls, for giving me the opportunity to work with her in Munich and for all her support and valuable feedback throughout this time. I also want to thank Ignacio Cirac for agreeing to be the supervisor of this thesis and for his advice.

I am grateful to my collaborators, Sattwik Deb Mishra, Luca Tagliacozzo, Rahul Trivedi, and Mingru Yang, for their enthusiasm, knowledge, and patience. A special thank you goes to Sofyan Iblisdir for introducing me to the field of tensor networks, for his guidance, and his optimism.

Thanks to the theory group secretaries Andrea Kluth, Andrea Angione, Regina Jasny, and Elena Wiggert for their incredible helpfulness. Thanks also to Sonya Gzyl for making IMPRS a pleasant environment and for her advice and listening when needed.

It was a pleasure to get to know, both academically and personally, the members of the theory group at MPQ. For the fun conversations, the inspiring discussions, and the good times, I would like to thank David Castells, Esther Cruz, Adrián Franco-Rubio, Albert Gasull, Guillermo González-García, Reinis Irmejs, Irene Papaefstathiou, Yorgos Styliaris, Bennet Windt, and many others.

Special thanks to my family, including its latest additions, who always support me. It is encouraging to know that I can always count on them. Finally, I am grateful to Nikol for all her love and support, and for all the beautiful memories we have built and will continue to.

References

[1]R. J. Baxter, *Exactly solved models in statistical mechanics* (Elsevier, 2016).

[2]I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, "Rigorous results on valence-bond ground states in antiferromagnets", Phys. Rev. Lett. **59**, 799–802 (1987).

[3]S. R. White, "Density matrix formulation for quantum renormalization groups", Phys. Rev. Lett. **69**, 2863–2866 (1992).

[4]T. Nishino and K. Okunishi, "Corner transfer matrix algorithm for classical renormalization group", Journal of the Physical Society of Japan **66**, 3040–3047 (1997).

[5]M. Fannes, B. Nachtergaele, and R. F. Werner, "Finitely correlated states on quantum spin chains", Communications in Mathematical Physics **144**, 443–490 (1992).

[6]A. Klümper, A. Schadschneider, and J. Zittartz, "Matrix product ground states for one-dimensional spin-1 quantum antiferromagnets", Europhysics Letters **24**, 293 (1993).

[7]G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, "Entanglement in quantum critical phenomena", Phys. Rev. Lett. **90**, 227902 (2003).

[8]F. Verstraete, D. Porras, and J. I. Cirac, "Density matrix renormalization group and periodic boundary conditions: a quantum information perspective", Phys. Rev. Lett. **93**, 227205 (2004).

[9]F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, "Matrix product density operators: simulation of finite-temperature and dissipative systems", Phys. Rev. Lett. **93**, 207204 (2004).

[10]M. Zwolak and G. Vidal, "Mixed-state dynamics in one-dimensional quantum lattice systems: a time-dependent superoperator renormalization algorithm", Phys. Rev. Lett. **93**, 207205 (2004).

[11]A. E. Feiguin and S. R. White, "Finite-temperature density matrix renormalization using an enlarged hilbert space", Phys. Rev. B **72**, 220401 (2005).

[12]S. R. White, "Minimally entangled typical quantum states at finite temperature", Phys. Rev. Lett. **102**, 190601 (2009).

[13]S. R. White and D. A. Huse, "Numerical renormalization-group study of low-lying eigenstates of the antiferromagnetic s=1 heisenberg chain", Phys. Rev. B **48**, 3844–3852 (1993).

[14]J. Haegeman, B. Pirvu, D. J. Weir, J. I. Cirac, T. J. Osborne, H. Verschelde, and F. Verstraete, "Variational matrix product ansatz for dispersion relations", Phys. Rev. B **85**, 100408 (2012).

[15]J. Cui, J. I. Cirac, and M. C. Bañuls, "Variational matrix product operators for the steady state of dissipative quantum systems", Phys. Rev. Lett. **114**, 220601 (2015).

[16]G. Vidal, "Classical simulation of infinite-size quantum lattice systems in one spatial dimension", Phys. Rev. Lett. **98**, 070201 (2007).

[17]J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, "Classical simulation of infinite-size quantum lattice systems in two spatial dimensions", Phys. Rev. Lett. **101**, 250602 (2008).

[18]G. Vidal, "Entanglement renormalization", Phys. Rev. Lett. **99**, 220405 (2007).

[19]F. Verstraete and J. I. Cirac, "Renormalization algorithms for quantum-many body systems in two and higher dimensions", arXiv preprint cond-mat/0407066 (2004).

[20]F. Verstraete and J. I. Cirac, "Continuous matrix product states for quantum fields", Phys. Rev. Lett. **104**, 190405 (2010).

[21]M. Motta, C. Genovese, F. Ma, Z.-H. Cui, R. Sawaya, G. K.-L. Chan, N. Chepiga, P. Helms, C. Jiménez-Hoyos, A. J. Millis, U. Ray, E. Ronca, H. Shi, S. Sorella, E. M. Stoudenmire, S. R. White, and S. Zhang, "Ground-state properties of the hydrogen chain: dimerization, insulator-to-metal transition, and magnetic phases", Phys. Rev. X **10**, 031058 (2020).

[22]N. Chepiga and F. Mila, "Floating phase versus chiral transition in a 1d hard-boson model", Phys. Rev. Lett. **122**, 017205 (2019).

[23]J. L. Jiménez, S. P. G. Crone, E. Fogh, M. E. Zayed, R. Lortz, E. Pomjakushina, K. Conder, A. M. Läuchli, L. Weber, S. Wessel, A. Honecker, B. Normand, C. Rüegg, P. Corboz, H. M. Rønnow, and F. Mila, "A quantum magnetic analogue to the critical point of water", Nature **592**, 370–375 (2021).

[24]B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, "Stripe order in the underdoped region of the two-dimensional hubbard model", Science **358**, 1155–1160 (2017).

[25]M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, "The hubbard model: a computational perspective", Annual Review of Condensed Matter Physics **13**, 275–302 (2022).

[26]O. Gauthé and F. Mila, "Thermal phase diagram of the square lattice ferro-antiferromagnetic $J_1 - J_2$ heisenberg model", arXiv preprint arXiv:2310.09344 (2023).

[27]M. Ljubotina, M. Žnidarič, and T. Prosen, "Spin diffusion from an inhomogeneous quench in an integrable system", Nature Communications **8**, 16117 (2017).

[28]M. T. Manzoni, D. E. Chang, and J. S. Douglas, "Simulating quantum light propagation through atomic ensembles using matrix product states", Nature Communications **8**, 1743 (2017).

[29]K. Vodenkova and H. Pichler, "Continuous coherent quantum feedback with time delays: tensor network solution", Phys. Rev. X **14**, 031043 (2024).

[30]Y. Zhou, E. M. Stoudenmire, and X. Waintal, "What Limits the Simulation of Quantum Computers?", Phys. Rev. X **10**, 041038 (2020).

[31]F. Pan, K. Chen, and P. Zhang, "Solving the sampling problem of the sycamore quantum circuits", Phys. Rev. Lett. **129**, 090502 (2022).

[32]F. Pan and P. Zhang, "Simulation of Quantum Circuits Using the Big-Batch Tensor Network Method", Phys. Rev. Lett. **128**, 030501 (2022).

[33]J. J. Sakurai, *Modern quantum mechanics; rev. ed.* (Addison-Wesley, Reading, MA, 1994).

[34]P Kramer, "A review of the time-dependent variational principle", Journal of Physics: Conference Series **99**, 012009 (2008).

[35]G. Carleo, F. Becca, M. Schiró, and M. Fabrizio, "Localization and glassy dynamics of many-body quantum systems", Scientific Reports **2**, 243 (2012).

[36]L. Hackl, T. Guaita, T. Shi, J. Haegeman, E. Demler, and J. I. Cirac, "Geometry of variational methods: dynamics of closed quantum systems", SciPost Phys. **9**, 048 (2020).

[37]T. Shi, E. Demler, and J. I. Cirac, "Variational approach for many-body systems at finite temperature", Phys. Rev. Lett. **125**, 180602 (2020).

[38]U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states", Ann. Phys. **326**, January 2011 Special Issue, 96–192 (2011).

[39]R. Orús, "A practical introduction to tensor networks: matrix product states and projected entangled pair states", Annals Phys. **349**, 117–158 (2014).

[40]L. Vanderstraeten, J. Haegeman, and F. Verstraete, "Tangent-space methods for uniform matrix product states", SciPost Phys. Lect. Notes, 7 (2019).

[41]R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement", Rev. Mod. Phys. **81**, 865–942 (2009).

[42]B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, "Matrix product operator representations", New J. Phys. **12**, 025012 (2010).

[43]A. Molnar, N. Schuch, F. Verstraete, and J. I. Cirac, "Approximating gibbs states of local hamiltonians efficiently with projected entangled pair states", Phys. Rev. B **91**, 045138 (2015).

[44]S. Liang and H. Pang, "Approximate diagonalization using the density matrix renormalization-group method: a two-dimensional-systems perspective", Phys. Rev. B **49**, 9214–9217 (1994).

[45]D. N. Page, "Average entropy of a subsystem", Phys. Rev. Lett. **71**, 1291–1294 (1993).

[46]D. Poulin, A. Qarry, R. Somma, and F. Verstraete, "Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space", Phys. Rev. Lett. **106**, 170501 (2011).

[47]J. Eisert, M. Cramer, and M. B. Plenio, "Colloquium: area laws for the entanglement entropy", Rev. Mod. Phys. **82**, 277–306 (2010).

[48]N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, "Entropy scaling and simulability by matrix product states", Phys. Rev. Lett. **100**, 030504 (2008).

[49]F. Verstraete and J. I. Cirac, "Valence-bond states for quantum computation", Phys. Rev. A **70**, 060302 (2004).

[50]F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, "Criticality, the area law, and the computational power of projected entangled pair states", Phys. Rev. Lett. **96**, 220601 (2006).

[51]D. Landau and B. Kurt, *A guide to monte carlo simulations in statistical physics* (Cambridge University Press, 2005).

[52]H. G. Katzgraber, "Introduction to monte carlo methods", arXiv preprint arXiv:0905.1629 (2009).

[53]J. Gubernatis, N. Kawashima, and P. Werner, *Quantum monte carlo methods: algorithms for lattice models* (Cambridge University Press, 2016).

[54]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines", The Journal of Chemical Physics **21**, 1087–1092 (1953).

[55]L. David, P. Yuval, and W. Elizabeth, "Markov chains and mixing times", Markov Chains and Mixing Times, `10.1090/mbk/058` (2008).

[56]W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications", Biometrika **57**, 97–109 (1970).

[57]R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in monte carlo simulations", Phys. Rev. Lett. **58**, 86–88 (1987).

[58]U. Wolff, "Collective monte carlo updating for spin systems", Phys. Rev. Lett. **62**, 361–364 (1989).

[59]K. Hukushima and K. Nemoto, "Exchange monte carlo method and application to spin glass simulations", Journal of the Physical Society of Japan **65**, 1604–1608 (1996).

[60]J. Houdayer, "A cluster monte carlo algorithm for 2-dimensional spin glasses", The European Physical Journal B - Condensed Matter and Complex Systems **22**, 479–484 (2001).

[61]Z. Zhu, A. J. Ochoa, and H. G. Katzgraber, "Efficient cluster algorithm for spin glasses in any space dimension", Phys. Rev. Lett. **115**, 077201 (2015).

[62]R. P. Feynman, "Simulating physics with computers", International Journal of Theoretical Physics **21**, 467–488 (1982).

[63]S. Lloyd, "Universal quantum simulators", Science **273**, 1073–1078 (1996).

[64]J. I. Cirac and P. Zoller, "Goals and opportunities in quantum simulation", Nat. Phys. **8**, 264–266 (2012).

[65]P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring", in Proceedings 35th annual symposium on foundations of computer science (1994), pp. 124–134.

[66]L. K. Grover, "A fast quantum mechanical algorithm for database search", in Proceedings of the twenty-eighth annual acm symposium on theory of computing, STOC '96 (1996), 212–219.

[67]A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations", Phys. Rev. Lett. **103**, 150502 (2009).

[68]M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th anniversary edition* (Cambridge University Press, 2010).

[69]D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, "Adiabatic quantum computation is equivalent to standard quantum computation", SIAM Review **50**, 755–787 (2008).

[70]T. Albash and D. A. Lidar, "Adiabatic quantum computation", Rev. Mod. Phys. **90**, 015002 (2018).

[71]R. Raussendorf and H. J. Briegel, "A one-way quantum computer", Phys. Rev. Lett. **86**, 5188–5191 (2001).

[72]A. Y. Kitaev, "Quantum computations: algorithms and error correction", Russian Mathematical Surveys **52**, 1191–1249 (1997).

[73]D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error rate", SIAM Journal on Computing **38**, 1207–1282 (2008).

[74]F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor", Nature **574**, 505–510 (2019).

[75]H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "Quantum computational advantage using photons", Science **370**, 1460–1463 (2020).

[76]Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, Q. Zhu, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, "Strong quantum computational advantage using a superconducting quantum processor", Phys. Rev. Lett. **127**, 180501 (2021).

[77]J. Preskill, "Quantum Computing in the NISQ era and beyond", Quantum **2**, 79 (2018).

[78]D. Bluvstein, A. A. Evered Simon J.and Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, "Logical quantum processor based on reconfigurable atom arrays", Nature **626**, 58–65 (2024).

[79]R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, N. Astrakhantsev, J. Atalaya, R. Babbush, D. Bacon, B. Ballard, J. C. Bardin, J. Bausch, A. Bengtsson, A. Bilmes, S. Blackwell, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, D. A. Browne, B. Buchea, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, A. Cabrera, J. Campero, H.-S. Chang, Y. Chen, Z. Chen, B. Chiaro, D. Chik, C. Chou, J. Claes, A. Y. Cleland, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, S. Das, A. Davies, L. De Lorenzo, D. M. Debroy, S. Demura, M. Devoret, A. Di Paolo, P. Donohoe, I. Drozdov, A. Dunsworth, C. Earle, T. Edlich, A. Eickbusch, A. Moshe Elbag, M. Elzouka, C. Erickson, L. Faoro, E. Farhi, V. S. Ferreira, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, S. Ganjam, G. Garcia, R. Gasca, É. Genois, W. Giang, C. Gidney, D. Gilboa, R. Gosula, A. Grajales Dau, D. Graumann, A. Greene, J. A. Gross, S. Habegger, J. Hall, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, F. J. H. Heras, S. Heslin, P. Heu, O. Higgott, G. Hill, J. Hilton, G. Holland, S. Hong, H.-Y. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, S. Jordan, C. Joshi, P. Juhas, D. Kafri, H. Kang, A. H. Karamlou, K. Kechedzhi, J. Kelly, T. Khaire, T. Khattar, M. Khezri, S. Kim, P. V. Klimov, A. R. Klots, B. Kobrin, P. Kohli, A. N. Korotkov, F. Kostritsa, R. Kothari, B. Kozlovskii, J. M. Kreikebaum, V. D. Kurilovich, N. Lacroix, D. Landhuis, T. Lange-Dei, B. W. Langley, P. Laptev, K.-M. Lau, L. Le Guevel, J. Ledford, K. Lee, Y. D. Lensky, S. Leon, B. J. Lester, W. Y. Li, Y. Li, A. T. Lill, W. Liu, W. P. Livingston, A. Locharla, E. Lucero, D. Lundahl, A. Lunt, S. Madhuk, F. D. Malone, A. Maloney, S. Mandrá, L. S. Martin, S. Martin, O. Martin, C. Maxfield, J. R. McClean, M. McEwen, S. Meeks, A. Megrant, X. Mi, K. C. Miao, A. Mieszala, R. Molavi, S. Molina, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. How Ng, A. Nguyen, M. Nguyen, C.-H. Ni, T. E. O'Brien, W. D. Oliver, A. Opremcak, K. Ottosson, A. Petukhov, A. Pizzuto, J. Platt, R. Potter, O. Pritchard, L. P. Pryadko, C. Quintana, G. Ramachandran, M. J. Reagor, D. M. Rhodes, G. Roberts, E. Rosenberg, E. Rosenfeld, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, A. W. Senior, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, S. Singh, V. Sivak, J. Skruzny, S. Small, V. Smelyanskiy, W. Clarke Smith, R. D. Somma, S. Springer, G. Sterling, D. Strain, J. Suchard, A. Szasz, A. Sztein, D. Thor, A. Torres, M. Mert Torunbalci, A. Vaishnav, J. Vargas, S. Vdovichev, G. Vidal, B. Villalonga, C. Vollgraff Heidweiller, S. Waltman, S. X. Wang, B. Ware, K. Weber, T. White, K. Wong, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, B. Ying, J. Yoo, N. Yosri, G. Young, A. Zalcman, Y. Zhang, N. Zhu, and N.

Zobrist, "Quantum error correction below the surface code threshold", arXiv e-prints, arXiv:2408.13687, arXiv:2408.13687 (2024).

[80]K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, "Noisy intermediate-scale quantum algorithms", Rev. Mod. Phys. **94**, 015004 (2022).

[81]M. M. Wilde, *Quantum information theory* (Cambridge University Press, 2013).

[82]A. Müller-Hermes, D. Stilck França, and M. M. Wolf, "Relative entropy convergence for depolarizing channels", Journal of Mathematical Physics **57**, 022202 (2016).

[83]D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, "Limitations of noisy reversible computation", arXiv preprint quant-ph/9611028 (1996).

[84]G. González-García, R. Trivedi, and J. I. Cirac, "Error Propagation in NISQ Devices for Solving Classical Optimization Problems", PRX Quantum **3**, 040326 (2022).

[85]D. Stilck França and R. García-Patrón, "Limitations of optimization algorithms on noisy quantum devices", Nat. Phys. **17**, 1221–1227 (2021).

[86]Z. Zhu and H. G. Katzgraber, "Do tensor renormalization group methods work for frustrated spin systems?", arXiv preprint arXiv:1903.07721 (2019).

[87]G Bhanot, "The metropolis algorithm", Reports on Progress in Physics **51**, 429–457 (1988).

[88]K. Ueda, R. Otani, Y. Nishio, A. Gendiar, and T. Nishino, "Snapshot observation for 2d classical lattice models by corner transfer matrix renormalization group", Journal of the Physical Society of Japan **74**, 111–114 (2005).

[89]K. Binder and A. P. Young, "Spin glasses: experimental facts, theoretical concepts, and open questions", Rev. Mod. Phys. **58**, 801–976 (1986).

[90]V. Murg, F. Verstraete, and J. I. Cirac, "Efficient evaluation of partition functions of inhomogeneous many-body spin systems", Phys. Rev. Lett. **95**, 057206 (2005).

[91]B. A. Cipra, "An introduction to the ising model", The American Mathematical Monthly **94**, 937–959 (1987).

[92]T. Horiguchi, "Fully Frustrated Ising Model on a Square Lattice", Progress of Theoretical Physics Supplement **87**, 33–42 (1986).

[93]H. G. Katzgraber, M. Palassini, and A. P. Young, "Monte carlo simulations of spin glasses at low temperatures", Phys. Rev. B **63**, 184422 (2001).

[94]D. H. Bailey, "Algorithm 719: multiprecision translation and execution of fortran programs", ACM Trans. Math. Softw. **19**, 288–319 (1993).

[95]X.-Z. Wang and J. S. Kim, "The critical line of an ising antiferromagnet on square and honeycomb lattices", Phys. Rev. Lett. **78**, 413–416 (1997).

[96]H. W. J. Blöte and M. P. M. den Nijs, "Corrections to scaling at two-dimensional ising transitions", Phys. Rev. B **37**, 1766–1778 (1988).

[97]Z. Y. Xie, J. Chen, M. P. Qin, J. W. Zhu, L. P. Yang, and T. Xiang, "Coarse-graining renormalization by higher-order singular value decomposition", Phys. Rev. B (2012).

[98]S. Wang, Z.-Y. Xie, J. Chen, B. Normand, and T. Xiang, "Phase transitions of ferromagnetic potts models on the simple cubic lattice", Chinese Physics Letters **31**, 070503 (2014).

[99]L. Vanderstraeten, B. Vanhecke, and F. Verstraete, "Residual entropies for three-dimensional frustrated spin systems with tensor networks", Phys. Rev. E **98**, 042145 (2018).

[100]A. M. Ferrenberg, J. Xu, and D. P. Landau, "Pushing the limits of monte carlo simulations for the three-dimensional ising model", Phys. Rev. E **97**, 043301 (2018).

[101]C Domb, "Ising model", Phase transitions and critical phenomena **3**, 357–484 (1974).

[102]Y. Meurice, R. Sakai, and J. Unmuth-Yockey, "Tensor lattice field theory for renormalization and quantum computing", Rev. Mod. Phys. **94**, 025005 (2022).

[103]J. V. José, L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, "Renormalization, vortices, and symmetry-breaking perturbations in the two-dimensional planar model", Phys. Rev. B **16**, 1217–1241 (1977).

[104]J. F. Yu, Z. Y. Xie, Y. Meurice, Y. Liu, A. Denbleyker, H. Zou, M. P. Qin, J. Chen, and T. Xiang, "Tensor renormalization group study of classical XY model on the square lattice", Phys. Rev. E **89**, 013308 (2014).

[105]L. Vanderstraeten, B. Vanhecke, A. M. Läuchli, and F. Verstraete, "Approaching the kosterlitz-thouless transition for the classical XY model with tensor networks", Phys. Rev. E **100**, 062136 (2019).

[106]R. G. Jha, "Critical analysis of two-dimensional classical xy model", Journal of Statistical Mechanics: Theory and Experiment **2020**, 083203 (2020).

[107]Y. Meurice, R. Sakai, and J. Unmuth-Yockey, "Tensor lattice field theory for renormalization and quantum computing", Rev. Mod. Phys. **94**, 025005 (2022).

[108]M. Levin and C. P. Nave, "Tensor renormalization group approach to two-dimensional classical lattice models", Phys. Rev. Lett. **99**, 120601 (2007).

[109]G. Evenbly and G. Vidal, "Tensor network renormalization", Phys. Rev. Lett. **115**, 180405 (2015).

[110]A. N. Chowdhury and R. D. Somma, "Quantum algorithms for gibbs sampling and hitting-time estimation", Quantum Info. Comput. **17**, 41–64 (2017).

[111]M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, "Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution", Nat. Phys. **16**, 205–210 (2020).

[112]L. Saloff-Coste, "Lectures on finite markov chains", Lectures on Probability Theory and Statistics **1665** (1997).

[113]T. Nishino and K. Okunishi, "Corner transfer matrix renormalization group method", Journal of the Physical Society of Japan **65**, 891–894 (1996).

[114]T. Nishino and K. Okunishi, "Corner transfer matrix algorithm for classical renormalization group", Journal of the Physical Society of Japan **66**, 3040–3047 (1997).

[115]M. Lubasch, J. I. Cirac, and M.-C. Bañuls, "Algorithms for finite projected entangled pair states", Phys. Rev. B **90**, 064425 (2014).

[116]P. Calabrese and J. Cardy, "Evolution of entanglement entropy in one-dimensional systems", Journal of Statistical Mechanics: Theory and Experiment **2005**, P04010 (2005).

[117]T. J. Osborne, "Efficient approximation of the dynamics of one-dimensional quantum spin systems", Phys. Rev. Lett. **97**, 157202 (2006).

[118]G. Vidal, "Efficient classical simulation of slightly entangled quantum computations", Phys. Rev. Lett. **91**, 147902 (2003).

[119]D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, "Matrix product state representations", Quantum Info. Comput. **7**, 401–430 (2007).

[120]M. B. Hastings, "Solving gapped hamiltonians locally", Phys. Rev. B **73**, 085115 (2006).

[121]T. Kuwahara, A. M. Alhambra, and A. Anshu, "Improved thermal area law and quasilinear time algorithm for quantum gibbs states", Phys. Rev. X **11**, 011047 (2021).

[122]M. J. Hartmann, J. Prior, S. R. Clark, and M. B. Plenio, "Density matrix renormalization group in the heisenberg picture", Phys. Rev. Lett. **102**, 057202 (2009).

[123]M. C. Bañuls, M. B. Hastings, F. Verstraete, and J. I. Cirac, "Matrix product states for dynamical simulation of infinite chains", Phys. Rev. Lett. **102**, 240603 (2009).

[124]T. Prosen and M. Žnidarič, "Is the efficiency of classical simulations of quantum dynamics related to integrability?", Phys. Rev. E **75**, 015202 (2007).

[125]D. Muth, R. G. Unanyan, and M. Fleischhauer, "Dynamical simulation of integrable and nonintegrable models in the heisenberg picture", Phys. Rev. Lett. **106**, 077202 (2011).

[126]C. D. White, M. Zaletel, R. S. K. Mong, and G. Refael, "Quantum dynamics of thermalizing systems", Phys. Rev. B **97**, 035127 (2018).

[127]T. Rakovszky, C. W. von Keyserlingk, and F. Pollmann, "Dissipation-assisted operator evolution method for capturing hydrodynamic transport", Phys. Rev. B **105**, 075131 (2022).

[128]C Krumnow, J Eisert, and Ö Legeza, "Towards overcoming the entanglement barrier when simulating long-time evolution", arXiv preprint arXiv:1904.11999 (2019).

[129]J. Surace, M. Piani, and L. Tagliacozzo, "Simulating the out-of-equilibrium dynamics of local observables by trading entanglement for mixture", Phys. Rev. B **99**, 235115 (2019).

[130]M. M. Rams and M. Zwolak, "Breaking the entanglement barrier: tensor network simulation of quantum transport", Phys. Rev. Lett. **124**, 137701 (2020).

[131]J. Lopez-Piqueres, B. Ware, S. Gopalakrishnan, and R. Vasseur, "Operator front broadening in chaotic and integrable quantum chains", Phys. Rev. B **104**, 104307 (2021).

[132]T. K. Kvorning, L. Herviou, and J. H. Bardarson, "Time-evolution of local information: thermalization dynamics of local observables", SciPost Phys. **13**, 080 (2022).

[133]A. Müller-Hermes, J. I. Cirac, and M. C. Bañuls, "Tensor network techniques for the computation of dynamical observables in one-dimensional quantum spin systems", New Journal of Physics **14**, 075003 (2012).

[134]M. B. Hastings and R. Mahajan, "Connecting entanglement in time and space: improving the folding algorithm", Phys. Rev. A **91**, 032306 (2015).

[135]M. C. Bañuls, J. I. Cirac, and M. B. Hastings, "Strong and weak thermalization of infinite nonintegrable quantum systems", Phys. Rev. Lett. **106**, 050405 (2011).

[136]L. Piroli, B. Bertini, J. I. Cirac, and T. Prosen, "Exact dynamics in dual-unitary quantum circuits", Phys. Rev. B **101**, 094304 (2020).

[137]M. Sonner, A. Lerose, and D. A. Abanin, "Influence functional of many-body systems: temporal entanglement and matrix-product state representation", Annals of Physics **435**, Special issue on Philip W. Anderson, 168677 (2021).

[138]A. Lerose, M. Sonner, and D. A. Abanin, "Influence matrix approach to many-body floquet dynamics", Phys. Rev. X **11**, 021040 (2021).

[139]E. Ye and G. K.-L. Chan, "Constructing tensor network influence functionals for general quantum dynamics", The Journal of Chemical Physics **155**, 044104 (2021).

[140]E. H. Lieb and D. W. Robinson, "The finite group velocity of quantum spin systems", Communications in Mathematical Physics **28**, 251–257 (1972).

[141]M. B. Hastings and T. Koma, "Spectral gap and exponential decay of correlations", Communications in Mathematical Physics **265**, 781–804 (2006).

[142]B. Nachtergaele and R. Sims, "Lieb-robinson bounds and the exponential clustering theorem", Communications in Mathematical Physics **265**, 119–130 (2006).

[143]M. B. Hastings, "Light-cone matrix product", Journal of Mathematical Physics **50**, 095207 (2009).

[144]T. Enss and J. Sirker, "Light cone renormalization and quantum quenches in one-dimensional hubbard models", New Journal of Physics **14**, 023008 (2012).

[145]E. Gillman, F. Carollo, and I. Lesanovsky, "Numerical simulation of quantum nonequilibrium phase transitions without finite-size effects", Phys. Rev. A **103**, L040201 (2021).

[146]V Zauner, M Ganahl, H. G. Evertz, and T Nishino, "Time evolution within a comoving window: scaling of signal fronts and magnetization plateaus after a local quench in quantum spin chains", Journal of Physics: Condensed Matter **27**, 425602 (2015).

[147]H. N. Phien, G. Vidal, and I. P. McCulloch, "Dynamical windows for real-time evolution with matrix product states", Phys. Rev. B **88**, 035103 (2013).

[148]A. Milsted, J. Haegeman, T. J. Osborne, and F. Verstraete, "Variational matrix product ansatz for nonuniform dynamics in the thermodynamic limit", Phys. Rev. B **88**, 155116 (2013).

[149]J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, "Quantum algorithm for simulating real time evolution of lattice hamiltonians", SIAM Journal on Computing **52**, FOCS18–250–FOCS18–284 (2023).

[150]G. Giudice, G. Giudici, M. Sonner, J. Thoenniss, A. Lerose, D. A. Abanin, and L. Piroli, "Temporal entanglement, quasiparticles, and the role of interactions", Phys. Rev. Lett. **128**, 220401 (2022).

[151]G. Vidal, "Efficient simulation of one-dimensional quantum many-body systems", Phys. Rev. Lett. **93**, 040502 (2004).

[152]S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, "Time-evolution methods for matrix-product states", Annals of Physics **411**, 167998 (2019).

[153]R. Hübener, V. Nebendahl, and W. Dür, "Concatenated tensor network states", New Journal of Physics **12**, 025004 (2010).

[154]B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, "Exponential thermal tensor network approach for quantum lattice models", Phys. Rev. X **8**, 031082 (2018).

[155]T. Barthel, U. Schollwöck, and S. Sachdev, "Scaling of the thermal spectral function for quantum critical bosons in one dimension", arXiv preprint arXiv:1212.3570 (2012).

[156]C. Karrasch, J. H. Bardarson, and J. E. Moore, "Finite-temperature dynamical density matrix renormalization group and the drude weight of spin-1/2 chains", Phys. Rev. Lett. **108**, 227206 (2012).

[157]T. Barthel, "Precise evaluation of thermal response functions by optimized density matrix renormalization group schemes", New Journal of Physics **15**, 073010 (2013).

[158]C. Sünderhauf, D. Pérez-García, D. A. Huse, N. Schuch, and J. I. Cirac, "Localization with random time-periodic quantum circuits", Phys. Rev. B **98**, 134204 (2018).

[159]B. Bertini, F. Heidrich-Meisner, C. Karrasch, T. Prosen, R. Steinigeweg, and M. Žnidarič, "Finite-temperature transport in one-dimensional quantum lattice models", Rev. Mod. Phys. **93**, 025003 (2021).

[160]R. Steinigeweg, H. Wichterich, and J. Gemmer, "Density dynamics from current auto-correlations at finite time- and length-scales", Europhysics Letters **88**, 10004 (2009).

[161]H. Kim and D. A. Huse, "Ballistic spreading of entanglement in a diffusive nonintegrable system", Phys. Rev. Lett. **111**, 127205 (2013).

[162]A. Lerose, M. Sonner, and D. A. Abanin, "Scaling of temporal entanglement in proximity to integrability", Phys. Rev. B **104**, 035137 (2021).

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

[164]A. M. Läuchli and C. Kollath, "Spreading of correlations and entanglement after a quench in the one-dimensional bose–hubbard model", Journal of Statistical Mechanics: Theory and Experiment **2008**, P05018 (2008).

[165]D. Rossini, S. Suzuki, G. Mussardo, G. E. Santoro, and A. Silva, "Long time dynamics following a quench in an integrable quantum spin chain: local versus nonlocal operators and effective thermal behavior", Phys. Rev. B **82**, 144302 (2010).

[166]K. Van Acoleyen, M. Mariën, and F. Verstraete, "Entanglement rates and area laws", Phys. Rev. Lett. **111**, 170501 (2013).

[167]E. Leviatan, F. Pollmann, J. H. Bardarson, D. A. Huse, and E. Altman, "Quantum thermalization dynamics with matrix-product states", arXiv preprint arXiv:1702.08894 (2017).

[168]A. Hallam, J. G. Morley, and A. G. Green, "The lyapunov spectra of quantum thermalisation", Nature Communications **10**, 2708 (2019).

[169]L. Pastori, M. Heyl, and J. C. Budich, "Disentangling sources of quantum entanglement in quench dynamics", Phys. Rev. Research **1**, 012007 (2019).

[170]M. Frías-Pérez and M. C. Bañuls, "Light cone tensor network and time evolution", Phys. Rev. B **106**, 115117 (2022).

[171]M. Srednicki, "Chaos and quantum thermalization", Phys. Rev. E **50**, 888–901 (1994).

[172]C. Gogolin and J. Eisert, "Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems", Reports on Progress in Physics **79**, 056001 (2016).

[173]L. Vidmar and M. Rigol, "Generalized Gibbs ensemble in integrable lattice models", Journal of Statistical Mechanics: Theory and Experiment **2016**, 64007 (2016).

[174]A. Mitra, "Quantum quench dynamics", Annual Review of Condensed Matter Physics **9**, 245–259 (2018).

[175]A. Zamora, J. Rodríguez-Laguna, M. Lewenstein, and L. Tagliacozzo, "Splitting a critical spin chain", Journal of Statistical Mechanics: Theory and Experiment **2014**, P09035 (2014).

[176]Y. Chen and G. Vidal, "Entanglement contour", J. Stat. Mech. **2014**, P10011 (2014).

[177]J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete, "Time-dependent variational principle for quantum lattices", Phys. Rev. Lett. **107**, 070601 (2011).

[178]G. Evenbly and G. Vidal, "Tensor network states and geometry", Journal of Statistical Physics **145**, 891–918 (2011).

[179]G. Evenbly, "A practical guide to the numerical implementation of tensor networks i: contractions, decompositions, and gauge freedom", Frontiers in Applied Mathematics and Statistics **8**, 806549 (2022).

[180]G. Evenbly and G. Vidal, "Algorithms for entanglement renormalization", Phys. Rev. B **79**, 144108 (2009).

[181]T. Kraft, C. Ritz, N. Brunner, M. Huber, and O. Gühne, "Characterizing genuine multilevel entanglement", Phys. Rev. Lett. **120**, 060502 (2018).

[182]J. Hauschild, E. Leviatan, J. H. Bardarson, E. Altman, M. P. Zaletel, and F. Pollmann, "Finding purifications with minimal entanglement", Phys. Rev. B **98**, 235163 (2018).

[183]G. Vidal and R. F. Werner, "Computable measure of entanglement", Phys. Rev. A **65**, 032314 (2002).

[184]M. B. Plenio, "Logarithmic negativity: a full entanglement monotone that is not convex", Phys. Rev. Lett. **95**, 090503 (2005).

[185]J. Surace and L. Tagliacozzo, "Fermionic Gaussian states: an introduction to numerical approaches", SciPost Phys. Lect. Notes, 54 (2022).

[186]L. Amico, R. Fazio, A. Osterloh, and V. Vedral, "Entanglement in many-body systems", Rev. Mod. Phys. **80**, 517–576 (2008).

[187]S. Gharibian, Y. Huang, Z. Landau, and S. W. Shin, "Quantum hamiltonian complexity", Foundations and Trends® in Theoretical Computer Science **10**, 159–282 (2015).

[188]F. Verstraete, M. M. Wolf, and J. Ignacio Cirac, "Quantum computation and quantum-state engineering driven by dissipation", Nat. Phys. **5**, 633–636 (2009).

[189]H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, "Quantum computational advantage using photons", Science **370**, 1460–1463 (2020).

[190]Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, "Evidence for the utility of quantum computing before fault tolerance", Nature **618**, 500–505 (2023).

[191]Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, Q. Zhu, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, "Strong Quantum Computational Advantage Using a Superconducting Quantum Processor", Phys. Rev. Lett. **127**, 180501 (2021).

[192]Y. Ge, J. Tura, and J. I. Cirac, "Faster ground state preparation and high-precision ground energy estimation with fewer qubits", Journal of Mathematical Physics **60**, 022202 (2019).

[193]Y. Ge, A. Molnár, and J. I. Cirac, "Rapid Adiabatic Preparation of Injective Projected Entangled Pair States and Gibbs States", Phys. Rev. Lett. **116**, 080503 (2016).

[194]M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms", Nat. Rev. Phys. **3**, 625–644 (2021).

[195]E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm", arXiv preprint arXiv:1411.4028 (2014).

[196]D. Wecker, M. B. Hastings, and M. Troyer, "Progress towards practical quantum variational algorithms", Phys. Rev. A **92**, 042303 (2015).

[197]M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming", J. ACM **42**, 1115–1145 (1995).

[198]M. F. Anjos, "On semidefinite programming relaxations for the satisfiability problem", Mathematical Methods of Operations Research **60**, 349–367 (2004).

[199]R. Oliveira and B. M. Terhal, "The complexity of quantum spin systems on a two-dimensional square lattice", Quantum Info. Comput. **8**, 900–924 (2008).

[200]D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, "A Polynomial-Time Classical Algorithm for Noisy Random Circuit Sampling", in Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023 (June 2023), pp. 945–957.

[201]R. Trivedi and J. I. Cirac, "Transitions in Computational Complexity of Continuous-Time Local Open Quantum Dynamics", Phys. Rev. Lett. **129**, 260405 (2022).

[202]G. De Palma, M. Marvian, C. Rouzé, and D. S. França, "Limitations of variational quantum algorithms: a quantum optimal transport approach", PRX Quantum **4**, 010309 (2023).

[203]A. Deshpande, P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans, "Tight Bounds on the Convergence of Noisy Random Circuits to the Uniform Distribution", PRX Quantum **3**, 040329 (2022).

[204]R. Trivedi, G. Angeris, L. Su, S. Boyd, S. Fan, and J. Vučković, "Bounds for Scattering from Absorptionless Electromagnetic Structures", Phys. Rev. Appl. **14**, 014025 (2020).

[205]G. Angeris, J. Vuckovic, and S. P. Boyd, "Computational bounds for photonic design", ACS Photonics **6**, 1232–1239 (2019).

[206]K. Schab, A. Rothschild, K. Nguyen, M. Capek, L. Jelinek, and M. Gustafsson, "Trade-offs in absorption and scattering by nanophotonic structures", Opt. Express **28**, 36584–36599 (2020).

[207]P. Chao, B. Strekha, R. Kuate Defo, S. Molesky, and A. W. Rodriguez, "Physical limits in electromagnetism", Nature Reviews Physics **4**, 543–559 (2022).

[208]S. Molesky, P. Chao, W. Jin, and A. W. Rodriguez, "Global $\mathbb{T}$ operator bounds on electromagnetic scattering: upper bounds on far-field cross sections", Phys. Rev. Res. **2**, 033172 (2020).

[209]S. Molesky, P. Chao, and A. W. Rodriguez, "Hierarchical mean-field $\mathbb{T}$ operator bounds on electromagnetic scattering: upper bounds on near-field radiative purcell enhancement", Phys. Rev. Res. **2**, 043398 (2020).

[210]H. Zhang, Z. Kuang, S. Puri, and O. D. Miller, "Conservation-Law-Based Global Bounds to Quantum Optimal Control", Phys. Rev. Lett. **127**, 110506 (2021).

[211]S. D. Mishra, R. Trivedi, A. H. Safavi-Naeini, and J. Vučković, "Control Design for Inhomogeneous-Broadening Compensation in Single-Photon Transducers", Phys. Rev. Applied **16**, 044025 (2021).

[212]M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum rényi entropies: a new generalization and some properties", Journal of Mathematical Physics **54**, 122203 (2013).

[213]S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge University Press, 2004).

[214]M. Slater, "Lagrange Multipliers Revisited", in *Traces and Emergence of Nonlinear Programming*, edited by G. Giorgi and T. H. Kjeldsen (Springer Basel, Basel, 2014), pp. 293–306.

[215]A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal, "Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces", J. Stat. Mech. **2004**, P04005 (2004).

[216]S. R. White and A. E. Feiguin, "Real-Time Evolution Using the Density Matrix Renormalization Group", Phys. Rev. Lett. **93**, 076401 (2004).

[217]D. Gottesman, "The heisenberg representation of quantum computers", arXiv preprint quant-ph/9807006 (1998).

[218]S. Aaronson and D. Gottesman, "Improved Simulation of Stabilizer Circuits", Phys. Rev. A **70**, 052328 (2004).

[219]Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, "Exponentially tighter bounds on limitations of quantum error mitigation", Nat. Phys. **20**, 1648–1658 (2024).

[220]T. Kuwahara, K. Kato, and F. G. S. L. Brandão, "Clustering of Conditional Mutual Information for Quantum Gibbs States above a Threshold Temperature", Phys. Rev. Lett. **124**, 220601 (2020).

[221]F. Pan, P. Zhou, S. Li, and P. Zhang, "Contracting Arbitrary Tensor Networks: General Approximate Algorithm and Applications in Graphical Models and Quantum Circuit Simulations", Phys. Rev. Lett. **125**, 060503 (2020).

[222]I. L. Markov and Y. Shi, "Simulating Quantum Computation by Contracting Tensor Networks", SIAM J. Comput. **38**, 963–981 (2008).

[223]M. Lubasch, J. I. Cirac, and M.-C. Bañuls, "Algorithms for finite projected entangled pair states", Phys. Rev. B **90**, 064425 (2014).

[224]C. Hirche, C. Rouzé, and D. Stilck França, "On contraction coefficients, partial orders and approximation of capacities for quantum channels", Quantum **6**, 862 (2022).

[225]C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation", Phys. Rev. A **80**, 012304 (2009).

[226]S. Beigi, N. Datta, and C. Rouzé, "Quantum reverse hypercontractivity: its tensorization and application to strong converses", Communications in Mathematical Physics **376**, 753–794 (2020).

[227]N. Schuch and B. Bauer, "Matrix product state algorithms for Gaussian fermionic states", Phys. Rev. B **100**, 245121 (2019).