Technische Universität München
Max-Planck-Institut für Quantenoptik

Methods for Quantum Dynamics, Localization and Quantum Machine Learning

Nicola Pancotti

Vollständiger Abdruck der von der Fakultät für Physik
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigten Dissertation.

Vorsitzender: Prof. Dr. Alexander Holleitner

Prüfer der Dissertation: 1. Hon.-Prof. Dr. Ignacio Cirac
2. Prof. Dr. Frank Pollmann

Die Dissertation wurde am 20.02.2020 bei der Technischen Universität München
eingereicht und durch die Fakultät für Physik am 29.06.2020 angenommen.

Abstract

In order to tackle the challenging problems that researchers are facing, it is of great importance in quantitative sciences to develop new methods. Theoretical physics offers an excellent playground in such respect, as it contributes to push the frontiers of mathematics and computer science beyond their limits. An exciting research task is to explore the extent to which these methods can be applied to other research fields. In the last decades, several solutions originally developed in theoretical physics have found applications in Chemistry, Biology, Social and Economic sciences, Neurology, Computer Science, just to mention a few of them. The goal of this work is twofold. On the one hand, we will develop numerical and analytical methods to the study of complex quantum systems. And on the other hand, we will use methods from theoretical quantum physics to address problems in machine learning.

The thesis is divided into three parts. The first part is concerned with the development and application of novel methods which may help us comprehending some features of quantum dynamics. Understanding how an ensemble of many quantum particles evolves in time is of great practical and theoretical interest, albeit poorly mastered. In Part II, we explore scenarios where such dynamics is slow and information is localized. We focus on two different mechanisms which may induce localization, namely disorder and kinetic constraints. In Chapter 1 we employ a technique which was recently introduced in order to construct slowly evolving operators for one-dimensional spin systems. By employing tensor-network representations of operators we study the crossover between to different quantum phases where on the one side of the transition the dynamics is fast and ergodic, and on the other it is slow and localized. We show that the statistics of these quantities can provide insightful information about the structure of the phase transition and that Extreme Value Theory — a well-established theory in statistics which describes the probability of occurrence of rare events — is very well suited for characterizing the probability distributions of certain quantities of interest. In Chapter 2 we turn our attention to classically-inspired quantum glasses: a class of quantum models inspired from classical kinetically-constrained "structural" glasses. We perform an in-depth analysis of the quantum East model by means of analytical and numerical tools such as tensor networks and exact diagonalization. By tuning a single parameter in the Hamiltonian we show that the system undergoes a transition throughout the spectrum where several eigenstates admit an efficient representation at arbitrary energy densities. Furthermore, we construct a large family of orthogonal states with surprisingly slow dynamics, which we can infer from rigorous bounds. All our results apply to one-dimensional lattices, however we provide evidence that similar effects might be present in higher dimensions as well.

In Part III we elaborate on an opposite scenario where, rather than impeding thermalization, we aim at enhancing it. By employing a number of techniques from the theory of open quantum systems and Gaussian states, we design thermodynamic

protocols by engineering time-dependent Hamiltonians of a system interacting with its bath. Our goal is to speed up thermodynamic protocols while the total dissipated work by the system is kept constant. We show that, by adiabatically bringing the system to interact strongly with its bath, we can design a family of protocols which outperform the standard classical bounds. This class of protocols interpolates between the Curzon-Ahlborn efficiency and the Carnot efficiency.

Finally, in Part IV, we examine the intersections between tensor-network methods for quantum mechanics and machine learning methods. In Chapter 4 we adopt neural network quantum states — a class of variational wave functions which was originally inspired from machine learning — as an asatz for approximating the ground state of a strongly interacting Hamiltonian. These states, in the form of a Restricted Boltzmann Machine, provide an excellent and flexible tool whose performances can be compared against standard and well-established variational wave functions. At the same time, we develop a mathematical mapping between the Restricted Boltzmann Machine and a particular class of tensor networks called String Bond States. It is often hard to extract physical insights from neural network quantum states, this mapping provides a simple way to rephrase the problem into the more familiar (and mathematically structured) framework of tensor networks. Motivated by these findings, in Chapter 5 and Chapter 6, we deploy tensor network techniques to solve tasks in machine learning. In particular, in Chapter 5, we generalize the mapping above to the broad class of probabilistic graphical models by showing that they can always be mapped onto some tensor-network factorization. We perform numerical simulations where we show that tensor networks can classify images and sound and, in particular, that they can be combined with modern neural-network architectures providing state-of-the-art performances. In Chapter 6, we use tensor networks for approximating probability distributions. Driven by the success enjoyed by tensor networks in describing the properties of one-dimensional quantum systems, we extract rigorous bounds on the resource requirements of a given tensor decomposition for describing certain classical probability distributions. We show that Hidden Markov models — a popular class of probabilistic graphical models — can be exactly mapped onto Matrix Product States with real and positive tensors. We consider different tensor-network models and we prove that there exist probability distributions for which there are unbounded separations between the resource requirements of some of these factorizations. Of particular interest, we show that using complex instead of real tensors can lead to an arbitrarily large reduction in the number of parameters of the network. Our findings can also be of practical interest, as we show that they can be straightforwardly applied to solve real-world problems for unsupervised learning.

Zusammenfassung

Die Theoretische Physik hat unter den quantitativen Wissenschaften eine besondere Stellung. So hat sie in den letzten Jahrzehnten hat die theoretische Physik neue Lösungsansätze in der Chemie, Biologie, Neurologie, Informatik und den Sozial- und Wirtschaftswissenschaften geliefert. Das Ziel der vorliegenden Arbeit besteht in der Entwicklung von Methoden, die zunächst auf komplexe Quantensystem angewendet und dann auf Probleme der künstlichen Intelligenz verallgemeinert werden.

Diese Arbeit besteht aus vier Teilen: In Teil II entwickeln und testen wir neue Methoden für die Zeitentwicklung von Quantensystemen, einem Problem von großem praktischen und theoretischem Interesse. In Teil II diskutieren wir Szenarien, in denen die Zeitentwicklung nicht zur Thermalisierung des Systems führt und in denen Information räumlich lokalisiert bleibt. Insbesondere beleuchten wir zwei Aspekte, die zur Lokalisierung führen können: Unordnung und kinetische Zwangsbedingungen. In Kapitel 1 benutzen wir eine Tensornetzwerke zur Konstruktion langsam evolvierender Operatoren für Spinketten und studieren den Phasenübergang von einer ergodischen in eine lokalisierte Phase. Wir zeigen, dass das statistische Verhalten dieser Operatoren Rückschlusse auf die Phase geben kann und dieses durch die Extremwerttheorie beschrieben werden kann. Inspiriert von klassischen Systemen mit Zwangsbedingungen, widmen wir in Kapitel 2 unsere Aufmerksamkeit einer Klasse von Quantenglasen. Wir analysieren das "Quantum East Model" durch exakte Diagonalisierung und mittels Tensornetzwerkmethoden. Wir zeigen die Existenz eines Phase, in der Eigenzustände bei beliebigen Energiedichten eine effiziente Tensornetzwerkdarstellung besitzen. Wir schätzen die Dynamik einer Familie von Eigenzuständen ab. Wir zeigen, dass unsere Resultate prinzipiell auch in höheren Dimensionen plausibel sind. In Teil III beschäftigen wir uns mit dem gegensätzlichen Szenario. Wir benutzen Methoden aus den Bereichen offene Quantensysteme und Gaussche Zustände, um thermodynamische Protokolle zu konstruieren, die die Thermalisierung beschleunigen, über das in der klassischen Physik gegebene Limit. Diese Protokolle interpolieren zwischen der Curzon-Ahlborn- und Carnotmaschine. In Teil IV schliessen wir mit der Begutachtung von Tensornetzwerkmethoden in der künstlichen Intelligenz ab. In Kapitel 4 adaptieren wir Neuronale Netzwerkzustände als variationelle Ansatzzustände für wechselwirkende Hamiltonians. Diese "Restricted Boltzmann Machines" bieten hervorragende Energien im Vergleich zu etablierten Methoden. Weiterhin zeigen wir die Äquivalenz dieser Zustände und der sogennanten "String Bond States", in der wir einen intuitiven Zugriff auf die physikalischen Eigenschaften des Systems bekommen. In Kapitel 5 und 6 verwenden wir Tensornetzwerke um Probleme in der künstlichen Intelligenz zu lösen. Wir verallgemeinern die oben genannte Äquivalenz auf "Probabilistic Graphical Models". Durch die Kombination von Tensornetzwerken mit modernen Neuronalen Netzwerken klassifizieren wir Bilder und Geräusche mit hoher Effizienz. In Kapitel 6 approximieren wir Wahrscheinlichkeitsverteilungen durch Tensornetzwerkzustände und geben präzise Fehlerabschätzungen. Wir zeigen,

dass “Hidden Markov Models” äquivalent zu Matrixproduktzuständen sind und wir zeigen, dass Wahrscheinlichkeitsverteilungen existieren, die in einer Faktorisierung einfach und in einer anderen unmöglich darzustellen sind. Insbesondere zeigen wir, dass komplexe Tensoren zu einer beliebig großen Reduktion der Parameter führen können. Wir wenden diese Erkenntnisse auf “unsupervised learning” Probleme an.

Acknowledgements

This thesis is the final result of one of the happiest period of my life where I grew up intellectually, personally and professionally. I had the opportunity to meeting world leading scientists and inspiring people from whom I learned numerous invaluable things.

Among all these people, a special place is reserved to Mari Carmen Bañuls. Your careful guidance and kind support have been two robust pillars on which I could rely throughout my entire PhD. All I know about tensor networks, quantum dynamics, simulations, and many other things, I learned from you.

Another important place is for Ignacio Cirac. The paternal way you take care of our group, and the way you strive to spend more time and to discuss science with us, made me understand what the priorities really are. I learned a lot from you - as a scientist, as as a mentor, as a leader.

Each and everyone of my present and former colleagues would deserve a personal acknowledgment. You are certainly the fuel of this vibrant and stimulating environment. The constant interaction and the diffusion of ideas is what makes this place different. Here, we understood that collaborating - rather than competing - is always more beneficial and rewarding.

I especially thank Ivan Glasser. Our office was a unique place, where we could discuss, learn and explore whatever would cross our minds. Our curiosities resonate so well to become extremely productive. Many achievements in this work would have never been possible without you.

The people who most contributed to make these years so unforgettable are - without any doubt - my booiz: the Leo Gang. A few words will not ever be enough to describe the magic of our flat. It will remain impressed in our memories, in our hearts, in our brains. *Thanks Henrik! Thanks Ruben! Thanks Andras! Thanks Cosimo!*

Together with them, I want to thank Elena, who was there whenever I would feel down. We overcame many hard challenges, often together. Thank you for you never complained! And thank you Benzo for teaching me the true sense of humor. We had some great times. And thanks for you never stopped - and hopefully will never stop - to make fun of Elena. To all my friends back home, I still find unbelievable that after so many years and so far apart, we still feel so close. By seeing you growing and developing in your careers and in your lives, I can feel the great incentive to always improve further.

And most importantly, my family. Everyone of you - in one way or the other - constructively supported me on every choice. When you encourage me and when you are proud of me, it makes me strive to do even better.

Contents

Part I

Introduction

"I see the structure of the world as a hierarchy with the levels separated by stages of emergence, which means that each is intellectually independent from its substrate. Reduction has real value in terms of unifying the sciences intellectually and strengthening their underpinnings, but not as a program for comprehending the world completely."

P. W. Anderson

Statistical Mechanics is one of the pillars of modern theoretical physics [1], and one of its most fascinating and active branches. The theory was originally developed in an effort to connect the macroscopic behavior of an ensemble of many particles ($\sim 10^{23}$) with the microscopic dynamics of each particle. In the 19-th century, such goal might have sounded paradoxical to many physicists, and the reason is simple. At microscopic level, systems of many particles obey the Newton's law of classical mechanics. Once the position and the velocity of each particle is known, the state of the system at any time is uniquely determined by Hamilton's equations of motion: the dynamics is deterministic and reversible. On the other hand, the macroscopic properties of an ensemble of many particles obey the laws of thermodynamics which are intrinsically irreversible and, perhaps even more surprisingly, their characterization requires only a small number of variables (such as temperature, pressure, volume, ...). Thus, a reductionist explanation of the laws of thermodynamics in terms of single-particle dynamics seems contradictory.

The revolutionary intuition of the pioneers of Statistical Mechanics was the introduction of a probabilistic interpretation of the underlying mechanisms [1]. What they first realized was that the time scales of macroscopic systems, or macroscopic measurements, are several orders of magnitudes longer than a typical microscopic dynamics. In other words, any macroscopic measurement is the result of a long time average of a microscopic dynamical system of many particles. Boltzmann's *ergodic hypothesis*, which casts these concepts into the framework of probability theory, is deeply rooted in this intuition. Specifically, for a macroscopic system at a given energy all the microscopic states with the same energy are equiprobable, or equivalently, any microscopic configuration of the particles can be reached in a finite time from any other configuration with the same energy. These considerations led Boltzmann to define the entropy as a function of microscopic properties of the system, via the celebrated formula

$$S = k_B \log(\Gamma), \tag{1}$$

where k_B is the universal Boltzmann's constant and Γ is the volume of the configuration space at a fixed small window of energy, volume and number of particles. Eq. (1) is of fundamental historical and theoretical importance. By introducing the new *method* of statistical analysis, it was suddenly possible to connect macroscopic observations to the microscopic dynamical theory. Since Boltzmann's original conceptual leap, the theory of Statistical Mechanics has experienced enormous advancements, with numerous ramifications even outside physics, the driving force being always the development of microscopic models capable of describing macroscopic emergent phenomena. Statistical Mechanics has provided accurate

description of phase transitions, collective and critical phenomena, it led to the elegant concept of universality, renormalization group and to the fervent theories of dynamical, chaotic and disordered systems.

Historically, quantum mechanics followed a similar path. Already by the first half of the 20-th century, all fundamental principles of the quantum theory had already been developed. As P. A. M. Dirac concluded at the Proceedings of the Royal Society in 1929 [2]:

> *"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."*

That is to say that most of the laws governing the atomic world were already known by 1929, and that most of condensed matter physics and quantum chemistry would have followed simply by their adoption. Nowadays however, almost one century later, we still witness vigorous research both in condensed matter and chemistry. And the reason was beautifully worded by Anderson [3]: "*More is different*". With three simple words, he rejected the presumptuous idea that the reductionist hypothesis implies a constructionist one: "*The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe*". He rather developed the view of emergence at every level: "*at each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other*". Being one of the fathers of modern condensed matter, Anderson soon recognized that his own research field was an excellent example of a new "level of complexity" which enjoys a broad range of striking emergent phenomena, unpredictable by the mere quantum theory. Such phenomena, which always arise from the interplay of many quantum degrees of freedom, include quantum phase transitions, superconductivity, topological phases, quantum localization, just to mention some.

"*Is the theory of everything the theory of anything?* " Anderson titled a chapter of his book [4]. And although the idea that each level of intellectual endeavor necessitates its peculiar theoretical treatment is certainly unquestionable, it seems there is always a golden thread that connects each of the hierarchy levels. It is very often the case that, in quantitative sciences, methods and tools get recycled and reused at different levels of complexity. In physics, *methods* that were initially developed with the scope of solving abstract problems, later found applications elsewhere and often very far from their original scope. Statistical mechanics is a good example. Rather than a theory which describes a particular physical setting, it provides a toolbox of powerful methods that have been massively expanded in the last decades. The Boltzmann distribution — originally invented for the description of thermal probability distributions of an ensemble of particles — is today used in the most diverse fields of science from quantum physics, chemistry, biology to socioeconomic sciences, computer science and machine learning [5]. Monte Carlo methods — first developed at Los Alamos National Laboratory for the estimation of expectation values in nuclear reactions — are today at the core of modern

quantitative finance, artificial intelligence, engineering, environmental sciences, and many others [6].

Theoretical physics — by constantly expanding its frontiers — offers an excellent playground for the development and testing of new quantitative methods. This is the main drive behind this thesis: this work attempts to contribute to the growing body of numerical and analytical methods for interacting quantum systems, as well as to the application of quantum-inspired methods to problems outside the physical realm.

Before diving into the details, in the following chapters, we will briefly outline the main concepts and results contained in this thesis.

> *"Ludwig Boltzmann, who spent much of his life studying statistical mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study statistical mechanics."*
>
> *David L. Goodstein — States of Matter*

Tensor networks in one dimension

Complementary discussions about tensor networks, their graphical notation, and their applications to machine learning can be found in chapter 5 and chapter 6.

Tensor networks [7–9] provide a method for the efficient description of certain quantum many-body states. From a purely mathematical point of view, this is a class of functions which map the contraction of a collection of tensors onto a large vector of complex numbers. In the quantum many-body setting each tensor is normally associated to some local degrees of freedom whereas the complex number represents the amplitude of a given configuration of all degrees of freedom of the system. More precisely, in quantum mechanics, the state of a closed quantum system can be described by a (normalized) vector belonging to a Hilbert space. In the case of an ensemble of N spin-1/2 particles, the Hilbert space can be spanned by the basis of the 2^N orthonormal vectors $|s_1 \cdots s_N\rangle$, where s_i can either take value zero or one and $s_1 \cdots s_N$ are all possible 2^N binary strings. A generic state $|\psi\rangle$ belonging to the Hilbert space can be written as a linear combination

$$|\psi\rangle = \sum_{s_1 \cdots s_N} \psi(s_1, \cdots, s_N) \, |s_1 \cdots s_N\rangle \, . \tag{2}$$

Clearly, accounting for all exponentially many amplitudes $\psi(s_1, \cdots, s_N)$, one would need 2^N free parameters. In order to avoid such exponential growth, researchers have developed efficient approximations which (sometimes) can approximate $|\psi\rangle$ with a smaller number of parameters. One prominent example is unquestionably tensor networks. For the sake of simplicity, let us assume the particles to be arranged on a one dimensional lattice. One of the most popular tensor network approximations is the Matrix Product State [10], where each amplitude is approximated by a product of tensors

$$\psi(s_1, \cdots, s_N) \sim v_1^{s_1} A_2^{s_2} A_3^{s_3} \cdots A_{N-1}^{s_{N-1}} v_N^{s_N} . \tag{3}$$

Once the configuration $s_1 \cdots s_N$ is chosen, the corresponding amplitude is determined by the multiplication of the $D \times D$ matrices $A_i^{s_i}$ and the two boundary vectors $v_1^{s_1}$ and $v_N^{s_N}$. The accuracy of the approximation is controlled by their dimension D (bond dimension), and in the limit of large D any state in the Hilbert space can be reached. In some cases, the matrix product approximation is provably good or even exact. One celebrated example is the ground state of one-dimensional local non-critical Hamiltonians [11]. This is a consequence of the *Area Law of entanglement* [12]. Given a quantum state $|\phi\rangle$ and two partitions A and B of its degrees of freedom, one can measure the amount of correlations between A and B via the von Neumann entanglement entropy $S = -\operatorname{Tr}(\rho_A \log(\rho_A)) = -\operatorname{Tr}(\rho_B \log(\rho_B))$, where $\rho_{A/B} = \operatorname{Tr}_{A/B}(|\phi\rangle \langle\phi|)$ are reduced density matrices and $\operatorname{Tr}_{B/A}(\cdot)$ is the partial trace[1].

A state fulfills the area law of entanglement if, as we increase the degrees of freedom in A and B, its entanglement entropy grows at most linearly with the size of the boundary between the partitions A and B.

[1] $\rho_{A/B} = \operatorname{Tr}_{A/B}(|\phi\rangle \langle\phi|) = \sum_i \langle i_{A/B}|\phi\rangle \langle\phi|i_{A/B}\rangle$ and $|i_{A/B}\rangle$ is a complete basis in A/B

In the case of one-dimensional systems, all the states that fulfill an area law can be efficiently approximated by matrix product states.

Equivalently, it is possible to represent (or approximate) operators [13] acting on the Hilbert space via the product of tensors

$$O(s_1, s_1', \cdots, s_N, s_N') \sim u_1^{s_1,s_1'} M_2^{s_2,s_2'} \cdots M_{N-1}^{s_{N-1},s_{N-1}'} u_N^{s_N,s_N'}, \tag{4}$$

which represent the operators,

$$\mathcal{O} = \sum_{s_1,s_1'\cdots s_N,s_N'} O(s_1, s_1', \cdots, s_N, s_N') \left|s_1 \cdots s_N\right\rangle \left\langle s_1' \cdots s_N'\right|. \tag{5}$$

All locally-interacting one-dimensional Hamiltonians can be written exactly as matrix product operators.

There are several calculations that can be carried out efficiently with tensor-network states and operators, corresponding to local (or MPO) operators – e.g. computing expectation values, amplitudes and correlations. For later discussions, it is worth mentioning the Density Matrix Renormalization Group (DMRG) [8, 14]: an optimization algorithm which tries to find an efficient approximation of the eigenvector with the smallest (or largest) eigenvalue of any operator written in terms of matrix product operators – e.g. the ground state of locally interacting one-dimensional Hamiltonians. Given a Hamiltonian H decomposed as a matrix product operator in Eq. (4), the goal of the algorithm is to find the tensors $v_1^{s_1} A_2^{s_2} A_3^{s_3} \cdots A_{N-1}^{s_{N-1}} v_N^{s_N}$ in Eq. (3) which minimize the expectation value $\langle H \rangle = \langle\psi| H |\psi\rangle / \langle\psi|\psi\rangle$. The normalization condition $\langle\psi|\psi\rangle = 1$ can be enforced by exploiting the gauge degree of freedom of Matrix Product States [8]. In the general case, minimizing $\langle H \rangle$ is a non-convex problem. Thanks to the matrix product form of states and operators, it can be relaxed onto a series of local convex problems where each tensor in Eq. (3) is individually (and repeatedly) optimized. Although this procedure does not guarantee convergence to the global minimum, it heuristically provides excellent results.

Matrix Product States and Operators can also be used to simulate time evolution [8, 15–18]. However, they often fail to deliver satisfactory results as the typical growth of entanglement which characterizes quantum dynamics, critically affects the tensor-network approximation. For models as simple as local one-dimensional spin systems neither analytical tools nor numerical algorithms have reached satisfactory results [19–23]. The main obstacles include the growth of quantum correlations, the spreading of information, and the highly entangled nature of the excited states which translates into the Eigenstate Thermalization Hypothesis (see next chapter) [24, 25]. More precisely, given a local time-independent Hamiltonian H, one can time evolve any state $|\psi\rangle$ by applying the unitary operator $U = e^{-itH}$ to the state $|\psi\rangle$. In the generic case, the operator U is highly non-local and one need to resort to approximations Aware of these strengths and limitations, in Chapter 1, in order to avoid explicit time evolution, we will address dynamical problems by mapping some dynamical feature onto a particular variational minimization.

Quantum localization

One of the few cases where tensor networks can deliver dynamical insights about quantum systems is when information is localized. Localization of quantum information is a central problem in modern quantum dynamics and quantum computing. One of the first proposals for quantum localization was put forward in 1957 by P. W. Anderson in his seminal work "Absence of Diffusion in Certain Random Lattices" [26] where he showed that non-interacting quantum particles moving in a random potential fail to explore the entire phase space and they remain exponentially localized in the vicinity of their initial position. In the recent years, after the ground-breaking paper "Metal-insulator transition in a weakly interacting many-electron system with localized single-particle states" by D.M. Basko, I.L. Aleiner, B.L. Altshuler there has been overwhelming attention on the so-called phenomenon of Many Body Localization (MBL) [27, 28], the interacting counterpart of Anderson's original idea. Before the discovery of MBL, it was believed that any generic isolated non-integrable system would thermalize under time evolution. If a system at a given temperature thermalizes for any possible initial state, then this implies that *all the many-body eigenstates of the Hamiltonian are thermal*. This is one possible fomulation of the Eigenstate Thermalization Hypothesis (ETH) [24, 25, 29]. More precisely, consider an eigenstate $H\left|e_n\right\rangle = e_n\left|e_n\right\rangle$, where e_n is the thermal equilibrium energy at a given temperature T_n: $e_n = \langle H\rangle_{T_n}$. The ETH asserts that in the thermodynamic limit for any compact partition of the state $\rho_A^{(n)} = \mathrm{Tr}_B\left[\left|e_n\right\rangle\left\langle e_n\right|\right]$ holds the equality $\rho_A^{(n)} = \rho_A^{eq}$, where $\rho_A^{eq} = \mathrm{Tr}_B\left[\exp\left(-\beta_n H\right)\right]$ and the temperature β_n is uniquely determined by the energy e_n [29].

The ETH is a hypothesis, and it is natural to ask when and if it fails. One the most celebrated examples was Many Body Localization (MBL). Recently, the investigation of MBL systems has attracted a lot of attention from the scientific community. Abundant numerical, experimental and analytical evidences have been developed in support of the claim that MBL does not fulfill ETH. One of the central problems for the understanding of the MBL phase is the question of how localization emerges from disorder which translates in the study of the MBL phase transition. For the single-particle Anderson localization, an infinitesimal small amount of disorder is enough in order to localize the non-interacting particles in one dimension. In the many-body interacting case, a phase transition seems to occur at finite value of the disorder strength [30, 31]. A great effort was invested in order to understand this atypical quantum phase transition which supposedly affects all the eigenstates of the system and not only the ground state.

In Part II, we study dynamical properties of quantum many-body systems. We will mostly focus on systems whose dynamics is slow, sometimes slow enough to localize quantum information or to impede thermalization. In particular, we show how tensor-network methods can be employed along with exact diagonalization and analytical techniques in order to explore localization properties of interacting quantum many-body systems — systems where quantum information does not propagate in space. Statements about localization often necessitate the study of the dynamics and thermalization of the models under analysis. As we discussed above, such tasks are extremely challenging and still attract the attention of physicists.

In Part II, we will take a different route, aware of the difficulties of simulating time evolution, we will explore the static properties of the Hamiltonian – e.g. conserved quantities and eigenstates – in order to understand its dynamical behavior.

We will study two different mechanisms that give rise to localization. In chapter 1, we will employ the slow operator method to study a one-dimensional disordered spin chain in the vicinity of the phase transition between the thermal and the localized phase.

An interesting open question is whether localization can arise in absence of disorder. In chapter 2, we will study a class of quantum systems inspired from classical structural glasses – systems where thermalization is impeded by kinetic constrains. In particular, we will show how these kinetic constrains may induce localization and slow dynamics in the quantum scenario.

Isothermal processes

Isothermal transformations play a fundamental role in thermodynamics, being the building block of optimal processes such as the Carnot engine [32]. In principle, however, they are infinitesimally slow; this means in practice that the total time of the process needs to be much larger than the timescale of thermalization, τ_{eq}, over which the system of interest equilibrates with its thermal environment. These processes can then be sped up by increasing the system-environment coupling, which naturally reduces τ_{eq}. However, modifying the interaction also induces additional dissipation (see Refs. [33–36] for recent discussions), which prohibits the non-physical possibility of performing an isothermal process arbitrarily quickly. Given this non-trivial trade-off, the goal of Part III is to develop quantum-thermodynamic protocols that smoothly modify the system-bath interaction in order to speed up an isothermal process while keeping the overall dissipation constant. This enables us to increase the power of heat engines without compromising their efficiency, a well-known challenge in thermodynamics [37–39].

The idea of speeding up different thermodynamic processes by external control has received a lot of attention in the last years. Particularly relevant are shortcuts to adiabaticity (STA), which speed up unitary (and hence closed-system) evolutions [40], making them suited to improve the adiabatic part of thermodynamic cycles [41–43]. For open quantum systems, speed-ups of the evolution to a particular target state [44–47], such as an equilibration or thermalization process [48–52], have also been developed. For classical systems, such equilibration speed-ups (the so-called Engineered Swift Equilibration [53–55]) have been experimentally tested [53, 54]. Furthermore, these ideas have been extended to full isothermal classical processes, so that the state remains in the desired Gibbs distribution along the whole process [56–58]. These ideas have also been recently applied to the optimisation of a finite-time Carnot cycle [59], Otto engines [60, 61], and refrigerators [62]. In general, such speed-ups are possible by adding a time-dependent term to the Hamiltonian which, in the presence of a thermal bath, leads to a new source of dissipation. Indeed, speed-ups of equilibration and thermalization generally come with an extra work cost [50, 53, 61] (see also the discussion in [63]

for thermodynamic computing).

As we discussed above, localization of quantum dynamical systems often relates to slow down or absence of thermalization. In Part III, we will study an alternative scenario where instead of impeding thermalization we aim at enhancing it. We will consider the situation of a small quantum system interacting with a large bath. Our goal is to engineer thermodynamic transformations between the system and the bath that are minimally dissipative. This kind of processes are normally slow as the system needs to remain close to thermal equilibrium along the protocol. Here, we explore the possibility of smoothly modifying the system-bath interaction in order to *speed-up* such transformations. Our goal is to engineer thermodynamic protocols where the total dissipated work decays faster than in standard isothermal processes.

Quantum machine learning

Recognizing complex patterns is a central problem that pervades all fields of science. The increased computational power of modern computers has allowed the application of advanced methods to the extraction of such patterns from humongous amounts of data and we are witnessing an ever increasing effort to find novel applications in numerous disciplines. This led to a line of research now called Quantum Machine Learning[64], which is divided in two main different branches. The first tries to develop quantum algorithms capable of learning, i.e. to exploit speed-ups from quantum computers to make machines learn faster and better. The second, that we will consider here, tries to use classical machine learning algorithms to extract insightful information about quantum systems.

The versatility of machine learning has allowed scientists to employ it in a number of problems which span from quantum control[65–67] and error correcting codes[68] to tomography [69]. In the last few years we are experiencing interesting developments also for some central problems in condensed matter, such as quantum phase classification/recognition[70–73], improvement of dynamical mean field theory[74], enhancement of Quantum Monte Carlo methods [75, 76] or approximations of thermodynamic observables in statistical systems[77].

On the other hand, tensor networks have reached state-of-the-art results for a number of problems, but they still suffer severe limitations in many particular settings — e.g. quantum dynamics, high-dimensional lattices. It is hence desirable to develop new methods in order to challenge them and to explore their limitations. The scope of Part IV is twofold. On the one hand we will explore the potential of machine-learning techniques for the study of strongly correlated systems and, in particular, their relationship with tensor networks. On the other hand, we speculate on the applicability of tensor networks for solving machine learning tasks. In Chapter 4 we employ machine-learning techniques along with tensor networks for approximating ground states of interacting quantum many-body Hamiltonians. Neural-network quantum states is a new class of variational wave functions inspired from machine learning. Their optimization comes with the cost of Monte Carlo sampling and nevertheless they have already performed extremely

well on traditionally difficult tasks for tensor networks. In Chapter 4 we compare the power of different variational wave functions based on either machine learning or tensor networks. In particular, we develop a number of exact mathematical mappings which connect some popular neural network quantum states to different classes of tensor networks.

The connection between tensor networks and neural-network quantum states highlights their intimate relation, and it paves the way for the application of tensor networks for machine-learning tasks. In Chapter 5 we build upon this idea. Namely, we formalize the connection between tensor networks and probabilistic graphical models — a popular class of parameterized probability distributions in machine learning. As for the case of probabilistic graphical models, we show that tensor networks can be relevant tools for supervised tasks such as image and sound recognition and, at the same time, they can provide with new mathematical insights.

Finally, motivated by the outstanding analytical results developed in the framework of tensor networks for one dimensional quantum systems and by the solid connection with probabilistic graphical models, in Chapter 6, we explore the expressive power of tensor networks for unsupervised learning tasks. Our goal is to quantify the resources required by a given model for describing a probability distribution over discrete variables. We consider particular probabilistic graphical models from machine learning, along with models used in the simulation of one-dimensional quantum systems in order to compare their relative expressive power. We perform rigorous analysis as well as numerical simulations to show that tensor-network factorizations and, in particular the ones with complex parameters, should be preferred over more traditional methods.

Most of the models considered in this part enjoy intimate connections with quantum circuits. Our contributions might provide a fertile ground for testing and extending classical machine-learning algorithms based on tensor networks to the framework of quantum computing.

Part II

Localization in Quantum Many-Body Systems

Chapter 1

Almost conserved operators in nearly many-body localized systems

Most of the content of this chapter is contained in:

- *"Almost conserved operators in nearly many-body localized systems"*
  Nicola Pancotti, Michael Knap, David A. Huse, J. Ignacio Cirac, and Mari Carmen Bañuls
  PRB 97, 094206, (2018)

In this chapter, we study localization of many-body quantum systems induced by disorder. We employ tensor-network techniques along with exact diagonalization and statistical tool in order to gain insights about the physics in the vicinity of the phase transition point. In particular, we study the dynamics of a disordered and interacting Heisenberg spin chain by constructing almost conserved local operators with finite support, which minimize the commutator with the system Hamiltonian.[78] Deep in the MBL phase the slow operators resemble the local integrals of motion (LIOMs).[79, 80] However, our procedure can be directly extended to the thermal phase as well. The slow operators can potentially be used to study dynamical properties, because the value of the commutator gives a lower bound on the thermalization times of the corresponding operator.[78] We compute the slowest operators for different support sizes and disorder strengths, using both exact diagonalization and tensor networks.[81, 82] This analysis, rather than focusing on the phase transition, aims at characterizing the sub-diffusive region in its vicinity. Specifically, we collect statistics from numerous disorder realizations and show that not only the mean value of the smallest commutator, but also the probability distributions are sensitive to the underlying phase. The *tails* of the distributions are affected by the appearance of rare Griffiths regions,[83, 84] which act as bottlenecks for transport and are argued to provide a simple description for the slow sub-diffusive dynamics in the thermal phase near the MBL transition. [83, 85, 86] We use extreme value theory (EVT) to analyze the tails of the distribution and find that it is well described by the generalized extreme value distribution. [87] Moreover, we discuss how the presence of rare regions may affect the asymptotic behavior of such distributions. Our results demonstrate that finding slow operators provides access to dynamical quantities without resorting to the time evolution of a particular state.

1.1 Model

We consider the non-integrable spin-1/2 Heisenberg Hamiltonian with random magnetic field,

$$H = \sum_i J \left(S_i^x S_{i+1}^x + S_i^y S_{i+1}^y + S_i^z S_{i+1}^z \right) + h_i S_i^z, \tag{1.1}$$

where S_i^α are spin-1/2 operators, and the values of the transverse field h_i are randomly chosen from the uniform distribution in the interval $[-h, h]$, where h is the disorder strength. In the following, we set $J = 1$. This model exhibits a phase transition from a thermal to a MBL phase at $h_c \sim 4$.[88, 89]

In the MBL phase the system can be described by LIOMs, which lead to an emergent integrability. The corresponding effective Hamiltonian can be expressed as [79, 80]

$$H = \sum_i \tau_i^z + \sum_{ij} J_{ij} \tau_i^z \tau_j^z + \sum_n \sum_{i,j,\{k\}} K_{i\{k\}j}^{(n)} \tau_i^z \tau_{k_1}^z ... \tau_{k_n}^z \tau_j^z. \tag{1.2}$$

The LIOMs, τ_i^z, commute with each other and with the Hamiltonian. They are exponentially localized when written in terms of the physical spins $\{\vec{S}_i\}$. Moreover, the coefficients J_{ij} and $K_{i\{k\}j}^{(n)}$ decay exponentially with their spatial range.

1.1.1 Methods

Slow operator technique

Almost conserved quasi-local operators can identify long thermalization time scales. A way to compute such time scales is to search for slow operators that minimize the commutator with the Hamilonian[78] (see also Refs. [90–92]). In particular, we may restrict the search to operators O_M with finite support on M consecutive sites. Defining $\mathcal{L}(O_M)$ to be the squared norm of the commutator between the Hamiltonian and the (normalized) operator, the problem reduces to solving the variational minimization

$$\lambda_M := \min_{O_M;\, \mathrm{tr} O_M = 0} \mathcal{L}(O_M) = \min_{O_M;\, \mathrm{tr} O_M = 0} \frac{\|[H, O_M]\|_F^2}{\|O_M\|_F^2}, \tag{1.3}$$

where $\|A\|_F^2 = \mathrm{tr}(A^\dagger A)$ is the Frobenius norm, and O_M is restricted to be traceless.[1] The operator that minimizes (1.3) is the one evolving the *slowest* in Heisenberg picture at time $t = 0$. Moreover, for a slightly perturbed infinite temperature state of the form $\rho(O_M) = \mathbb{1}/Z + \epsilon O_M$ the thermalization time is lower bounded as $t_{\mathrm{th}} \geq 1/\sqrt{\lambda_M}$.[78]

If the system has a conserved extensive quantity, such as the total energy, it naturally gives rise to slow operators. For a translationally invariant system, the Hamiltonian commutes with the sum of all translations of the corresponding local

[1] The traceless condition ensures the solution will have no overlap with the trivially commuting identity.

operator, but not with translations restricted to a finite window M. Nevertheless, the norm of the latter commutator seems to decrease polynomially with M.[78]

In our particular case of Hamiltonian (1.1), the total polarization in the z direction, $\sum_i \sigma_i^z$, is conserved for any value of the disorder strength. The restriction of the sum to M sites, $Z_M = \sum_{i=0}^{M} \sin\left(\frac{\pi k}{M}\right) \sigma_i^z$, is thus a natural slow operator for any given support M. It is easy to check that as the support increases, the corresponding commutator decays as $\mathcal{L}(Z_M) \sim 1/M^2$.[78] This sets a reference for comparison with the commutators found by numerically solving the variational problem, Eq. (1.3). [2]

For small support size M the optimization can be solved via exact diagonalization. This has been used to show that for model (1.1) there exists an extensive number of exponentially localized LIOM in the MBL phase which are not present in the thermal phase [93]. To reach larger supports the search can be restricted to operators O_M of the form of a matrix product operator (MPO). Then standard tensor network techniques can be applied to solve this optimization, and the bond dimension can then be systematically increased until convergence is achieved.

In this chapter, we use exact and approximate MPO solutions in order to collect information on the minimal commutators and the corresponding operators. For different values of the support M and the disorder strength h, we compute λ_M for several configurations of the random magnetic field (ranging from $2 \cdot 10^3$ for $M = 12$ to 10^5 for $M = 4$). In order to simplify the statistical analysis, we choose independent configurations over $M + 2$ sites. This is equivalent to choosing non-overlapping windows over an infinite chain.

To solve the problem numerically, the minimization Eq. (1.3) can be interpreted as an eigenvalue problem. For fixed support size M and disorder strength h and for a particular disorder realization, there is an effective operator $\mathcal{H}_{\text{eff}}^{(M)}$ acting on the d^{2M}-dimensional space with support on the chosen window, such that the (vectorized) operator that minimizes Eq. (1.3) corresponds to its eigenvector with lowest eigenvalue. Since the optimization is restricted to operators supported on a certain window of size M, only terms in the Hamiltonian that overlap with that region will contribute to the commutator. For our nearest-neighbor Hamiltonian this means that $M + 2$ sites of the Hamiltonian contribute, which we denote by H_{M+2}. The matrix representation of the corresponding commutator, acting on vectorized operators, can be written as $\mathcal{C}_{M+2} = H_{M+2} \otimes \mathbb{1} - \mathbb{1} \otimes H_{M+2}^T$, and the effective operator can be constructed as

$$\begin{aligned} \langle O_M | \, \mathcal{H}_{\text{eff}}^{(M)} \, | O_M \rangle &:= \operatorname{tr}\left([H_{M+2}, \tilde{O}_M]^\dagger [H_{M+2}, \tilde{O}_M] \right) \\ &= \langle \tilde{O}_M | \, \mathcal{C}_{M+2} \mathcal{C}_{M+2}^\dagger \, | \tilde{O}_M \rangle \,, \end{aligned} \tag{1.4}$$

where $\tilde{O}_M = \mathbb{1} \otimes O_M \otimes \mathbb{1}$ and the trace is taken over the space of all $M + 2$ sites.

[2] In general we expect that a numerical minimization for any given support, will find operators slower than the polarization fluctuations, even in the clean system. [78]

1.1.2 Extreme value theory

The statistical analysis of the smallest commutator contains relevant information about the phases of the model and the critical region, both in the average scaling of λ_M with the support size M and in the probability distributions. In particular, the probability density function (PDF) $p(\lambda_M)$ and the corresponding cumulative density function (CDF), $F(\lambda_M) = \int_{-\infty}^{\lambda_M} p(x)dx$, will be sensitive to the presence of rare regions of atypically large disorder which emerge in the proximity of the critical point.

As we show in this chapter, these PDFs can be described and analyzed within the mathematical framework of Extreme Value Theory (EVT), a branch of statistics concerned with the description of rare events,[87] which is applied to floods, earthquakes or risk in finance and insurance, as well as to several branches of physics, including statistical mechanics and critical phenomena. [94–96]

In particular, EVT deals with the asymptotic behavior of the extreme values of a sample, i.e., with the tails of the distributions. Let us consider a random variable x governed by a CDF $F(x)$. Then EVT ensures (under some assumptions on the distribution) that the maxima (equivalently the minima) of samples of $F(x)$, properly normalized and centered, will be governed by a CDF[87]

$$G_\zeta(y) = \exp\left(-(1+\zeta y)^{-1/\zeta}\right), \quad (1+\zeta y)^{-1/\zeta} > 0 \tag{1.5}$$

for $y = ax + b$ with a, $b \in \mathbb{R}$ and $a > 0$. The Fisher-Tippett-Gnedenko theorem (Thm. 1.1.3 in [87]) states that this single-parameter family, called generalized extreme value distribution (GEV), includes all possible limiting distributions for the extreme values of a sample of i.i.d. random variables. The family contains three subclasses which exhibit quite different behavior, the Gumbel ($\zeta = 0$), Fréchet ($\zeta > 0$) and Weibull ($\zeta < 0$) distributions. Qualitatively, the deciding criterion for a PDF to belong to the basin of attraction of one family or another (i.e. for the extrema of the samples to be described by the corresponding family) is the form of its tails.[87]

In our particular problem, as we try to find the minimum λ_M for a certain configuration, we are effectively sampling from the left tail of $p(\Lambda_M)$, which is the distribution of eigenvalues Λ_M of $\mathcal{H}_{\text{eff}}^{(M)}$. Typically the eigenvalues of a matrix are not uncorrelated, and thus GEV is *a priori* not expected to describe the probability distribution of extreme eigenvalues[95, 97]. Nevertheless, our results indicate that the distribution (1.5) provides indeed a very good description for our data, with the particular form depending only on the asymptotic behavior of $p(\Lambda_M)$ for small Λ_M.

1.2 Structure of the slow operators

We will first study the structure of the slowest operators. In the strong disorder regime, we expect that the slow operators correspond to some LIOMs, or more precisely to a truncated version of them since our support size is fixed. It is worth noticing here that although the slow operator method does not directly target the

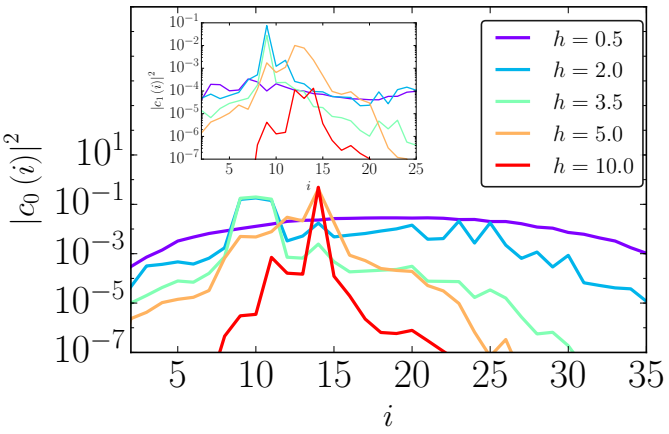


Figure 1.1: **Local structure of the slowest operator with support $M = 40$.** The weight of single-site $|c_0(i)|^2$ contribution is shown as a function of the position i for various disorder strengths for a single disorder realization. At weak disorder, the local contributions are spread over the whole support and has the shape of a sine as expected from diffusive dynamics. For strong disorder, the operator becomes increasingly localized and exhibits exponential tails. The optimization has been performed for MPO bond dimension $D = 100$. **Inset:** For larger support contributions such as $|c_1(i)|^2$, we observe similar tails with considerably smaller values.

(truncated) LIOMs, as was done in Refs. [98, 99], in the localized regime, truncated LIOMs and their combinations are good candidates to attain (exponentially) small commutators.

To address this question, we can analyze to which extent the slowest operators are local by examining their spatial structure in different regimes. This can be done by studying their decomposition as a sum of tensor products of single-site Pauli matrices, $O_M = 2^{-M/2} \sum_{\{\alpha_j\}=0}^{3} C_{\alpha_1 \dots \alpha_M} \sigma_1^{\alpha_1} \dots \sigma_M^{\alpha_M}$. In principle, for an operator O_M found by the minimization, either exactly or as a MPO approximation, we can efficiently evaluate any single coefficient $C_{\alpha_1 \dots \alpha_M} = 2^{-M/2} \mathrm{tr}(\sigma_1^{\alpha_1} \dots \sigma_M^{\alpha_M} O_M)$. Nevertheless, due to the exponential growth of the basis dimension with the support, already for moderate values of M it is unfeasible to inspect all the individual coefficients. Instead, a more physical quantity for exploring the localized nature of the operators is the combined weight of all terms with a fixed range which are supported on a certain subregion. We define a range-k operator as a product of Pauli matrices that act non trivially on sites i to $i + k$. Formally, this corresponds to the operator $\Theta_{i,k}^{a,b,\{\alpha_m\}} = \sigma_1^0 \dots \sigma_{i-1}^0 \sigma_i^a \sigma_{i+1}^{\alpha_1} \dots \sigma_{i+k-1}^{\alpha_{k-1}} \sigma_{i+k}^b \sigma_{i+k+1}^0 \dots \sigma_M^0$, where $a, b \in 1, 2, 3$ to ensure Pauli matrices at the edges (thus imposing that the support is strictly $i \to i + k$) and $\{\alpha_m\} \in 0, 1, 2, 3$.

The weight of these range-k terms can be written as

$$|c_k(i)|^2 := \frac{1}{2^M} \sum_{a,b=1}^{3} \sum_{\{\alpha_m\}=0}^{3} \left| \mathrm{Tr}\left(\Theta_{i,k}^{a,b,\{\alpha_m\}} O_M \right) \right|^2 . \tag{1.6}$$

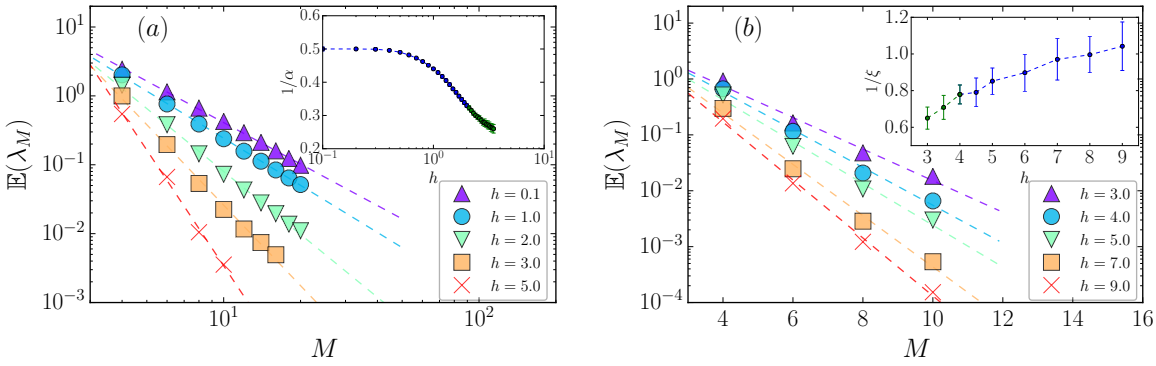


Figure 1.2: **Disorder average of the smallest commutator $\mathbb{E}(\lambda_M)$. (a)** Decay of $\mathbb{E}(\lambda_M)$ at weak disorder. We observe that the average of λ_M scales polynomially $M^{-\alpha}$ with an exponent that increases as the transition is approached. **Inset:** The inverse exponent $1/\alpha$ as a function of the disorder strength h is shown. **(b)** Decay of $\mathbb{E}(\lambda_M)$ at strong disorder. For strong disorder, the decay of λ_M is compatible with an exponential form, $e^{-M/\xi}$. **Inset:** The inverse length scale $1/\xi$ decays as the transition point is approached.

We thus solve the optimization (1.3) and compute the weights, $|c_k(i)|^2$, along the chain of terms with fixed range k. Even if the support is large, as O_M is written as an MPO, this quantity can be computed efficiently, i.e., with a cost that only scales polynomially in the range k and the support M.[3]

We find clear differences in the structure of the operator depending on the disorder strength. This is illustrated in Fig. 1.1 for the MPO, that minimizes Eq. (1.3) for a particular disorder realization $h_i = h \cdot r_i$, where r_i are fixed random numbers for each value of h. We have chosen a support of $M = 40$ sites and a relatively large bond dimension of $D = 100$, in order to ensure that truncation errors are negligible compared to the effects we observe. The figure shows the different spatial profile of the single-site contributions $|c_0(i)|^2 = 2^{-M} \sum_{a=1}^{3} |\mathrm{Tr}(\sigma_i^a O_M)|^2$ and two-site contributions $|c_1(i)|^2$ (Eq.(1.6) with $k = 1$) for different values of the disorder strength. For strong disorder, where the LIOMs are exponentially localized, at least one of the $|c_0(i)|^2$ is expected to be dominant. Indeed, we observe that for large disorder the operator is well localized around a single site, with weights that decay exponentially around this point. As the disorder strength decreases, the weights decay slower with distance, and the tails are no longer exponential. Finally for very small h the profile becomes flat and has the shape of a sine as expected in the diffusive regime. The single-site terms shown in the figure always accumulate most of the weight. We found that higher order terms Fig. 1.1 (inset), which have smaller contributions, exhibit a similar spatial decay. The observed profile at strong disorder is in good agreement with what one could expect from the decomposition of the LIOMs in the canonical operator basis, what

[3]The reason for this efficient computation is that the weight $|c_i(k)|^2$ can be expressed in terms of the vectorized operator O_M as the expectation value of a superoperator with tensor product structure and can then be evaluated with MPS primitives.

would in principle allow the extraction of a localization length scale of the LIOM.

1.3 Average commutator

The precise value of the minimum λ_M for fixed support M and disorder strength h will depend on the disorder realization. Hence, the average over realizations $\mathbb{E}(\lambda_M)$ will provide information about the underlying phase and also indirectly about transport.

In the following, we will optimize Eq. (1.3) over the family of MPOs with finite bond dimension $D = 10$. This allows us to reach large support sizes M and collect a significant amount of statistics. The average of the smallest commutator $\mathbb{E}(\lambda_M)$ is shown as a function of the support size M for different disorder strength h in Fig. 1.2. At very weak disorder $h \lesssim 0.4$ our data recovers the diffusive scaling $\mathbb{E}(\lambda_M) \sim M^{-2}$ of conserved quantities.

When increasing the disorder strength, the decay remains a powerlaw $\mathbb{E}(\lambda_M) \sim M^{-\alpha}$, however, with an increasing exponent α indicating that the transport becomes slower due to the presence of disorder, see inset of Fig. 1.2 (a) which shows the inverse exponent $1/\alpha$. This is consistent with the observation of sub-diffusive transport on the thermal side of the MBL transition.[83, 85, 86] The longer relaxation times in this regime can be interpreted as follows: rare insulating regions with larger than typical disorder act as bottlenecks for the conventional diffusive transport.[83, 100, 101] Upon approaching the MBL transition at $h \sim 4$, [88] we expect the exponent to diverge. Our data indeed shows that the power law decay becomes very steep. Due to the finite support sizes M and due to the slow convergence of the algorithm for large disorder, we, however, find that the exponent levels off at a finite value. For practical purposes, we adopt the strategy of limiting the maximum computational effort for each search. Therefore, the result of the algorithm provides an upper bound for the true minimum $\lambda_M(h)$, and the corresponding α is biased towards smaller values.

When crossing the MBL transition, an extensive number of exponentially localized conserved quantities, corresponding to the LIOMs τ_i^z in Eq. (1.2), emerges.[4] Therefore, provided the window M is sufficiently large to capture most of the support of any LIOM, the smallest commutator should decay exponentially as $\lambda_M \sim e^{-M/\xi}$. Such an exponential decay is supported by our numerical simulations for strong disorder, see Fig. 1.2 (b). The length scale ξ is related (in a possibly non-trivial way) to the localization length of the LIOMs. In the inset, we show that with increasing disorder h, the inverse length scale $1/\xi$ increases monotonically. The decays of α^{-1} and ξ^{-1} as the transition is approached are both compatible with previous estimates of the critical region, see e.g. Refs. [88, 98].

[4] There will actually be exponentially many conserved quantities, which can be constructed by any product or combination of LIOMs τ_i^z.

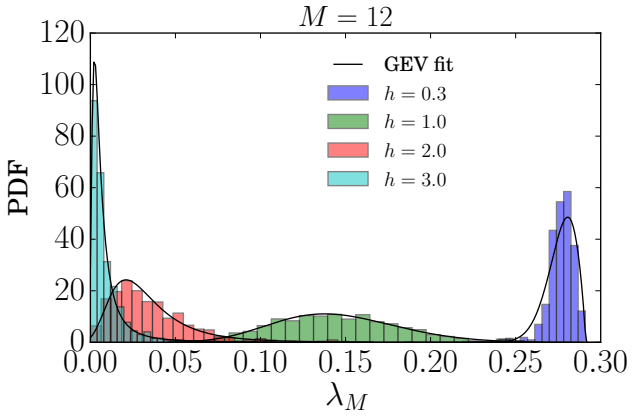


Figure 1.3: **Probability density function of λ_M.** Using the maximum likelihood method we fit the numerical data for λ_M with the GEV distribution (1.5). The shape of the distribution function is strongly influenced by the presence of disorder.

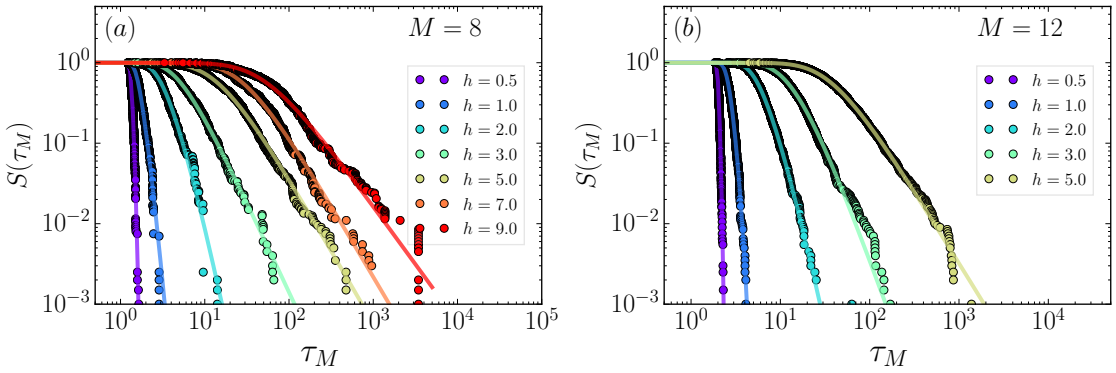


Figure 1.4: **Survival function of $\tau_M = \sqrt{1/\lambda_M}$.** We fit the survival function of τ_M to the GEV, Eq. (1.5), which decays super-exponentially for weak disorder and polynomially for large disorder. The form of these tails is sensitive to rare Griffiths effects.

1.4 Extreme value theory

1.4.1 Numerical results

We now study the probability distributions of the slow commutators λ_M, which can be sensitive to rare events. In particular, we show that EVT provides useful mathematical tools to characterize the tails of the probability density and cumulative distribution functions. We first illustrate in Fig. 1.3, that the GEV distribution, obtained from differentiating Eq. (1.5), provides a very good fit to the collected data over a range of disorder strengths. Moreover, we find that the shape of the distribution strongly depends on the disorder value and that the peak of the distribution shifts from a finite, relatively large value for weak disorder, toward zero as h increases. In between those limits, in a regime where one expects sub-diffusive dynamics, the distribution broadens considerably.

In the following we will analyze the tails of this distribution quantitatively. To this end, we introduce a new variable, with dimensions of time, as $T_M = (\Lambda_M)^{-1/2}$. The minimal eigenvalue λ_M thus corresponds to the maximum of T_M, that we call τ_M, and has the physical interpretation of bounding the thermalization time of the operator $\rho(O_M)$ from below. This transformation of variables moves the left tails of $p(\lambda_M)$ that arise due to rare events, to right tails of $p(\tau_M)$. The corresponding survival functions $S(\tau_M) = 1 - F(\tau_M)$ are shown in Fig. 1.4 on logarithmic scale. We observe that for small disorder, the survival function approaches zero super-exponentially fast describing weak tails, due to the small probability of large τ_M. As the disorder increases, the survival function decays with a power-law tail. The value of the disorder at which the tails change is in agreement with the crossover from diffusive to sub-diffusive dynamics.

To make this observation more quantitative, we study the variation of the shape parameter ζ from the fitted GEV, defined in Eq. (1.5), with the disorder strength. Our data shows that ζ shifts from values close to zero for weak disorder, to clearly positive values as the MBL phase is approached. The ζ parameter determines the type of distribution. In particular, the deviation of ζ from zero describes the crossover from a peaked distribution to one with polynomial tails. In terms of the GEV families, this corresponds to a change from a Gumbel ($\zeta = 0$) to a Fréchet ($\zeta > 0$) distribution. The observed qualitative change can be explained by an intermediate regime in which atypically slow operators appear for any given support M, leading to strong tails in the probability density function of τ_M. The value of the disorder strength at which the shape parameter ζ starts being clearly positive is again consistent with the expected sub-diffusive regime of the thermal phase.[86]

1.4.2 Interpretation in terms of Griffiths effects

Although the eigenvalues of the effective operators in our study are not expected to be uncorrelated, the results in the previous section show that GEV does indeed describe our findings accurately. Hence, in this section we use EVT arguments to show how the observed distributions can be qualitatively explained in terms of the existence of rare Griffiths regions.

We first consider the effect of rare localized regions in the thermal phase but close enough to the transition. In this situation, given a support size M, the *typical* values of λ_M will be polynomial in M. Nevertheless, if rare regions are present that (partly) support an exponentially localized operator, they can give rise to exponentially small values of λ_M. More concretely, within a fixed window M, such a rare configuration of size $\ell < M$ will occur with an exponentially small probability $p(\ell) \sim c^\ell$, for some $c < 1$. This patch can support a localized operator, with localization length ξ, with a correspondingly small commutator $\Lambda \sim e^{-\ell/\xi}$ (or equivalently[96] $T \sim e^{\ell/2\xi}$). Strictly speaking, this will be detected as the slowest operator only if this commutator is smaller than the (polynomial) typical commutator for the complementary region of size $M - \ell$. Since we are interested in the probability for very small commutators, and there is an exponential separation between the scaling of both terms, we may assume that small enough

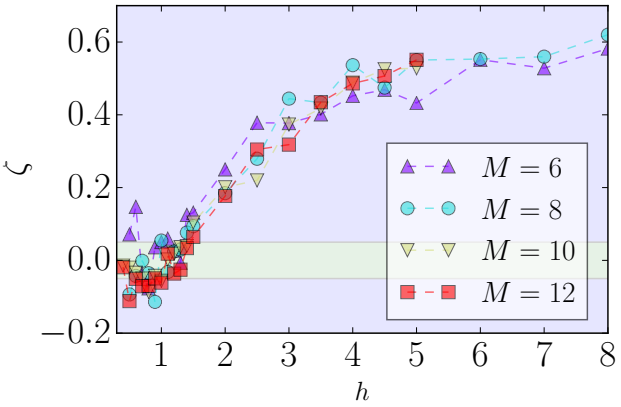


Figure 1.5: **Shape parameter ζ of the generalized extreme value distribution as a function of the disorder strength.** The values of the shape parameter ζ range from ~ 0 (corresponding to a Gumbel distribution) to $\zeta \sim 1/2$ (corresponding to Fréchet distribution) as the disorder h is increased.

commutators will always come from such rare patches. Evidently, this can only be true for commutators below a certain (M and h dependent) threshold. With this assumption, the probability of some very small value of Λ is determined by the probability of finding a rare patch such that $\Lambda \sim e^{-\ell/\xi}$, and thus we can identify $p(\ell)|d\ell| = p\,(T)\,|dT|$. From that we find that, for the range of T that correspond to rare regions of length $\ell < M$, the probability for the largest T values is polynomial,

$$p\,(T) \sim 2\xi T^{-2\xi|\ln(c)|-1}. \tag{1.7}$$

For PDFs with polynomial tails, EVT predicts that the extreme values will be governed by a Fréchet distribution with $\zeta > 0$ (see below).

For very weak disorder, instead, the same rare configurations of the field will not support an exponentially localized operator, so that, even if a large effective disorder on some region may give rise to commutators that are below the typical one, they still decay at most polynomially. Consequently, the argument above does not apply and we expect the probability of the smallest values to decay faster (even exponentially). For a PDF with such properties, EVT predicts a Gumbel distribution with $\zeta = 0$ (see below). This is in agreement with the shape parameter ζ shown in Fig. 1.5 where, for weak disorder, we obtain values of ζ very close to zero which correspond to a Gumbel distribution.

Beyond the localization transition we expect the typical regions to (partly) support localized operators, and thus give rise to exponentially small values of λ_M (correspondingly exponentially large values of τ_M), as we explicitly observed in the average values shown in Section 1.3. In our data for strong disorder, we have also found broad tails of the PDF $p(T)$ which are consistent with a power law. Thus, the function describing our data resembles the Fréchet distribution. This might be explained by the fact that matrix elements in the MBL phase have been found to exhibit broad tails. [102, 103]

As we approach the transition from the MBL side by lowering the disorder,

rare thermal inclusions can appear that potentially correspond to larger than typical commutators. Yet, our method only looks for the smallest commutator in each given window. Thus, if a support M encloses one such thermal subregion, a competition arises between the values of the commutator for the inclusion and that of the (typically) exponentially localized complement. Because we expect that a thermal inclusion gives rise to only polynomially decaying commutators, their value can only be the smallest one when the inclusion is sufficiently large in relation to the size M of the support. This causes our method to be less sensitive to rare thermal regions in the localized side of the transition.

In contrast to other numerical studies, where the presence of Griffiths effects was inferred from averaged observables,[83, 85, 86] with our method we may directly locate rare regions. In particular, we can obtain the disorder potential for the eigenvalues λ_M that contribute to the tail of the distribution and analyze the microscopic configuration of the random field $\{h_i\}$ in real space. Following this procedure, we could, however, not unambiguously determine an obvious correlation between strong fluctuations of the field in real space and small commutators. It remains an open question whether it is possible to predict the location of the rare regions from the disorder landscape using a more direct method than the optimization.

1.4.3 Extreme value theory

A given PDF is said to belong to the basin of attraction of one of the extreme values distributions, namely Gumbel, Fréchet or Weibull, when the extrema are distributed according to the corresponding function. The von Mises conditions[87] establish simple criteria to determine whether $p(x)$ belongs to one of them. In this section we show how the conditions apply to the particular cases of the distributions discussed in section 1.4.2.

Strong disorder implies a Fréchet distribution. Rare regions in the thermal phase near the MBL transition may support localized operators. As shown in Eq. (1.7), the corresponding probability distribution is expected to decay as $p(T) \sim 2\xi T^{-2\xi|\ln c|-1}$. A sufficient condition[87] for a PDF $p(T)$ to belong to the basin of attraction of the Fréchet distribution is that the corresponding CDF $F(T) = \int_{-\infty}^{T} p(T')dT'$ satisfies the condition

$$\lim_{T\to\infty} \frac{TF'(T)}{1-F(T)} = \frac{1}{\zeta}, \tag{1.8}$$

with $\zeta > 0$.

From Eq. (1.7) we obtain

$$\frac{p(T)}{1-F(T)} = \frac{p(T)}{\int_T^\infty p(t)dt} = \frac{2\xi|\ln c|}{T}, \tag{1.9}$$

so that

$$\frac{p(T)}{1-F(T)} = \frac{2\xi|\ln c|}{T}, \tag{1.10}$$

and, asymptotically,

$$\lim_{T\to\infty} \frac{Tp(T)}{1-F(T)} = 2\xi|\ln c\,| > 0, \tag{1.11}$$

which ensures the condition above with $\zeta^{-1} = 2\xi|\ln c\,|$, and thus implies a limiting Fréchet distribution.

Exponentially decaying tails imply a Gumbel distribution. In order to prove that a PDF belongs to the basin of attraction of the Gumbel distribution it is sufficient to check the following condition [87]

$$\lim_{T\uparrow T_{max}} \frac{d}{dT}\left(\frac{1-F(T)}{F'(T)}\right) = 0, \tag{1.12}$$

We assume the simplest exponential decay for the right tail of the distribution $p(T) \sim e^{-kT}$. From the corresponding cumulative function we get the survival probability,

$$1 - F(T) = \int_T^\infty p(t)dt \propto \frac{1}{k}e^{-kT}, \tag{1.13}$$

so that

$$\frac{1-F(T)}{p(T)} \sim \frac{1}{k}. \tag{1.14}$$

Thus, the derivative vanishes, which ensures Eq.(1.12).

1.5 Discussion

We have constructed slow operators with finite support by minimizing their commutator with the Hamiltonian of the system using both exact diagonalization and tensor network techniques. In particular, we have considered the Heisenberg spin chain with random magnetic field, which displays a dynamical transition from the thermal to the many-body localized phase. The scaling of the minimal commutator with support size provides information on the localization transition as well as on transport in the system without resorting to a specific initial state.

Furthermore, we have demonstrated that the tails of the probability distributions are sensitive to rare insulating regions in the thermal phase near the many-body localization transition. We have shown that the statistics of the smallest commutators can be analyzed within the mathematical framework of extreme value theory. [87] In particular, we have found that the distributions are well described by generalized extreme value functions whose shape depends on the disorder strength. By extreme value theory arguments, the observed behavior in the tails can be connected to the appearance of rare, strongly disordered regions, that give rise to atypically small minimal commutators. In particular, the disorder strength at which the distribution functions obtain power-law tails is consistent with the appearance of sub-diffusive transport. [83, 85, 86]

We conclude that the slow operator technique combined with extreme value theory, constitutes a valuable tool for exploring microscopic mechanisms of the

MBL transition. Further developments may include tailoring the optimization technique to target explicitly the rare thermal regions on the localized side, which could provide new insights into this less explored aspect of MBL physics. Another intriguing question would be how coupling an MBL system to an external bath [104–108] would change the structure of slow operators.

Chapter 2

The Quantum East model

Most of the content of this chapter is contained in:

- *"Quantum East model: localization, non-thermal eigenstates and slow dynamics"*
  Nicola Pancotti, Giacomo Giudice, J. Ignacio Cirac, Juan P. Garrahan, Mari Carmen Bañuls
  arXiv:1910.06616 (2019)

In the previous chapter, we studied the physics in the vicinity of the phase transition of a many body-localized system. In that scenario localization was induced by disorder, here we consider a different mechanism where information remains localized due to constrains in the dynamics which are manifested by the form of the Hamiltonian. In particular, we study the properties of the *quantum East model*, introduced in Ref. [109] as a candidate disorder-free system displaying breakdown of ergodicity at long times, and further studied with and without disorder in the context of MBL in Ref. [110, 111]. This model is inspired by the classical stochastic East model [112], a prototypical kinetically constrained model (KCM) of classical glasses (For reviews on classical KCMs and their application to the glass transition problem see [113, 114]). The numerical simulations of Ref. [109] suggested a possible transition in the quantum East model from a thermalizing phase where relaxation is fast, to a phase of slow relaxation where dynamics retains memory of initial conditions for long times indicating the possible absence of ergodicity. However, as it is often the case with numerics for the small systems accessible to exact diagonalization, it is difficult to make convincing extrapolations from the results of [109] for the asymptotic behavior for large system sizes in the quantum East model.

Here we provide what we believe is robust evidence of non-thermal behavior in the quantum East model that goes beyond that of other quantum constrained systems such as the PXP model [115] or quantum dimers [116, 117]. For technical convenience we consider the case of open boundary conditions. We employ a combination of analytical arguments, exact diagonalization and tensor network methods to show that the model displays a *fast-to-slow phase transition* throughout its spectrum, by which we mean a change from a dynamical phase where thermalization is fast to a phase where dynamics is slow and even non-ergodic depending on

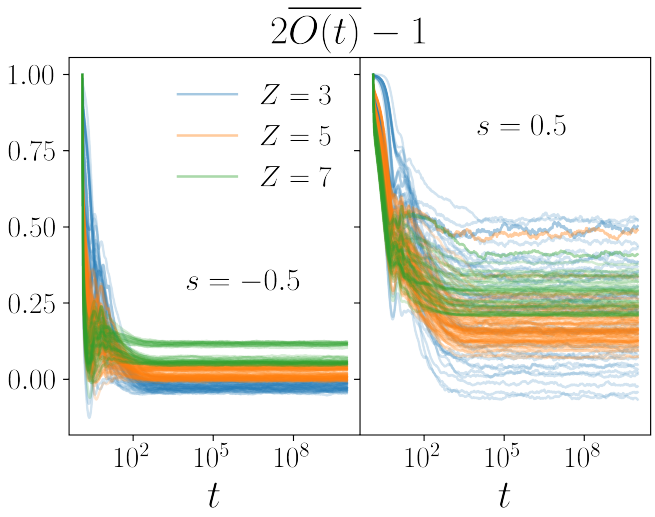


Figure 2.1: **Fast vs. slow dynamics in the quantum East model.** Relaxation to equilibrium of (time-averaged) density autocorrelator (2.2) starting from all possible product initial states for $N = 10$. For $s < 0$ (left) equilibration is fast and memory of the initial conditions is rapidly lost. For $s > 0$ (right) relaxation is slow and memory of initial conditions is preserved throughout the simulation window.

initial conditions. The transition occurs when changing the parameter that controls the balance between kinetic and potential energies in the Hamiltonian across a "Rokhsar–Kivelson" (RK) point [118, 119]. In particular, we demonstrate that the slow dynamical phase is characterized by the following: (i) the ground state is exponentially localized at one of the boundaries; (ii) there is number - scaling exponentially with system size - of eigenstates at finite-energy density that are non-thermal, which we show how to construct analytically exploiting the localization of the ground state and the kinetically constrained nature of the Hamiltonian; (iii) of these, at least a number which is linear in size has area-law entanglement, while for the rest their bipartite entanglement is spatially heterogeneous; (iv) these non-thermal eigenstates have large overlap with product states and can be approximated well by matrix product states (MPS) at arbitrary energy densities; (v) it is possible to construct an even larger number of area-law states, each with an extensive number of localized *super-spins*, that are guaranteed to display very slow growth of entanglement and long term memory of initial conditions. The remarkable range of non-thermal behavior that we uncover in the quantum East model underlines the potential richness of non-equilibrium behavior of quantum KCMs with appropriately tailored constraints.

2.1 Quantum East model

The quantum East model was originally introduced in [109] in order to consider slow quantum dynamics akin to (quasi-)MBL in the absence of disorder with kinetic constraints as the mechanism behind slow evolution. The model is defined in terms

of spin-1/2 degrees of freedom on a one dimensional lattice with Hamiltonian [109],

$$H = -\frac{1}{2}\sum_{i=0}^{N} n_i(e^{-s}\sigma_{i+1}^{x} - \mathbb{1}). \tag{2.1}$$

where the operator $n_i = (\mathbb{1} - \sigma_i^z)/2$ is a projector onto the state $|1\rangle$ in the local z-basis, and σ_i^α is the Pauli-α operator at site i. When $s = 0$, the operator in Eq. (2.1) is (up to a sign) the same as the continuous-time Markov generator of the classical East model, a stochastic KCM much studied in the context of the classical glass transition [112, 120–123]. For $s \neq 0$, it corresponds to the "tilted" generator studied in the context of the dynamical large deviations of the stochastic model, see e.g. [114, 124, 125]. When considered as a quantum Hamiltonian, $s = 0$ is a so-called RK point [118, 119], where the ground state corresponds to the equilibrium probability of the stochastic model.

When interpreted as a stochastic generator, the operator Eq. (2.1) corresponds to the "infinite temperature" classical East model. Note that this terminology does not refer to the temperature of the quantum system, but to the characteristics of the equilibrium probability, i.e., the ground state of Eq. (2.1) at the stochastic point $s = 0$, which is given by the equal superposition of all configurations. Such as maximum entropy probability is that of an "infinite temperature" equilibrium state. For the "finite temperature" version of the East model, i.e. one where the equilibrium state is not the equal weight combination of all configurations, see e.g. [125].

The factor n_i at the front of each term in the Hamiltonian (2.1) is the *kinetic constraint*. It represents an operator valued rate, which in the case of H above makes the action of the local Hamiltonian at site $i + 1$ non-trivial only when n_i projects into the state $|1\rangle$. In the KCM jargon, when this constraint is satisfied, the site i is said to "facilitate" dynamics of its $i + 1$ neighbour (i.e., the one to the East, thus the name of the model) [113, 114]. In contrast to Ref. [109], here we will study the properties of the Hamiltonian (2.1) with open boundary conditions. We do this for technical convenience, as we do not expect the physics we uncover below to be very different for the case with periodic boundaries.

The key numerical observation in Ref. [109] was the change in the dynamical behavior when the parameter s is changed from one side of the RK point, that is from $s < 0$, to the other side, that is $s > 0$. In Fig. 2.1 we reproduce this observation for the case of open boundaries: we show the relaxation to equilibrium of the normalized two-time density autocorrelator $2\overline{O(t)} - 1$, defined as the time average $\overline{O(t)} = \frac{1}{t}\int_0^t O(t')dt'$ of

$$O(t) \equiv \frac{1}{Z}\sum_i \langle n_i(t)n_i(0)\rangle\,, \tag{2.2}$$

where $n_i(t)$ is the occupation operator in the Heisenberg picture under unitary evolution generated by the Hamiltonian Eq. (2.1), and $Z \equiv \sum_i \langle n_i(0)\rangle$ is a normalization factor for the initial occupation. The figure shows results for initial states which are product states in the occupation basis (i.e., local z-basis) at different initial fillings (note that magnetization is not conserved in this model). Notice that,

for finite systems, the energy is determined not only by the initial polarization Z, but also by the occupation of the last site. This is the reason why we observe two different thermal values for the same polarization. This effect vanishes in the thermodynamic limit.

We observe two fundamentally different behaviors of the autocorrelator depending on the sign of s. For $s < 0$ dynamics is *fast* and most of the information about the initial state is quickly erased, as expected from thermalization and compliance with ETH [25]. In contrast, for $s > 0$ dynamics is *slow* and for a large class of initial product states, memory of the initial conditions is retained at arbitrarily long times. This is indicative of a transition in the quantum dynamics of the system.

Motivated by these results, in the following we will analyze the structure of the eigenstates of the Hamiltonian in order to collect information about the dynamical properties of the model both for finite system sizes and in the thermodynamic limit.

2.1.1 Symmetries of the quantum East model

Since the Hamiltonian is identically zero on the empty string $|0\dots0\rangle$, for open boundary conditions the Hilbert space splits in blocks that are not connected by the dynamics. Each block is determined by the position of the first occupied site, i.e. the k-th block corresponds to the subspace spanned by all classical configurations that start with a string of $k-1$ zeroes followed by a 1.

In the following, we will mostly focus on the dynamics of a single block, with N (dynamical) sites to the right of the first occupied one. The position of the latter naturally introduces an edge, and the effective Hamiltonian on the N dynamical sites to its right reads

$$H^N = -\frac{1}{2}(e^{-s}\sigma_1^x - \mathbb{1}) - \frac{1}{2}\sum_{i=1}^{N-1} n_i(e^{-s}\sigma_{i+1}^x - \mathbb{1}). \tag{2.3}$$

Since $[H^N, \sigma_N^x] = 0$, the Hamiltonian in Eq. (2.3) can be further divided in the sum of two commuting terms $H^N = H_+^{N-1} \otimes \Pi_N^+ + H_-^{N-1} \otimes \Pi_N^-$, where $\Pi^\pm = (\mathbb{1} \pm \sigma^x)/2$ are single site projectors onto $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, the eigenstates of σ^x, and

$$\begin{aligned} H_\pm^{N-1} = &-\frac{1}{2}(e^{-s}\sigma_1^x - \mathbb{1}) - \frac{1}{2}\sum_{i=1}^{N-2} n_i(e^{-s}\sigma_{i+1}^x - \mathbb{1}) \\ &- \frac{1}{2}n_{N-1}(\pm e^{-s} - 1). \end{aligned} \tag{2.4}$$

In what follows, we will study and discuss the properties of the Hamiltonians in Eq. (2.1), (2.3) and (2.4).

2.1.2 The special case $s = 0$

At the RK point, $s = 0$, the Hamiltonian (2.1) has an additional symmetry. It can be written as a sum of projectors $H = \sum_i n_i \otimes \Pi_{i+1}^-$ which, in addition to the

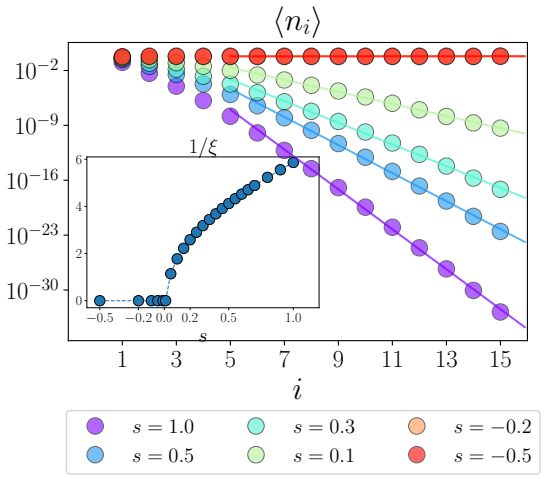


Figure 2.2: **Localization of the ground state for $N = 15$.** The main plot shows the single site occupation $\langle n_k \rangle$ as a function of the position in the chain k. For positive values of s the probability of finding an occupied spin is exponentially suppressed as the distance from the left edge increases. **Inset:** We fit the function (2.6) and plot the inverse of the localization length ξ as a function of the control parameter s.

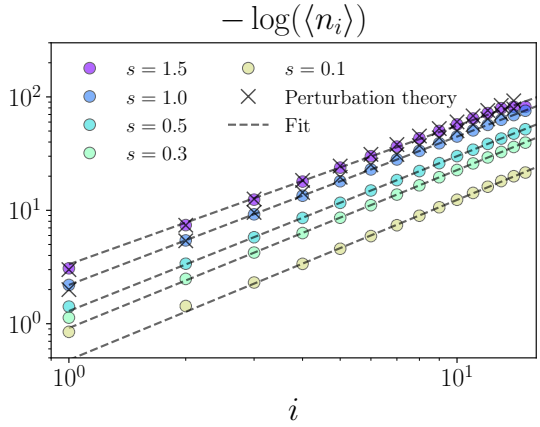


Figure 2.3: **Super-exponential localization of the ground state.** We plot the behavior of $\langle n_k \rangle$ in double-log scale versus log scale. A fit of the form $\log\langle n_k \rangle \propto k^\alpha$ (dashed lines) yields $\alpha \sim 1.3$, indicating a super-exponential decay. The result from a perturbation theory calculation at large s (crosses), without free parameters, shows good agreement with the numerical data.

empty string, annihilates also a string of $|+\rangle$ states. Thus the Hilbert space splits further in blocks determined by the lengths m and n of, respectively, the leading empty string and the trailing string of $|+\rangle$. Hence the eigenstates have the form $|0\rangle^{\otimes m} |1\rangle |\psi^{N_B}\rangle |-\rangle |+\rangle^{\otimes n}$, where N_B is the length of the dynamical part of the (m, n)-block, and $|\psi^{N_B}\rangle$ is an eigenstate of the corresponding effective Hamiltonian,

$$H^{N_B}_{s=0} = \Pi^-_1 + \sum_{i=1}^{N_B-1} n_i \otimes \Pi^-_i + n_{N_B}. \tag{2.5}$$

2.2 Ground state localization phase transition

We now show that the ground state of the quantum East model (2.4) is localized when $s > 0$. Namely, in the ground state of a block of (dynamical) size N, the probability of finding an occupied site decreases exponentially with the distance to the left edge introduced by the first occupied site. The localization length ξ can be extracted already at small sizes, accessible by exact diagonalization, by analyzing the expectation value in the ground state of the local operator n_k as a function of the position k. This is shown in Fig. 2.2 for the ground state of H^N_+, with $N = 15$. For $s < 0$ we observe an almost homogeneous occupation, independent of the system size and the value of s. For $s > 0$, in contrast, the occupation decays

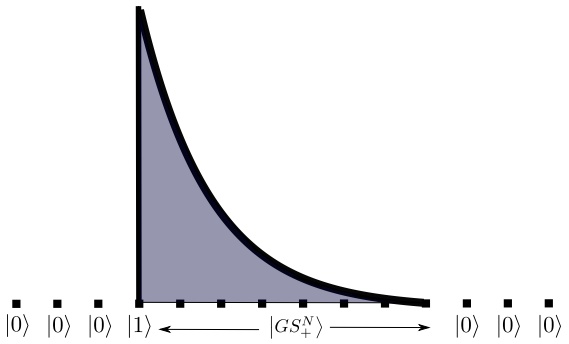


Figure 2.4: **Sketch of a super-spin.** For positive values of s the ground state is exponentially localized near the first occupied spin. The empty sites on the left of $|1\rangle$ are not dynamical. If we complete the state with $|0\rangle$s on the right, we obtain a good approximation of the ground state of the Hamiltonian (2.1) in the thermodynamic limit.

fast with the distance to the edge, with faster decay as we increase s. We find that the results can be fitted assuming an exponential decay,

$$\langle n_i \rangle \sim e^{-i/\xi}, \tag{2.6}$$

and the localization length ξ from the fit captures the phase transition at $s = 0$. Indeed, we find that the value of ξ diverges as $s = 0$ is approached from the positive side, according to $\xi \propto s^{-\nu}$ with $\nu = 0.533 \pm 0.006$ (see inset of Fig. 2.2). These results hold for the ground state of H_+^N in Eq. 2.4. We observe the same qualitative behavior for the ground state of H_-^N. Indeed, both Hamiltonians differ only in the last site, $H_+^N - H_-^N = -e^{-s} n_N$, with the difference decreasing fast for $s > 0$.

The form in Eq. (2.6) provides a reasonable fit of the numerical data for the occupation, but a more detailed look at our numerical results suggests in fact a faster-than-exponential decay, as shown in Fig. 2.3. Indeed, in Sec. 2.6 we show that for large s perturbation theory provides an approximate decay of the form $\langle n_i \rangle \sim (e^{-is}/i!)^2 \sim e^{-i \log(i)}$. As can be seen in figure 2.3, this is in good agreement with the numerical data.

In Fig. 2.4 we provide a cartoon picture of the ground state which for $s > 0$ is localized near the edge. The spatial structure of the GS revealed by these studies can be understood in light of the adiabatic theorem. Away from the phase transition, which happens at $s = 0$ [1], the system is gapped, and we can apply the adiabatic theorem to connect the ground state to the non-interacting one at $s \to \infty$. The latter corresponds to the product state with only the first site occupied, $|10 \ldots 0+\rangle$. Within the gapped region, the evolution with the adiabatically changing Hamiltonian will dress the initial site with an exponential tail like the one shown in our numerical results and depicted in Fig. 2.4; see Sec. 2.6 (and also Ref. [110, 111]).

[1] For finite systems, the transition is actually shifted to a small $s_c > 0$, which converges to 0 faster than $1/N$ [125].

2.3 Eigenstates for large system sizes

Given the localization properties of the ground state discussed above, and the peculiar form of the Hamiltonian, in this section we provide an ansatz for the ground state and some excited states of finite energy density at arbitrarily large system sizes.

2.3.1 The ground state for large system sizes

Consider the normalized state

$$|\Psi_0(L;N)\rangle = |\mathrm{GS}_+^{N-1}\rangle \otimes |0\rangle^{\otimes(L-N+1)}\,, \tag{2.7}$$

where $|\mathrm{GS}_+^{N-1}\rangle$ is the ground state of the Hamiltonian H_+^{N-1} (2.4), supported on $N-1$ sites, $H_+^{N-1}\,|\mathrm{GS}_+^{N-1}\rangle = E_+^0\,|\mathrm{GS}_+^{N-1}\rangle$. We want to show that, in the localized phase, $|\Psi_0(L;N)\rangle$ is close to the ground state of H^L in Eq. (2.3), supported on L sites. In Sec. 2.7, we demonstrate that the only contribution to the energy variance comes from the boundary term between $|\mathrm{GS}_+^{N-1}\rangle$ and the string of empty sites. By using $H_+^{N-1} - H_-^{N-1} = -e^{-s} n_{N-1}$ from Eq. (2.4), it can be easily seen that neither the mean value, nor the variance of the energy evaluated in $|\Psi_0(L;N)\rangle$ depend on L, and they take the simple form

$$\langle H^L\rangle_{\Psi_0} = E_+^0 + \frac{1}{2}e^{-s}\delta, \tag{2.8}$$

$$\langle \Delta H^L\rangle_{\Psi_0} = \frac{e^{-2s}}{2}\left[\delta - \frac{1}{2}\delta^2\right], \tag{2.9}$$

where we have defined $\delta = \langle \mathrm{GS}_+^{N-1}|\, n_{N-1}\,|\mathrm{GS}_+^{N-1}\rangle$. Eqs. (2.8), (2.9) show that both the mean energy and the variance of the state $|\Psi_0(L;N)\rangle$ (supported on L sites) can be estimated from the knowledge of $|\mathrm{GS}_+^{N-1}\rangle$ (supported on $N-1<L$ sites). For small values of δ, namely when the last spin of $|\mathrm{GS}_+^{N-1}\rangle$ is close to $|0\rangle$, the state $|\Psi_0(L;N)\rangle$ is close to an eigenstate of H^L for any L. As can be seen in Fig. 2.2, this is precisely the case when $s>0$. Eq. (2.9) also shows that the quantity δ fully quantifies the energy variance of the extended state. Accordingly, as long as the variance is smaller that the gap (which is sizable already for small positive values of s and for all system sizes), we expect that the state $|\Psi_0(L;N)\rangle$ approximates the ground state of the Hamiltonian, independently of L.

Notice that the form of Eqs. (2.8), (2.9) are also valid (with the E_+^0 in (2.8) replaced by the appropriate energy) if the factor $|\mathrm{GS}_+^{N-1}\rangle$ in Eq. (2.7) is replaced by any other eigenstate of H_+^{N-1}. In Sec. 2.5 we will use δ as a figure of merit for quantifying the number of eigenstates that admit an extension as the one in Eq. (2.7), with small variance. We will show that, for positive values of s, the property above is shared by several eigenstates of the model and not only by the ground state.

2.3.2 Excited states for large system sizes

As we have shown above, by combining the ground state of small systems and strings of empty sites, it is possible to approximate ground states for large system sizes.

The construction utilizes two particular ingredients: the localization properties of the ground state, and the fact that the Hamiltonian annihilates a string of empty sites. In this section we will construct an ansatz for excited states based on similar ideas. Suppose $|\phi_\epsilon^M\rangle$ is an excited state of H^M in Eq. (2.3) supported on M sites, such that $H^M |\phi_\epsilon^M\rangle = E_\epsilon^M |\phi_\epsilon^M\rangle$. The state

$$|\Psi_\epsilon(L;N)\rangle = |\mathrm{GS}_+^{N-1}\rangle \otimes |0\rangle \otimes |1\rangle \otimes |\phi_\epsilon^M\rangle\,, \tag{2.10}$$

(such that $L = N + M + 1$) exhibits similar properties as the one defined in Eq. (2.7). More precisely, as in the previous case, the only contribution to the energy variance comes from the boundary term between the ground states and the empty site and is given by Eq. (2.9). The corresponding expectation value of the energy is $E = E_+^0 + E_\epsilon^M + e^{-s}\delta/2$.

Notice that the states in Eq. (2.10) can be arbitrarily close to an eigenstate of H^L in Eq. (2.3) as long as δ is small enough. Since the typical energy gap between two neighboring eigenstates in the middle of the spectrum for a generic Hamiltonian supported on L sites scales as 2^{-L}, in order to provide accurate approximations, δ needs to decrease at least as fast. As illustrated by Fig. 2.3, δ decays super-exponentially, $\delta \sim \exp(-N \log N)$, which implies that $N \log N \gtrsim L$ will be enough to satisfy that condition. For very large system sizes ($L \to \infty$) this can be achieved if the ground state occupies a fraction of the sites N/L approaching zero. Therefore, the fraction M/L of sites that can be occupied by an excited state approaches one as we increase the system size. As M becomes larger, the states $|\phi_\epsilon^M\rangle$ can reach higher energies leading to *any* finite energy density for the states $|\Psi_\epsilon(L;N)\rangle$.

It is worth stressing that the approximate eigenstates $|\Psi_\epsilon(L;N)\rangle$ are non-thermal and, as long as $M = \mathcal{O}(L)$, they are exponentially many in system size L. More precisely, for any given N, there are $2^{L-(N+1)}$ states of that form: a fraction $2^{-(N+1)}$ of the total number of states in the Hilbert space.

Exploiting the maximally excited state The construction we just described provides an explicit way of addressing excited states at large system sizes by using eigenstates from smaller sizes. In general, nevertheless, states of the form (2.10) do not need to fulfill an area law of entanglement, even if the leftmost N sites are always in a product state with respect to the rightmost $M + 1$ sites of the system, because a highly excited eigenstate $|\phi_\epsilon^M\rangle$ may have volume law entanglement. Thus, the description of $|\Psi_\epsilon(L;N)\rangle$ may require exponential resources. However, there is at least one interesting exception to this situation, when the excited state $|\phi_\epsilon^M\rangle$ corresponds to the maximally excited state of the Hamiltonian H^M in Eq. (2.3), or equivalently, the ground state of $-H^M$ which also admits a MPS approximation.

If we choose $|\phi_\epsilon^M\rangle$ in Eq. (2.10) to be the maximally excited state $|\phi_{\max}^M\rangle$, we obtain an area-law state $|\Psi_{\max}(L;N)\rangle$, with energy $E = E_+^0 + E_{\max}^M + e^{-s}\delta/2$. Since we expect $E_{\max}^M \sim \mathcal{O}(M)$, as long as $M = \mathcal{O}(L)$, the resulting $|\Psi_{\max}(L;N)\rangle$ has finite energy density. Moreover, its energy variance is $\langle \Delta H^L \rangle < \delta$, so that in the localized phase it can be made arbitrarily small by increasing N, and the construction can provide approximate eigenstates.

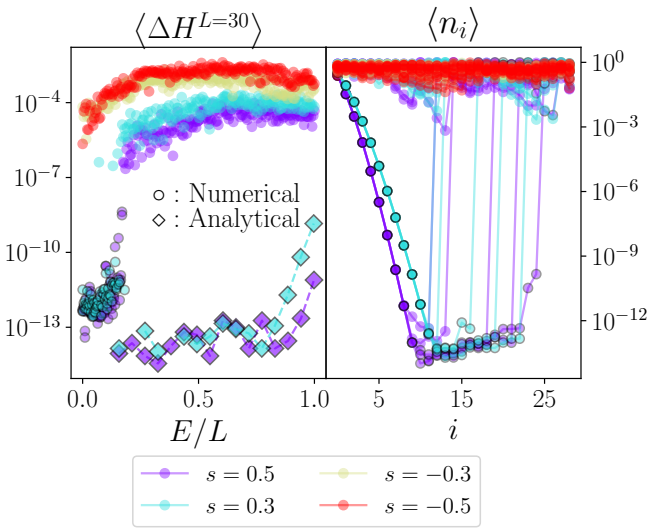


Figure 2.5: **Energy variance(left) and single-site occupation expectation(right) of MPS approximations ($D = 50$) to the excited states for $L = 30$ sites.** (left) For small values of the energy densities and $s > 0$, the MPS approximation are close to machine precision. A black edge indicates the MPS with variance below 10^{-8}. For such states (right) we observe that the spatial distribution of the single site occupation corresponds to the profile of the analytical construction Eq. (2.10).

From the exact diagonalization results above we know that even for small system sizes δ quickly reaches machine precision at least exponentially fast in N. This means that even for modest N its value becomes negligible in the construction above. The construction above immediately suggests an efficient numerical algorithm to construct quasi-exact highly excited eigenstates for system sizes much larger than the ones allowed by exact diagonalization, since we can use variational MPS methods to find the ground states of H^N and $-H^M$ for chains of several hundred sites with extremely good precision [125].

Fig. 2.5 illustrates the construction for a chain of size $L = 30$. In particular, we show the energy variance and occupation distribution of MPS approximations to excited states, found numerically as described in Sec. 2.5.4. For $s > 0$ and small energy densities, for which the MPS provide almost exact eigenstates, we observe that their spatial profile indeed agrees with that of the analytical construction presented in this section. Moreover, for $s > 0$ the construction yields energy variances close to machine precision over practically the whole range of energies, where the direct MPS search is far from reaching an exact eigenstate.

2.4 The super-spin picture

Here we exploit the results from previous sections to engineer a large class of states with small variance. The basic idea is concatenating several blocks of $N + 2$ sites,

each of them in one of two mutually orthogonal states,

$$|\tilde{0}\rangle := |0\rangle^{\otimes N+2}, \quad |\tilde{1}\rangle := |1\rangle \otimes |\mathrm{GS}_+^N\rangle \otimes |0\rangle . \tag{2.11}$$

We identify the subspace spanned by these two vectors with the Hilbert space of a *super-spin*. For a system of size L, we can thus construct $2^{L/(N+2)}$ orthogonal states $|\tilde{s}\rangle \in \{|\tilde{1}\rangle, |\tilde{0}\rangle\}^{\otimes L/(N+2)}$. All such states fulfill an area law and can be approximated as a MPS insofar as $|\mathrm{GS}_+^N\rangle$ does.

The states in this set retain long memory of their initial conditions and stay weakly entangled under time evolution, as we will see in the following subsections. The key dynamical property that we exploit is their energy variance, which can be easily computed using the same procedure as in section 2.3.1. Since $|\tilde{0}\rangle$ blocks do not contribute to the variance, the only contributions come from blocks in $|\tilde{1}\rangle$, and correspond to the value computed in Eq. (2.9)

$$\langle \Delta H \rangle_{\tilde{s}} = \sum_{k=1}^{\mathcal{M}} \langle \Delta H_k \rangle_{\Psi_0^N} = \mathcal{M} \langle \Delta H^N \rangle_{\Psi_0^N} < \mathcal{M}\delta, \tag{2.12}$$

where the index k runs over the positions of the occupied super-spins and $\mathcal{M}$ is the Hamming weight of $|\tilde{s}\rangle$. It is important to stress that $\mathcal{M}$ is potentially unbounded in the thermodynamic limit, in which case the variance becomes unavoidably large.

The energy can also be easily computed,

$$\langle H \rangle_{\tilde{s}} = \mathcal{M}(E_+^0 + e^{-s}\delta/2). \tag{2.13}$$

Equations (2.12) and (2.13) show that if $\mathcal{M}$ is chosen appropriately we can construct states with high energy and exponentially small variance in N. Notice however that, as we want states with small energy variance, we need to introduce limitations on the values of $\mathcal{M}$.

From Eq. (2.12), note that the variance of the super spins cannot exceed the value $\delta L/(N+2) \lesssim 2^{-N\log(N)} L/(N+2) < 2^{-N} L/(N+2)$, since for any given super spin we can accommodate at most $\mathcal{M} = L/(N+2)$ occupied sites. Clearly, we have the freedom of choosing N at will. However, the choice will affect the variance of the super spins, and the dimension of the Hilbert space spanned by them. It is illustrative to mention a few interesting cases. (*i*) If $N \sim \log L^\beta$ with $\beta > 1$ then, the state with $\mathcal{M}$ occupied super spins can have a large variance, exponentially larger than 2^{-L}: $\langle \Delta H \rangle_{\tilde{s}} < L^{1-\beta}/\log L^\beta$. The dimension of the corresponding Hilbert space is of the order $2^{L/\log L^\beta}$. (*ii*) An opposite scenario is when $N \sim L/\log L$. In this case, the variance is small $\langle \Delta H \rangle_{\tilde{s}} < 2^{-L/\log L} \log L$, but the dimension of the Hilbert space is linear in L. (*iii*) An interesting intermediate example consists in $N \sim L^\alpha$, with $0 < \alpha < 1$. Here the variance is $\langle \Delta H \rangle_{\tilde{s}} < L^{1-\alpha} 2^{-L^\alpha}$ and the dimension of the Hilbert space scales as $2^{L^{1-\alpha}}$, which is sub-exponential in system size. In the following section we will show how the variance of an initial state can be use to quantify its slowness. The super-spin picture provides a flexible platform where one can choose the appropriate trade-off between the dimension of the Hilbert space and the dynamical activity of the super-spin vectors that span it.

2.4.1 Dynamical properties of the super-spin states

The memory of the initial state during time evolution admits a general bound based on the initial energy variance. For an initial state $|\psi_0\rangle$, we define the overlap [126]

$$a(t) = |\langle\psi_0|\psi_t\rangle|^2 = \mathrm{Tr}\left(\rho_0\rho_t\right), \tag{2.14}$$

where $|\psi_t\rangle$ is the state at time t and $\rho_t = |\psi_t\rangle\langle\psi_t|$ the corresponding density matrix. Using the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\frac{d^2a(t)}{dt^2}\right| \leq & \| \left[\rho_0, H\right] \|_F \| \left[\rho_t, H\right] \|_F \\ = & \sqrt{4 \left\langle \Delta H \right\rangle_{\psi_0} \left\langle \Delta H \right\rangle_{\psi_t}} = 2 \left\langle \Delta H \right\rangle_{\psi_0},\end{aligned} \tag{2.15}$$

where $\| \cdot \|_F$ denotes the Frobenius norm. In the second line we used the fact that for the commutator with the Hamiltonian this norm does not depend on time. Exploiting Eq. (2.15) we can compute the memory of the initial state as

$$\begin{aligned}|a(t) - a(0)| = & \left| \int_0^t \int_0^\tau \frac{d^2a(\tau')}{d\tau'^2} d\tau d\tau' \right| \\ \leq & 2 \left\langle \Delta H \right\rangle_{\psi_0} \int_0^t \int_0^\tau d\tau d\tau',\end{aligned} \tag{2.16}$$

which leads to the bound

$$a(t) \geq 1 - \left\langle \Delta H \right\rangle_{\psi_0} t^2, \tag{2.17}$$

where we used $a(0) = 1$, $\frac{da}{dt}|_{t=0} = 0$ and $a(t) \leq a(0)$. Eq. (2.17) is a general bound on the memory of a time evolved state based on the energy variance of the corresponding initial state.

The bound in Eq. (2.17) can be used to bound the growth of the entanglement entropy of an arbitrary subsystem. According to the Fannes inequality, for any pair of density matrices M_1, M_2, of dimensions $\mathcal{D} \times \mathcal{D}$ [127],

$$|S(M_1) - S(M_2)| \leq 2T \log(\mathcal{D}) - 2T \log(2T), \tag{2.18}$$

where $T = \frac{1}{2}\|M_1 - M_2\|_1 = \frac{1}{2}\,\mathrm{Tr}[\sqrt{(M_1 - M_2)(M_1 - M_2)^\dagger}]$ is the trace distance between both matrices, and $S(M) = -\,\mathrm{Tr}\left[M \log(M)\right]$ is the von Neumann entropy. We can apply Eq. (2.18) to the reduced density matrix of a subsystem at the initial time and after evolution [2]. Given some partition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of the Hilbert space, let us define the (time-dependent) reduced density matrix $\sigma_\tau = \mathrm{Tr}_B(\rho_\tau)$. Contractivity of the trace norm ensures

$$\frac{1}{2}\|\sigma_t - \sigma_0\|_1 \leq \frac{1}{2}\|\rho_t - \rho_0\|_1 \leq t\sqrt{\langle \Delta H \rangle_{\psi_0}}, \tag{2.19}$$

[2]Note that Eq. 2.18 holds if $T \leq 1/2e$. This does not spoil the results at larger times, since we can use the weaker relation $\Delta S \leq 2T \log \mathcal{D} + 1/(e \log 2)$, which qualitatively gives the same scaling.

where in the second inequality we used

$$\frac{1}{2}\|\rho_t - \rho_0\|_1 = \sqrt{1 - |\langle\psi_0|\psi_t\rangle|^2} = \sqrt{1 - a(t)}, \tag{2.20}$$

and Eq. (2.17). By plugging Eq. (2.19) in Eq. (2.18), we can bound the growth of the entanglement entropy as

$$|S(\sigma_t) - S(\sigma_0)| \leq 2t\sqrt{\langle\Delta H\rangle_{\psi_0}} \log\left(\frac{\mathcal{D}}{2t\sqrt{\langle\Delta H\rangle_{\psi_0}}}\right). \tag{2.21}$$

Eq. (2.21) provides a general bound on the growth of the entanglement entropy of a subsystem based on the energy variance of the initial extended pure state, and the dimension $\mathcal{D}$ of the subsystem.

The bounds on the memory of the initial state in Eq. (2.17) and the growth of the entropy in Eq. (2.21), can be straightforwardly applied to the super-spins $|\tilde{s}\rangle$, defined in Sec. 2.4. In the particular case when $\rho_0 = |\tilde{s}\rangle\langle\tilde{s}|$ (supported on L sites) we can bound the memory of the initial conditions by using Eq.(2.17) and $\langle\Delta H\rangle_{\tilde{s}} < \mathcal{M}\delta$. Namely,

$$|\langle\tilde{s}(t)|\tilde{s}(0)\rangle|^2 \geq 1 - t^2\mathcal{M}\delta. \tag{2.22}$$

Accordingly, if we take σ_0 to be the corresponding reduced density matrix for a region of $N \ll L$ sites, $\mathcal{D} = 2^N$. The bound in Eq. (2.21) then reads

$$|S(\sigma_t) - S(\sigma_0)| \leq 2t\sqrt{\mathcal{M}\delta}\left(N\log(2) - \log\left(2t\sqrt{\mathcal{M}\delta}\right)\right). \tag{2.23}$$

In the previous sections we showed that in the localized region δ decreases exponentially with N. As a consequence, if $\mathcal{M}$ is sufficiently small, Eq. (2.22) and Eq. (2.23) provide tight bounds on the dynamics of $|\tilde{s}\rangle$. Specifically, in order to erase half of the memory of the initial state, i.e. $|\langle\tilde{s}(t^*)|\tilde{s}(0)\rangle|^2 \leq 1/2$, the dynamics needs at least exponentially long times in N, $t^* \gtrsim (\mathcal{M}\delta/2)^{-1/2}$. At the same time, for entangling a sub-region of size N, i.e. $2t\sqrt{\mathcal{M}\delta}N \sim 1$, the time evolution necessitates exponential times of the order $t^* \sim (2N\sqrt{\mathcal{M}\delta})^{-1}$. We conclude that the dynamics of the states $|\tilde{s}\rangle$, in order to entangle a sub-region, requires exponentially long time in the subsystem size.

The states $|\tilde{s}\rangle$ can then be seen as an orthonormal set of quasi-conserved area-law vectors, and any superposition of them will result in a state whose dynamics at short times is governed by dephasing only. The super-spin picture thus provides an effective description of a subset of the Hilbert space in the thermodynamic limit which evolves slowly in time, is weakly entangled, and efficiently simulable.

The results in Ref. [109] can be reinterpreted from a super-spin point of view. It was numerically argued there, for the case of periodic boundary conditions, that for certain product states the dynamics exhibits a slow growth of the entanglement entropy, exponential in system size. The *slowness* of the state was quantified by the number of empty sites following an occupied one. Since the previous statements about the energy variance of $|\tilde{s}\rangle$ do not depend on the boundary conditions and, as argued in section 2.2, the block ground state for $s > 0$ is very close to the product state $|1000\ldots\rangle$, the bound in Eq. (2.23) gives a rigorous interpretation of the previous numerical observations.

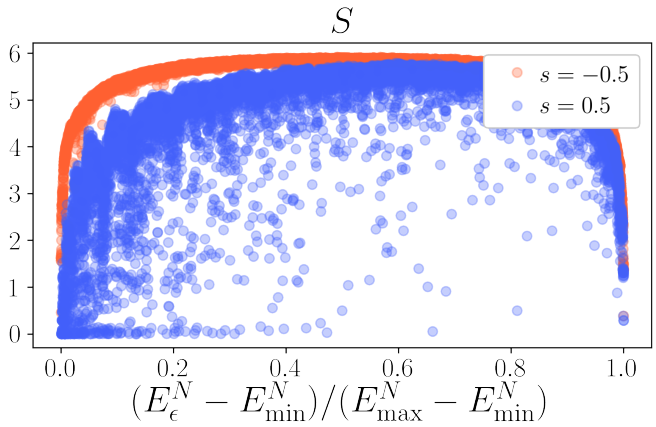


Figure 2.6: **Entanglement Entropy of the eigenstates vs. energy density, $N = 14$.** In the two regimes of positive and negative s, the entanglement entropy of the eigenstates in the middle of the chain shows abrupt changes from *normal* eigenstates at $s = -0.5$, to *anomalous* eigenstates, where many of them have small entropy in the middle of the spectrum at $s = 0.5$.

Extensions The super-spin construction described above can be made more general in several ways. On the one hand, a larger set of states can be constructed by allowing not only the ground state, but also (sufficiently localized) excited states as building blocks $|\phi^M_\epsilon\rangle$. In Sec. 2.5 we show that such excited states do actually exist. On the other hand, by combining the super-spin picture with the excited state construction in Sec. 2.3.2 we can also construct states with finite energy density. Namely, we can construct states $|\mathcal{S}\rangle = |\tilde{s}\rangle \otimes |\phi^M_{\max}\rangle$ with energy $\langle H\rangle_{\mathcal{S}} = \mathcal{M}(E^0_+ + e^{-s}\delta/2) + E^M_{\max}$ and energy variance $\langle \Delta H\rangle_{\mathcal{S}} < \mathcal{M}\delta$. By increasing M the energy density can be increased, but at the cost of reducing the dimension of the super-spin subspace to $2^{(L-M)/(N+2)}$.

2.5 Non-thermal excited states in small system sizes

In the following we explore the properties of the whole Hamiltonian spectrum using exact diagonalization for small system sizes. The results indicate a substantial change in the properties of eigenstates across the spectrum in the region $s > 0$. In particular, in this region many localized eigenstates can be found, beyond the ground state, which can be used in the constructions of the previous sections.

2.5.1 Entropy of the exact eigenstates

Since eigenstates of the Hamiltonian incorporate the whole information about the dynamics of the system, their entanglement entropy is often used as an indicator of the associated dynamical behavior. In Fig. 2.6, we show the entanglement entropy in the middle of the chain of $N = 14$ spins from exact diagonalization, for two values of s. For negative s, the entanglement entropy of eigenstates exhibits behavior compatible with a thermalizing system — apart from the extremes of

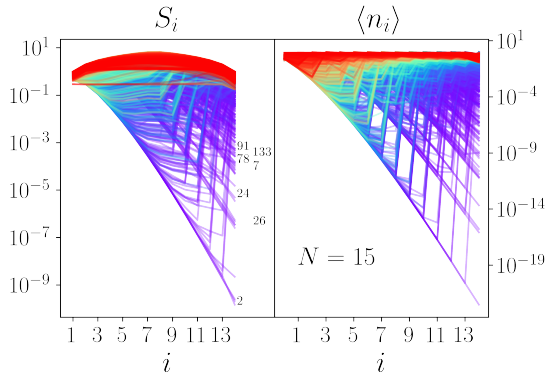


Figure 2.7: **Entanglement entropy and local occupation as a function of the position on the chain.** For several eigenstates both quantities decay exponentially with the distance to the left edge. The labels in the entropy plot indicate the indices of (some of) the eigenvectors ordered by increasing energy. Colors indicate the energy of the eigenstates from purple (low energy) to red (high energy). The figures show data for $s = 0.5$ and $N = 15$.

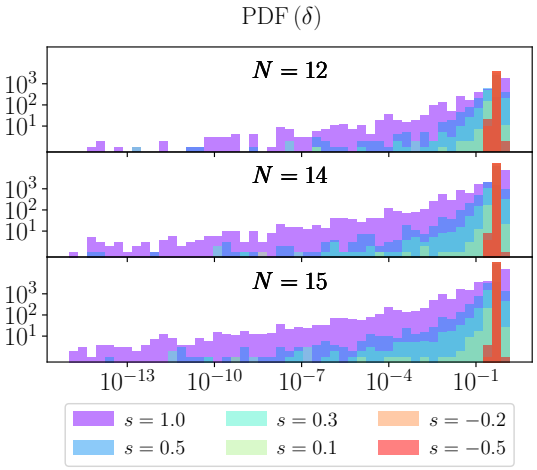


Figure 2.8: **Probability density function of δ, log-log scale.** For any given eigenstate we compute the expectation value of the projector onto occupied site n_N. As in the previous cases, the distribution is strongly peaked for negative values of s, and it develops very long tails for positive values.

the spectrum, most of the eigenstates have large entanglement, almost saturating the upper bound given by system size. In contrast, for positive values of s a considerable number of excited eigenstates have low entanglement. This is an indication of non-thermal eigenstates and reminiscent of the quantum scars found in the PXP model [115], but here we observe this behavior for a much larger number of states.

In order to collect detailed information about the distribution of the entanglement along the spin chain, we compute, for each eigenstate, the entanglement entropy with respect to all possible cuts of the chain, $S_i = S(\rho_i)$, where ρ_i is the reduced density matrix obtained when tracing out all but the leftmost i spins. In Fig. 2.7, we plot the entanglement entropy S_i and single site occupation $\langle n_i \rangle$ as a function of the position of the cut (respectively the site) i for all eigenstates in the case $s = 0.5$ and $N = 15$. The figure suggests a peculiar *heterogeneous entanglement* structure of a significant number of eigenstates, for which both quantities decay exponentially as the cut moves to the right. In other words, for many eigenstates, the spins far from the left edge are almost in a product state with the rest of the system, and the corresponding sites are almost empty. These results are qualitatively similar to the ones discussed in Sec. 2.2 where we analyzed the localization properties of the ground state.

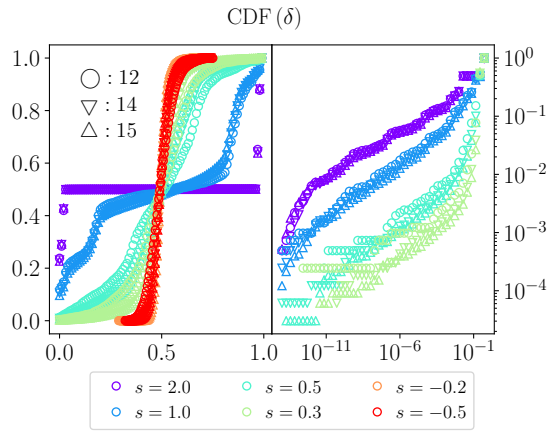


Figure 2.9: **Cumulative distribution function of δ.** Same data as in Fig. 2.8. The CDF shows a steep curve for negative values of s, where most of the eigenstates have large values of $\langle n_N \rangle$. When s is positive the CDF shows fat tails extending to values close to machine precision. For large enough values of s, the fraction of eigenstates with small $\langle n_N \rangle$ seems to be size independent.

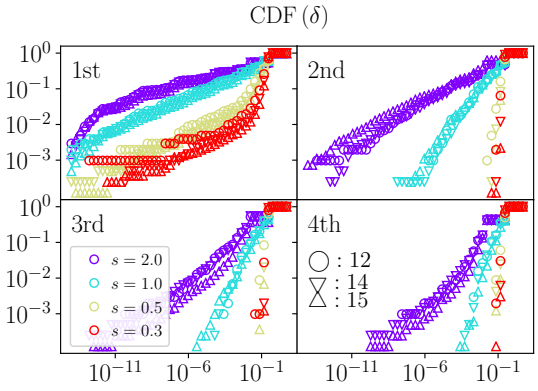


Figure 2.10: **Energy-resolved cumulative distribution function of δ.** We split the eigenstates in four equal-size energy intervals, ordered by growing energy. Most of the eigenstates with small δ are concentrated at low energy. For the largest $s = 2$, the count for eigenstates with small δ increases at all energies.

2.5.2 Small-δ eigenstates

We diagonalize the Hamiltonian H_+^N in Eq. (2.4) for different system sizes N and values of s. Given the set of eigenvectors, we consider the probability distribution of the last site occupation $\delta = \langle n_N \rangle$, the parameter which, as discussed in Sec. 2.3.1, quantifies the variance of the extended eigenstates $|\phi_\epsilon^N\rangle \otimes |0 \dots 0\rangle$. Fig. 2.8 shows the histogram of the corresponding probability density function PDF(δ). Notice that, for positive values of s, many eigenstates exhibit surprisingly small values of δ. Namely, there are several eigenstates $|\phi_\epsilon^N\rangle$ such that the energy variance of the state $|\phi_\epsilon^N\rangle \otimes |0 \dots 0\rangle$ can be bounded by extremely small values.

In order to quantify the number of eigenstates with small δ, in Fig. 2.9 and 2.10 we consider the cumulative distribution function CDF(δ). In particular, in Fig. 2.9 we observe an abrupt change from negative values of s, where most of the eigenstates have large values of δ to positive s, where more and more eigenstates have very small values. For the sizes accessible by exact diagonalization, the fraction of eigenstates with small δ does not seem to depend on the size of the system. In Fig. 2.10 we show the energy-resolved CDF. In particular, we divide the spectrum in four intervals of equal energy width, which we number in order of increasing energy. The figure shows that most of the small-δ eigenstates are concentrated in the lower part of the spectrum, in agreement with the results in Fig. 2.7. As s increases, we observe that the number of eigenstates with small δ values grows for all energy regions, as we indeed expect from the discussion in the previous sections and the smaller localization length.

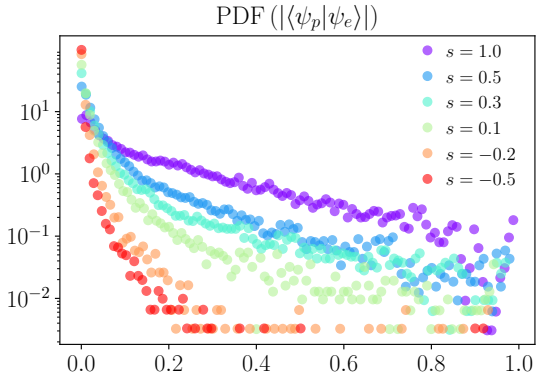


Figure 2.11: **Probability density function of the estimated maximal overlap with a product state.** For negative values of s most of the eigenstates have small overlap with product states. For positive values of s the distribution develops long tails: many eigenstates have considerably large overlap with product states. Data is shown for $N = 15$.

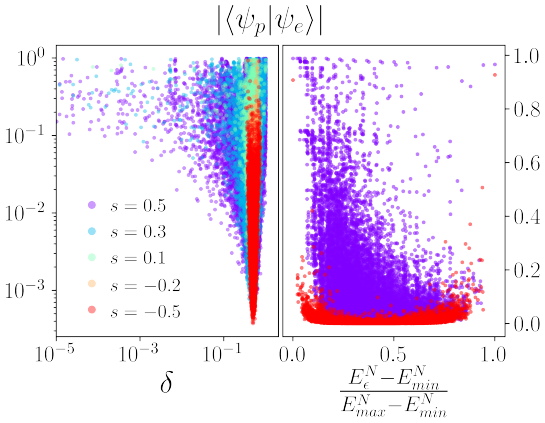


Figure 2.12: **Scatter plot of the overlap with product states vs. δ (left), and energy density (right).** The states with large overlaps are mostly concentrated at small energy densities, although large values can be found everywhere along the spectrum. We show data for $N = 15$. On the right, only data points for $s = \pm 0.5$ are shown.

2.5.3 Geometric entanglement

The geometric entanglement of a state, defined as its minimum distance to a product state, also provides interesting insights about the properties of the eigenstates. Given a pure state, the geometric entanglement can be found by maximizing its overlap with a product state. Although it is possible to solve this optimization problem with exact or approximate numerical algorithms, in our case this is unpractical, since we need to repeat the calculation for each eigenstate. Instead, we apply a simpler *one-sweep* truncation strategy to construct an approximation to the closest product state. Namely, for each eigenstate, we sequentially perform a singular value decomposition with respect to each cut of the chain and keep only the largest singular value for each of them. The resulting product state, once normalized, provides a lower bound to the maximum overlap.

In Fig. 2.11 we plot the probability density function (over all energy eigenstates) of this estimate for the maximum overlap. For negative values of s, most of the eigenstates have a small overlap with product states (as expected for an ergodic system). For positive values of s, the distributions develop a fat tail towards small values of δ, indicating that many eigenstates have a large overlap with product states.

An alternative view of this feature is demonstrated in Fig. 2.12, which shows the value of the overlap for each eigenstate, as a function of the corresponding δ (left) or energy density (right). Small values of δ are strongly correlated with large — order $\mathcal{O}(1)$ — overlaps with product states. They are mostly concentrated at small energy densities, but Fig. 2.12 shows that large overlaps can actually be

found at arbitrary energy densities.

2.5.4 Numerical approximation of non-thermal excited states for large sizes

The discussion in Sec. 2.3.2 indicates the existence of highly excited states with small entanglement, that can be written as MPS. We can thus try to find them with numerical methods. There are several possible variations of DMRG to try and target excited states [128]. The simplest one attempts to find the MPS that minimizes the expectation value of the operator $W = (H - \lambda)^2$, where λ is the target energy of the desired states. Since H is a matrix product operator, also W has that form and the minimization can be run efficiently with standard MPS algorithms. We use this tool to probe the whole energy spectrum for eigenstates that can be approximated by MPS.

In the numerical study we fix the system size to $L = 30$ and the bond dimension to $D = 50$. For several values of s, we then collect 300 data points, uniformly distributed in energy (excluding the edges of the spectrum). In Fig. 2.5 (left) we show the energy variance as a function of the energy density. We observe that, for $s > 0$ and low energy, the algorithm produces MPS with variance close to machine precision. Fig. 2.5 (right) shows the profile of the expectation value of the single-site occupation number $\langle n_i \rangle$ as a function of the site i. At low energy densities, and positive s, the optimization finds states with a structure that resembles $|\Psi_{\max}\rangle$, with exponentially decreasing occupation from the left edge to the right until a certain site, where the occupation abruptly increases to stay close to one until the right edge. In Fig. 2.5 we marked with a black circle the states with energy variance smaller than 10^{-8}. We find that all the states with small energy variance have an exponential tail which starts from the left edge. According to our analytical construction in Sec. 2.3, the position of the jump should correspond to the energy of the state. For large energies such construction becomes harder to capture, and the optimization is forced to search for a trade-off between accurate target energy or small energy variance. Notice that our optimization is not tailored to search for this specific construction, as each run starts from a random initial MPS.

2.6 Perturbation theory

We split Eq. (2.4) into a bare Hamiltonian and an interaction term

$$H_{\pm}^{N} = \frac{1}{2}\left(H_0 + \varepsilon V_{\pm}\right), \tag{2.24}$$

where

$$\begin{aligned} H_0 &= \sum_{i=1}^{N-1} n_i \\ V_{\pm} &= -\sigma_1^x - \sum_{i=1}^{N-1} n_i \sigma_{i+1}^x + (1 \mp e^{-s}) n_N . \end{aligned} \tag{2.25}$$

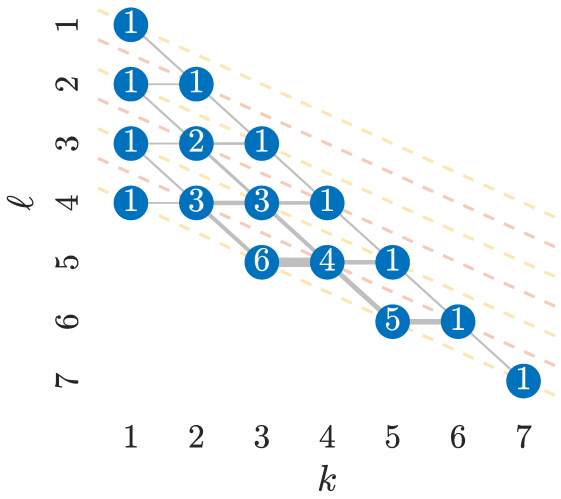


Figure 2.13: **Transition graph of states in perturbation theory.** The states contained in the first seven orders of perturbation theory are illustrated as a graph, where each nodes groups states with the same k (the number of occupied sites) and ℓ (the position of the rightmost occupied site). The number of states for fixed (k, ℓ) is indicated on each node. The thickness of each edge represents the number of transitions between states contained in the nodes via an application of RV. We notice that the graph has the structure of a Pascal triangle, where its shallow diagonals are highlighted with a dashed line. Notice that with an even (odd) number of hops along the edges from the topmost node $|1000\ldots\rangle$ we end up on an even (odd) shallow diagonal, colored in yellow (red).

We treat $\varepsilon = e^{-s}$ as a small parameter, so we can look at the first order corrections in ε of the ground state deep in the localized phase, i.e. $s \to \infty$. In the following, we consider only the perturbation $V = V_+$, since the following calculations are the same for the other case. The bare Hamiltonian H_0 is diagonal in the computational basis and has as unique ground state $|\chi^{(0)}\rangle = |000\ldots\rangle$ with a unit gap above it. The excited states are labeled by $|k_\alpha\rangle$, where $k = 1, \ldots, N$ counts the number of occupied sites and α labels the degeneracy: $H_0 |k_\alpha\rangle = k |k_\alpha\rangle$.

Using the notation in Ref. [129] we can express the corrections to the ground state as a Taylor expansion in the parameter ε

$$\begin{aligned} H_+^N |\chi\rangle &= E_0 |\chi\rangle , \\ |\chi\rangle &= |\chi^{(0)}\rangle + \varepsilon |\chi^{(1)}\rangle + \varepsilon^2 |\chi^{(2)}\rangle \ldots , \\ E_0 &= \frac{1}{2} \left(E^{(0)} + \varepsilon E^{(1)} + \varepsilon^2 E^{(2)} + \ldots \right) . \end{aligned} \tag{2.26}$$

By defining the resolvent matrix

$$R = -\sum_{k_\alpha} \frac{|k_\alpha\rangle \langle k_\alpha|}{k}, \tag{2.27}$$

the n-th order terms can be computed as

$$\begin{aligned} |\chi^{(n)}\rangle &= (RV)^n |\chi^{(0)}\rangle , \\ E^{(n)} &= \langle \chi^{(0)} | V | \chi^{(n-1)} \rangle . \end{aligned} \tag{2.28}$$

From a more physical point of view, the perturbation should be seen as the introduction of a local impurity that has an effect that is localized at the boundary, rather than a global perturbation giving an extensive energy contribution. This can be seen by looking explicitly the first orders in Eq. (2.28):

$$\begin{aligned} |\chi^{(1)}\rangle &= |1000\ldots\rangle , \\ |\chi^{(2)}\rangle &= \frac{1}{2} |1100\ldots\rangle , \\ |\chi^{(3)}\rangle &= \frac{1}{2} \left(|1000\ldots\rangle + |0100\ldots\rangle + \frac{1}{3} |1110\ldots\rangle \right) . \end{aligned} \tag{2.29}$$

The corresponding energy is straightforward to compute

$$E_0 = -\frac{\varepsilon^2}{2} - \frac{\varepsilon^4}{4} + \mathcal{O}\left(\varepsilon^6\right) . \tag{2.30}$$

We now comment on the convergence radius of the perturbation expansion. Firstly, we notice that in a Banach space, if the series $\sum_{n=0} |\varepsilon^n| \|\chi^{(n)}\|$ converges, then the series for $|\chi\rangle$ converges too. We shall now prove that the norm $\|\chi^{(n)}\|^2$ grows as the n-th power of the golden ratio φ. In order to understand the structure of each perturbation theory order, it is instructive to construct a graph of transition between the states, as shown in Fig. 2.13. The states are grouped in cluster labeled by the number k of occupied sites and the length ℓ of the non-trivial string. Each cluster contains $\binom{\ell}{k-1}$ states, and applying RV to a state $|k, \ell\rangle$

generates up to $k+1$ states in at most 4 distinct ways: $|k \pm 1, \ell\rangle$ or $|k \pm 1, \ell \pm 1\rangle$. Then the n-th perturbation term can contain at most the first $\lfloor \frac{n+1}{2} \rfloor$ even or odd "shallow" diagonals, depending on whether n is even or odd. In other words, the number of states generated by applying RV to the n-th perturbation order is equal to the number of states in the order $(n-1)$ plus the number of states in the n-th shallow diagonal, which is the corresponding Fibonacci number F_n, since $\sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n-j-1}{j} = F_n$. Hence the total number of states $\#^{(n)}$ in the n-th order is

$$\#^{(n)} = \begin{cases} \sum_{j=0}^{\lfloor n/2 \rfloor} F_{2j+1} = F_{n+1} & n \text{ odd} \\ \sum_{j=0}^{n/2} F_{2j} = F_{n+1} - 1 & n \text{ even} \end{cases} \tag{2.31}$$

which asymptotically tends to φ^{n+1}. The amplitude of each state will be the sum of all the paths of length $(n-1)$ leading to it from $|1000\ldots\rangle$, weighted by $\prod_k 1/k$ of all the visited states $|k_\alpha\rangle$ along the path. The weighted sum cannot exceed $(1/2)^{\lfloor (n-1)/2 \rfloor}$, since this is the path connecting $|\chi^{(1)}\rangle$ and $|\chi^{(2)}\rangle$. Furthermore, the number of distinct paths cannot exceed 4^{n-1}, since in principle we can move in the 4 directions for $(n-1)$ moves, as shown in Fig 2.13. While these estimates are not particularly tight, they are sufficient to bound the norm of $|\chi^{(n)}\rangle$ as

$$\|\chi^{(n)}\|^2 < \left(\frac{2^{2(n-1)}}{2^{\lfloor (n-1)/2 \rfloor}} \right)^2 \varphi^{n+1} \sim (2^3 \varphi)^n \tag{2.32}$$

Thus the perturbation series in Eq. (2.26) is at least convergent in the radius

$$\varepsilon < \frac{1}{2\sqrt{2\varphi}} \approx 0.278 \tag{2.33}$$

or equivalently $s \gtrsim 1.28$. While this result is used to prove a positive radius of convergence, in practice we can monitor the rate of convergence of the norm of each order to have confidence in the method. Indeed, numerically, we observe a rapidly converging norm in the regime $s \gtrsim 0.9$.

The perturbation theory picture not only gives us a method to compute the first order corrections to the ground state, but also an understanding of its structure. Indeed, looking at the first terms in the expansion, we notice that the terms remain localized close to the left boundary. As we apply powers of V, higher order terms in Eq. (2.28) become progressively delocalized, but these contributions get damped at least by a factor ε^n.

Let us define the domain-wall state $|\Theta_m\rangle = |1\rangle^{\otimes m} |000\ldots\rangle$. From Eq. (2.29), we notice that $\langle \Theta_n | \chi^{(n)} \rangle = 1/n!$ and that $\langle \Theta_{m>n} | \chi^{(n)} \rangle = 0$. At each order n there is a contribution $|\Theta_n\rangle$ coming from a $|\Theta_{n-1}\rangle$ in the previous order. In Fig. 2.13, these states correspond to the rightmost diagonal, i.e. $k = n$. One can think of this term as the fastest possible "excitation" created by applying V repeatedly. Since all other terms in $|\chi^{(m)}\rangle$ have a smaller support, clearly $n_m |\chi^{(m)}\rangle = 1/m!$, and more generally, at the first order

$$n_r |\chi\rangle = \frac{\varepsilon^r}{r!} |\Theta_r\rangle + \ldots. \tag{2.34}$$

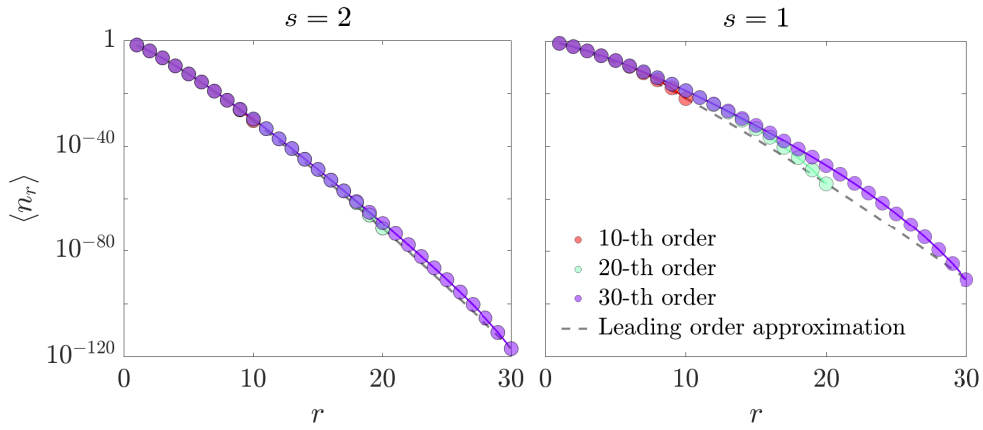


Figure 2.14: **Perturbation theory results for the occupation profile.** The results for the occupation obtained using finite order perturbation theory and the extracted leading order contribution are illustrated for different values of the perturbation. We notice that, in the limit of weak perturbation, the two methods give similar results. Close to the boundary of the convergence radius, higher order term contributions become relevant, giving corrections only at larger distances.

We can then conclude that

$$\langle n_r \rangle = \frac{\langle \chi | n_r | \chi \rangle}{\langle \chi | \chi \rangle} = \left(\frac{\varepsilon^r}{r!} \right)^2 + \mathcal{O}\left(\varepsilon^{2r+1} \right) . \tag{2.35}$$

We have dropped the normalization factor, since $\langle \chi | \chi \rangle = 1 + \mathcal{O}(\varepsilon^2)$ leads to sub-leading corrections. Checking with finite-order expansions, Eq 2.35 gives a qualitatively accurate prediction of the decay of the occupation number, as shown in Fig. 2.14. Using Stirling's formula, the asymptotic behavior of Eq. (2.35) at large distances is $\langle n_r \rangle \sim \exp[-2r \log r + 2r(1-s)]$. Therefore, in the perturbative regime, the ground state exhibits super-exponential localization around the first site.

These perturbation calculations can equivalently be connected to the adiabatic theorem [130]. By introducing a time-dependent coupling $\varepsilon(t)$, the different perturbation orders correspond to the expansion of the evolution operator. Similar perturbation calculations were performed on the quantum East model in the case of periodic boundary conditions [110], in order to derive an effective Hamiltonian describing the hopping of domain wall in a given magnetization sector.

2.7 Computation of the variance

Consider the Hamiltonian defined in Eq. (2.1) on a one dimensional lattice of N sites, and the eigenstates,

$$\begin{aligned} |\Phi_\pm\rangle &= |\phi_\pm(N-1)\rangle \otimes |\pm\rangle , \\ H^N |\Phi_\pm\rangle &= H_\pm^{N-1} \otimes \Pi_\pm |\phi_\pm(N-1)\rangle \otimes |\pm\rangle = E_\pm |\Phi_\pm\rangle . \end{aligned} \tag{2.36}$$

We consider the normalized state

$$\begin{aligned} |\Psi_+\rangle =& |\phi_+(N-1)\rangle \otimes |0\rangle^{\otimes(L-N+1)} \\ =& \sqrt{2}\Pi_0^N |\Phi_+\rangle \otimes |0\rangle^{\otimes(L-N)} , \end{aligned} \tag{2.37}$$

in the $|+\rangle$ sector, similar considerations hold in the $|-\rangle$ sector. The state $|\Psi_+\rangle$ admits a simple computation of its energy expectation, and its variance. In order to compute the variance, we need two ingredients. The expectation value of the energy

$$\langle\Psi_+|H^L|\Psi_+\rangle = \left\langle \Psi_+ \right| \left[H_+^{N-1}\otimes\Pi_+ + H_-^{N-1}\otimes\Pi_-\right] \left| \Psi_+ \right\rangle , \tag{2.38}$$

which can be computed from

$$\begin{aligned} \langle\Psi_+|H_+^{N-1}\otimes\Pi_+|\Psi_+\rangle =& \frac{1}{2}E_+ \\ \langle\Psi_+|H_-^{N-1}\otimes\Pi_-|\Psi_+\rangle =& \frac{1}{2}\left(E_+ + e^{-s}\left\langle n_{N-1}\right\rangle_{\phi_+}\right), \end{aligned} \tag{2.39}$$

and takes the form

$$\langle H^L\rangle = \langle\Psi_+|H^L|\Psi_+\rangle = E_+ + \frac{1}{2}e^{-s}\left\langle n_{N-1}\right\rangle_{\phi_+} . \tag{2.40}$$

The expectation value of the square of the Hamiltonian is

$$\begin{aligned} \langle H^L H^L\rangle =& \left\langle (H_+^{N-1})^2\otimes\Pi_+\right\rangle \\ &+ \left\langle (H_-^{N-1})^2\otimes\Pi_-\right\rangle , \end{aligned} \tag{2.41}$$

where we can explicitly calculate

$$\begin{aligned} \langle (H_+^{N-1})^2\otimes\Pi_+\rangle =& \frac{1}{2}E_+^2 \\ \langle (H_-^{N-1})^2\otimes\Pi_-\rangle =& \frac{1}{2}\left(E_+^2 + e^{-2s}\left\langle n_{N-1}\right\rangle_{\phi_+}\right) \\ &+ E_+ e^{-s}\left\langle n_{N-1}\right\rangle_{\phi_+} . \end{aligned} \tag{2.42}$$

We have now all the ingredients to compute the variance

$$\langle H^L H^L\rangle - \langle H^L\rangle^2 = \frac{e^{-2s}}{2}\left[\left\langle n_{N-1}\right\rangle_{\phi_\pm} - \frac{1}{2}\left\langle n_{N-1}\right\rangle_{\phi_\pm}^2\right] . \tag{2.43}$$

2.8 Conclusions and Outlook

We provided an in-depth analysis of the dynamical properties of the quantum East model. The model is known [124, 125] to have a first-order quantum phase transition at the critical point $s = 0$. Here we showed that, in correspondence to the phase transition point, the ground state undergoes a localization transition from completely delocalized to super-exponentially localized ($s > 0$). We showed how this ground state transition leads to a sharp change throughout the spectrum from a *fast* dynamical phase at $s < 0$, where ergodicity is established quickly under unitary evolution, to a *slow* dynamical phase at $s > 0$ where thermalization is impeded. We provided rigorous results about the dynamical consequences of this transition focusing on the behavior in the slow non-thermalizing side.

Summary of the results. By combining analytical arguments, exact diagonalization and tensor network methods, we made three key findings.

(i) The localized nature of the ground state for $s > 0$ allows for a systematic construction of the ground state for arbitrary system sizes. This construct is very simple, that of a tensor product of the ground state of a small system with a completely empty state. Since the second factor is annihilated due to the constraints, all cost is concentrated at the juncture, which the localization in the first factor makes vanishingly small in the large size limit.

(ii) This procedure can be extended to obtain exact large-size eigenstates of finite energy density. By replacing the right factor by an excited state, one can systematically construct a large number of non-thermal excited states. If the right factor is that of the eigenstate of maximal energy, the ensuing large-size eigenstate has area law entanglement. This means that there are (at least) polynomially (in system size) many area law eigenstates of finite energy density. Exact diagonalization of small systems suggests furthermore that many other eigenstates have small entanglement, so repeating the construction with these as the rightmost factor could, in principle, allow us to obtain an even larger number of area law eigenstates.

(iii) By generalizing the tensor product construction to many junctions we can define an even larger class of non-thermal states in terms of what we call super-spins. A state composed of super-spins is the tensor product of several ground states for a finite system of a fixed size, possibly separated by empty blocks of the same size, and thus corresponds to a dressed occupied spin localized at each occupied juncture. From the arguments above, if the number of super-spins scales sub-extensively with system size, and the distance between junctions is large enough, such states become area-law eigenstates in the large size limit. States with extensive number of super-spins in contrast, while may still have small energy variance, are not guaranteed to be eigenstates. Nevertheless, it is easy to prove that under unitary evolution they retain memory of initial conditions and take exponentially long times to entangle a small sub-region. These are the states that underpin the slow dynamics of the model.

We supported our claims with extensive numerical results for small systems obtained with exact diagonalization, as well as for large systems using tensor networks. In particular, we considered several quantities of interest, such as the entanglement entropy of the eigenstates, and the distributions of their last site occupation and of their maximal overlap with a product state. The statistical analysis shows that many eigenstates exhibit atypical behavior, signaling the presence of non-thermal dynamical properties. These properties confirm for small sizes the (apparent) singular change throughout the whole spectrum as one changes the parameter s from negative to positive. Although all our analytical constructions rely on the localization of the ground state, our numerical studies indicate that many other eigenstates have similar localization properties. Our results suggest the existence of an emergent non-ergodicity of the quantum East model in its slow dynamical phase, stemming from the localization properties of a subset of its eigenstates, which is already evident at small system sizes. This provides a robust analytical framework for the investigation of the breakdown of ergodicity in this

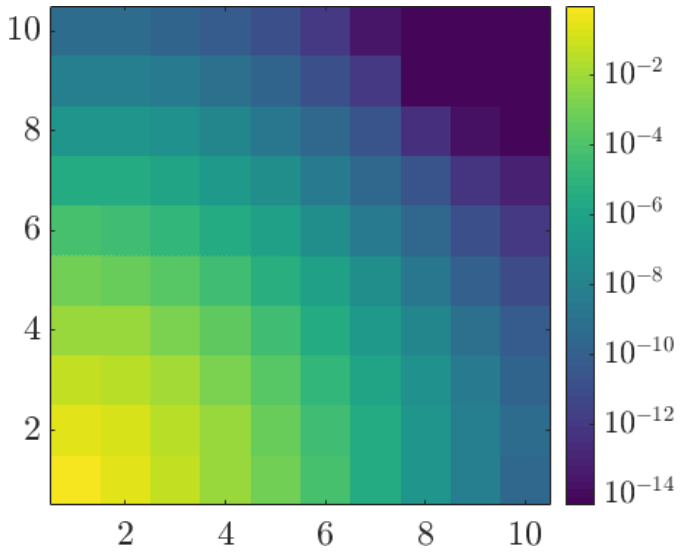


Figure 2.15: **Single-site occupation of the quantum North-or-East model ground state with $s = 0.01$ on a 10×10 lattice.** The low-entanglement structure allow to perform accurate DMRG calculations [131]. Even for small values of s, the ground state exhibits strong localization.

system.

Comparison with other constrained models. It is natural to compare the large number of non-thermal finite-energy-density eigenstates we found for the quantum East model to *quantum scars* [115] in the PXP model. The PXP [132, 133] is a constrained system known to thermalize [134] for typical states, but possessing a number of low entanglement excited eigenstates across the spectrum: the scar states. These states — whose number grows linearly in the system size — are responsible for non-thermalization when starting from particular initial conditions [115].

There is one key difference with the PXP model. The PXP constraint allows spin-flips only when two nearest neighbors are in the down state, in contrast to the "1-spin facilitation" (albeit directional) of the East model. The PXP constraint is thus stronger than that of the East model, and as a consequence, in one dimension the state space breaks into exponentially many dynamically disconnected blocks. In contrast, as we explained above, in the East model (with appropriate boundary conditions) all states are connected. Thus a *weaker* constraint in the East model gives rise to *stronger* dynamical features, associated with the fact that regions devoid of occupied spins are locally frozen — yet still dynamically connected — a feature we exploited in the construction of non-thermal states.

More generally, it is important to distinguish the mechanisms for the emergence of non-thermal excited states — and concomitant slow dynamics and potential non-ergodicity — we have uncovered here from those based on what recently has been dubbed "shattering of Hilbert space" [135–140]. For the quantum East model and the boundary conditions we considered, see Sec. 2.1, the dynamics connects all the states in the Hilbert space. Furthermore, the non-thermal excited states that we find become exact *only in the limit* $N \to \infty$. That is to say that the change

throughout the spectrum from $s < 0$ to $s > 0$ has the character of a true phase transition, not occurring at finite N but in the thermodynamic limit.

Classically, the issue of fragmentation of configuration space for Markov generators of stochastic dynamics with constraints is well understood [113]. Before establishing whether a KCM dynamics is not ergodic, it is necessary to understand if the constraints make the generator *reducible*. Namely, whether there are regions of configuration space that are disconnected by the dynamics *at any finite system size*. When a dynamical generator is reducible, there can be an apparent breakdown of ergodicity simply by starting with an initial condition which has weight on disconnected sectors. However, reducibility should not be confused with non-ergodicity, which deals with diverging relaxation times within *a single connected irreducible sector*, and which occurs in the large size limit only. For the quantum case, similar considerations may apply. Our results for the quantum East model show a similar emergence of non-thermal behaviour within an irreducibile sector which becomes exact in the large size limit. (For further discussion on these issues, see also [141, 142]).

Future directions. The slow dynamics of the East model is a *first-order* phenomenon — cf. the transition in the ground state. It is a consequence of having spatial coexistence of two very different kinds of dynamics. That is, since a region with no occupied spins is locally stable, it can only be relaxed starting from the interface with an active region. While here we studied explicitly the spectral properties of the East model with open boundaries, we expect to find similar slow characteristics in the case of periodic boundaries, and for other one-dimensional models with similar constraints, such as the quantum "2-spin facilitated" Fredrickson-Andersen model (FA) [113, 143] (a 1-spin facilitated model but with a symmetric constraint).

Finding new and broader classes of area-law eigenstates might give insightful information about the interplay between localization in disorder-free models and ergodicity breaking. Along the lines of our constructions, one possible direction includes the characterization of excited states for smaller sizes in order to promote them to fundamental building blocks for eigenstates at larger sizes in other systems. Their characterization may as well contribute to the understanding of the dynamical properties of their classical counterparts.

Perhaps the most interesting research direction is the extension of these ideas to higher dimensions. From classical KCMs [114] we know that qualitative features are not very dependent on dimensionality. It would be natural to study two quantum generalizations of KCMs in particular, the quantum North-or-East model (see e.g. Ref. [23]), which extends the East model constraint to two dimensions, and the quantum 2-spin facilitated FA model (see Ref. [113] for the classical version), both of which have the essential feature of local "inactive" regions annihilated by the constraint, which is an essential building block in the constructions above. Furthermore, a DMRG calculation for the quantum North-or-East model [144] on a 10×10 lattice shows that the ground state for $s = 0.01$ is strongly localized, too (see Fig. 2.15). This suggests that these two-dimensional KCMs should also display slow dynamics and a prominence of non-thermal states like the ones uncovered here for the quantum East model, therefore being candidates for non-ergodic and

non-disordered quantum systems beyond one dimension.

Part III

Quantum Thermodynamics

Chapter 3

Shortcuts to Isothermality

Most of the content of this part is contained in:

- *"Speed-ups to isothermality: Enhanced quantum heat engines through control of thesystem-bath coupling"*
  Nicola Pancotti, Matteo Scandi, Mark T. Mitchison, and Martí Perarnau-Llobet
  arXiv:1911:XXX (2019)

In this chapter, our aim is to design speed-ups to isothermal processes which do not come at the price of higher dissipation or work cost. As a consequence, our speed-ups to isothermality (STI) can be readily used to maximize the power of finite-time Carnot engines [145, 146] while keeping their efficiency constant. Due to the extra control of the system-bath interaction we find that the dissipation $W_{\rm diss}$ for optimal STI can asymptotically decay as

$$W_{\rm diss} \propto \frac{1}{\tau_{\rm tot}^{2\alpha+1}} \tag{3.1}$$

where $\tau_{\rm tot}$ is the total time of the process, and different $\alpha > 0$ can be obtained by a suitable STI. In particular, we provide two explicit examples where $\alpha = 1/2$. The decay in Eq. (3.1) can substantially outperform the standard scaling for large $\tau_{\rm tot}$, $W_{\rm diss} \propto \tau_{\rm tot}^{-1}$, commonly found in protocols where no control on the system-bath interaction is possible [37, 38, 147–149]. Furthermore, we show how the scaling in Eq. (3.1) leads to a new family of efficiencies at maximum power that interpolate between the Curzon-Ahlborn efficiency (for $\alpha = 0$) and the Carnot efficiency (for $\alpha \to \infty$).

These results are obtained by working in the regime of slow driving (i.e. the timescale of the driving is slower than the time-dependent equilibration timescale), and by assuming that the (time-dependent) coupling g remains weak along the whole process. In particular, we assume that the timescale of thermalization $\tau_{\rm eq}(g)$ scales as $\tau_{\rm eq}(g) \propto 1/g^2$, where g is the strength of the interaction. This scaling is expected in Markovian open (quantum) systems, where our findings are expected to be most relevant. Yet, we will also argue that some of our considerations also work for beyond weak coupling, in particular for strongly correlated systems that thermalize [150].

The above heuristic approach is supported by explicit calculations for two specific yet illustrative models. Firstly, we study quantum Brownian motion, where a quantum harmonic oscillator with time-dependent frequency interacts with a time-dependent coupling to a large (but finite) set of quantum harmonic oscillators [151, 152]. By exploiting the quadratic nature of the Hamiltonian we simulate the full unitary dynamics, which allows us to account for all sources of dissipation, including the one induced by changing the interaction strength. Secondly, we consider the resonant-level model [153–157], where a single fermionic level with a time-dependent energy couples to an infinite bath of fermionic modes via a time-dependent interaction. We employ a quantum Langevin approach to derive analytical results valid in the limit of large yet finite reservoir bandwidth. In both models, we explicitly evaluate all sources of dissipation, including those introduced by the time-dependent coupling. These complementary analyses confirm our analytical findings based on heuristic assumptions.

3.1 Isothermal processes

Consider a driven Hamiltonian

$$H(t) = H^{(S)}(t) + g(t)V + H^{(B)} \tag{3.2}$$

where $H^{(S)}(t)$ is the Hamiltonian of the system S, on which one has experimental control, while $H^{(B)}$ is the Hamiltonian of the bath B and V is the interaction between the two, whose strength is governed by the parameter g. The whole information of system and bath together (SB) is contained in the density matrix ρ.

Consider a transformation between an initial Hamiltonian $H(0) = H^i$ and final one $H(\tau_{\rm tot}) = H^f$. Without loss of generality we can normalise the parameter t to the unit interval by introducing the compact notation $X_t \equiv X(t\tau)$ with $t \in [0, 1]$, τ the duration of the process under consideration and $X = H$, $H^{(S)}$, ρ etc. The average work associated to this transformation is given by the expression

$$W = \int_0^1 {\rm d}t \ {\rm Tr}\left(\rho_t \dot{H}_t\right), \tag{3.3}$$

where ρ_t describes the instantaneous state of SB.

Suppose first that the integrand is well described by the equilibrium value at all times, i.e. ${\rm Tr}\left(\rho_t \dot{H}_t\right) = {\rm Tr}\left(\rho_t^{\rm th} \dot{H}_t\right)$ with

$$\rho_t^{\rm th} \equiv \frac{e^{-\beta H_t}}{\mathcal{Z}}, \tag{3.4}$$

and $\mathcal{Z} = {\rm Tr}(e^{-\beta H_t})$. It follows that

$$W = \int_0^1 {\rm d}t \ {\rm Tr}\left(\rho_t^{\rm th} \dot{H}_t\right) = \frac{1}{\beta} \ln \frac{\mathcal{Z}_i}{\mathcal{Z}_f} =: \Delta F, \tag{3.5}$$

where $\mathcal{Z}$ is the partition function, $\mathcal{Z}_{i/f} = {\rm Tr}(e^{-\beta H^{i/f}})$, and $F = -\frac{1}{\beta}\log(\mathcal{Z})$ is the free energy of SB. Eq. (3.5) is fulfilled in the limit $\tau_{\rm tot} \to \infty$ and when the

driven observables $\dot{H}$ thermalize (as expected for local observables). Note that the quantities in Eq. (3.5) depend on the Hamiltonian (3.2) of the system and bath together, in general.

In the slow driving limit, i.e. for large but finite $\tau_{\rm tot}$, Eq. (3.5) no longer holds as some work is dissipated into the bath. In order to quantify the dissipated work, one introduces $W_{\rm diss} \equiv W - \Delta F \geq 0$, which tends to zero as $\tau_{\rm tot} \to \infty$. Expanding $W_{\rm diss}$ around $1/\tau_{\rm tot}$, one obtains

$$W_{\rm diss} = \frac{1}{\tau_{\rm tot}} \int_0^1 {\rm d}t\, G_{\rho_t^{\rm th}}\Big(\dot{H}_t, \dot{H}_t\Big) + \mathcal{O}\left(\frac{1}{\tau_{\rm tot}^2}\right) \tag{3.6}$$

where $G_{\rho_t^{\rm th}}$ is a bilinear form evaluated at equilibrium $\rho_t^{\rm th}$. The form $G_{\rho_t^{\rm th}}$ was previously studied in different contexts. It was obtained through linear-response theory [38, 158–160], by master equation approaches [149, 161–163], or directly from the partition function [164, 165]. For clarity of the exposition, here we focus on the latter, but our (heuristic) arguments can be extended to more general $G_{\rho_t^{\rm th}}$ (see Sec. 3.5). Furthermore, for this work it is enough to consider time-dependent Hamiltonians satisfying $\dot{H}_t = \dot{\lambda}_t \tilde{H}$, where $\tilde{H}$ is some (time-independent) observable and λ_t is the control parameter. In this case, we can write [162, 164, 165]

$$W_{\rm diss} = \frac{\tau_{\rm eq}\beta}{\tau_{\rm tot}} \int_0^1 {\rm d}t\, \dot{\lambda}_t^2 {\rm cov}_{\rho_t^{\rm th}}\left(\tilde{H}, \tilde{H}\right) + \mathcal{O}\left(\frac{\tau_{\rm eq}^2}{\tau_{\rm tot}^2}\right), \tag{3.7}$$

where $\tau_{\rm eq}$ is the time-scale of relaxation (associated to ${\rm Tr}(\rho_t \tilde{H})$), and

$${\rm cov}_{\rho_t^{\rm th}}\left(\tilde{H}, \tilde{H}\right) = \frac{1}{\beta^2} \frac{\partial^2 \ln \mathcal{Z}}{\partial \lambda^2}, \tag{3.8}$$

which can be expressed in terms of the generalised covariance defined as

$${\rm cov}_{\rho_t^{\rm th}}(A, B) = {\rm Tr}\left(A \int_0^1 {\rm d}s\, (\rho_t^{\rm th})^{1-s} \left(B - {\rm Tr}(\rho_t^{\rm th} B)\mathbb{1}\right) (\rho_t^{\rm th})^s\right). \tag{3.9}$$

Eq. (3.8) (and hence Eq. (3.7)) gives the standard notion of a thermodynamic metric commonly used to describe dissipative systems near equilibrium [164, 165], and it provides us with a simple analytical form which depends on a single timescale $\tau_{\rm eq}$.

In this chapter we are interested in modifications of the system-bath interaction strength, assuming initially weak coupling. In this regime, we can expand around $g = 0$, corresponding to replacing the thermal state of the interacting system $\rho_t^{\rm th}$ by the non interacting one $\rho_0^{\rm th}$. In particular, for ${\rm cov}_{\rho_t^{\rm th}}(A, A)$ we have

$${\rm cov}_{\rho_t^{\rm th}}(A, A) = c_A^{(0)} + c_A^{(1)} g + c_A^{(2)} g^2 + ..., \tag{3.10}$$

where we note that a similar expansion can be performed for the more general $G_{\rho_t^{\rm th}}$ in Eq. (3.6). Due to weak coupling between S and B, we also assume that the thermalization time $\tau_{\rm eq}(g)$ is related to the strength of the interaction g introduced in Eq. (3.2) via

$$\tau_{\rm eq}(g) \propto \frac{1}{g^2}, \tag{3.11}$$

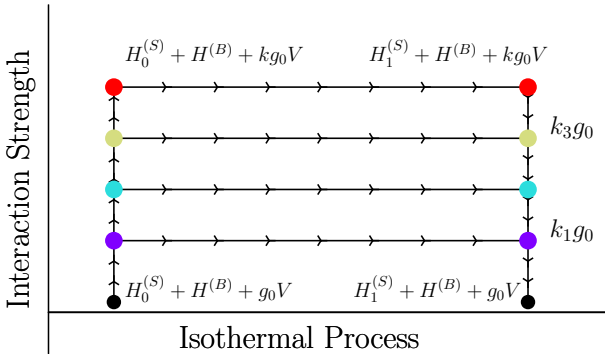


Figure 3.1: **Schematic representation of the protocol.** We consider the family of thermodynamic protocols from an initial Hamiltonian $H_0 = H_0^{(S)} + H^{(B)} + g_0V$ to a final Hamiltonian $H_1 = H_1^{(S)} + H^{(B)} + g_0V$. For each protocol we choose an interaction strength $g_1 = k_ig_0$.

which is expected in Markovian dissipative evolutions [166].

Given Eqs. (3.7) and (3.11), it is clear that the dissipated work W_{diss} may be reduced by increasing g and hence decreasing the thermalization timescale $\tau_{\text{eq}}(g)$. However, any modification of the Hamiltonian will require additional work to be performed, leading to a non-trivial trade-off between speed and dissipated work. In what follows we develop strategies to optimally modulate g in order to speed up the process while keeping the overall dissipation constant.

3.2 Speed-ups to isothermality

Consider thermodynamic protocols governed by the time-dependent Hamiltonian,

$$H_t = H_t^{(S)} + g_tV + H^{(B)}. \tag{3.12}$$

Note that the time dependence in (3.12) affects both the Hamiltonian of the system $H_t^{(S)}$ and the interaction strength g_t. We focus on protocols comprising the following three steps:

1. The interaction between system and bath is increased from g_0 to g_1 in a time τ_{on}, keeping the system Hamiltonian constant.

2. An isothermal transformation $H_0^{(S)} \to H_1^{(S)}$ is performed in a time τ_{iso}, while the interaction strength is kept constant at g_1.

3. The interaction between the system and the bath is reduced to the initial value g_0 in a time τ_{off}, again holding $H^{(S)}$ constant.

In Fig. 3.1 we give a schematic representation of the thermodynamic protocol for different coupling strengths.

For simplicity we assume $\tau_{\text{off}} = \tau_{\text{on}}$. Since both τ_{on} and τ_{iso} are finite, work is dissipated during each step of the protocol. We call $W_{\text{diss}}^{\text{on}}$, $W_{\text{diss}}^{\text{iso}}$ and $W_{\text{diss}}^{\text{off}}$ the

dissipated work in steps 1, 2 and 3, respectively. The total dissipation reads $W_{\rm diss} = W_{\rm diss}^{\rm on} + W_{\rm diss}^{\rm iso} + W_{\rm diss}^{\rm off}$ and the total duration of the protocol is given by

$$\tau_{\rm tot} = \tau_{\rm on} + \tau_{\rm iso} + \tau_{\rm off} = 2\tau_{\rm on} + \tau_{\rm iso}. \tag{3.13}$$

Our goal is to optimize $\tau_{\rm on}$, $\tau_{\rm iso}$ and the interaction strength g_1 such that $\tau_{\rm tot}$ is reduced and the dissipated work $W_{\rm diss}$ stays approximately constant.

3.2.1 Steps 1 and 3 : Taming the dissipation when the interaction is increased or decreased

We consider a family of protocols where the system-bath interaction strength changes polynomially in time according to

$$g_t = g_0 + (g_1 - g_0)t^\alpha, \tag{3.14}$$

for $\alpha > 1$ and $t \in (0, 1)$. Throughout this section, we take $g_1 = kg_0$, with $g_0 > 0$. The assumption of a small, non-zero initial interaction strength is technically necessary to ensure that we remain in the slow-driving regime, i.e. $\dot{g}_t \tau_{\rm eq}(g_t) \ll 1$ even for small t.

In order to quantify the dissipation during the transformation we make use of an expansion analogous to Eq. (3.7),

$$W_{\rm diss}^{\rm on} = \frac{\beta}{\tau_{\rm on}} \int_0^1 {\rm d}t\, \tau_{\rm eq}(g_t)\, \dot{g}_t^2 {\rm cov}_{\rho_t^{\rm th}}(V, V), \tag{3.15}$$

where we have introduced a time-dependent equilibration time-scale $\tau_{\rm eq}(g_t)$. Furthermore, through Eq. (3.11) we have

$$\tau_{\rm eq}(g_t) = \frac{\tau_{\rm eq}(g_0)}{(1 + (k-1)t^\alpha)^2}. \tag{3.16}$$

In order to evaluate Eq. (3.15), we approximate the covariance as in (3.10). We assume that $c_V^{(1)} = 0$, since ${\rm Tr}(V\rho_0^{\rm th}) = 0$ holds exactly in a broad class of relevant open quantum systems, such as the examples discussed in Sec. 3.4. Let us now consider two cases separately: keeping only the lowest-order term ($c_V^{(0)}$) or retaining also the second-order one ($c_V^{(2)}$).

Zeroth order

In this case, by replacing ${\rm cov}_{\rho_t^{\rm th}}(V, V)$ by $c_V^{(0)}$ in (3.15) we obtain

$$W_{\rm diss}^{(1)} = \frac{\beta g_0^2 \tau_{\rm eq}(g_0)\, c_V^{(0)}}{\tau_{\rm on}} F^{(1)}(\alpha, k) \tag{3.17}$$

with

$$F^{(1)}(\alpha, k) = \int_0^1 {\rm d}t\, \frac{\alpha^2 (k-1)^2 t^{2(\alpha-1)}}{(1 + (k-1)t^\alpha)^2}, \tag{3.18}$$

an integral that admits a solution in terms of the incomplete beta function. For large k, we can approximate $F(\alpha, k)$ as

$$F^{(1)}(\alpha, k) \approx \pi(\alpha - 1) \sin^{-1}\left(\frac{\pi}{\alpha}\right) k^{\frac{1}{\alpha}} \tag{3.19}$$

with $\alpha > 1$, whereas $F^{(1)}(1, k) = (k-1)^2/k$.

Now, let us introduce a time-scale $\tau_{\text{on}}^{\text{weak}}$, corresponding to the time of turning on the interaction for $k = 1$. Then, note that by choosing,

$$\tau_{\text{on}} = F^{(1)}(\alpha, k)\tau_{\text{on}}^{\text{weak}}, \tag{3.20}$$

the dissipation (3.17) becomes independent of k. Namely, Eq. (3.20) indicates how to scale up τ_{on} with k in such a way that the dissipation stays constant at leading order in g when the interaction is increased.

Second order

One can also consider a more conservative choice than the one in Eq. (3.20) by accounting for $c_V^{(2)}$ in Eq. (3.10). The dissipation $W_{\text{diss}}^{(2)}$ induced by the second order term corresponds to

$$W_{\text{diss}}^{(2)} = \frac{\beta g_0^4 \tau_{\text{eq}}(g_0) c_V^{(2)}}{\tau_{\text{on}}} F^{(2)}(\alpha, k) \tag{3.21}$$

with $F^{(2)}(\alpha, k) = (k-1)^2\alpha^2/(2\alpha - 1)$. So, for large k, we can assume for simplicity,

$$F^{(2)}(\alpha, k) \approx \frac{\alpha^2}{2\alpha - 1} k^2. \tag{3.22}$$

Thus, by taking

$$\tau_{\text{on}} = F^{(2)}(\alpha, k)\tau_{\text{on}}^{\text{weak}}, \tag{3.23}$$

we ensure that $W_{\text{diss}}^{(2)}$ is independent of k. Note that this choice is more conservative since $F^{(2)}(\alpha, k) \geq F^{(1)}(\alpha, k)$ for $k \geq 1$.

Higher orders

In principle, one can extend the previous considerations to find more conservative choices of τ_{on} as a function of k by accounting for higher orders in Eq. (3.10). However, for stronger couplings a more useful approach is to use the fact that $\text{cov}_{\rho_t^{\text{th}}}(V, V) \leq 2||V||^2$, in order to bound the (exact) dissipation (3.15) as

$$W_{\text{diss}} \leq \frac{\beta g_0^2 \tau_{\text{eq}}(g_0)\, 2||V||^2}{\tau_{\text{on}}} F^{(1)}(\alpha, k). \tag{3.24}$$

Hence, in models where $g_0^2||V||^2$ is finite (and possibly small), it appears plausible that the choice (3.20) is in fact already sufficient to keep the dissipation controlled (note that with (3.20) the upper bound becomes independent of k). Importantly,

the bound (3.24) also works for strongly correlated and non-Markovian systems, suggesting that our considerations in principle also apply beyond weak coupling for strongly correlated systems that thermalize [150]. In particular, the bound (3.24) is tight for finite-dimensional and locally interacting systems, such as fermionic or spin models, where $\|V\|$ is of the order of the system-bath boundary and independent of the size of the bath. In such cases, the scaling of the equilibration time might differ from Eq. (3.11), but our framework can be easily adapted to account for that.

Discussion

In this section, we showed how to scale up τ_{on} with k to ensure that W_{diss} does not increase as we increase the interaction. We followed two complementary approaches. First, taking a perturbative expansion of W_{diss} for weak coupling, we derived two possible choices: Eq. (3.20) and Eq. (3.23). The former ensures stays W_{diss} stays constant at leading order in the expansion (zeroth order), whereas the latter ensures that W_{diss} does not increase with k up to second order in g. Secondly, we showed that one can also upper bound W_{diss} by a k-independent bound by combining (3.24) and (3.20), a bound which holds at arbitrary strong coupling (i.e. large g) as long as $||V||$ is finite. In Sec. 3.4 we will test these choices for fermionic and bosonic baths (see Figs. 3.7), showing that these generic considerations work well in relevant physical models.

3.2.2 Step 2: Isothermal part of the process

Now we focus on the isothermal part of the protocol. The protocol consists of modifying the Hamiltonian of the system $H_t^{(S)}$ whilst keeping the coupling strength g constant. Analogously to the previous section, we introduce $\tau_{\text{iso}}^{\text{weak}}$ as the time spent to perform the isothermal part of the protocol for $k = 1$, i.e., in absence of modulations of the interaction. By assuming the scaling in Eq. (3.11) and by using the expansion in Eq. (3.7), we can choose the time τ_{iso} of the isothermal process for $k > 1$ as

$$\tau_{\text{iso}} = \frac{\tau_{\text{iso}}^{\text{weak}}}{k^2} \tag{3.25}$$

in order to keep the dissipation constant for any value of k. Similar to the previous section, this is strictly valid at leading order in g_0, i.e. when keeping only the first term in (3.10). This appears justified in the dissipative systems we consider in this work (see Sec. 3.4).

3.2.3 Full protocol

Collecting all the considerations above, we have devised choices of τ_{on} and τ_{iso} as a function of k which guarantee an overall constant dissipation in the weak coupling regime. The total time of the protocol reads,

$$\tau_{\text{tot}} = 2F^{(i)}(\alpha, k)\tau_{\text{on}}^{\text{weak}} + \frac{\tau_{\text{iso}}^{\text{weak}}}{k^2}, \tag{3.26}$$

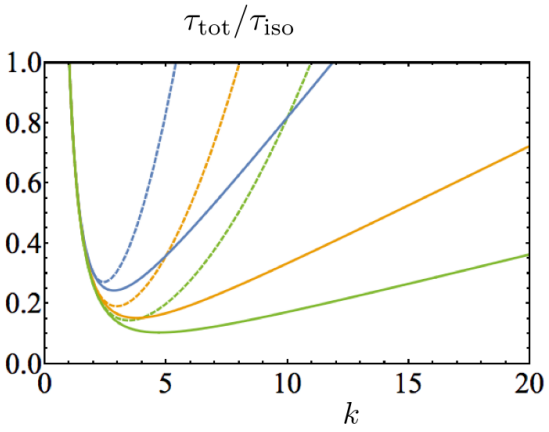


Figure 3.2: The total time $\tau_{\rm total}$ in (3.26) as a function of k, for different values of $\mathcal{T} = \tau_{\rm iso}^{\rm weak}/\tau_{\rm on}^{\rm weak}$: $\mathcal{T} = 20$ (blue), $\mathcal{T} = 50$ (orange), $\mathcal{T} = 100$ (green); and $\alpha = 1$. Dashed line corresponds to $F^{(2)}$ and solid line to $F^{(1)}$.

where $F^{(i)}(\alpha, k)$ is given by either Eq. (3.18) or Eq. (3.22), depending on the specific model (see Sec. 3.4 for examples). In Fig. 3.2 we show the behaviour of Eq. (3.26) for different values of the dimensionless ratio $\mathcal{T} = \tau_{\rm iso}^{\rm weak}/\tau_{\rm on}^{\rm weak}$, and for both choices (3.18) and (3.22). Note that for large $\mathcal{T}$, as in realistic situations (normally the isothermal process is much longer than the time spent switching the system-bath interaction on and off), we obtain substantial improvements on the time of protocol. Our proposal hence provides a way of substantially speeding up isothermal processes through control of the system-bath interaction, which crucially does not come at the price of increased dissipation.

3.2.4 Optimal protocols and decay of dissipation

Above, we designed a family of protocols in which the dissipation remains constant, while the total time of the process can be adjusted as a function of k (see also Fig. 3.2). Let us now minimize the expression in Eq. (3.26) to find the fastest isothermal process for a given dissipation.

We first consider the zeroth-order expansion from Sec. 3.2.1. To obtain an analytical expression, we use the large-k approximation in Eq. (3.19) to obtain

$$\tau_{\rm tot} = 2D_\alpha k^{\frac{1}{\alpha}} \tau_{\rm on}^{\rm weak} + \frac{\tau_{\rm iso}^{\rm weak}}{k^2}, \tag{3.27}$$

where $D_\alpha = \pi(\alpha - 1)\sin^{-1}(\pi/\alpha)$. This expression can be minimized with respect to k, yielding

$$k = \left(\frac{\alpha \tau_{\rm iso}^{\rm weak}}{D_\alpha \tau_{\rm on}^{\rm weak}}\right)^{\frac{\alpha}{2\alpha+1}}, \tag{3.28}$$

and the corresponding minimal time

$$\tau_{\rm tot} = C_\alpha \tau_{\rm iso}^{\rm weak} \left(\frac{\tau_{\rm on}^{\rm weak}}{\tau_{\rm iso}^{\rm weak}}\right)^{\frac{2\alpha}{2\alpha+1}}$$

where C_α is the constant

$$C_\alpha = (2\alpha+1)\left(\frac{D_\alpha}{\alpha}\right)^{\frac{2\alpha}{2\alpha+1}}. \tag{3.29}$$

For a standard isothermal process at $k = 1$, in which the interaction is not modified, at leading order in $1/\tau_{\rm iso}^{\rm weak}$ the dissipated work can be expressed as [145, 162, 167]

$$W_{\rm diss}^{\rm weak} = \frac{\Sigma}{\tau_{\rm iso}^{\rm weak}} \tag{3.30}$$

where $\Sigma > 0$ can be obtained from the integral expression in Eq. (3.7). By construction, the family of protocols in Eq. (3.29) will dissipate the same $W_{\rm diss} = W_{\rm diss}^{\rm weak}$. If we combine this observation with Eq. (3.29) and (3.30), we obtain that

$$W_{\rm diss} = \Sigma C_\alpha^{2\alpha+1}\frac{(\tau_{\rm on}^{\rm weak})^{2\alpha}}{\tau_{\rm tot}^{2\alpha+1}}, \tag{3.31}$$

with $\alpha > 0$. For constant $\tau_{\rm on}^{\rm weak}$, the dissipation decays as $\tau_{\rm tot}^{-(2\alpha+1)}$ in the total time $\tau_{\rm tot}$ of the process, which can greatly outperform the standard decay in Eq. (3.30).

Naively, the decay in Eq. (3.31) may suggest that one can make the dissipation arbitrary small simply by increasing α. This is not the case, however, due to the contribution of the constant C_α, which diverges exponentially as α increases. As a consequence one can show that for any τ there exists an optimal α, which scales logarithmically in τ. Hence, one needs exponentially long protocols in order to choose larger α.

Next, we discuss the case where $\tau_{\rm on}$ is scaled as in Eq. (3.22), in order to account for contributions to the dissipated work at second order in g. Using the large-k approximation in Eq. (3.22), the total time now reads as

$$\tau_{\rm tot} = 2B_\alpha \tau_{\rm on}^{\rm weak} k^2 + \frac{\tau_{\rm iso}^{\rm weak}}{k^2}, \tag{3.32}$$

where $B_\alpha = \alpha^2/(2\alpha-1)$. Following the same steps as before, we find that the total time is minimized when we choose $k^4 = \tau_{\rm iso}^{\rm weak}/(2B_\alpha\tau_{\rm on}^{\rm weak})$, yielding $\tau_{\rm tot} = \sqrt{8B_\alpha\tau_{\rm iso}^{\rm weak}\tau_{\rm on}^{\rm weak}}$. This leads to a decay of the efficiency given by

$$W_{\rm diss} = \frac{8\alpha^2\Sigma\tau_{\rm on}^{\rm weak}}{(2\alpha-1)\tau_{\rm tot}^2}. \tag{3.33}$$

Therefore, for a fixed $\tau_{\rm on}^{\rm weak}$, the dissipation decays with the total time as $\tau_{\rm tot}^{-2}$, in contrast to the standard decay (3.30).

3.3 Efficiency at maximum power through speed-ups to isothermality

In this section, we study the implications of optimal shortcuts to isothermality for thermodynamic cycles. In Sec. 3.2, we presented different possible choices for

speeding up an isothermal process. We discussed two possible scenarios arising from perturbation theory around weak coupling. Here, we will use a general form for the decay of the dissipated work which encompasses all regimes considered in Sec. 3.2.

Let us assume that the dissipation decays as

$$W_{\rm diss} = \frac{\Sigma_\gamma}{\tau_{\rm tot}^\gamma} \tag{3.34}$$

with $\gamma \geq 1$ and $\tau_{\rm tot}$ is the time of the process. For the optimal shortcuts to isothermality, we have that $\gamma = 2\alpha + 1$ with $\alpha > 0$ and $\Sigma_\gamma = \Sigma C_\alpha^{2\alpha+1} (\tau_{\rm on}^{\rm weak})^{2\alpha}$, whereas for the more conservative choice in Eq. (3.22) we have $\gamma = 2$ and Σ_γ given in Eq. (3.33).

We consider a finite-time Carnot-like cycle between two thermal baths at different temperatures T_h and T_c [145, 146, 167]. Furthermore, when the (finite-time) isothermal part of the cycle is carried out, we assume a decay as in Eq. (3.34). Using $Q + W = \Delta E_S$, the heat exchanged between the system and each of the two thermal baths reads

$$\begin{aligned} Q_c &= T_c \left(-\Delta S - \frac{\Sigma_\gamma}{\tau_c^\gamma} + ... \right) \\ Q_h &= T_h \left(\Delta S - \frac{\Sigma_\gamma}{\tau_h^\gamma} + ... \right) \end{aligned} \tag{3.35}$$

where $\tau_{c,h}$ are the times of the isothermal processes (with the cold, hot bath, respectively), and we have assumed a symmetric cycle such that the constants Σ_γ are equal for each isothermal process [145, 146, 167]. The efficiency of the engine is given by,

$$\eta = 1 + \frac{Q_c}{Q_h} \tag{3.36}$$

whereas the power reads,

$$P = \frac{Q_h + Q_c}{\tau_h + \tau_c}. \tag{3.37}$$

In the case of $\gamma = 1$, i.e. $W_{\rm diss} \propto \tau_{\rm tot}^{-1}$, the efficiency at maximum power η^* is given by the Curzon-Ahlborn efficiency [145]. We want to compute η^* for a generic value of γ. The maximum power is obtained by imposing the two conditions:

$$\frac{\partial P}{\partial \tau_c} = 0, \quad \frac{\partial P}{\partial \tau_h} = 0. \tag{3.38}$$

The system has a unique real and positive solution for $\tau_{c,h}$ given by:

$$\begin{aligned} \tau_c &= \frac{\theta^{\frac{1}{\gamma+1}}}{\theta^{\frac{1}{\gamma+1}} + 1} \left(\frac{\Sigma_\gamma (\gamma+1) \theta \left(\theta^{-\frac{1}{\gamma+1}} + 1 \right)^{\gamma+1}}{\Delta S (1-\theta)} \right)^{\frac{1}{\gamma}} \\ \tau_h &= \theta^{\frac{1}{\gamma+1}} T_c, \end{aligned} \tag{3.39}$$

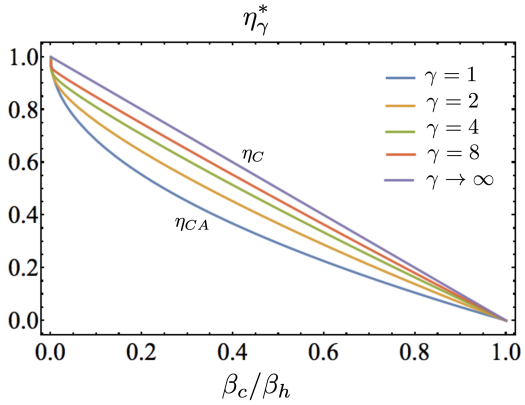


Figure 3.3: **Efficiency at maximum power.** The figure shows η^*_γ for $\gamma = 1, 2, 4, 8, \infty$; with $\gamma = 1$ and $\gamma = \infty$ corresponding to Curzon-Ahlborn (η_{CA}) and Carnot efficiency (η_C), respectively.

where we used the notations $\theta := T_c/T_h$. The corresponding efficiency at maximum power reads

$$\eta^*_\gamma = 1 - \frac{(\gamma\theta + 1)\theta^{\frac{1}{\gamma+1}} + (\gamma + 1)\theta}{(\gamma + 1)\theta^{\frac{1}{\gamma+1}} + \gamma + \theta} \tag{3.40}$$

which depends only on the ratio of temperatures θ and γ. This formula has two interesting limits: for $\gamma \to 1$, one obtains the Curzon-Ahlborn efficiency $\eta^*_1 = 1 - \sqrt{\theta} \equiv \eta_{CA}$, while for $\gamma \to \infty$ we regain the Carnot efficiency $\eta^*_\infty = 1 - \theta \equiv \eta_C$. The efficiency at maximum power interpolates between these two regimes as γ varies, as illustrated in Fig. 3.3. If we expand Eq. (3.40) around $\theta \to 1$ (i.e. $\eta_C \to 0$), we obtain

$$\eta^*_\gamma = \frac{\gamma}{\gamma + 1}\eta_C + \frac{\gamma}{2(\gamma + 1)^2}\eta_C^2 + \mathcal{O}(\eta_C^3). \tag{3.41}$$

The expansion in Eq. (3.41) neatly shows how η^*_γ approaches η_C as γ increases. Notice that for the optimal STI that we defined in the previous section, the time of the process in Eq. (3.39) is proportional to C_γ, and hence tends to infinity as $\eta^*_\gamma \to \eta_C$, preventing the possibility of achieving a Carnot cycle with finite power.

3.4 Numerical results

In the previous sections we have combined heuristic and rigorous arguments to show that the time of an isothermal process in Eq. (3.26) can be considerably reduced by suitably modifying the coupling between system and bath. The goal of this section is to illustrate these considerations for exactly solvable models. Specifically, we consider two complementary examples: a bosonic environment described by the Caldeira-Leggett model and a fermionic bath described by the resonant-level model. The quadratic nature of their corresponding Hamiltonians allows us to efficiently simulate large or infinite bath sizes such that our statements about thermalization remain meaningful.

With the Caldeira-Leggett model, we study a problem with bosonic degrees of freedom using exact calculations but with a finite, discretized bath. In this context, we quantitatively demonstrate that the heuristic assumptions underlying our analytical results hold to an excellent approximation, even with relatively fast driving. Then we move to a resonant-level model with fermionic degrees of freedom, which is analysed using an approximate analytical approach. This allows us to study the slow-driving regime in order to demonstrate a genuine speed-up for isothermal processes at strong coupling.

3.4.1 The Caldeira-Leggett model

We start by illustrating our results with the Caldeira-Leggett (CL) model [151, 152, 166], prototypical example of a quantum Brownian motion. The CL model describes a Brownian quantum particle of mass m in a harmonic potential. The full Hamiltonian consists of four terms,

$$H = H^{(S)} + H^{(B)} + V + H^{(R)}. \tag{3.42}$$

where the Hamiltonian of the system S reads

$$H^{(S)} = \frac{1}{2}\left(m\omega_S^2 x^2 + \frac{p^2}{m}\right) \tag{3.43}$$

where x and p are the position and momentum operators; the Hamiltonian of the bath B is

$$H^{(B)} = \frac{1}{2}\sum_{n=0}^{N}\left(\frac{p_n^2}{m_n} + m_n\omega_n^2 x_n^2\right), \tag{3.44}$$

where $\omega_n = \frac{n}{N}(\omega_{\max} - \omega_{\min}) + \omega_{\min}$ are the frequencies of the modes in the bath, and we defined $\omega_{\max} = 2\omega_S$ and $\omega_{\min} = \omega_S/N$. The interaction V between the system and the bath is defined as

$$V = x\sum_n \gamma_n x_n, \tag{3.45}$$

where γ_n are the coupling constant between system and bath. In the remainder we will assume all the masses $m, m_n = 1$, and that the couplings satisfy: $\gamma_n = g\omega_n\sqrt{2\omega_{\max}/(\pi N\omega_S)}$, which tends to an Ohmic spectral density with hard cutoffs in the continuum limit ($N \to \infty$). The last term $H^{(R)}$ in Eq. (3.42) is a renormalization term which ensures the positivity of H,

$$H^{(R)} = x^2\sum_n \frac{\gamma_n^2}{m_n\omega_n^2}, \tag{3.46}$$

which may be absorbed within $H^{(S)}$.

The CL Hamiltonian in Eq. (3.42) is quadratic. This enables us to diagonalize it efficiently and to describe the time-evolved state by covariance matrices (of size $2N \times 2N$ for systems composed of N particles), allowing us to reach large but finite baths. Thus, the dynamics induced by the CL Hamiltonian in Eq. (3.42) can be simulated without making any assumption on the coupling strength g (see e.g. [168] for details).

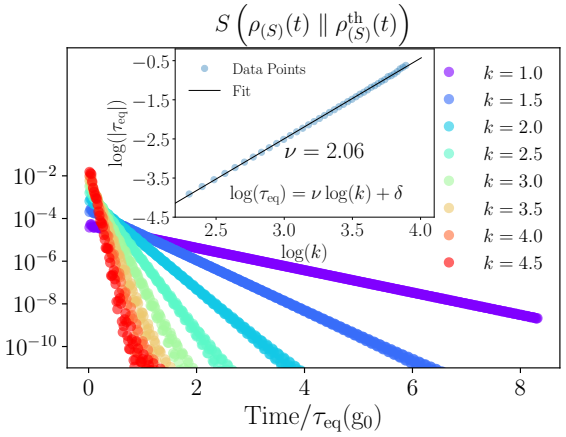


Figure 3.4: **Relaxation to the thermal state.** We compute the relative entropy $S(\rho_{(S)}(t) \parallel \rho^{\mathrm{th}}_{(S)}(t))$ between the reduced state $\rho_{(S)}(t) = \mathrm{Tr}_B(\rho(t))$ of the total time-dependent density matrix and thermal state of the system $\rho^{\mathrm{th}}_{(S)}(t)$. For a wide range of values of k the time evolved state approaches the thermal equilibrium exponentially $S(\rho_{(S)}(t) \parallel \rho^{\mathrm{th}}_{(S)}(t)) \sim e^{-t/\tau_{\mathrm{eq}}(g)}$. **Inset:** We extrapolate the decay of the relaxation time with a power law $\tau_{\mathrm{eq}}(g) \sim \tau_{\mathrm{eq}}(g_0)k^{-\nu}$. The optimal parameters correspond to $\nu = 2.06$ and $\tau_{\mathrm{eq}}(g_0) = 58.5$.

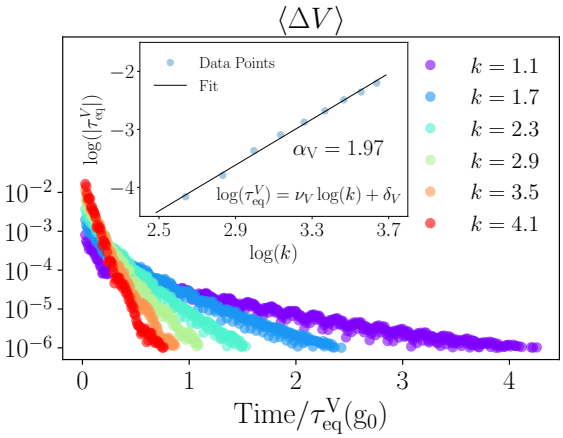


Figure 3.5: **Thermalization of the potential.** Moving average of the expectation value of the potential as a function of time for different coupling strengths. The value of $\langle \Delta V(t) \rangle$ decays exponentially as a function of time. The slope becomes increasingly steeper for stronger coupling $g_{\max}(k)$, i.e. the thermalization is faster. **Inset:** We extrapolate the decay of the relaxation time with a power law $\log(\tau^V_{\mathrm{eq}}(g)) = \log(\tau^V_{\mathrm{eq}}(g_0)) - \nu_V \log(k)$. The optimal parameters correspond to $\nu_V = 1.97$ and $\tau^V_{\mathrm{eq}}(g_0) = 113$.

Thermalization in the CL model.

We first study the dependence of the thermalization time on g for observables on the system. In the simulation, we take as an initial state the thermal state of the non-interacting Hamiltonian $\rho_\beta^{\text{th}}(H(t=0)) = \rho_\beta^{\text{th}}(H_S) \otimes \rho_\beta^{\text{th}}(H_B)$, and then perform a quench to a finite interaction strength $g = kg_0$, and consider the corresponding relaxation to the new equilibrium state (the number of particles in the bath is $N = 300$, and $g_0 = 0.1 \cdot \omega_S$).

In particular, in Fig. 3.4, we compute the relative entropy $S(\rho \parallel \sigma) = \text{Tr}\left(\rho\left(\log\rho - \log\sigma\right)\right)$ between the reduced of the time evolved state $\rho_{(S)}(t) = \text{Tr}_B(\rho(t))$ and the thermal state of the system $\rho_{(S)}^{\text{th}}(t)$. The relative entropy decays exponentially in time

$$S(\rho_{(S)}(t) \parallel \rho_{(S)}^{\text{th}}(t)) \sim e^{-t/\tau_{\text{eq}}(g)}, \tag{3.47}$$

where $\tau_{\text{eq}}(g)$ is the relaxation time-scale for a given coupling strength g. As we expect, the slope becomes increasingly steeper for stronger couplings. In order to understand the behaviour of $\tau_{\text{eq}}(g)$ as a function of g, we assume a power law decay $\tau_{\text{eq}}(g) = \tau_{\text{eq}}(g_0)k^{-\nu}$, where $\tau_{\text{eq}}(g_0)$ corresponds to the relaxation time for $k = 1$ and ν quantifies the scaling with interaction strength. In Fig. 3.4 (inset) we fit the function $\log(\tau_{\text{eq}}(g)) = \log(\tau_{\text{eq}}(g_0)) - \nu\log(k)$ with a straight line, which confirms the scaling predicted by Eq. (3.11) with $\nu \approx 2$ even for rather large coupling strengths up to $g/\omega_S \approx 0.5$.

Similarly, we need to verify that the interaction energy thermalizes and satisfies Eq. (3.11). This is shown in Fig. 3.5, where we plot $\Delta V = \langle V(t)\rangle - \langle V_{\text{eq}}\rangle$ for different values of the coupling $g_{\max}(k)$, and where $\langle V(t)\rangle$ is the exact value of the interaction energy for the unitary-evolved state and V_{eq} is its thermal equilibrium value (with respect to the global thermal state). By performing a extrapolation as the one of Fig. 3.5, we confirm the scaling in Eq. (3.16) up to the relatively large interaction strength of $g/\omega_S \approx 0.5$.

Generalised covariance.

In Fig. 3.6 we show the behavior of the covariance from Eq. (3.9) for the relevant quantities $H^{(S)}$ and V as a function of the interaction strength g. One observes that $\text{cov}(H^{(S)}, H^{(S)})$ stays essentially constant, which means that only $c^{(0)}$ in (3.10) contributes, hence also justifying (3.25). On the other hand, $\text{cov}(V, V)$ does vary with g, suggesting that higher order terms in the expansion in (3.10) can play a role.

Shortcuts to isothermality.

In order to confirm the intuition given by the generalised covariance, as a last step we simulate the full thermodynamic protocol and compute the associated dissipation. First, in Fig. 3.7 (a) we show the total dissipation W_{diss} for increasing and decreasing the interaction between system and bath according to (3.23) (in order to account for higher order corrections in g). One observes that W_{diss} either decreases or stays constant with k, as expected from our analytic reasoning. Furthermore, in

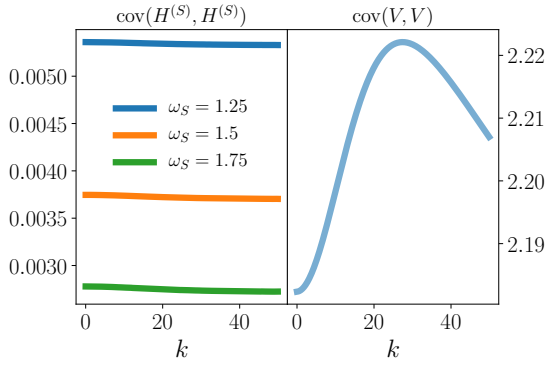


Figure 3.6: **Bounds on the dissipated work: covariance matrices.** is the In Eq. (3.9) and (3.7) we bounded the dissipated work by employing the covariance $\mathrm{cov}_{\omega_t}\left(\tilde{H}, \tilde{H}\right)$. Notice that the covariance of H_s and V stays bounded for very large values of k.

Fig. 3.7 (b) we show the total dissipated work for the full thermodynamic protocol depicted in Fig. 3.1 and with (3.13). As we increase the interaction strength, the dissipation remains constant or drops close to zero, and as shown in Fig. 3.2 the time substantially decreases. Hence, we have obtained the desired speed-ups. According to the general considerations of Sec. 3.2, we vary the time of the isotherm as $\tau_{\mathrm{iso}} \propto 1/k^2$, while the system-bath coupling is changed as $\tau_{\mathrm{on}} \propto k^2$. These findings are in good agreement with the analytical results in the previous sections. Notice that the times shown in Fig. 3.7 are comparable to the thermalization times in Fig. 3.4 and 3.5.

3.4.2 The resonant-level model

In this section, we benchmark our predictions using the analytically tractable resonant-level model. Specifically, the system of interest comprises a single distinguished fermionic mode coupled to an infinite collection of reservoir modes, also fermionic. The total system-bath Hamiltonian reads as $H = H^{(S)} + H^{(B)} + H^{(SB)}$, with

$$H^{(S)} = \varepsilon(t) a^\dagger a, \tag{3.48}$$

$$H^{(B)} = \sum_k \omega_k b_k^\dagger b_k, \tag{3.49}$$

$$H^{(SB)} = g(t) \sum_k \gamma_k \left(a^\dagger b_k + b_k^\dagger a \right). \tag{3.50}$$

Here, a annihilates a fermion with time-dependent energy $\varepsilon(t)$, while b_k annihilates a fermion in the bath with energy ω_k. The relevant bath properties are characterised by the spectral density $\mathfrak{J}(\omega) = 2\pi \sum_k \gamma_k^2 \delta(\omega - \omega_k)$, where we take

$$\mathfrak{J}(\omega) = \Gamma \Theta(\Lambda - |\omega|), \tag{3.51}$$

with Γ a characteristic dissipation rate, Λ a high-frequency cutoff and $\Theta(z)$ the Heaviside step function.

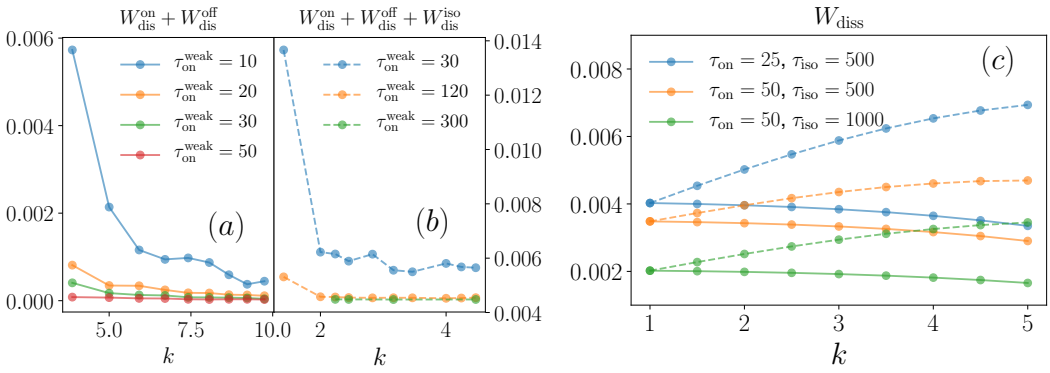


Figure 3.7: **(a) Dissipated Work for turning on and off the interaction.** For different values of $\tau_{\rm on}^{\rm weak}$ the error induced by going to the strong coupling regime decreases to zero ($N = 300$, $g_0 = 0.1 \cdot \omega_S$, $k \in [0, 10]$, $\beta = 1.2$, and $\omega_S = 1$). **(b) Dissipated Work for full protocols.** The total dissipation decays as the interaction increases. For different values of $\tau_{\rm on}^{\rm weak}$, the error induced by going to the stronger coupling regime decreases to zero. The protocol consists on modifying the frequency of the systems from $\omega_S^i = \omega_S$ to $\omega_S^f = 2\omega_S$ ($N = 300$, and $g_0 = 0.1 \cdot \omega_S$, $k \in [0, 5]$, $\tau_{\rm iso}^{\rm weak} = 50$, $\beta = 1.2$, and $\omega_S = 1$). **(c) Dissipated work $W_{\rm diss}$ over the entire protocol for various switching times and values of k.** As a function of k, the isotherm time is reduced as $\tau_{\rm iso} = \tau_{\rm iso}^{\rm weak}/k^2$, while the switch-on and switch-off times are increased either as $\tau_{\rm on} = k\tau_{\rm on}^{\rm weak}$ (dashed lines) or $\tau_{\rm on} = k^2\tau_{\rm on}^{\rm weak}$ (solid lines). The other parameters are $\Gamma = 0.1$, $\varepsilon_i = 1$, $\varepsilon_f = 2$, $\beta = 1$ and $\Lambda = 100$. Times are shown in units of the frequency of the system: ω_S and ε_i, respectively.

Solution for the dynamics

Exact solutions for the resonant-level model have recently been presented in the context of a debate regarding heat in strongly coupled open quantum systems, with particular emphasis on the wide-band limit $\Lambda \to \infty$ [153–157]. Note, however, that the system-bath interaction energy is proportional to Λ, and thus formally divergent in this limit (this can be seen easily using the reaction-coordinate representation [169], for example). We thus take Λ to be finite but much larger than all other energy scales.

Under this assumption, we use a quantum Langevin approach to solve for the open-system evolution, detailed in Sec. 3.6. Our approximate analysis requires that the dynamics proceeds much more slowly than the inverse cutoff scale Λ^{-1}, but otherwise allows for arbitrary driving protocols and strong system-bath coupling. Taking a factorised system-bath density matrix at the initial time, $\rho(0) = \rho^{(S)}(0)\omega_\beta(H^{(B)})$, we find the level occupation $n(t) = \langle a^\dagger a\rangle$ and the system-bath correlations $v(t) = \sum_k \gamma_k \langle a^\dagger b_k + \text{h.c.}\rangle$ to be given by

$$n(t) = \frac{1}{2} + |K(t,0)|^2 n(0) \tag{3.52}$$
$$- \frac{1}{2}\int_0^t \mathrm{d}s \int_0^t \mathrm{d}s'\, K(t,s)g(s)\phi(s-s')g(s')K^*(t,s'),$$
$$v(t) = \mathrm{Im}\int_0^t \mathrm{d}s\, K^*(t,s)\phi(t-s)g(s). \tag{3.53}$$

These expressions are written in terms of the propagator

$$K(t,t') = \exp\left[\int_{t'}^t \mathrm{d}s \left(-\mathrm{i}\varepsilon(s) - \frac{\Gamma}{2}g(s)^2\right)\right], \tag{3.54}$$

and the noise correlation function

$$\phi(t) = \frac{\Gamma}{\mathrm{i}\beta}\left[\frac{1}{\sinh(\pi t/\beta)} - \frac{\cos(\Lambda t)}{\pi t/\beta}\right]. \tag{3.55}$$

Note that the second, cutoff-dependent term is essential to regulate the divergence of the integrand in Eq. (3.53) as $s \to t$, but plays essentially no role in Eq. (3.52) for large Λ.

It follows immediately from Eq. (3.54) that the relaxation timescale is given by $\tau_{\text{eq}} \sim 1/\Gamma g^2$, in agreement with Eq. (3.11), even though the evolution is non-Markovian, in general. Note also that, since Eq. (3.53) contains one propagator while Eq. (3.52) includes two, the relaxation timescale of the interaction energy is twice as long as that of $n(t)$. This is in accordance with the relaxation behaviour of the CL model shown in Figs. 3.4 and 3.5.

Dissipated work

We now compute the dissipated work during an isothermal protocol, during which the level energy $\varepsilon(t)$ is linearly ramped from an initial to a final value $\varepsilon_i \to \varepsilon_f$, while interacting with the bath. At the start and end of the protocol, the system-bath

interaction energy is switched on according to Eq. (3.14) with $\alpha = 1$, $g_0 = 0$ and $g_1 = k$, and switched off via the reverse procedure. Note that here, unlike in Sec. 3.2, we completely decouple the system from the bath at the initial and final time. Nevertheless, we find that the same scaling of dissipation with the final coupling strength $g_1 \propto k$.

The dissipated work is given by $W_{\text{diss}} = W - \Delta F$, where W is found from Eqs. (3.3), (3.52) and (3.53), while $\Delta F = \Delta\varepsilon + \beta^{-1} \ln[f(\varepsilon_f)/f(\varepsilon_i)]$, with $f(\omega) = (\mathrm{e}^{\beta\omega} + 1)^{-1}$ the Fermi-Dirac distribution. In Fig. 3.7 (c), we plot W_{diss} for several different parameters as a function of k. We see that the dissipation grows sublinearly with k for $\tau_{\text{on}} = k\tau_{\text{on}}^{\text{weak}}$, while the dissipation strictly decreases for the more conservative choice of $\tau_{\text{on}} = k^2\tau_{\text{on}}^{\text{weak}}$. This confirms that control over the system-bath interaction can indeed reduce the time taken by an isothermal process without incurring additional dissipation (c.f. Fig. 3.2 showing the time of the isothermal process).

3.5 Extensions to more general dissipations

In this section we consider the case in which the dissipation is of the form as in Eq. (3.6):

$$W_{\text{diss}} = \frac{1}{\tau} \int_0^1 dt\, G_{\rho_t^{\text{th}}}\Big(\dot{H}_t, \dot{H}_t\Big) + \mathcal{O}\left(\frac{\tau^2}{\tau^2}\right), \tag{3.56}$$

where $G_{\rho_t^{\text{th}}}$ is a bilinear form which depends only on the base point ρ_t^{th}. This expression is generic, and it arises in the expansion of the entropy production rate $\dot{\sigma}_t$ in the quasi-static limit [162, 164, 165]:

$$\dot{\sigma}_t = G_{\rho_t^{\text{th}}}\Big(\dot{H}_t, \dot{H}_t\Big) + \mathcal{O}\left(\frac{\tau^2}{\tau^2}\right). \tag{3.57}$$

In particular, if the dynamics is described by the time dependent Liouvillian equation:

$$\rho = L_t[\rho], \tag{3.58}$$

where L_t has for every t only one thermal steady state and, moreover, the real part of all its eigenvalues is negative (this two conditions are sufficient to ensure thermalisation), then the integrand in Eq. (3.56) is given at first order by [162]:

$$G_{\rho_t^{\text{th}}}\Big(\dot{H}_t, \dot{H}_t\Big) = -\beta\ \mathrm{Tr}\left[\dot{H}_t L_t^{+}\Big[\mathbb{J}_{\rho_\beta^{\text{th}}(H_t)}[\dot{H}_t]\Big]\right], \tag{3.59}$$

where we defined the two operators:

$$\mathbb{J}_\rho[A] := \int_0^1 \mathrm{d}s\ \rho^{1-s}\,(A - \mathrm{Tr}[\rho A]\mathbb{1})\,\rho^s, \tag{3.60}$$

$$L_t^{+}[A] := \int_0^\infty \mathrm{d}\nu\ e^{\nu L_t}\big(\rho_\beta^{\text{th}}(H_t)\,\mathrm{Tr}\,[A] - A\big). \tag{3.61}$$

Before going on, it should be noticed that the operator $\mathbb{J}_\omega$ is related to the generalised covariance through the equality:

$$\mathrm{cov}_\rho\left(A, B\right) = \mathrm{Tr}\left[A\, \mathbb{J}_\rho[B]\right]. \tag{3.62}$$

Moreover, carrying out the integral in Eq. (3.61) in the eigenbasis of L_t shows that the eigenvalues of L^+ are directly connected with the different thermalisation time-scales in the system. In particular, in the case in which all the observables thermalises at the same rate, Eq. (3.59) reduces to Eq. (3.7).

Considering again the simplified case in which the derivative of the Hamiltonian is given by $\dot{H} = \dot{\lambda}_t X$, we have the chain of inequalities:

$$\begin{aligned} |W_{\mathrm{diss}}| &= \frac{\beta}{\tau}\left| \int_0^1 dt\, \dot{\lambda}_t^2 \;\; \mathrm{Tr}\left[X L_t^+\left[\mathbb{J}_{\rho_\beta^{\mathrm{th}}(H_t)}[X]\right]\right]\right| \leq \\ &\leq \frac{\beta}{\tau} \sup_{t\in[0,1]} \mathrm{cov}_{\rho_t^{\mathrm{th}}}(X,X) \int_0^1 dt\, \dot{\lambda}_t^2 \tau_{g(t)}^{\max}, \end{aligned} \tag{3.63}$$

where we indicate with $\tau_{g(t)}^{\max}$ the biggest eigenvalue of L_t^+. Since during the turning on and off procedure we want to keep track of the dependence of the thermalisation time-scale on the interaction strength, we keep this term inside the integral. This expression should be compared with Eq. (3.15) and (3.17) above.

As a final remark, the bound in Eq. (3.17) on the covariance can be improved to [170]:

$$\sup_{t\in[0,1]} \mathrm{cov}_{\rho_t^{\mathrm{th}}}(F,F) \leq 2 \sup_{t\in[0,1]} \left(\left\langle V^2\right\rangle_{\rho_t^{\mathrm{th}}} - \left\langle V\right\rangle_{\rho_t^{\mathrm{th}}}^2\right). \tag{3.64}$$

This quantity is expected to be finite even in the limit in which $||V|| \to \infty$.

3.6 Solution of the resonant-level model

In this section, we detail our approach to solve the resonant-level model described in Sec. 3.4.2. Starting from the Hamiltonian given in Eqs. (3.48)–(3.50), we derive the Heisenberg equations

$$\dot{a}(t) = -\mathrm{i}\varepsilon(t)a(t) - \mathrm{i}g(t)\sum_k \gamma_k b_k(t), \tag{3.65}$$

$$\dot{b}_k(t) = -\mathrm{i}\omega_k b_k(t) - \mathrm{i}g(t)\gamma_k a(t). \tag{3.66}$$

The second equation can be formally solved to give

$$\sum_k \gamma_k b_k(t) = \xi(t) - \mathrm{i}\int_{t_0}^t \mathrm{d}t'\, \chi(t-t')g(t')a(t'), \tag{3.67}$$

where we defined the noise operator

$$\xi(t) = \sum_k \gamma_k \mathrm{e}^{-\mathrm{i}\omega_k(t-t_0)} b_k(t_0), \tag{3.68}$$

whose Gaussian statistics with respect to the initial state define the memory kernel $\chi(t-t') = \langle\{\xi(t), \xi^\dagger(t')\}\rangle$ and the noise correlation function $\phi(t-t') = \langle[\xi(t), \xi^\dagger(t')]\rangle$. These are given explicitly by

$$\chi(t) = \int \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{-\mathrm{i}\omega t}\mathfrak{J}(\omega), \tag{3.69}$$

$$\phi(t) = \int \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{-\mathrm{i}\omega t}\mathfrak{J}(\omega)\tanh[\beta(\omega-\mu)/2], \tag{3.70}$$

where the spectral density is defined by Eq. (3.51). Note that here, for completeness, we allow for a finite chemical potential μ. In the wide-band limit $\Lambda \to \infty$, the chemical potential can be set to zero without loss of generality by simply redefining all energies relative to μ, which justifies our choice of $\mu = 0$ in the main text.

To obtain a tractable description, we approximate the memory kernel as

$$\chi(t) = \frac{\Gamma \sin(\Lambda t)}{\pi t} \approx \Gamma\delta(t). \tag{3.71}$$

This is an exact equality (in the distributional sense) in the limit $\Lambda \to \infty$, and is a good approximation for finite Λ so long as slowly varying functions and large times relative to the cut-off scale Λ^{-1} are considered. The noise correlation function is approximated as

$$\begin{aligned}
\phi(t) &= \Gamma \int_{-\Lambda}^{\Lambda} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{-\mathrm{i}\omega t}\tanh[\beta(\omega-\mu)/2] \\
&\approx \Gamma \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\, \mathrm{e}^{-\mathrm{i}\omega t}\tanh[\beta(\omega-\mu)/2] \\
&\quad - \Gamma \int_{\Lambda}^{\infty} \frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{-\mathrm{i}\omega t} + \Gamma \int_{-\infty}^{-\Lambda} \frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{-\mathrm{i}\omega t} \\
&= \frac{\Gamma}{\mathrm{i}\beta}\left[\frac{\mathrm{e}^{-\mathrm{i}\mu t}}{\sinh(\pi t/\beta)} - \frac{\cos(\Lambda t)}{\pi t/\beta}\right].
\end{aligned} \tag{3.72}$$

On the second line, the integration domain is partitioned into three parts, and the approximation $\tanh(z) \approx \pm 1$ for $\pm z \gg 1$ is made. The first integral is essentially the Fourier transform of $\tanh(z)$, which is calculated by a standard contour integration, resulting in a geometric sum over Matsubara frequencies that evaluates to the first term in Eq. (3.72). The remaining two integrals yield the second term in Eq. (3.72) with the help of the Sokhotski–Plemelj theorem. Note that this second term regulates the $1/t$ divergence as $t \to 0$ but is negligible (in the distributional sense) for time scales $t \gg \Lambda^{-1}$. It can be shown that, within these approximations, the fluctuation-dissipation relation $\tilde{\phi}(\omega) = \tilde{\chi}(\omega)\tanh[\beta(\omega-\mu)/2]$ between the Fourier components of the memory kernel $\tilde{\chi}(\omega)$ and the noise spectrum $\tilde{\phi}(\omega)$ holds for all $|\omega| < \Lambda$.

As a consequence of Eq. (3.71), Eq. (3.65) reduces to a time-local differential equation

$$\dot{a}(t) = \left(-\mathrm{i}\varepsilon(t)a(t) - \frac{\Gamma}{2}g(t)^2\right)a(t) - \mathrm{i}g(t)\xi(t), \tag{3.73}$$

which can be easily solved to find

$$a(t) = K(t, t_0)a(t_0) - \mathrm{i}\int_{t_0}^{t} \mathrm{d}t'\, K(t, t')g(t')\hat{\xi}(t'), \tag{3.74}$$

where the propagator is given by Eq. (3.54). Combining this with Eqs. (3.67), (3.71) and (3.72), and the fact that $\langle a^\dagger(t_0)b(t_0)\rangle = 0$ for a factorized initial condition at $t_0 = 0$, we deduce Eqs. (3.52) and (3.53).

Our analysis relies on two approximations, given by Eqs. (3.71) and (3.72). The former assumes that the dynamics is much slower than Λ^{-1}, while the latter requires that the temperature and chemical potential are much smaller than Λ. In particular, we require that $\varepsilon(t), \Gamma g(t)^2, \beta^{-1}, |\mu|$ and $|\mu \pm \beta^{-1}|$ are all much smaller than Λ.

3.7 Conclusions

We have put forward the idea of a speed-up to isothermality (STI), where an isothermal process is sped up by smoothly increasing (and decreasing) the system-bath interaction. This leads to faster isothermal processes while keeping the overall thermodynamic dissipation constant. As a consequence, our proposal allows for increasing the power of a finite-time Carnot cycles [145, 146, 167] without compromising their efficiency.

The proposed STI are based upon two main assumptions:

1. We work in the slow-driving (or adiabatic) regime, allowing for an expansion of the dissipation as in Eq. (3.6).

2. We assume that the time-scale of thermalization satisfies $\tau_{\mathrm{eq}} \propto g^{-2}$, where g quantifies the strength of the system-bath interaction. This behaviour, which is expected in Markovian open system dynamics [166], has been confirmed in our models of interest (bosonic and fermionic bath) even for reasonably large coupling strengths.

These assumptions hold in a wide range of relevant open quantum systems. Furthermore, we have also argued that some of our ideas can be extended beyond weak system-reservoir coupling to strongly correlated many-body systems that thermalize [150] (see, for example, the discussion after Eq. (3.24)). In this case, the relation $\tau_{\mathrm{eq}} \propto g^{-2}$ might change, but it is worth pointing out that the developed framework is rather flexible and that a different dependence can be easily accounted for. This will naturally change the quantitative scaling of the speed-ups but not the essence of the method, namely to speed-up isothermal processes without increasing the dissipation.

Under these assumptions, we have shown that STI can decrease the time of a given isothermal process by several orders of magnitude, see Eq. (3.26) and Fig. 3.2. This leads to faster decays of the dissipation with time (Sec. 3.2.4) and higher efficiencies at maximum power of finite-time Carnot engines (Sec. 3.3). These new thermodynamic behaviours have been illustrated for a fermionic and bosonic

baths, where we computed exactly the system-bath evolution at strong coupling in order to account for all sources of dissipation, and to control all the different assumptions underlying STI.

An interesting future direction is to combine these ideas with open systems techniques to deal with strong, time-dependent coupling such as the reaction-coordinate mapping [33, 36, 169, 171–173], or more sophisticated tensor-network methods [174–176]. Another interesting direction is to characterise the work fluctuations due to such STI, which have been characterised in e.g. STA [177], and other tradeoffs between thermodynamic cost and time enhancements [178, 179]. Indeed, because STIs allow for accessing larger energy scales, one expects that they shall generate higher work fluctuations [177]. In this sense, we note that one expects a competing effect in the work fluctuations generated by a STI: because we are accessing stronger coupling and hence larger energy scales, one expects stronger fluctuations; however, for a fixed time, STI allow for decreasing dissipation, and in the quasistatic regime the minimisation of dissipation comes together with the minimisation of fluctuations, at least for commuting protocols [163]. A further interesting possibility is to combine these considerations with geometric optimal paths [38, 161–163].

Part IV

Quantum Machine Learning

Chapter 4

Machine Learning for Quantum Systems

Most of the content of this chapter is contained in:

- *"Neural-Network Quantum States, String-Bond States, and Chiral Topological States"*
  Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, and J. Ignacio Cirac
  PRX 8, 011006, (2018)

In this chapter, we study the low temperature physics of interacting quantum systems by employing ideas from machine learning. An idea which received a lot of attention from the scientific community consists in using neural networks as variational wave functions to approximate ground states of many-body quantum systems[180]. These networks are trained/optimized by the standard Variational Monte Carlo (VMC) method and while a few different neural-network architectures have been tested[180–182], the most promising results so far have been achieved with Boltzmann Machines[183]. In particular, state of the art numerical results have been obtained on popular models with Restricted Boltzmann Machines (RBM) and recent effort has demonstrated the power of Deep Boltzmann Machines to represent ground states of many-body Hamiltonians with polynomial-size gap and quantum states generated by any polynomial size quantum circuits[184, 185].

Other seemingly unrelated classes of states that are widely used in condensed matter physics are Tensor Networks States. In 1D, Matrix Product States (MPS) can approximate ground states of physical Hamiltonians efficiently [11, 186] and their structure has led to both analytical insights over the entanglement properties of physical systems as well as efficient variational algorithms for approximating them[187–189]. The natural extension of MPS to larger dimensional systems are Projected Entangled Pair States (PEPS)[190], but their exact contraction is $\#P$ hard[191] and algorithms for optimizing them need to rely on approximations. Another approach to define higher dimensional Tensor Networks consists in first dividing the lattice into overlapping clusters of spins. The wave function of the spins in each cluster is then described by a simple Tensor Network. The global wave function is finally taken to be the product of these Tensor Networks, which

introduces correlations among the different clusters. This construction for local clusters parametrized by a full tensor gives rise to Entangled Plaquette States (EPS)[192–194], while taking one dimensional clusters of spins each described by a MPS leads to a String-Bond States (SBS) Ansatz[195, 196]. These states can be variationally optimized using the VMC method[195, 197] and have been applied to 2D and 3D systems.

All these variational wave functions have been successful in describing strongly correlated quantum many body systems, including topologically ordered states. The Toric code[198] is a prototypical example which can be written exactly as a PEPS[199], an EPS[194], a SBS[195] or a short-range RBM[200]. This shows that in some cases Tensor Networks and Neural-Network Quantum States can be related. Indeed it was recently shown that local Tensor Networks can be represented efficiently by Deep Boltzmann Machines[184, 185, 201]. Not every topological state can however easily be represented by local Tensor Networks. A class of states for which this is challenging are chiral topological states breaking time-reversal symmetry. Such states were first realized in the context of the Fractional Quantum Hall (FQH) effect[202] and significant progress has since been made towards the construction of lattice models displaying the same physics, either in Hamiltonians realizing fractional Chern insulators[203–208] or in quantum anti-ferromagnets on several lattices[209–212]. One approach to describe the wave function of these anti-ferromagnets is to use parton constructed wave functions[213–216]. It has also been suggested to construct chiral lattice wave functions from the FQH continuum wave functions, the paradigmatic example being the Kalmeyer-Laughlin wave function[217]. Efforts to construct chiral topological states with PEPS have been undertaken recently[218–222], but the resulting states are critical. In the non-interacting case it has moreover been proven that the local parent Hamiltonian of a chiral fermionic Gaussian PEPS has to be gapless[219].

In this chapter we show that there is a strong relation between Restricted Boltzmann Machines and Tensor Network States in arbitrary dimension. We demonstrate that short-range RBM are a special subclass of EPS, while fully-connected RBM are a subclass of SBS with a flexible non-local geometry and low bond dimension. This relation provides additional insights over the geometric structure of RBM and their efficiency. We discuss the advantages and drawbacks of RBM and SBS and provide a way to combine them together. This generalization in the form of non-local String-Bond States takes leverage of both the entanglement structure of Tensor Networks and the efficiency of RBM. It allows for the description of states with larger local Hilbert space and has a flexible geometry. It can moreover be combined with more traditional Ansatz wave functions that serve as an initial approximation of the ground state.

We then apply these methods to the challenging problem of approximating chiral topological states. We prove that any Jastrow wave function, and thus the Kalmeyer-Laughlin wave function, can be written exactly as a RBM. We moreover show that a remarkable accuracy can be achieved numerically with much less parameters than is required for an exact construction. We numerically evaluate the power of EPS, SBS and RBM to approximate the ground state of a chiral spin liquid for which the Laughlin state is already a good approximation[209]

and find that RBM and non-local SBS are able to achieve lower energy than the Laughlin wave function. By combining these classes of states with the Laughlin wave function, we are able to reach even lower energies and to characterize the properties of the ground state of the model.

4.1 Variational Monte Carlo with Tensor Networks and Neural-Network States

4.1.1 The Variational Monte Carlo method

Given a general Hamiltonian H, one of the main challenges of quantum many-body physics is to find its ground state $|\psi_0\rangle$ satisfying the Schrödinger equation $H|\psi_0\rangle = E_0|\psi_0\rangle$. This eigenvalue problem can be mapped to an optimization problem through the variational principle, stating that the energy of any quantum state is higher than the energy of the ground state. A general pure quantum state on a lattice with N spins can be expressed in the basis spanned by $|s_1, \ldots, s_N\rangle$, where s_i are the projection of the spins on the z axis, as

$$|\psi\rangle = \sum_{s_1,\ldots,s_N} \psi(s_1, \ldots, s_N)|s_1, \ldots, s_N\rangle. \tag{4.1}$$

Finding the ground state amounts to finding the exponentially many parameters $\psi(s_1, \ldots, s_N)$ minimizing the energy, which can only be done exactly for small sizes. Instead of searching for the ground state in the full Hilbert space, one may restrict the search to an Ansatz class specified by a particular form for the function $\psi_w(s_1, \ldots, s_N)$ depending on polynomially many variational parameters w. The Variational Monte Carlo method [223, 224] (VMC) provides a general algorithm for optimizing the energy of such a wave function. One can compute the energy by expressing it as

$$E_w = \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} = \sum_{\mathbf{s}} p(\mathbf{s}) E_{\text{loc}}(\mathbf{s}), \tag{4.2}$$

where $\mathbf{s} = s_1, \ldots, s_N$ is a spin configuration, $p(\mathbf{s}) = \frac{|\psi_w(\mathbf{s})|^2}{\sum_{\mathbf{s}}|\psi_w(\mathbf{s})|^2}$ is a classical probability distribution and the local energy $E_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'}\langle\mathbf{s}|H|\mathbf{s}'\rangle\frac{\psi_w(\mathbf{s}')}{\psi_w(\mathbf{s})}$ can be evaluated efficiently for Hamiltonians involving few-body interactions. The energy is therefore an expectation value with respect to a probability distribution p that can be evaluated using Markov Chain Monte Carlo sampling techniques such as the Metropolis-Hastings algorithm [225, 226]. The second ingredient required to minimize the energy with respect to the parameters w is the gradient of the energy, which can be expressed in a similar form since

$$\frac{\partial E_w}{\partial w_i} = 2\sum_{\mathbf{s}} p(\mathbf{s})\Delta_{w_i}(\mathbf{s})^* \left(E_{\text{loc}}(\mathbf{s}) - E_w\right), \tag{4.3}$$

where we have defined $\Delta_{w_i}(\mathbf{s}) = \frac{1}{\psi_w(\mathbf{s})}\frac{\partial\psi_w(\mathbf{s})}{\partial w_i}$ as the log-derivative of the wave function with respect to some parameter w_i. This is also an expectation value

with respect to the same probability distribution p and can therefore be sampled at the same time, which allows for the use of gradient-based optimization methods. At each iteration, the energy and its gradient are computed with Monte Carlo, the parameters w are updated by small steps in the direction of negative energy derivative ($w_i \leftarrow w_i - \alpha \frac{\partial E_w}{\partial w_i}$) and the process is repeated until convergence of the energy. The VMC method, in its simplest form, only requires the efficient computation of $\frac{\psi_w(\mathbf{s}')}{\psi_w(\mathbf{s})}$ for two spin configurations s and s', as well as the log-derivative of the wave function $\Delta_w(\mathbf{s})$. More efficient optimization methods can be used, such as conjugate-gradient descent, Stochastic Reconfiguration[227, 228], the Newton method[229] or the linear method[230–232].

At this point one has to choose a special form for the wave function ψ_w. One of the traditional variational wave functions for a many-body quantum system is a Jastrow wave function[223, 233], which consists in its most general form of a product of wave functions for all pairs of spins:

$$\psi_w(\mathbf{s}) = \prod_{i<j} f_{ij}(s_i, s_j), \tag{4.4}$$

where each f_{ij} is fully specified by its four values $f_{ij}(s_i, s_j),\ s_i, s_j \in \{-1, 1\}$. Such an Ansatz does not presuppose a particular local geometry of the many-body quantum state: in general this Ansatz can be non-local due to the correlations between all pairs of spins (Fig. 4.1a). A local structure can be introduced by choosing a form for f_{ij} which decays with the distance between position i and j.

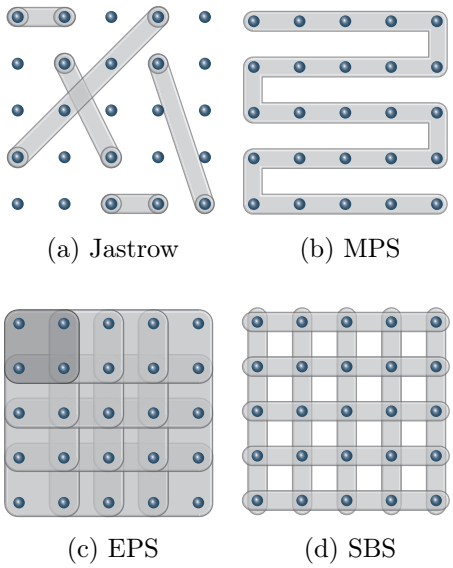
4.1.2 Variational Monte Carlo method with Tensor Networks

In condensed matter physics, important assets to simplify the problem are the geometric structure and locality of physical Hamiltonians. In 1D, it has been proven that ground states of gapped local Hamiltonians have an entanglement entropy of a subsystem which grows only like the boundary of the subsystem[186]. States satisfying such an area-law can be efficiently approximated by Matrix Product States (MPS)[11]. Matrix Product State are one dimensional Tensor Network States whose wave function for a spin configuration reads

$$\psi_w(\mathbf{s}) = \mathrm{Tr}\left(\prod_{j=1}^{N} A_j^{s_j}\right). \tag{4.5}$$

For each spin and lattice site, the matrix $A_i^{s_i}$ of dimension $D \times D$, where D is called the bond dimension, contains the variational parameters. Matrix Product States can be efficiently optimized using the Density Matrix Renormalization Group (DMRG)[234], but the previously described VMC method can also be applied[195, 197] by observing that the ratio of two configurations is straightforward to compute, and that the log-derivative with respect to some matrix $A_k^{s'_k}$ is given by

$$\Delta_{A_k^{s_{k'}}}(\mathbf{s}) = \frac{\delta_{s_k, s_{k'}} \left(A_{k+1}^{s_k} \cdots A_N^{s_N} A_1^{s_1} A_{k-1}^{s_{k-1}}\right)^\top}{\mathrm{Tr}(A_1^{s_1} \cdots A_N^{s_N})}. \tag{4.6}$$



(a) Jastrow (b) MPS

(c) EPS (d) SBS

Figure 4.1: Geometry of Ansatz wave functions: (a) Jastrow wave function include correlations within all pairs of spins. (b) Matrix Product States (MPS) in 2D cover the lattice with one snake. (c) Entangled Plaquette States (EPS) include all spin correlations within each plaquette (2x2 on the figure) and mediate correlations between distant spins through overlapping plaquettes. (d) String-Bond States (SBS) cover the lattice with many 1D strings on which the interactions within spins are captured by a MPS.

In some cases, this method is less likely to be trapped in a local minimum than DMRG, since all coefficients can be updated at once. In addition, the cost only scales as $O(D^3)$ in the bond dimension for periodic boundary conditions.

In higher dimensions, Matrix Product States can be defined by mapping the system to a line (Fig. 4.1b). The problem of this construction is evident from Fig. 4.1b. Spins which sit close to each other might be separated by a long distance on the line, the Ansatz thus fails to reproduce the local structure of the state, which leads to an exponential scaling of the computing resources needed with the system size[235]. The natural extension of MPS to 2D systems are Projected Entangled Pair States (PEPS)[190], for which the wave function can be written as a contraction of local tensors on the 2D lattice. While PEPS have been successful in describing strongly correlated quantum many body systems, their exact contraction is $\#P$ hard[191] and their optimization cannot rely on the standard VMC method without approximations. In the following we will instead consider other classes of tensor-network states in more than one dimension for which the exact computation of the wave function is efficient, which allows for the direct use of the VMC method.

One approach consists in cutting a lattice in P small clusters of n_p spins, or plaquettes, and construct the wave function exactly on each plaquette. The wave function of the full quantum system is then taken to be the product of the wave functions in each plaquette, in a mean-field fashion. Choosing overlapping plaquettes allows one to go beyond mean-field and include correlations between different plaquettes (Fig. 4.1c). The wave function of such an Entangled Plaquette

State (EPS, also called a Correlated Product State) is written as[192–194]:

$$\psi_w(\mathbf{s}) = \prod_{p=1}^{P} C_p^{\mathbf{s}_p}, \tag{4.7}$$

where a coefficient $C_p^{\mathbf{s}_p}$ is assigned to each of the 2^{n_p} (for spin-1/2 particles) configurations $\mathbf{s}_p = s_{a_1}, \ldots, s_{a_{n_p}}$ of the spins on the plaquette p. Each C_p can be seen as the most general function on the Hilbert space corresponding to the spins in plaquette p. The accuracy can be improved by enlarging the size of the plaquettes and the Ansatz is exact once the size of the plaquettes reaches the size of the lattice (which can only be achieved on small lattices). Moreover, once the spin configuration $\mathbf{s}_p$ is fixed, the log-derivative of the wave function with respect to the variational parameters is simply

$$\Delta_{C_p^{\mathbf{s}_p}}(\mathbf{s}) = \frac{1}{C_p^{\mathbf{s}_p}}, \tag{4.8}$$

which is efficient to compute.

EPS are limited to small plaquettes since for each plaquette the number of coefficients scales exponentially with the size of the plaquette. However one can generalize this Ansatz by describing the state of clusters of spins by a MPS, avoiding the exponentially many coefficients needed. The lattice is now cut in overlapping 1D strings which can mediate correlations on longer distances compared to local plaquettes (Fig. 4.1d). The resulting Ansatz is a String-Bond State (SBS)[195] defined by a set of strings $i \in S$ (each string i is an ordered subset of the set of spins) and a MPS for each string:

$$\psi_w(\mathbf{s}) = \prod_i \mathrm{Tr}\left(\prod_{j \in i} A_{i,j}^{s_j}\right). \tag{4.9}$$

The descriptive power of this Ansatz is highly dependant on the choice of strings: for example, by using small strings covering small plaquettes and a large bond dimension it includes EPS; whereas a single long string in a snake pattern includes MPS in 2D. In 3D, it has been used by choosing strings parallel to the axes of the lattice[196]. Since the form of the wave function is a product of MPS, the log-derivative with respect to some elements present in one of the MPS is simply the log-derivative for the corresponding MPS (Eq. (4.6)). The VMC procedure for optimizing SBS and MPS thus have the same cost. In addition, the ratio of two configurations which differ only by a few spins can be computed by considering only the strings including these spins, which speeds up the computation considerably. Let us note that a SBS can be mapped analytically to a MPS, but that the resulting MPS would have a bond dimension exponential in the number of strings.

4.1.3 Variational Monte Carlo method with Neural Networks

Recently, it was realized that the VMC method can be viewed as a form of learning, which motivated the use of another class of seemingly unrelated states

for describing the ground state of many-body quantum states: Neural-Network Quantum States[180] are quantum states for which the wave function has the structure of an artificial neural network. While a few different networks have been investigated[69, 180–182], the most promising results so far have been obtained with Boltzmann Machines[183]. Boltzmann Machines are a kind of generative stochastic artificial neural networks that can learn a distribution over the set of their inputs. In quantum many-body physics, the inputs are spin configurations and the wave function is interpreted as a (complex) probability distribution that the networks tries to approximate. Boltzmann Machines consist of two sets of binary units (classical spins): the visible units $v_i,\ i \in \{1, \ldots, N\}$, corresponding to the configurations of the original spins in a chosen basis, and hidden units $h_j,\ j \in \{1, \ldots, M\}$ which introduce correlations between the visible units. The whole system interacts through an Ising interaction which defines a joint probability distribution over the visible and hidden units as the Boltzmann weight of this Hamiltonian:

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{\mathcal{H}(\mathbf{v},\mathbf{h})}, \tag{4.10}$$

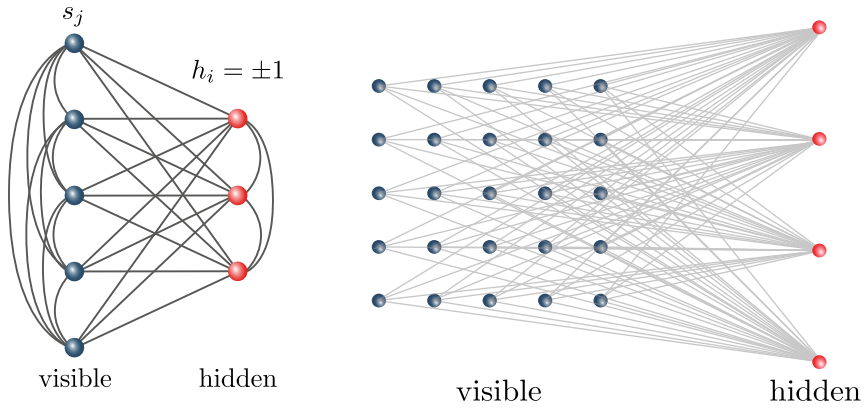
where the Hamiltonian $\mathcal{H}$ is defined as

$$\begin{aligned}\mathcal{H} = &\sum_j a_j v_j + \sum_i b_i h_i + \sum_{i<j} c_{ij} v_i v_j \\ &+ \sum_{i,j} w_{ij} h_i v_j + \sum_{i<j} d_{ij} h_i h_j,\end{aligned}$$

and Z is the partition function. The marginal probability of a visible configuration is then given by summing over all possible hidden configurations:

$$P(\mathbf{v}) = \sum_{\mathbf{h}} \frac{1}{Z} e^{\mathcal{H}(\mathbf{v},\mathbf{h})}, \tag{4.11}$$

and we take this quantity as Ansatz for the wave function: $\psi_w(\mathbf{s}) = P(\mathbf{s})$. The variational parameters are the complex parameters of the Ising Hamiltonian. In the case where there are interactions between the hidden units (Fig. 4.2a), the Boltzmann Machine is called a Deep Boltzmann Machine. It has been shown that Deep Boltzmann Machines can efficiently represent ground states of many-body Hamiltonians with polynomial-size gap, local tensor-network state and quantum states generated by any polynomial size quantum circuits[184, 185, 201]. On the other hand, computing the wave function $\psi_w(\mathbf{s})$ of such a Deep Boltzmann Machine in the general case is intractable, due to the exponential sum over the hidden variables, so the VMC method cannot be applied to Deep Boltzmann Machines without approximations. We therefore turn to the investigation of Restricted Boltzmann Machines (RBM), which only include interactions between the visible and hidden units (as well as the one-body interaction terms which correspond to biases). In this case, the sum over the hidden units can be performed analytically and the resulting wave function can be written as (here we take the hidden units



(a) Deep Boltzmann Machine (b) Restricted Boltzmann Machine in 2D

Figure 4.2: (a) Boltzmann Machines approximate a probability distribution by the Boltzmann weights of an Ising Hamiltonian on a graph including visible units (corresponding to the spins s_j) and hidden units h_i which are summed over. (b) Restricted Boltzmann Machines (here in 2D) only include interactions between the visible and the hidden units.

to have values ± 1):

$$\psi_w(\mathbf{s}) = e^{\sum_j a_j s_j} \prod_i \cosh\left(b_i + \sum_j w_{ij} s_j\right). \tag{4.12}$$

RBM can represent many quantum states of interest, such as the toric code[200], any graph state, cluster states and coherent thermal states[184]; the possibility of computing efficiently $\psi_w(\mathbf{s})$ prevents it however to approximate all PEPS and ground states of local Hamiltonians[184]. On the other hand, since computing $\psi_w(\mathbf{s})$ and its derivative is very efficient, RBM can be optimized numerically via the VMC method.

4.2 Relationship between Tensor-Network and Neural-Network states

While the machine learning perspective which leads to the application of Boltzmann Machines to quantum many-body systems seems quite different from the information-theoretic approach to the structure of tensor-network states, we will see that they are in fact intimately related. It was recently shown that while fully connected RBM can exhibit volume-law entanglement, contrary to local tensor networks, all short-range RBM satisfy an area law[236]. Moreover short-range and sufficiently sparse RBM can be written as a MPS[201], but doing so for a fully-connected RBM would require an exponential scaling of the bond dimension with the size of the system. In this section we show that there is a tighter connection between RBM and the previously introduced tensor networks in arbitrary dimension.

4.2.1 Jastrow wave functions, RBM and the Majumdar-Gosh model

Before turning to tensor networks, let us first consider the simple case of the Jastrow wave function (Eq. (4.4)). Boltzmann Machines including only interactions between the visible units lead to a wave function

$$\psi_w(\mathbf{s}) = \prod_k e^{a_k s_k} \prod_{i<j} e^{c_{ij} s_i s_j}, \tag{4.13}$$

which has the form of a product between functions of pairs of spins, and is thus a Jastrow wave function. More generally, semi-restricted Boltzmann Machines including interactions between visible units as well as between hidden and visible units are a product of a RBM and a Jastrow factor.

Nevertheless, one may ask whether a RBM alone is enough to describe a Jastrow factor. We first rewrite the RBM as

$$\psi_w(\mathbf{s}) = \prod_j A_j^{s_j} \prod_i \left(B_i \prod_j W_{ij}^{s_j} + \frac{1}{B_i \prod_j W_{ij}^{s_j}} \right), \tag{4.14}$$

where we have redefined the parameters with uppercase letters as the exponential of the original parameters, thus removing the exponentials in the hyperbolic cosine. This form will be convenient for the numerical simulations presented later. Since Jastrow wave functions are a product of functions of all pairs of spins, let us show that a RBM with one hidden unit can represent any function of two spins. It then follows that a RBM with $M = N(N-1)/2$ hidden units, each representing a function of one pair of spins, can represent a Jastrow wave function with polynomial resources. We thus have to solve for a system of four non-linear equations with $s_1, s_2 \in \{-1, 1\}$ and f the most general function of two spins : $\psi_w(s_1, s_2) = f(s_1, s_2)$. This system is solved in Sec. 4.4, providing an analytical solution for the parameters of the RBM to represent the Jastrow wave function exactly, or to arbitrary precision if $f(s_1, s_2) = 0$ for some spins.

As an application, we use this result to write the ground state of the Majumdar-Gosh model[237] exactly as a RBM. The Majumdar-Ghosh model is defined by the following spin-1/2 Hamiltonian:

$$H = J \sum_{i=1}^{N-1} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{J}{2} \sum_{i=1}^{N-2} \mathbf{S}_i \cdot \mathbf{S}_{i+2} \tag{4.15}$$

The ground state wave function is a product of singlets formed by neighboring pairs of spins:

$$|\psi\rangle \propto \prod_{n=1}^{N/2} |\uparrow_{2n-1}\rangle |\downarrow_{2n}\rangle - |\downarrow_{2n-1}\rangle |\uparrow_{2n}\rangle, \tag{4.16}$$

This wave function can also be expanded in the computational basis as

$$\psi(s_1,\ldots,s_N) \propto \prod_{n=1}^{N/2}(-1)^{(s_{2n-1}+3)/2}\delta_{s_{2n-1}\neq s_{2n}}, \tag{4.17}$$

$$\propto \prod_{n=1}^{N/2} f(s_{2n-1},s_{2n}). \tag{4.18}$$

Using the previous result, each function of two spins f can be written as a RBM using one hidden unit, which leads to a RBM representation of the ground states with $M = N/2$ hidden units. We also find numerically on small systems that a RBM using less than $M = N/2$ has higher energy than the ground state, which suggests that $M = N/2$ could be optimal.

4.2.2 Short-range RBM are EPS

Let us now turn to the specific case of RBM with short-range connections (sRBM). This encompasses all quantum states that have previously been written exactly as a RBM, such as for example the toric code or the 1D symmetry-protected topological cluster state[200]. Such states have weights connections between visible hidden units that are local. Each hidden unit is connected to a local region with at most d neighboring spins. If we divide the lattice into M subsets $p_i,\ i \in \{1,\ldots,M\}$, the wave function can be rewritten as (we omit here the biases a_j which are local one-body terms):

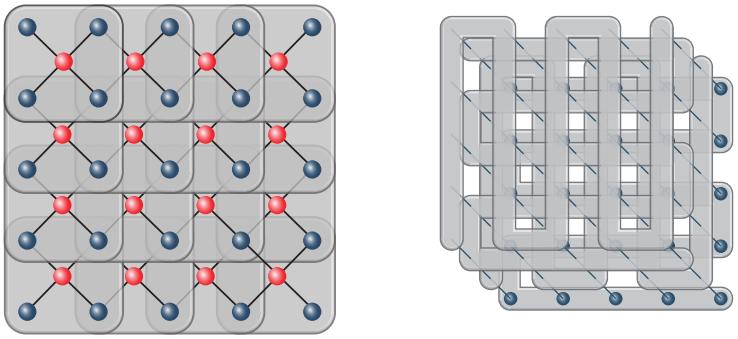
$$\psi_w(\mathbf{s}) = \prod_{i=1}^{M} \cosh\left(b_i + \sum_{j\in p_i} w_{ij}s_j\right) \tag{4.19}$$

$$= \prod_{i=1}^{M} C_i^{\mathbf{s}_i}, \tag{4.20}$$

where $\mathbf{s}_i$ is the spin configuration in the subset p_i. This is the form (Eq.(4.7)) of an EPS (Fig. 4.3a). For translational invariant systems, the short-range RBM becomes a convolutional RBM, which corresponds to a translational invariant EPS. The main difference between a short-range RBM and an EPS is that the RBM considers a very specific function among all possible functions of the spins inside a plaquette, hence EPS are more general than short-range RBM. This also directly implies that the entanglement of short-range RBM follows an area law. The main advantage of short-range RBM over EPS is that due to the exponential scaling of EPS with the size of the plaquettes, larger plaquettes can be used in short-range RBM than in EPS. Since in practice for finite systems it is possible to work directly with fully-connected RBM, we argue that EPS or fully-connected RBM should be preferred to short-range RBM for numerical purposes.

4.2.3 Fully-connected RBM are SBS

Fully-connected RBM, on the other hand, do not always satisfy an area law[236] and hence cannot always be approximated by local tensor networks. Nevertheless,



(a) Local RBM as an EPS (b) RBM as a non-local SBS

Figure 4.3: (a) A locally connected RBM is an EPS where each plaquette encodes the local connections to a hidden unit. (b) Once expressed as a SBS a fully-connected RBM can be represented by many strings on top of each other. Enlarging the RBM by using non-commuting matrices to non-local SBS induces a geometry in each string.

one can express the RBM wave function as (here we also omit the bias a_j):

$$\psi_w(\mathbf{s}) = \prod_i \cosh\left(b_i + \sum_j w_{ij} s_j\right) \tag{4.21}$$

$$\propto \prod_i \left(e^{b_i + \sum_j w_{ij} s_j} + e^{-b_i - \sum_j w_{ij} s_j}\right) \tag{4.22}$$

$$\propto \prod_i \mathrm{Tr}\begin{pmatrix} e^{b_i + \sum_j w_{ij} s_j} & 0 \\ 0 & e^{-b_i - \sum_j w_{ij} s_j} \end{pmatrix} \tag{4.23}$$

$$\propto \prod_i \mathrm{Tr}\left(\prod_{j \in i} A_{i,j}^{s_j}\right), \tag{4.24}$$

where

$$A_{i,j}^{s_j} = \begin{pmatrix} e^{b_i/N + w_{ij} s_j} & 0 \\ 0 & e^{-b_i/N - w_{ij} s_j} \end{pmatrix} \tag{4.25}$$

are diagonal matrices of bond dimension 2. This shows that RBM are String-Bond States, as the wave function can be written as a product of MPS over strings, where each hidden unit corresponds to one string. The only difference between the SBS as depicted in Fig. 4.1d and the RBM is the geometry of the strings. In a fully-connected RBM, each string goes over the full lattice, while SBS have traditionally been used with smaller strings and with at most a few strings overlapping at each lattice site.

4.2.4 Generalizing RBM to non-local SBS

In the SBS language, RBM consists in many strings overlapping on the full lattice. The matrices in each string in the RBM are diagonal, hence commute, so they

can be moved in the string up to a reordering of the spins. This means that each string does not have a fixed geometry and can adapt to stronger correlations in different parts of the lattice, even over long distances. This motivates us to generalize RBM to SBS with diagonal matrices in which each string covers the full lattice (Fig. 4.3b). In the following we denote these states as non-local dSBS. This amounts to relaxing the constraints on the RBM parameters to the most general diagonal matrix and enlarging the bond dimension of the matrices. For example taking the matrices

$$A_{i,j}^{s_j} = \begin{pmatrix} a_{i,j}^{s_j} & 0 & 0 \\ 0 & b_{i,j}^{s_j} & 0 \\ 0 & 0 & c_{i,j}^{s_j} \end{pmatrix}, \tag{4.26}$$

with different parameters $a_{i,j}^{s_j}$ for each string, lattice site and spin direction, leads to the wave function (here $D = 3$):

$$\psi_w(\mathbf{s}) = \prod_i \left(\prod_j a_{i,j}^{s_j} + \prod_j b_{i,j}^{s_j} + \prod_j c_{i,j}^{s_j} \right). \tag{4.27}$$

Note that even for 2×2 matrices, the non-local dSBS is more general than a RBM since the coefficients in each of the two matrices corresponding to one spin are independent from each other, which is not the case in the RBM.

Generalizing such a wave function to larger spins than spin-1/2 is straightforward, since the spin s_i is just indexing the parameters. This provides a way of defining a natural generalization of RBM which can handle systems with larger physical dimension. For instance this can be applied to spin-1 systems, while a naive construction for a RBM with spin-1 visible and hidden units leads to additional constraints, as well as to approximate bosonic systems by truncating the local Hilbert space of the bosons.

A further way to extend this class of states is to include non-commuting matrices. This fixes the geometry of each string by defining an order and also enables to represent more complicated interactions. In the following we will refer to SBS in such a geometry as non-local SBS. The advantage of this approach is that it can capture more complex correlations within each string, while introducing additional geometric information about the problem at hand. It comes however at a greater numerical cost than non-local dSBS or RBM due to the additional number of parameters. In practice, one can use an already optimized RBM or dSBS as a way of initializing a non-local SBS.

In some cases, the SBS representation is more compact than the RBM/dSBS representation. Let us consider again the ground state of the Majumdar-Gosh Hamiltonian, which we previously wrote as a RBM with $M = N/2$ hidden units. The ground state of the Majumdar-Gosh Hamiltonian can also be written as a simple MPS with bond dimension 3 and periodic boundary conditions, with matrices [188]

$$A_n^{s_n=-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 0 \end{pmatrix}, A_n^{s_n=1} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{4.28}$$

or for open boundary conditions with

$$A_{2n}^{s=-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, A_{2n}^{s=1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{4.29}$$

$$A_{2n-1}^{s=-1} = \begin{pmatrix} 1 & 0 \end{pmatrix}, A_{2n-1}^{s=1} = \begin{pmatrix} 0 & 1 \end{pmatrix}. \tag{4.30}$$

Since this state is a MPS, it is also a SBS with 1 string. The RBM representation of the same state requires $N/2$ strings. In practice the number of non-zero coefficients are comparable, since in both cases the representation is sparse, but for numerical purposes a fully-connected RBM needs of the order $O(N^2)$ parameters before finding the exact ground state, while a MPS or SBS with one string will need $O(N)$ parameters for both open and periodic boundary conditions.

Another example is the AKLT model[238] defined by the following spin-1 Hamiltonian in periodic boundary conditions:

$$H = \sum_{i=1}^{N} \left[\frac{1}{2} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{6} \left(\mathbf{S}_i \cdot \mathbf{S}_{i+1} \right)^2 + \frac{1}{3} \right]. \tag{4.31}$$

Its ground state has a simple form as a MPS of bond dimension 2. It can also be written as an exact RBM by mapping the system to a spin-1/2 chain, but the number of hidden units needed for an exact representation scales as $O(N^2)$ in the system size[239]. We have numerically optimized the spin-1 extension of a RBM with form Eq. (4.27) (see Sec. 4.5 for the details of the numerical optimization) and found that already for small sizes of the chain a much higher number of parameters is required to approach the ground state energy as compared to a SBS with non-commuting matrices, which is exact with one string of bond dimension 2 (Fig. 4.4). We will also show in Section 4.3 that in some other cases the RBM needs less parameters than a SBS to obtain a similar energy. This demonstrates that both RBM and SBS have advantages and that their efficiency depends on the particular model that is investigated. It remains an open question whether there exist MPS or SBS which can provably not be efficiently approximated by a RBM (for which the RBM would need exponentially many parameters).

To be able to use both the advantages of RBM (efficient to compute, few parameters) and of SBS (complex representation, geometric interpretation), one can use the flexibility of SBS by including some strings that have a full MPS over the whole lattice, some strings which include only local connections and that will ensure that the locality of the system is preserved, and some strings that have the form of an RBM and that can easily capture large entanglement and long-range correlations. In many cases of interest, an initial approximation of the ground state can be obtained, either by optimizing simpler wave functions or by first applying DMRG to optimize a MPS. This initial approximation can then be used in conjunction with the previous Ansatz classes by multiplying an Ansatz wave function with the initial approximation. For the resulting wave function

$$\psi_w(\mathbf{s}) = \psi_w^{\text{init}}(\mathbf{s}) \psi_w^{\text{SBS}}(\mathbf{s}), \tag{4.32}$$

the ratio of the wave function on two configurations as well as the log-derivatives depend only on the respective ratio and log-derivatives of each separate wave

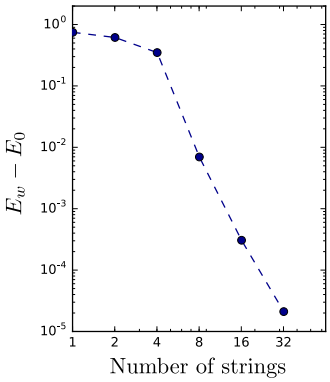


Figure 4.4: Energy difference with the exact ground state energy of a spin-1 extension of a RBM (Eq. (4.27)) with $D = 2$ and different number of strings for the AKLT model on a spin-1 chain with 8 spins. A non-local SBS with non-commuting matrices and one string is exact within numerical accuracy.

function, making the application of the VMC method straightforward. This procedure has the advantage of reducing the number of parameters necessary for obtaining a good approximation to the ground state and making the optimization procedure more stable, since the initial state is not a completely random state. Such a procedure provides a generic way to enhance the power of more specific Ansatz wave functions tailored to particular problems, as we will demonstrate in the next section. A similar technique has been used to construct tensor-product projected states with tensor networks in Ref. [240] and more generally it can be used to project the wave function of an initial reference state in a Fock space and is thus also suitable to describe fermionic systems.

4.3 Application to chiral topological states

In this section we turn to a practical application on a challenging problem for traditional tensor-network methods, namely the approximation of a state with chiral topological order. While chiral topological PEPS have been constructed, the resulting states are critical. Moreover the local parent Hamiltonian of a chiral fermionic Gaussian PEPS has to be gapless[219]. In the following we investigate if this obstruction carries on to the tensor-network and neural-network states that we have introduced previously.

4.3.1 RBM can describe a Laughlin state exactly

Let us consider a lattice version of the Laughlin wave function at filling factor 1/2 defined for a spin-1/2 system as

$$\psi_{\text{Laughlin}}(\mathbf{s}) = \delta_{\mathbf{s}} \prod_k \chi_k^{s_k} \chi \prod_{i<j} (z_i - z_j)^{\frac{1}{2} s_i s_j}, \tag{4.33}$$

where $\delta_{\mathbf{s}}$ fixes the total spin to 0, the z_i are the complex coordinates of the positions of the lattice sites and the phase factor are defined as $\chi_k^{s_k} = e^{i\pi(k-1)(s_k+1)/2}$, ensuring

that the state is a singlet. This wave function is equivalent to the Kalmeyer-Laughlin wave function in the thermodynamic limit and has been shown to describe a lattice state sharing the topological properties of the continuum Laughlin states on several lattices[241–243]. In addition, it can be written as a correlator from conformal fields, which has enabled the exact derivation of parent Hamiltonians for this state on any finite lattice[244].

The Laughlin wave function has the structure of a Jastrow wave function and we have shown in Section 4.2.1 that any Jastrow wave function can be written as a RBM with $M = N(N-1)/2$ hidden units. It follows that RBM and non-local SBS can represent a gapped chiral topological state exactly. This is in sharp contrast to local tensor-network states for which there is no exact description of a (non-critical) chiral topological state known. This difference is due to the non-local connections in the RBM and Jastrow wave function which allow them to easily describe a Laughlin state. We note that a chiral p-wave superconductor is another example of a gapped chiral topological state which has been recently written as a (fermionic) quasi-local Boltzmann Machine[185].

The previous construction is however not satisfactory in the sense that the RBM requires a number of hidden units scaling as $O(N^2)$, which is too high for numerical purposes on lattices which are not extremely small. We thus turn to the approximate representation of the Laughlin wave function using a RBM.

4.3.2 Numerical approximation of a Laughlin state

The lattice Laughlin wave function we consider has an exact parent Hamiltonian on a finite lattice[244] defined as

$$H_{\text{parent}} = \frac{2}{3}\sum_{i\neq j} |w_{ij}|^2 \boldsymbol{S}_i \cdot \boldsymbol{S}_j + \frac{2}{3}\sum_{i\neq j\neq k} \bar{w}_{ij} w_{ik} \boldsymbol{S}_j \cdot \boldsymbol{S}_k - \frac{2i}{3}\sum_{i\neq j\neq k} \bar{w}_{ij} w_{ik} \boldsymbol{S}_i \cdot (\boldsymbol{S}_j \times \boldsymbol{S}_k), \tag{4.34}$$

where $w_{ij} = \frac{z_i + z_j}{z_i - z_j}$ and $\boldsymbol{S}_j = (S_j^x, S_j^y, S_j^z)$ is the spin operator at site j. We specialize to the square lattice with open boundary conditions and minimize the energy of different wave functions with respect to this Hamiltonian by applying the VMC method presented in Section 4.1.2 with a Stochastic Reconfiguration optimization which is equivalent to the natural gradient descent[227, 245, 246] (details of the numerical optimization can be found in Sec. 4.5). Results are presented in Table 4.1.

We find that EPS with plaquettes of size up to 3×3 have an energy difference with the Laughlin state of the order 10^{-2}, which is better than a short-range RBM (denoted sRBM) on 3×3 plaquettes and up to $M' = 4$ hidden units per plaquette, while the energy of a fully connected RBM with $M = 2N$ hidden units is within 10^{-5} of the energy of the ground state. The resulting RBM uses much less hidden units than would be required for it to be exact, yet reaches an overlap of 99.99% with the Laughlin wave function. This result shows that the fully-connected structure of the RBM is an advantage to describe this state and that EPS can be used instead of short-range RBM. We have moreover found that

EPS are easier to optimize numerically than a short-range RBM: they are more stable, since each coefficient is considered separately, no exponentials or products that lead to unstable behavior are present and the derivatives have a very simple form (Eq. (4.8)).

Ansatz	$(E_w - E_0)/N$	$\lvert\langle\psi_w\lvert\psi_{\text{Laughlin}}\rangle\rvert$
EPS 2×2	4.3×10^{-2}	46.10%
EPS 3×3	2.2×10^{-2}	75.79%
sRBM $M' = 1$	8.3×10^{-2}	0.01%
sRBM $M' = 2$	3.1×10^{-2}	46.32%
sRBM $M' = 4$	2.5×10^{-2}	59.07%
RBM $M = N$	5.8×10^{-4}	99.7%
RBM $M = 2N$	1.1×10^{-5}	99.99%

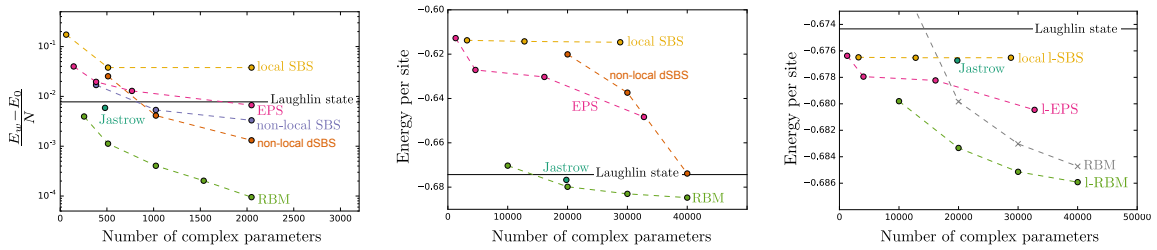
Table 4.1: Energy per site difference with the ground state energy and overlap with the Laughlin state of different Ansatz wave functions optimized with respect to the Hamiltonian H_{parent} on a 6×6 square lattice with open boundary conditions. sRBM have M' hidden units connected to all spins in each plaquette of size 3×3, while RBM have M hidden units connected to all spins of the lattice.

4.3.3 Numerical approximation of a chiral spin liquid

The previous results indicate that RBM might be useful for approximating chiral topological states numerically, but are limited to relatively small sizes due to the non-local nature of the parent Hamiltonian, which includes interactions between all triplets of spins on the lattice. In Ref. [209] a local Hamiltonian stabilizing a state in the same class as the Laughlin state was obtained by restricting H_{parent} to local terms and setting the long-range interactions to zero. This leads to the Hamiltonian

$$H_l = J \sum_{<i,j>} \boldsymbol{S}_i \cdot \boldsymbol{S}_j + J_\chi \sum_{<i,j,k>_\circlearrowleft} \boldsymbol{S}_i \cdot (\boldsymbol{S}_j \times \boldsymbol{S}_k), \tag{4.35}$$

where $< i, j >$ indicates indices of nearest neighbours on the lattice and $< i, j, k >_\circlearrowleft$ indicates indices of all triangles of neighboring spins, with vertices labelled in the counter clockwise direction. We focus on the case $J = 1, J_\chi = 1$ for which the ground state of H_l has above 98% overlap with the Laughlin wave function (Eq.(4.33)) on a 4×4 lattice. We minimize the energy of different classes of states on a 4×4 and 10×10 square lattice with open boundary conditions. For optimizing wave functions with tens of thousands of parameters we use a batch version of Stochastic Reconfiguration which optimizes a random subset of the parameters at each iteration (see Sec. 4.5). We consider several Ansatz wave functions including EPS with plaquettes of size 2×2, 3×2, 4×2 and 3×3, local SBS covering the lattice with horizontal, vertical and diagonal strings and increasing bond dimension, RBM with increasing number of hidden units, non-local SBS with diagonal matrices (denoted dSBS) or with non-commuting matrices of bond dimension 2 and different



(a) 4x4 lattice (b) 10x10 lattice (c) 10x10 lattice

Figure 4.5: Energy of H_l per site for different optimized Ansatz wave functions on a square lattice. The number of parameters (N_p) is modified by increasing the bond dimension D (local SBS, $N_p \propto D^2$), the size of the plaquettes (EPS, $N_p \propto M_P 2^P$, where M_P is the number of plaquettes and P is the number of spins in one plaquette), the number of strings M_S (non-local SBS and dSBS, $N_p \propto M_S$) or the number of hidden units M_h (RBM, $N_p \propto M_h$). (a) 4x4 lattice for which the energy difference with the exact ground state energy is plotted. (b) 10x10 lattice for which the exact ground state energy is unknown and the reference energy of the Laughlin state is indicated as a black line. (c) Optimization of wave functions that have been multiplied by the Laughlin wave function on a 10x10 lattice. The original RBM results are indicated for reference as grey crosses.

number of strings covering the full lattice. We observe that while the optimization of EPS and SBS is particularly stable, the optimization of RBM can lead to numerical instabilities that are resolved by writing the RBM in the form presented in Eq.(4.14). Since we use the same optimization procedure for all Ansatz wave functions and since the required time (and memory) to perform the optimization is mainly a function of the number of parameters and of the accuracy, we can compare the Ansatz classes by comparing how many parameters are needed to obtain a similar energy.

We first focus (Fig. 4.5a) on a 4×4 lattice for which the exact ground state can be obtained using exact diagonalization. Local SBS have an energy higher than the Laughlin state and the energy is saturated with increasing bond dimension, which means that the pattern of horizontal, vertical and diagonal strings is not enough to capture all correlations in the ground state. While a large 4×4 plaquette would make EPS exact on this small lattice, this would require 2^{16} parameters. The energy of the Laughlin state is already reached for 3×3 plaquettes. RBM with a number of hidden units larger than N and non-local SBS with a corresponding number of strings have lower energy than the Laughlin state or the Jastrow wave function. When the number of strings grows, the energy decreases even further. On a larger 10×10 lattice (Fig. 4.5b) the exact ground state energy is unknown but we can compare the energy of the different Ansatz wave functions and observe similar results. Only the Jastrow wave function, non-local SBS and RBM have an energy comparable to the Laughlin state. Notice that non-local SBS have a constant factor more parameters than a RBM for the same number of strings. On the one side this allows SBS to achieve better energy than RBM with the same number of strings. On the other side this comes with the drawback than we can

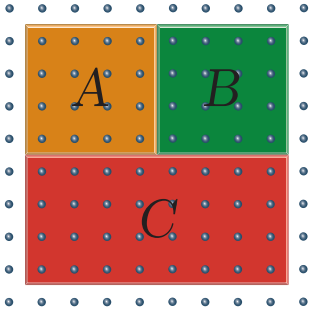


Figure 4.6: Partition of the lattice used to compute the topological entanglement entropy.

only optimize fewer strings and on the large lattice we are numerically limited to non-local dSBS with up to N strings. We can conclude that RBM are particularly efficient in this example since they require significantly less parameters than SBS for attaining the same energy. This has to be contrasted with the previous examples of the Majumdar-Gosh and AKLT models where the opposite was true. Therefore each class of states has advantages and drawbacks depending on the model we are looking at. We note in addition that a non-local SBS can be initialized with the results of a previous optimization with a RBM, which could provide a way of minimizing the difficulties of optimizing large number of parameters.

Ansatz	TEE
Laughlin	$-0.339(3)$
l-EPS 3×3	$-0.36(1)$
RBM $M = 4N$	$-0.34(1)$
l-RBM $M = 4N$	$-0.34(1)$

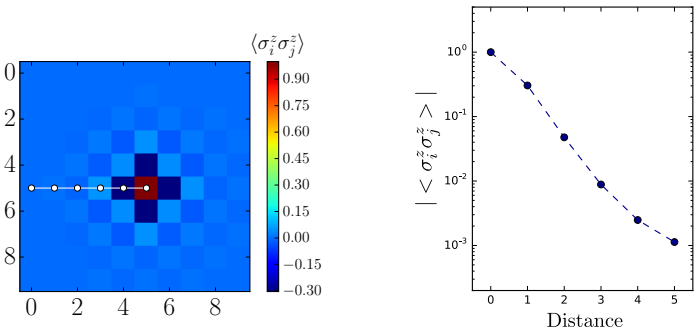
Table 4.2: Topological entanglement entropy (TEE) of the analytical Laughlin state and optimized l-EPS, RBM and l-RBM.

As we have previously noticed, we can also use an initial approximation of the ground state in combination with the previous Ansatz classes. In the case of the Hamiltonian H_l, the analytical Laughlin wave function can be used as our initial approximation in Eq. 4.32. We denote l-EPS (resp. l-SBS, l-RBM) a wave function that consists in a product of the Laughlin wave function and an EPS (resp. SBS, RBM) and minimize the energy of the resulting states. This allows us to obtain lower energies for each Ansatz class (Fig. 4.5c). Once the wave functions are optimized, their properties can be computed using Monte Carlo sampling. To check that the ground state is indeed in the same class as the Laughlin state, we compute the topological entropy of some of the optimized states by dividing the lattice into four regions (Fig. 4.6) and computing the Renyi entropy $S_A^{(2)} = -\ln \mathrm{Tr}\ \rho_A^2$ of each subregion using the Metropolis-Hastings Monte Carlo algorithm with two independent spin chains [247, 248]. The topological entanglement entropy is then

defined as[249, 250]

$$S_{\text{topo}} = S_A^{(2)} + S_B^{(2)} + S_C^{(2)} - S_{AB}^{(2)} - S_{AC}^{(2)} - S_{BC}^{(2)} + S_{ABC}^{(2)}, \tag{4.36}$$

and is expected to be equal to $-\ln 2 \approx -0.347$ for the Laughlin state[251]. The results we obtain are presented in Table 4.2 and provide additional evidence that the ground state of H_l has the same topological properties as the Laughlin state. The Hamiltonian H_l was recently investigated on an infinite lattice using infinite-PEPS[252] and further evidence was provided that the ground state is chiral. The PEPS results suggest the presence of long-range algebraically decaying correlations that may be a feature of the model or a restriction of PEPS to study chiral systems. The correlations on short distances agree with the correlations that we can compute on our finite system (Fig. 4.7a) but our method does not allow us to make claims about the long-distance behavior of the correlation function. We also observe that fully-connected RBM cannot be defined directly in the thermodynamic limit without a truncation of the distance of the interaction between visible and hidden units, thus transforming the RBM into a short-range RBM (albeit of larger range than an EPS). In Ref. [236] it was observed that the entanglement entropy of some specific short-range RBM can be computed analytically from the weights of the RBM. The method we use here works in the general case and also for a fully-connected RBM, but requires Monte Carlo sampling of the wave function. The optimized RBM weights encode every information about the wave function, it would thus be interesting to understand more precisely which quantities can be extracted directly from them. Whether direct information about the phase of the system can be obtained in this way without requiring Monte Carlo sampling remains an interesting open problem for future work.



(a) (b)

Figure 4.7: (a) The spin-spin correlation function between one lattice site (in red) and all other spins on the lattice measured on the optimized l-RBM with lowest energy reveals the antiferromagnetic behavior of the correlations. (b) Decay of the correlations with the distance across the direction indicated in (a) as a white solid line. The error bars are within dot size and finite size effects can already be seen for the last point.

4.4 Jastrow wave functions are Restricted Boltzmann Machines

Let us show that a RBM with one hidden unit can represent any function f of two spins. It then follows that a RBM with $M = N(N-1)/2$ hidden units, each representing a function of one pair of spins, can represent a Jastrow wave function. We parametrize f by its four values on two spins $s_1, s_2 \in \{-1, 1\}$ and solve for a system of four non-linear equations:

$$F_{11} = A_1 A_2 \left(W_1 W_2 + \frac{1}{W_1 W_2} \right) \tag{4.37}$$

$$F_{-1-1} = \frac{1}{A_1 A_2} \left(W_1 W_2 + \frac{1}{W_1 W_2} \right) \tag{4.38}$$

$$F_{1-1} = \frac{A_1}{A_2} \left(\frac{W_1}{W_2} + \frac{W_2}{W_1} \right) \tag{4.39}$$

$$F_{-11} = \frac{A_2}{A_1} \left(\frac{W_2}{W_1} + \frac{W_1}{W_2} \right), \tag{4.40}$$

where we have set $B_1 = B_2 = 1$. The RBM is well defined when all parameters are non-zero and we change of variables by defining $X = W_1 W_2$, $Y = \frac{W_1}{W_2}$, $A = A_1 A_2$, $B = \frac{A_1}{A_2}$, obtaining a new set of equations:

$$F_{-1-1} A^2 = F_{11} \tag{4.41}$$

$$F_{-11} B^2 = F_{1-1} \tag{4.42}$$

$$X^2 - \frac{1}{A} X + 1 = 0 \tag{4.43}$$

$$Y^2 - \frac{1}{B} Y + 1 = 0. \tag{4.44}$$

We first suppose that the values $F_{s_i s_j}$ are non-zero. These quadratic equations all have non-zero analytical solutions in the complex plane, that we denote A_0, B_0, X_0, Y_0. The original parameters are then the solutions of

$$W_1^2 = X_0 Y_0 \tag{4.45}$$

$$W_2^2 = X_0 / Y_0 \tag{4.46}$$

$$A_1^2 = A_0 B_0 \tag{4.47}$$

$$A_2^2 = A_0 / B_0, \tag{4.48}$$

which is again a set of quadratic equations with non-zero analytical solutions. If $F_{11} = F_{-1-1} = 0$ (resp. $F_{1-1} = F_{-11} = 0$), the exact solution is given directly by $A_0 = 1, X_0 = i$ (resp. $B_0 = 1, Y = i$). In the remaining cases where some $F_{s_i s_j}$ are zeros, the equations do not always have an exact solution, but the function can still be approximated to arbitrary precision. This case corresponds to strong restrictions on the part of the Hilbert space which is used to write the wave function and these constraints can also be imposed on the states directly by adding a delta function to the wave function which is equal to 1 only when the constraints on the spins are satisfied. Having a Markov Chain Monte Carlo sampling which does not visit these states then allows for a more efficient sampling.

4.5 Optimization procedure

The goal is to minimize the energy $E_{\mathbf{w}}$ depending on some vector of parameters $\mathbf{w}$. We define $\mathbf{f}$ to be the energy gradient vector at $\mathbf{w}$. Expanding the energy to first order around $\mathbf{w}$ leads to the steepest gradient descent, where the variational parameters are updated at each iteration according to $\mathbf{w}' = \mathbf{w} + \boldsymbol{\gamma}$, with a change of parameters given by $\boldsymbol{\gamma} = -\alpha\mathbf{f}$. Here α is a small step size. Expanding the energy to second order instead would result in the Newton method with a change of parameters given by:

$$\boldsymbol{\gamma} = -\alpha\mathbf{H}^{-1}\mathbf{f}, \tag{4.49}$$

where $\mathbf{H}$ is the Hessian of the energy. Small changes of the variational parameters may however lead to big changes in the wave function, especially in the case of compact non-local representations like RBM in which each parameter affects each part of the wave function. Taking into account the metric of changes of the wave function leads to the Stochastic Reconfiguration[227] method, which is equivalent to the natural gradient descent[246] and replaces the Hessian in Eq. (4.49) by the covariance matrix of the derivatives of the wave function, avoiding the computation of the second-order derivatives of the energy.

The Stochastic Reconfiguration method can also be viewed as an approximate imaginary-time evolution in the tangent space of the wave function. Consider the normalized wave function $|\bar{\psi}_0\rangle$ and its derivatives

$$|\bar{\psi}_0\rangle = \frac{|\psi_0\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}}, \tag{4.50}$$

$$|\bar{\psi}_i\rangle = \frac{|\psi_i\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}} - \frac{\langle\psi_0|\psi_i\rangle}{\langle\psi_0|\psi_0\rangle}\frac{|\psi_0\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}}, \tag{4.51}$$

defining a non-orthogonal basis set Ω. Expanding the wave function to linear order around some parameters $\mathbf{w}$ leads to

$$|\bar{\psi}(\mathbf{w}+\boldsymbol{\gamma})\rangle = \sum_{i=0}^{N_w} \gamma_i |\bar{\psi}_i\rangle. \tag{4.52}$$

To minimize the energy, one can apply the imaginary-time evolution operator $e^{-\alpha H}$, which expanded to first order for small α is $1 - \alpha H$. The change of coefficients $\boldsymbol{\gamma}$ is found by applying this operator to $|\bar{\psi}(\mathbf{w}+\boldsymbol{\gamma})\rangle$ and projecting in the set Ω, which leads to the equation

$$-\alpha\langle\bar{\psi}_i|H|\bar{\psi}_0\rangle = \sum_{j=1}^{M}\langle\bar{\psi}_i|\bar{\psi}_j\rangle\gamma_j, \tag{4.53}$$

which can be rewritten as

$$\boldsymbol{\gamma} = -\alpha\mathbf{S}^{-1}\mathbf{f}, \tag{4.54}$$

where $S_{ij} = \langle \bar{\psi}_i | \bar{\psi}_j \rangle$ and $f_i = \langle \bar{\psi}_i | H | \bar{\psi}_0 \rangle$. If we expand these expressions as expectation values over the probability distribution $p(\mathbf{s}) = \frac{|\psi_w(\mathbf{s})|^2}{\sum_{\mathbf{s}} |\psi_w(\mathbf{s})|^2}$, we obtain

$$f_i = \langle \Delta_i^* E_{\text{loc}} \rangle - \langle \Delta_i^* \rangle \langle E_{\text{loc}} \rangle, \tag{4.55}$$

$$S_{ij} = \langle \Delta_i^* \Delta_j \rangle - \langle \Delta_i^* \rangle \langle \Delta_j \rangle, \tag{4.56}$$

where the local energy is defined as $E_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} \langle \mathbf{s} | H | \mathbf{s}' \rangle \frac{\psi_w(\mathbf{s}')}{\psi_w(\mathbf{s})}$ and the log-derivative of the wave function as $\Delta_w(\mathbf{s}) = \frac{1}{\psi_w(\mathbf{s})} \frac{\partial \psi_w(\mathbf{s})}{\partial w}$. Finally, the complete algorithm is as follows:

1. Using a Metropolis-Hastings algorithm, generate samples of the probability p and compute stochastic estimates for the expectation values $\langle \Delta_j \rangle$, $\langle E_{\text{loc}} \rangle$, $\langle \Delta_i^* E_{\text{loc}} \rangle$, $\langle \Delta_i^* \Delta_j \rangle$
2. Construct the vector $\mathbf{f}$ and matrix $\mathbf{S}$,
3. Update the parameters according to $\mathbf{w} \leftarrow \mathbf{w} - \alpha \mathbf{S}^{-1} \mathbf{f}$,
4. Repeat the full procedure until convergence of the energy.

In practice we repeat the full procedure 1000 to 20000 times until the energy is converged. To optimize a large number of parameters we randomly select a subset of the parameters of size up to 10000 at each iteration of the algorithm and update only these parameters. This reduces the computational cost associated with the operations dealing with $\mathbf{f}$ and $\mathbf{S}$. Moreover we can avoid forming the full matrix $\mathbf{S}$ by instead solving Eq. (4.54) with a conjugate-gradient solver [245]. Numerical stability can be achieved by adding a small constant ϵ to the diagonal elements of the matrix $\mathbf{S}$, rotating the direction of change towards the steeped descent direction. We find that a step size α of the order $1/\sqrt{i}$, where i is the iteration step, works well in conjunction with a large stabilization at the beginning, while a fixed step size can also be chosen in conjunction with a small stabilization of the order $10^{-4} - 10^{-8}$ by performing several optimizations. At the later stages of the optimization, the step size is lowered to ensure that the energy is converged. Further improvements are achieved by projecting the wave functions in a subset of fixed total spin when it is conserved by the Hamiltonian we consider[253]. The spin-flip symmetry can be enforced in a RBM by choosing the bias $b_i = 0$.

4.6 Conclusion

We have shown that there is a strong connection between Neural-Network Quantum States in the form of Boltzmann Machines and some Tensor-Network states that can be optimized using the Variational Monte Carlo method : while short-range Restricted Boltzmann Machines are a subclass of Entangled Plaquette States, fully connected Restricted Boltzmann Machines are a subclass of String-Bond States. These String-Bond States are however different from traditional String-Bond States due to their non-local structure which connects every spin on the lattice to every string. This enabled us to generalize Restricted Boltzmann Machines by introducing

non-local (diagonal or non-commuting) String-Bond States which can be defined for larger local Hilbert space and with additional geometric flexibility. We compared the power of these different classes of states and showed that while there are cases where String-Bond States require less parameters than fully-connected Restricted Boltzmann Machines to describe the ground state of a many-body Hamiltonian, there are also cases where the additional parameters in each string make String-Bond States less efficient to optimize numerically. We applied these methods to the challenging problem of describing states with chiral topological order, which is hard for traditional Tensor Networks. We showed that every Jastrow wave function, and thus a Laughlin wave function, can be written as an exact Restricted Boltzmann Machine. In addition we gave numerical evidence that a Restricted Boltzmann Machine with a much smaller number of hidden units can still give a good approximation to the Laughlin state. Finally we turned to the approximation of the ground state of a chiral spin liquid and showed that Restricted Boltzmann Machines achieve a lower energy than the Laughlin state and the same topological entanglement entropy. We argued that combining different classes of states allows to take advantage of the initial knowledge of the model and of the particularities of each class. This was demonstrated by combining a Jastrow wave function to Tensor Networks and Restricted Boltzmann Machines, which allowed us to get lower energies than the initial states and characterize the ground state.

Our work sheds some light on the representative power of Restricted Boltzmann Machines and establish a bridge between their optimization and the optimization of Tensor Network states. On the one hand, the methods developed in this work can be used to target the ground state of other Hamiltonians and it would be interesting to know whether similar results can be achieved for example for non-Abelian chiral spin liquids[254, 255] or generalized to fermionic systems of electrons in the continuum displaying the Fractional Quantum Hall effect. On the other hand, we also showed that some tools used in machine learning can be rephrased in Tensor Network language, thus providing additional physical insights about the systems they describe. Matrix Product States have already been used as a tool for supervised learning[256, 257] and our work opens up the possibility of using not only Restricted Boltzmann Machines, but also String-Bond States to represent a probability distribution over some data while encoding additional information about its geometric structure.

Chapter 5

Supervised Learning with Quantum Techniques

Most of the content of this chapter is contained in:

- *"Supervised learning with generalized tensor networks"*
  Ivan Glasser, Nicola Pancotti, and J. Ignacio Cirac
  arXiv:1806.05964, (2018)

Some connections between neural-network quantum states and tensor networks were discussed in the previous chapter, in this chapter we try to employ tensor networks for tackling machine learning tasks. In particular, we explore the relationship between tensor networks and more common machine learning architectures, as probabilistic graphical models[201, 258–261]. We define generalized tensor networks which connect the two frameworks. These networks rely on the copy and reuse of local tensor information. Unlike regular tensor networks, they can be defined in complex geometries while remaining efficient to contract as long as an appropriate hierarchical order can be defined. We apply several variants of generalized tensor networks to image classification and environmental sound recognition and compare their performance, concluding that generalized tensor networks typically perform better than tensor networks alone. We also prove that generalized tensor networks are exponentially more efficient at describing some functions than regular tensor networks.

Generalized tensor networks share some structure with convolutional neural networks (CNN)[262, 263], while having direct connections to restricted Boltzmann machines (RBM)[260, 264, 265]. Examples of such networks that have been used in quantum physics include String-Bond States (SBS)[195, 196], which generalize Matrix Product States (MPS) (also known as Tensor Trains[266]), as well as Entangled Plaquette States (EPS)[192–194], which can be seen as a 1-layer CNN defining all possible convolutional filters over discrete inputs. Generalized tensor networks with tree structures have also been used to study the expressivity of deep learning models[267, 268].

We introduce an algorithm for performing supervised learning with generalized tensor networks, which combines stochastic gradient descent with previously introduced approaches for tensor networks. This framework generalizes works

based on regular tensor networks such as MPS[256, 257, 269, 270] or tree tensor networks[271]. It has the advantage that more complex structures can be formed while keeping the computation efficient. This is especially useful for data that possesses some geometrical structure in more than one dimension, such as images. We emphasize that the algorithm does not need to rely on any Monte Carlo techniques. This is unlike in quantum physics, where generalized tensor networks can only be optimized in combination with computationally expensive Monte Carlo sampling. In particular the cost of optimizing a SBS is only a constant factor times the cost of optimizing a MPS, but SBS are much more flexible in higher dimensions and can interpolate between a MPS and a restricted Boltzmann machine.

We discuss how real-valued data can be used in conjunction with tensor networks and suggest to learn the relevant tensor features of real data as part of the network. Inspired by deep network architectures, we also propose two ideas to combine neural networks and tensor networks. In the first case we use a neural network to extract features from the data in order to feed them into a tensor network, in the second we combine generalized tensor networks and neural networks in the same deep network architecture.

We benchmark our algorithms for several generalized tensor network architectures on different datasets[272–274]. For image classification, we find that generalized tensor networks outperform previously introduced tensor network algorithms based on MPS or trees while keeping a small dimension of the tensors. In the context of environmental sound recognition, we find that MPS and SBS deliver comparable performance. This shows that SBS should also be considered along with MPS when considering one-dimensional data, especially in the presence of long-range correlations, and may be applied in other settings such as natural language processing.

Some of the architectures we consider can be realized on a quantum computer, and serve as a reminder that the copy of classical input data may also be useful in quantum machine learning algorithms that cannot be simulated classically.

5.1 Graphical Models and Generalized Tensor Networks

We first review definitions of probabilistic graphical models and tensor networks, discuss their relationship and show that the two frameworks can be connected through the definition of generalized tensor networks in which parts of the network can be copied and reused. Examples of generalized tensor networks which have been successfully used in quantum physics are introduced, and their connection to more common machine-learning architectures is discussed.

5.1.1 Graphical models

Let us consider a set of discrete random variables $\mathbf{X} = \{X_1, \ldots, X_N\}$ taking values $\mathbf{x} = (x_1, \ldots, x_N)$ and a data set of samples from these variables $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_{|D|}\}$. Inferring the underlying probability distribution $p(\mathbf{x})$ can be done by maximizing

the log-likelihood

$$\mathcal{L} = \sum_{i=1}^{|D|} \log p(\mathbf{d}_i). \tag{5.1}$$

A common choice of parametrized models for p are graphical models [275], which correspond to a factorization of the probability distribution over a graph. Consider a graph $G = (V, E)$, where V is a set of vertices, E a set of edges between these vertices (each $e \in E$ is a pair of elements in V) and cl(G) is the set of maximal cliques of the graph. An undirected graphical model or a Markov random field defines a factorization of the joint probability of all random variables as

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathrm{cl}(G)} \phi_C(\mathbf{x}_C), \tag{5.2}$$

where $\mathbf{x}_C$ are the values of the variables in clique C, ϕ_C are the clique potentials which are positive functions and Z is the partition function that ensures normalization of the probability (Fig. 5.1a). Graphical models can be converted to factor graphs [276] defined on a bipartite graph of factors and variable vertices: one factor node f_C is created for each maximal clique and the factor is connected to the variables in the corresponding clique (Fig. 5.1b). The factorization of the probability distribution still reads

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_{C} f_C(\mathbf{x}_c), \tag{5.3}$$

and inference can be performed through belief propagation and the sum-product algorithm on factor graphs. To increase the set of distributions which can be represented we can add additional dependencies by introducing ancillary hidden variables (which are unobserved, that is their values are not supplied in the data) $\mathbf{H} = \{H_1, \ldots, H_M\}$ (Fig. 5.1c). The resulting probability distribution is obtained by marginalizing these hidden variables, giving

$$p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \sum_{\mathbf{h}} \prod_{C} f_C(\mathbf{x}_C, \mathbf{h}_C). \tag{5.4}$$

5.1.2 Tensor networks

Tensor networks are factorizations of tensors (multi-dimensional arrays) into a network of smaller tensors. This network admits a graphical notation, depicted in Fig. 5.2, in which boxes represent tensors and legs coming out of these boxes represent tensor indices. A leg joining two tensors represents a sum (contraction) over a joint index, such that the resulting tensor is a sum over joint indices of products of tensors. A simple example is a matrix factorization, in which a matrix is factorized in a product of two matrices. More formally, we consider a graph $G = (V, E)$ where we add to E edges containing only one vertex. These will

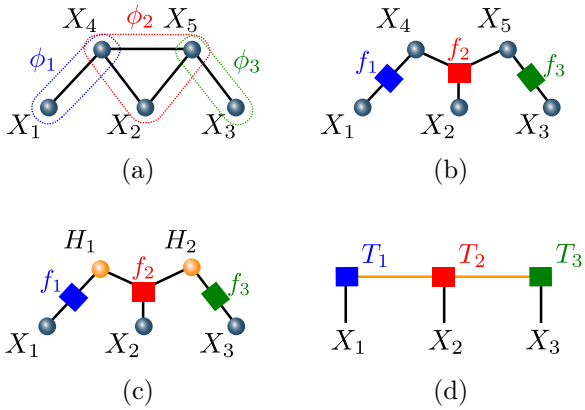


Figure 5.1: (a) Undirected graphical model with three maximal cliques depicted in colors (b) Corresponding factor graph (c) Factor graph with hidden units in orange that are marginalized (d) Equivalent tensor network, which is a Matrix Product State.

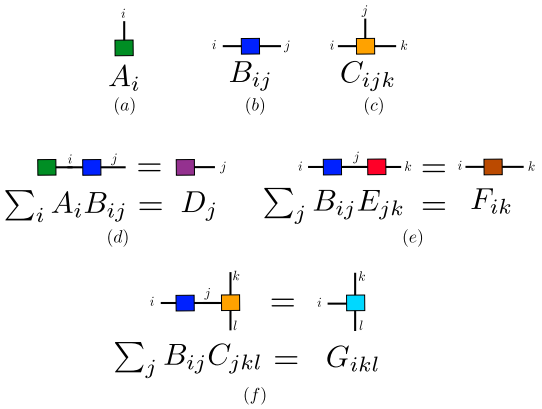


Figure 5.2: Graphical notation for tensor networks : (a) vector, (b) matrix, (c) order 3 tensor, (d) vector-matrix multiplication (e) matrix-matrix multiplication (f) matrix-tensor contraction.

correspond to open legs in a tensor-network graphical representation. We denote E' the subset of E containing edges that connect two vertices. We associate an integer size D_e called the bond dimension to each edge and define a tensor $T_v \in \otimes_{e \in v} \mathbb{R}^{D_e}$ for each vertex $v \in V$, with indices associated with the edges of this vertex. A tensor-network state is defined by summing over indices on all edges in the graph that connect two tensors. The resulting contracted tensor network is a tensor indexed by the indices of the open legs, denoted as $\mathbf{x} = (x_1, \ldots, x_N)$:

$$T_{\mathbf{x}} = \sum_{e \in E'} \prod_{v} T_v. \tag{5.5}$$

A particular case are one-dimensional matrix product states (MPS) (Fig. 5.1d),

also known as tensor-trains [266], which decompose a tensor as

$$T_{x_1,\ldots,x_N} = \sum_{e\in E'} A^{x_1}_{e_1} A^{x_2}_{e_2,e_3} A^{x_3}_{e_3,e_4} \cdots A^{x_N}_{e_N}, \tag{5.6}$$

where, for fixed value of $\mathbf{x}$, A^{x_1} and A^{x_N} are vectors, and A^{x_j}, $j = 2, \ldots, N-1$ are matrices. On a closed chain, also known as tensor ring [277], the corresponding decomposition is

$$T_{x_1,\ldots,x_N} = \mathrm{Tr}\left(\prod_{j=1,\ldots,N} A^{x_j}\right), \tag{5.7}$$

where all A^{x_j} are matrices, and the dimension of these matrices is the bond dimension. Generalizations to trees and lattices in higher dimensions have also been studied.

5.1.3 Tensor networks and graphical models duality

Consider a graphical model given as a factor graph in the special case where each hidden variable is connected to two factors, and each visible variable connected to one factor, as is the case in Fig. 5.1c. Because all the variables are discrete, each factor f_C is a non-negative function of discrete variables to which it is connected. It can therefore be written as a tensor $F_{C,\mathbf{x}_C,\mathbf{h}_C}$ with non-negative entries indexed by the variables $\mathbf{x}_C$ and $\mathbf{h}_C$ to which it is connected. The factor graph then defines a probability mass function over the visible variables that can be represented by a non-negative tensor $P_{\mathbf{x}}$ satisfying

$$P_{\mathbf{x}} = \sum_{\mathbf{h}} \prod_{C} F_{C,\mathbf{x}_C,\mathbf{h}_C}. \tag{5.8}$$

By comparing this expression with (5.5), we see that this probability mass function can be interpreted as a tensor-network state. Marginalization of a hidden variable corresponds to contracting the indices of the different factors connected to this variable, as indicated by the orange lines in Fig. 5.1d, and the visible variables correspond to the open legs of the tensor network. In the following we will refer to these open legs that correspond to visible variables or data inputs as the inputs of the tensor network. Probabilistic graphical models with discrete variables are therefore tensor networks of non-negative tensors. This connection has been previously observed in particular models [201, 258–260], and [261] provides a more detailed analysis of this duality on hypergraphs. The fact that graphical models rely on non-negative tensors, whereas tensor networks are usually studied in the context of real (or complex) elements, has important consequences for the optimization algorithms. Graphical models can be used in conjunction with expectation-maximization algorithms, which rely on the computation of conditional probabilities over a subset of the variables. Tensor networks do not have a probabilistic interpretation of the hidden variables (although they have a quantum interpretation), but powerful optimization algorithms such as the density matrix renormalization group (DMRG)

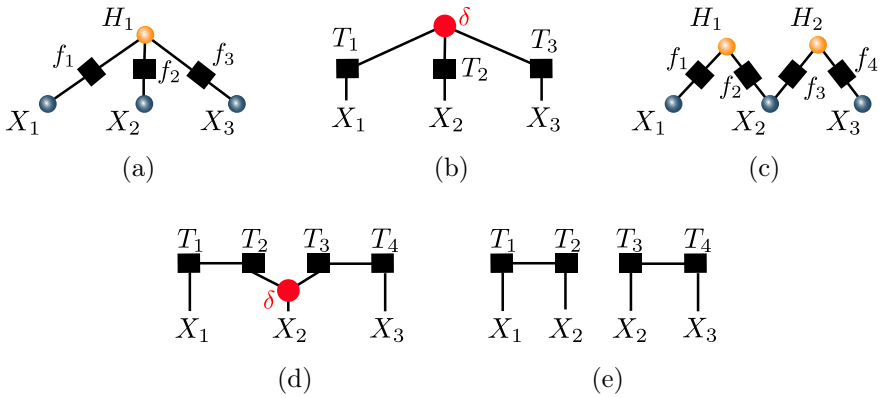


Figure 5.3: (a) Factor graph with one hidden variable connected to several factors (b) Equivalent tensor network with copy tensor (c) Factor graph with visible variables connected to several factors (d) Equivalent tensor network (e) Corresponding tensor network if the visible variables have fixed input values.

can rely on the singular-value decomposition of real and complex matrices. A comparison of the expressive power of non-negative tensor networks and real or complex tensor networks in the context of probabilistic modeling has been performed in [278].

5.1.4 Duality with copy tensors

Consider now the case of a probabilistic graphical model where one hidden variable is connected to several factors, as in Fig. 5.3a. To write the resulting probability mass function as a tensor network, we need to sum over the values of the hidden variable, which corresponds to summing over the value of the corresponding index which appears in several tensors. To represent this operation as a simple tensor network, we introduce a copy tensor δ with a number of indices corresponding to the number of factors connected to the hidden variable (each index can take as many values as the hidden variable), and such that δ is equal to one if all indices take the same value, and zero otherwise [279, 280]. In the case of Fig. 5.3a, δ is an order-3 tensor such that $\delta_{ijk} = 1$ if $i = j = k$ and $\delta_{ijk} = 0$ otherwise. We represent this tensor as a red dot in graphical representation. By inserting this copy tensor in the tensor network and connecting it to the corresponding factors, we obtain that the probability mass function defined by the graphical model can be represented by a tensor network, as in Fig. 5.3b.

The same procedure can be used if a visible variable is connected to several factors, as in Fig. 5.3c and Fig. 5.3d. Note that when computing the probability of inputs $\mathbf{x}$, the visible variables take specific given values. In this case contracting an input with a copy tensor will give the same result as fixing the value of all legs connected to the copy tensor to the input value and contracting the rest of the tensor network, as in Fig. 5.3e.

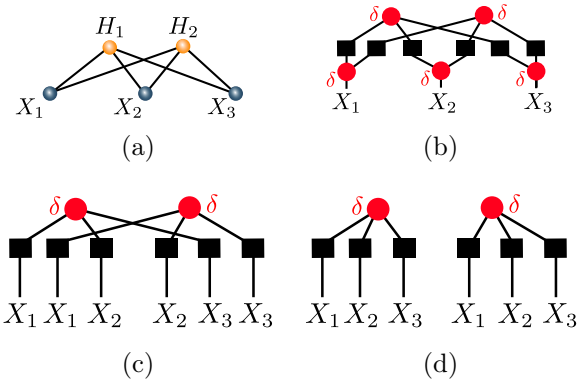


Figure 5.4: (a) Restricted Boltzmann machine (RBM) (b) Corresponding tensor network (c) Corresponding tensor network when the input values are fixed and the input copy tensors contracted (d) Same tensor network with a reordering of the inputs and tensors.

5.1.5 Application to restricted Boltzmann machines

One class of probabilistic graphical model which has the property that visible and hidden variables are connected to several factors are restricted Boltzmann machines (RBM) [264, 265]. They are defined on a bipartite graph with visible variables $\mathbf{X}$ and hidden variables $\mathbf{H}$ (Fig. 5.4a), so each visible variable is connected to each hidden variable. The connections between variables on this graph take the form of Ising interactions and the probability distribution of joint variables is

$$p(\mathbf{x}, \mathbf{h}) = \frac{1}{Z} e^{\mathcal{H}(\mathbf{x},\mathbf{h})}, \tag{5.9}$$

where the Hamiltonian $\mathcal{H}$ is a classical Ising Hamiltonian defined as (we omit here the bias terms for simplicity)

$$\mathcal{H} = \sum_{i,j} w_{ij} h_i x_j. \tag{5.10}$$

In the case where both visible and hidden variables are binary valued, the resulting probability distribution once the hidden variables have been marginalized is

$$p(\mathbf{x}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{\mathcal{H}(\mathbf{x},\mathbf{h})}, \tag{5.11}$$

$$= \frac{1}{Z} \prod_i (1 + e^{\sum_j w_{ij} x_j}). \tag{5.12}$$

Let us represent the corresponding factor graph, and associated tensor network. The cliques of the graph are all pairs of visible and hidden variables, for which there is an associated factor, which is a non-negative tensor, $f(x_j, h_i) = e^{w_{ij} h_i x_j}$. By inserting a copy tensor at the position of the hidden variables, as well as of the input variables, we obtain the tensor network representation of the RBM in Fig. 5.4b.

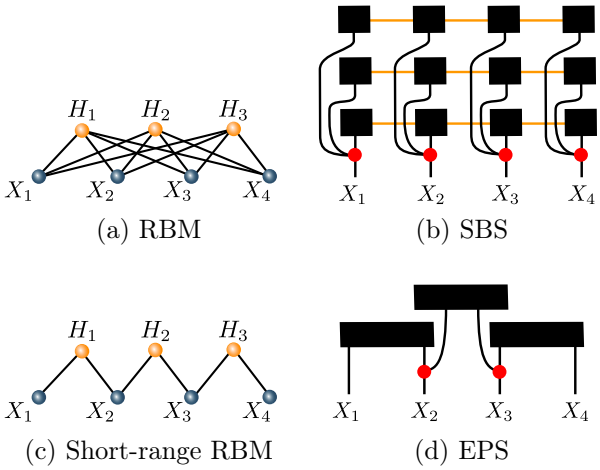


Figure 5.5: (a) Restricted Boltzmann machine (RBM) consisting of visible and hidden variables (b) String-bond state with 1D geometry generalizing RBM. The legs corresponding to contracted indices in each MPS are depicted in orange for visibility. (c) Short-range RBM with local connections between visible and hidden variables (d) Entangled plaquette state (EPS) generalizing the short-range RBM.

5.1.6 Relationship with models used in quantum physics

If we look at the graph for the tensor network equivalent to the restricted Boltzmann machine in Fig. 5.4b, we observe that this graph has many loops. Nevertheless, one can efficiently and exactly compute $p(\mathbf{x})$ for arbitrary size using (5.12). This comes from the fact that if we fix the values of the input variables and contract the copy tensors connected to the inputs, then the corresponding tensor network has no loops, as depicted in Fig. 5.4d.

Particular architectures of tensor networks satisfying the same property are tree tensor networks with copied inputs [267, 268]. Other examples of such networks have been used in the quantum physics community. The simplest example are entangled plaquette states (EPS) [192–194], also known as correlator product states, in which the tensor network is defined as a product of smaller tensors on overlapping clusters of variables:

$$T_{x_1,\ldots,x_N} = \prod_{p=1}^{P} T_p^{\mathbf{x}_p}, \tag{5.13}$$

where a coefficient $T_p^{\mathbf{x}_p}$ is assigned to each of the 2^{n_p} (for binary variables) configurations $\mathbf{x}_p$ of the variables in cluster p. Because the clusters overlap, the value of each variable is copied to all the tensors in which it is included (Fig. 5.5d). A sparse or short-range RBM is a special case of EPS [260], in which the tensor $T_p^{\mathbf{x}_p}$ takes the particular form $(1 + e^{\sum_j w_{pj} x_j})$, where the sum is limited to the variables in each cluster.

Another example are string-bond states (SBS) [195, 196], defined by placing MPS over strings (each string s is an ordered subset of the set of variables) on a graph which does not need to be a one-dimensional lattice. The resulting tensor

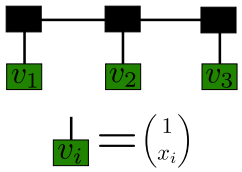


Figure 5.6: MPS with feature vectors as inputs.

network is

$$T_{x_1,\ldots,x_N} = \prod_s \mathrm{Tr}\left(\prod_{j\in s} A_{s,j}^{x_j}\right). \tag{5.14}$$

The value of each visible variable is copied and sent to different MPS (Fig. 5.5b). It was shown in [260] that a RBM is a special case of SBS for which each string is associated with a hidden variable and covers all visible variables, and the matrices are taken to be

$$A_{s,j}^{x_j} = \begin{pmatrix} 1 & 0 \\ 0 & e^{w_{sj}x_j} \end{pmatrix}. \tag{5.15}$$

SBS thus provide a generalization of RBM that is naturally defined for discrete variables of any dimension and can introduce different correlations through the use of higher dimensional and non-commuting matrices. Since SBS also include a MPS as a particular case, they provide a way to interpolate between a MPS (large bond dimension, only one string) and a RBM (bond dimension 2, diagonal matrices, many strings). Note that a product of S MPS of bond dimension D is an MPS with bond dimension D^S, so restricted Boltzmann machines can be written as MPS of large bond dimension but with diagonal matrices.

5.1.7 Generalized Tensor Networks

Input features for tensor networks

In many cases of interest, data is not discrete, but instead given in the form of real numbers. In order to train tensor networks to perform machine learning tasks on such data, it has been suggested to define feature vectors from the data, and then to contract these feature vectors with a tensor network to define a function of the inputs [256, 257]. Consider for example input data given as real numbers $\{x_1, \ldots, x_N\}$ and define feature vectors

$$v_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}. \tag{5.16}$$

Now it is possible to use a tensor network to define a function of the inputs by contracting these vectors with the open legs of a tensor network, as depicted in Fig. 5.6 in the case of a MPS.

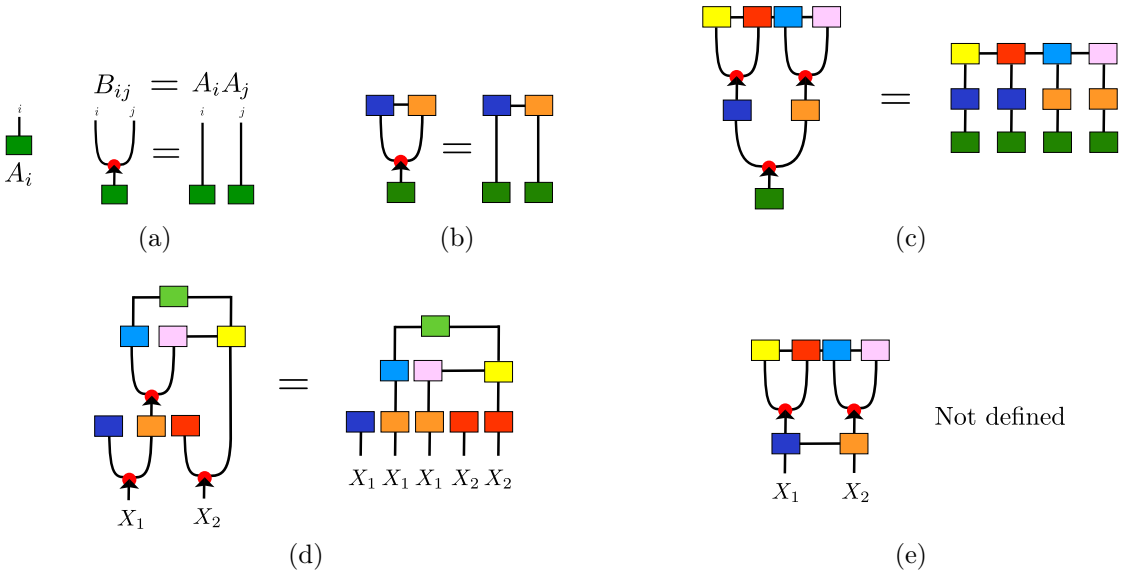


Figure 5.7: (a) Copy operation of a vector input A_i, resulting in a new tensor $B_{ij} = A_i A_j$ (b), (c) and (d) : Mapping of generalized tensor networks with copy operation to tensor networks with weight sharing (indicated by the same colors of the tensors) and copied input vectors or discrete inputs. (e) This representation is not defined because the copy operation only applies to vectors.

Copy operation with vector inputs

We would like to extend the previous results that use copy tensors to copy the inputs of the tensor network, allowing to efficiently contract complex tensor networks when the inputs are given. This idea that input data may be copied is also used in other machine learning architectures, such as for example Convolutional Neural Networks (CNN): a convolutional layer applies several filters to local clusters of input variables. In order to apply different filters to the same input data, this data needs to be copied and fed to the different parts of the neural network. In the case of tensor networks, we face the problem that copy tensors can only be used to copy discrete inputs [268]. Placing copy tensors as the input of the tensor network will therefore not have the expected effect.

For these reasons we introduce a different notion of copy, that we call the copy operation, that allows to copy real vectors. We will call the models that use this feature generalized tensor networks, to distinguish them from tensor networks with copy tensors, and we will restrict ourselves to the special cases in which these networks can be efficiently contracted as long as the inputs have fixed values. In the special case where the copy only takes place at the level of the input variables, and that these variables are discrete, then these networks will coincide with tensor networks that use a copy tensor at the level of the input legs.

We graphically depict the copy operation through a red dot between edges of the graph and an arrow which marks the incoming edge. The copy operation takes a vector as input, and it outputs two identical copies of it (Fig. 5.7). This operation is equivalent to having two copies of the input contracted with the rest

of the tensor network, and we depict this fact by using the same colors for the same tensors. In the case were the input is a discrete variable, we define the copy operation so that it has the same effect as the copy tensor that simply copies this variable.

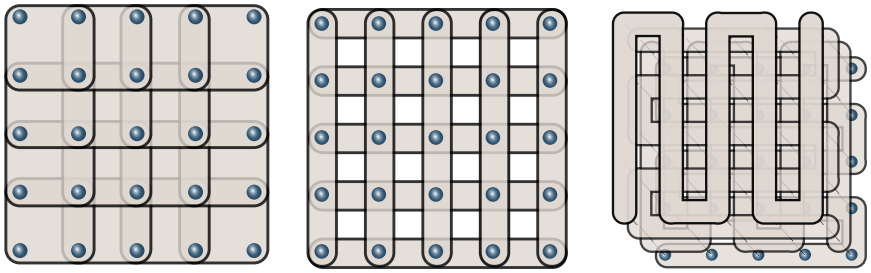
More generally, we can use this property to define complex generalized tensor networks where the copy operation is used at different places in the network. The resulting tensor networks are well defined if they can be expressed as a tensor network using several times the same tensor. Generalized tensor networks are therefore tensor networks that use weight sharing between some tensors, and such that several copies of the inputs may be used. In Fig. 5.7 we provide examples of such a mapping to clarify the definition of generalized tensor networks.

Architectures used in this work

The previous examples of EPS and SBS can be straightforwardly extended to generalized tensor networks by replacing the copy tensors as inputs by copy operations, so that they can be used with feature vector inputs. More generally, one can think of complex networks built using the copy operation for gluing different networks together. In practice, we will consider the following network architectures, in the case where the inputs have a two-dimensional geometry, such as images:

- EPS with 2×2 overlapping plaquettes with weight sharing such that the tensor for each plaquette is the same (Fig. 5.8a).
- SBS defined with horizontal and vertical strings covering the 2D lattice (Fig. 5.8b). We will denote this kind of SBS as 2D-SBS. Correlations along one of the two dimensions can be captured in the corresponding MPS, and more complex correlations are included through the overlap of the different strings.
- SBS consisting of 4 strings, each covering the whole lattice in a snake pattern, but in a different order (Fig. 5.8c). We denote these SBS as snake-SBS. They have the advantage, compared to a MPS, that two nearest neighbours variables always appear next to each other in one of the 4 strings, thus rendering the capture of strong local correlations efficient. More complex string geometries could be considered, and the choice of string could be itself learned with a RBM.
- EPS with an extra output leg in each plaquette, and such that the outputs are copied and taken as input into a SBS (Fig. 5.9). The input variables are first copied and fed into overlapping clusters parametrized by tensors. The output leg of each of these tensors is a vector which is copied a few times. Each of these copies can then be contracted with the open legs of a different MPS, forming together a SBS. In 2D, we choose 2x2 overlapping plaquettes in the first layer, and 4 strings forming a snake-SBS in the second layer. In the following we will call this generalized tensor network EPS-SBS.

These are simple examples of generalized tensor networks inspired from models invented in quantum physics to capture correlations on 2D lattice systems. More



(a) 2D-EPS (b) 2D-SBS (c) Snake-SBS

Figure 5.8: Possible geometries of generalized tensor networks in 2D: (a) EPS. (b) 2D-SBS consisting of horizontal and vertical overlapping strings. (c) Snake-SBS consisting of 4 overlapping strings in a snake pattern.

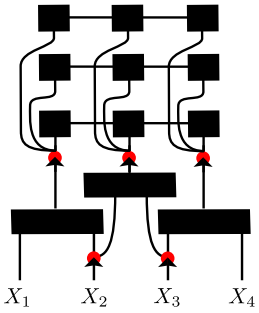


Figure 5.9: EPS-SBS consisting of a first layer of EPS, followed by a copy operation and a second layer of SBS.

complex networks based on trees or hierarchical designs with more than two layers can also be constructed in the same way.

We note that there is a relationship between EPS and one layer of a CNN : in the case of discrete inputs, there is a finite number of possible different filters that are applied in a CNN. If we represent the local convolutional operation as a function from discrete inputs, then this function can be represented as a tensor. The fact that convolutional filters are applied with a small stride over an input image corresponds to the overlap of the plaquettes in an EPS, and so EPS correspond to the first layer of a CNN applying all possible filters on discrete inputs.

These generalized tensor networks have the advantage, compared to standard tensor networks, that they can be easily defined in arbitrary dimension and geometry while remaining efficient to contract, as long as the input is fixed. This is in contrast to a 2D tensor network previously introduced in physics known as projected entangled pair states [190], which is naturally defined in higher dimensions, but cannot be contracted exactly efficiently. In particular, 2D-SBS form a subclass of projected entangled pair states that remains efficient to contract.

Another advantage of generalized tensor networks is that they can represent some functions with fewer parameters than regular tensor networks without copy tensors. Indeed, tensor networks such as MPS or tree tensor networks satisfy a constraint known as area law, which implies that they will require a number of

parameters that is exponential in the number of variables to represent functions not satisfying this constraint, whereas generalized tensor networks such as SBS or RBM can represent some of these functions with a polynomial number of parameter [236, 268].

5.2 Supervised Learning Algorithm

Graphical models are often used in conjunction with unsupervised learning algorithms, since they are designed to represent probability distributions. In particular cases it is possible to compute the normalization Z, which gives exact access to the likelihood and makes maximum likelihood estimation tractable. This is possible for graphical models and tensor networks on trees and has led to an algorithm for generative modeling with MPS [269, 278]. In the more general case, which includes RBM, the normalization Z cannot be computed efficiently. Approximate algorithms relying on Monte Carlo sampling can then be used, such as contrastive divergence [183, 265]. Generalized tensor networks suffer from the same issue, which makes unsupervised learning computationally expensive. Since these networks correspond to quantum states, it might be possible to implement them on a quantum computer and sample from them efficiently. In this work we focus instead on supervised learning, where access to the normalization Z is not necessary. In this section we discuss how RBM can be used for supervised learning, and generalize the corresponding algorithm to tensor networks.

5.2.1 Supervised Learning with Restricted Boltzmann Machines

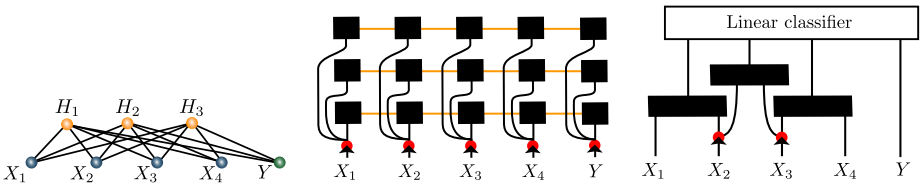
We first review how RBM can be used to perform supervised learning, in a classification setting [281, 282]. Given labelled training data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, where the y_i take discrete values corresponding to different classes, a RBM can be used to approximate the joint probability distribution of the variables and labels:

$$p(\mathbf{x}, y) = \frac{1}{Z} \sum_{\mathbf{h}} e^{\mathcal{H}(\mathbf{x},\mathbf{h},y)} \tag{5.17}$$

In such a model, the label is seen as an additional visible variable (Fig. 5.10a), possibly encoded in a one-hot representation to use only binary units. Training such a generative model can be done by maximizing the log-likelihood of the data. Since the likelihood is intractable, because the partition function Z cannot be efficiently computed, such a training can be done through approximate algorithms such as contrastive divergence. In supervised learning, one is interested in computing the conditional distribution

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{\sum_{y_j} p(\mathbf{x}, y_j)}, \tag{5.18}$$

which can be computed exactly when the number of classes is small enough, since the two partition functions cancel. The label predicted by the model for new data



(a) Discriminative RBM (b) Discriminative SBS (c) Discriminative EPS

Figure 5.10: (a) A classification RBM turns the label into an additional visible unit. (b) The same procedure can be defined for a SBS by adding a node corresponding to the label, and corresponding tensors which connect it to the rest of the tensor network. (c) Generalized tensor networks can be combined with additional layers of neural networks. For example an EPS output is a tensor that can be combined with a linear classifier.

$\mathbf{x}_i$ is the label maximizing $p(y|\mathbf{x}_i)$. Since one is ultimately interested in classification performance, it can be advantageous to directly optimize $p(y_i|\mathbf{x_i})$, which leads to a cost function to minimize

$$\mathcal{L}_{\text{discriminative}} = -\sum_{i=1}^{|D|} \log p(y_i|\mathbf{x}_i), \tag{5.19}$$

whose gradient can be computed analytically. A RBM can moreover be trained in a semi-supervised way, by using a combination of discriminative and generative training.

5.2.2 Supervised Learning with Generalized Tensor Networks

To generalize the discriminative training of RBM to generalized tensor networks, we would like to approximate the joint probability distribution of the variables and labels as a tensor network:

$$p(\mathbf{x}, y) \propto \text{GTN}(\mathbf{x}, y), \tag{5.20}$$

where $\text{GTN}(\mathbf{x}, y)$ is the function resulting of the contraction of a generalized tensor network with the inputs features and with the discrete label variable. More specifically, in the case of discrete variables the variables $\mathbf{x}$ and y fix the values of the open legs of the generalized tensor network, and the whole network is contracted, resulting in a scalar. In the case of input variables that are feature vectors, each vector is contracted with the corresponding open leg of the tensor network. In order for the cost function to be well defined, the result of the network contraction should be positive. This can be done by ensuring that the tensor elements are positive, as in a RBM or graphical model, or by choosing instead $p(\mathbf{x}, y) \propto (\text{GTN}(\mathbf{x}, y))^2$, which corresponds to a Born machine [278, 283, 284], or $p(\mathbf{x}, y) \propto e^{\text{GTN}(\mathbf{x},y)}$. In the following we will adopt this last choice, for which training is found to be more efficient.The label is now seen as the index of one tensor. Since it is discrete, there is no need to use a one-hot representation and one can simply enlarge the dimension of the leg of a tensor to accommodate for

the number of possible classes (Fig. 5.10b). We then define, by analogy with the graphical model case,

$$q(y_k, \mathbf{x_i}) = \frac{e^{\mathrm{GTN}(\mathbf{x_i}, y_k)}}{\sum_{y_j} e^{\mathrm{GTN}(\mathbf{x_i}, y_j)}}, \tag{5.21}$$

which can be seen as a softmax function applied to the different outputs of the tensor network contraction, and the cost function is again chosen to be a cross-entropy loss

$$\mathcal{L}_{\mathrm{discriminative}} = -\sum_{i=1}^{|D|} \log q(y_i, \mathbf{x}_i). \tag{5.22}$$

The gradient of the cost function can then be expressed as a function of $\mathrm{GTN}(\mathbf{x_i}, y_j)$ and its derivatives, since

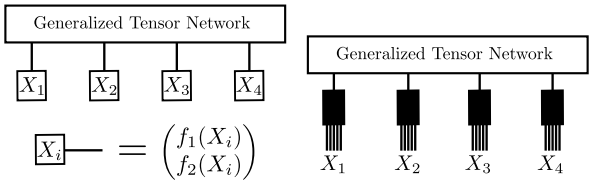
$$\frac{\partial \log q(y_i, \mathbf{x_i})}{\partial w} = \frac{\partial \mathrm{GTN}(\mathbf{x_i}, y_i)}{\partial w} - \sum_{y_j} q(y_j, \mathbf{x_i}) \frac{\partial \mathrm{GTN}(\mathbf{x_i}, y_j)}{\partial w}. \tag{5.23}$$

$\mathrm{GTN}(\mathbf{x_i}, y_j)$ can be computed exactly by fixing the value of the input units and labels and contracting the network. We note that in general we can contract the whole network without the labels, and perform the contraction for the different labels as a last step. Contraction of the whole network for different labels thus only adds a small cost (which depends on the shape of the network) to the contraction of the network without labels. We observe that from the point of view of supervised learning there is no essential difference between SBS and MPS in terms of the optimization algorithms: the cost of optimizing a SBS is only a constant factor (the number of strings) more expensive than for a MPS. This is unlike in quantum physics where Monte Carlo sampling is necessary to optimize a SBS. To further regularize the tensor network, we randomly drop tensor elements to 0 with probability δ during training, following the procedure in [256].

So far we have constructed tensor networks which, when an input and a label is given, have no open legs. We can also construct networks with open legs and use tensor networks in combination with other machine learning techniques. In this case the tensor network maps the input to a tensor which can for example be used as input in a neural network. In the case of EPS where each tensor over overlapping plaquettes has an open leg, an input is mapped to a tensor with an extra dimension as output. The simplest way to combine EPS with other neural networks is to place a linear classifier on top of the EPS (Fig. 5.10c). The backpropagation algorithm used to compute derivatives of the neural network is in this case combined with the algorithm for computing derivatives of a tensor network, and the joint network can be optimized using stochastic gradient descent. More complex combinations of tensor networks and neural networks may be defined in the same way.

5.3 Learning Feature Vectors of Data

In this section we explore several strategies that can be used to deal with data that is not discrete. We suggest to learn relevant tensor features as part of the



(a) (b)

Figure 5.11: (a) Real inputs X_i are mapped to a feature vector (here with length two). This vector can then be used as input to a generalized tensor network by contracting it with the open legs of the generalized tensor network. (b) Feature tensors can compress the discretized representation of the inputs X_i to a smaller dimensional space. These tensors can share weights and can be learned as part of the tensor network.

tensor network and discuss how tensor features can also be learned as part of a deep learning architectures which combines a neural network extracting features with a tensor network.

A naive way of applying tensor networks with real data would be to discretize data or use its binary representation. This is not a suitable approach, because that would amount to dramatically increasing the size of the data, rendering learning very slow, and it would also lead to large tensor networks prone to overfitting. Another approach, as suggested by [257] and [256], is to map the real data to a higher dimensional feature space. Each variable is first independently mapped to a vector of length (at least) two in order to be contracted with the open legs of the tensor network (Fig. 5.11a). Choices of feature maps that have been used include

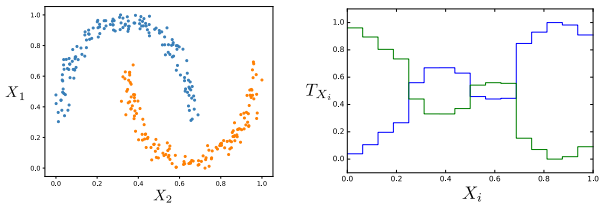
$$x \to \begin{pmatrix} 1 \\ x \end{pmatrix} \text{ or } \begin{pmatrix} \cos(\frac{\pi}{2}x) \\ \sin(\frac{\pi}{2}x) \end{pmatrix}, \tag{5.24}$$

and generalizations to higher dimensions. A choice which is suitable with our algorithm, assuming that the data is normalized between 0 and 1, is to use

$$x \to \begin{pmatrix} \cos^2(\frac{\pi}{2}x) \\ \sin^2(\frac{\pi}{2}x) \end{pmatrix}, \tag{5.25}$$

because this ensures that the vectors are positive and the normalization prevents numerical instabilities.

These choices however put severe limitations to the functions that can be learned. Indeed, the data set with just two variables presented in Fig. 5.12a cannot be separated by a MPS of bond dimension 2 with one of these feature choices, since the boundary decision will be a polynomial of degree two of the features. Nevertheless, a different feature choice could distinguish the two classes, even with bond dimension 2. We therefore suggest to learn the appropriate features as part of the learning algorithm. This can be done by parametrizing the feature functions and learning them at the same time as the rest of the network. To be able to use a purely tensor network algorithm, we can parametrize these functions using a tensor network. In the simplest case, we discretize the real data and use a tensor to



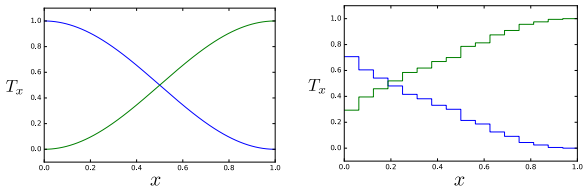
(a) (b)

Figure 5.12: (a) Dataset with two features X_1 and X_2 and two classes (depicted in different colors) that cannot be learned by a MPS of bond dimension 2 with the feature choice in (5.25). (b) Two normalized features learned by a tensor while classifying the previous data set with a MPS of bond dimension 2. The features have been discretized in 16 intervals. Using this choice of features the MPS can classify the data set perfectly.

compress the large dimensional input into a smaller dimensional vector of suitable length. This tensor can be learned as part of the whole tensor network and prevents the size of the rest of the tensor network to increase when the discretization size changes. The feature tensor can be the same for all variables, for example image pixels, but can be different in the case where the variables are of different nature. Using this procedure, a MPS of bond dimension 2 is able to get perfect accuracy on the data set presented in Fig. 5.12a. The two features that the network has learned are presented in Fig. 5.12b. We note that starting from random features on more complex data sets makes learning difficult, but the feature tensor can be pretrained using a linear classifier, before being trained with the rest of the network.

In comparison, we also show in Fig. 5.13b the features learned while classifying MNIST with greyscale pixels and a snake-SBS (see Section 5.4). These features are not very different from the choice in 5.25, and we could not distinguish performance with this choice or with learned features on this data set. We expect however that this procedure will be necessary for more complex data sets which are not easily approximated by a binary function. Moreover the size of the feature vector provides a regularization of the model, and higher sizes might be necessary for more complex data sets. More generally this tensor could be itself represented with a small tensor network, to prevent the number of parameters to increase too much with a very small discretization interval. It is interesting to note that the features learned in our examples are almost continuous even if we use smaller discretization intervals. This means that two real inputs that are close to each other will lead to the same predictions by the network, a property which is in general not true if we simply discretize the inputs and use a larger tensor network. Our approach of learning the features as part of the tensor network may be especially relevant in the context of quantum machine learning, where the tensor network is replaced by a quantum circuit and it might be suitable to have the full network as part of the same quantum machine learning architecture.

As an alternative way of choosing the features, we can combine the feature choice with other machine learning techniques. If the input data represents images,



(a) (b)

Figure 5.13: (a) Choice of two features in (5.25) for an input taking real values between 0 and 1. (b) Two normalized features learned by a tensor with output dimension 2 combined with a snake-SBS classifying the MNIST data set. The input features x are the greyscale value of pixels, normalized between 0 and 1 and discretized in 16 intervals.

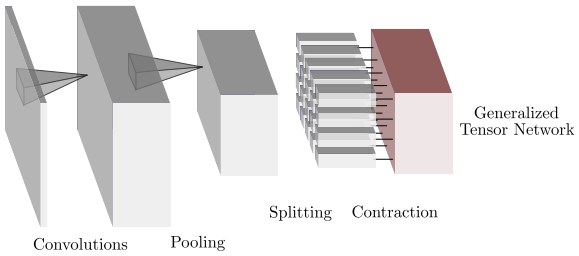


Figure 5.14: Using Convolutional Neural Networks as feature vector extractors from real data: the output of the CNN is seen as an image with a third dimension collecting the different features. For each pixel of this image, the vector of features is contracted with the open legs of a tensor network.

it is a natural choice to use Convolutional Neural Networks as feature extractors, since these have been highly successful for image classification. CNN consist in convolution filters, which use convolutional kernels to transform an image into a set of filtered images, and pooling layers which downsize the images (Fig. 5.14). The resulting features preserve a form of locality. Therefore it is natural to consider the vector of applied filters associated with each location in the image as a feature vector that can be used in conjunction with generalized tensor networks. The CNN and the tensor network can be trained together, since the derivatives of the tensor network can be used in the backpropagation algorithm which computes the gradient of the cost function.

5.4 Numerical Experiments

We test the generalized tensor network approach on the task of image classification, where a natural two-dimensional geometry that can be reflected in the architecture of the tensor network is present, as well as on the task of urban sound recognition, where the time dimension provides a one-dimensional geometry.

(a) (b)

Figure 5.15: Examples of images from the MNIST (a) and fashion MNIST (b) data set.

5.4.1 Image Classification

We first consider the MNIST data set [273], which consists of 28×28 greyscale images of digits. There are 10 classes and we adopt a multiclass classification procedure in which one tensor of the tensor network is parametrized by the ten possible labels. The original training set is split into training and validation sets of 55000 and 5000 examples and the performance of the different models is evaluated on the test set of 10000 examples. We consider the following generalized tensor networks: a snake-SBS with 4 strings (Fig. 5.8c), a 2D-SBS (Fig. 5.8b), an EPS with a 2×2 translational-invariant plaquette combined with a linear classifier, (Fig. 5.10c), an EPS-SBS with translational-invariant plaquette combined with a snake-SBS (Fig. 5.9) and a CNN-snake-SBS which uses a 1-layer CNN as input features (Fig. 5.14). The CNN considered here uses a convolutional layer applying 6 5×5 filters (stride 1) with ReLU activation function and a pooling layer performing max pooling with a 2×2 filter. All other networks use the choice of features presented in (5.25) and the greyscale values are normalized between 0 and 1. We compare the performance of these networks with a MPS [257] and a RBM (the number of hidden units of 250, 500, 750 or 1000 is taken as a hyperparameter). All networks use a batch size of 20 examples and hyperparameters such as the learning rate α and the regularization rate δ are determined through a grid search while evaluating the performance on the validation set. Best performance is typically achieved with $\alpha = 10^{-4}$, $\delta = 0.95$ and a hundred epochs of training.

The test set accuracy, presented in Fig. 5.16, shows that even with a very small bond dimension generalized tensor network are able to accurately classify the data set. Their performance is significantly better than that of a tree tensor network [271] or a MPS trained in frequency space[270], and while a MPS can also achieve 99.03% accuracy with a bond dimension of 120 [257], the cost of optimizing very large tensors has prohibited the use of this method for larger problems so far. The snake-SBS with bond dimension larger than 6 has also better performance than a RBM. Since the snake-SBS provides an interpolation between RBM and MPS, the choice of number of strings and geometry can be seen as additional parameters which could be tuned further to improve over the performance of both methods. All networks have a training set accuracy very close to 100% when

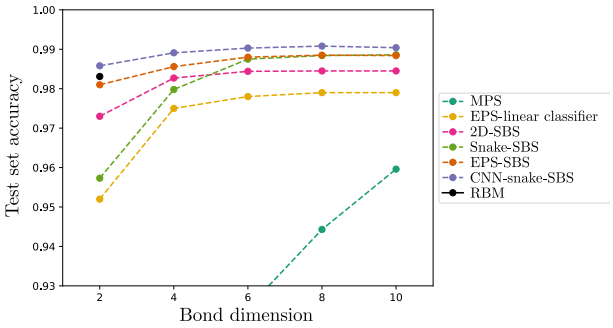


Figure 5.16: Test set accuracy of different generalized tensor networks on the MNIST data set.

the bond dimension is larger than 6, and we expect that better regularization techniques or network architectures have to be developed to significantly increase the test set performances obtained here. We also optimized a snake-SBS with positive elements (by parametrizing each element in a tensor as the exponential of the new parameters), which is a graphical model. Using the same algorithm, we were not able to achieve better performance than 93% classification accuracy with bond dimensions up to 10. This shows that while having a structure closely related to graphical models, tensor networks may provide different advantages.

Method	Accuracy
Support Vector Machine	84.1%
EPS + linear classifier	86.3%
Multilayer perceptron	87.7%
EPS-SBS	88.6%
Snake-SBS	89.2%
AlexNet	89.9%
CNN-snake-SBS	92.3%
GoogLeNet	93.7%

Table 5.1: Test set accuracy of generalized tensor networks and other approaches[272] on the fashion MNIST data set.

We then turn to the fashion MNIST data set [272], consisting of 28×28 greyscale images of clothes. While having the same size as the original MNIST data set, it is significantly harder to classify. We report the best accuracy obtained with different generalized tensor networks with bond dimension up to 10 in Table 5.1. It is found that these networks, while not state-of-the-art, are competitive with other approaches such as Support Vector Machines, AlexNet and GoogLeNet Convolutional Neural Networks or a multilayer perceptron neural network, which is encouraging considering the potential improvements in terms of network architecture or training algorithms.

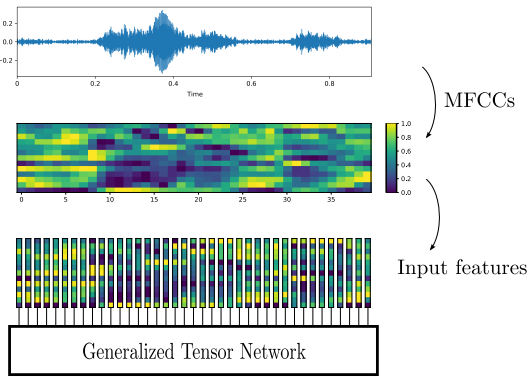


Figure 5.17: From the raw audio signal, Mel-frequency cepstral coefficients (MFCCs) are extracted over short overlapping windows, resulting in a sequence of high dimensional vectors. These vectors are taken as input to a generalized tensor network.

5.4.2 Environmental Sound Classification

So far we have considered black and white images, but it is also interesting to study how generalized tensor networks could be used for other types of data. In the following we consider the task of classifying environmental sounds. The UrbanSound8K data set [285] is a collection of 8732 audio clips (4s or less) divided into 10 classes of urban sounds: air conditioner, car horn, children playing, dog barking, drilling, engine idling, gun shot, jackhammer, siren and street music. The data set is divided into 10 folds and we use folds 1-9 for training and fold 10 for testing. The one-dimensional structure of sounds allows us to compare MPS and SBS with the same 1D string geometry. Preprocessing of the data takes place as follows : clips shorter than 4s are repeated to reach a fixed length of 4s. The first 13 Mel-frequency cepstral coefficients (MFCCs) are extracted for each clip (sampled at 22050Hz) using a window size of 2048 and hop length of 512, resulting in a sequence of length 173 and dimension 13 (Fig. 5.17). The corresponding 13-dimensional vectors are used as input feature vectors for the tensor network, and the time dimension of the sequence corresponds to the 1-dimensional structure of the MPS, or the strings of the SBS. Note that we do not perform any data augmentation nor split the training examples to enlarge the size of the data set, since we are interested in comparing MPS and SBS, rather than achieving the best possible accuracy on this data set. The training and testing accuracies are reported in Fig. 5.18 for a MPS with bond dimension up to 10 and a SBS with 4 strings and bond dimension up to 5. Since we are interested in comparing the expressivity of the different networks, no regularization is used and training is performed until the training accuracy does not improve anymore. Note that a MPS with bond dimension D has as many variational parameters as a SBS with 4 strings and bond dimension $D/2$.

We observe that the SBS has slightly higher training accuracy than a MPS with larger bond dimension and the same number of parameters. The test set performance is not significantly different between distinct architectures and in both cases we find that a lot of overfitting has taken place, which is not surprising given

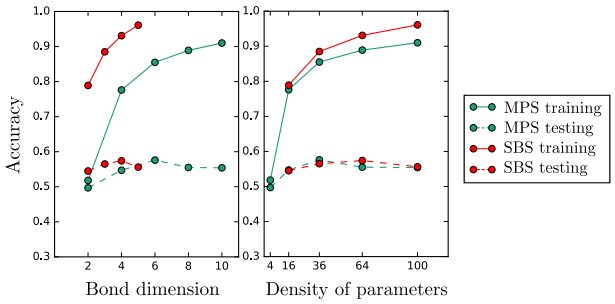


Figure 5.18: Training and testing accuracy of a MPS and a SBS with 4 strings on the UrbanSound8K data set. The density of parameters is the total number of parameters divided by 174 (the length of the strings).

the small number of training examples. Higher accuracies have been reported with other methods on the same data set. For example Convolutional Neural Networks can reach above 70% test set accuracy [274], however they use many more input features and rely on data augmentation. Nevertheless our results show that SBS should also be considered along with MPS when considering one-dimensional data, and may be applied in other settings such as natural language processing [286, 287].

5.5 Conclusion

We have introduced generalized tensor networks, which enlarge the class of tensor networks by introducing a reuse of information taking the form of a copy operation of tensor elements. The resulting networks are related to graphical models and we have discussed the strong relations that exist between particular graphical models and tensor network structures, such as restricted Boltzmann machines and string-bond states. We provided an algorithm to train these models to perform a supervised learning task and discussed several strategies to use tensor networks in conjunction with real-valued data. We showed that generalized tensor networks that can be contracted exactly can perform accurate image classification with much smaller bond dimension than regular tensor networks, that they can be used in other settings such as sound recognition and that they can be combined with neural-network architectures. Tensor networks can also be seen as a tool to simulate quantum circuits, and there is much research trying to understand how quantum circuits can be used in machine learning. Quantum circuits corresponding to MPS or tree tensor networks have been studied in the context of quantum machine learning [257, 288]. To implement the function corresponding to a generalized tensor network, one would need to copy the data at the input of the quantum circuit. Our results thus show that quantum machine learning circuits should take as input several copies of each data input and not just a single one, as well as share parameters between the different unitaries in the circuit. Generalized tensor networks which originate from the classical simulation of quantum states may thus serve as a testing and benchmarking platform of near-term quantum machine learning architectures.

Chapter 6

Unsupervised Learning with Quantum Techniques

Most of the content of this chapter is contained in:

- *"Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning"* Ivan Glasser, Ryan Sweke, Nicola Pancotti, Jens Eisert, J. Ignacio Cirac **arXiv:1907.03741, (2019)**

In the previous chapter, we showed that tensor networks are in natural correspondence with probabilistic graphical models [201, 258–261, 289]. Motivated by this correspondence, and with the goal of enhancing the toolbox for deriving analytical results on the properties of machine-learning algorithms, we study the expressive power of various tensor-network models of discrete multivariate probability distributions. The models we consider, defined in Section 6.1, fall into two main categories:

- **Non-negative tensor networks**, which decompose a probability mass function as a network of non-negative tensors [290], as in a probabilistic graphical model [276].
- **Born machines**, which model a probability mass function as the absolute value squared of a real or complex function, which is itself represented as a network of real or complex tensors. While Born machines have been previously employed for probabilistic modeling [269, 283, 291–294], they have additional potential applications in the context of quantum machine learning [284, 288, 295, 296], since they arise naturally from the probabilistic interpretation of quantum mechanics.

These models are considered precisely because they represent non-negative tensors by construction. In this chapter we focus on tensor networks which are based on tensor-trains/MPS and generalizations thereof, motivated by the fact that these have tractable likelihood, and thus efficient learning algorithms, while lending themselves to a rigorous theoretical analysis. In this setting non-negative tensor networks encompass hidden Markov models, while Born machines include models that arise from local quantum circuits of fixed depth.

Our main result is a characterization of the expressive power of these tensor networks. Interestingly, we prove that there exist families of probability distributions for which there are unbounded separations between the resource requirements of some of these tensor-network factorizations. This allows us to show that neither HMM nor Born machines should be preferred to each other in general. Moreover, we prove that using complex instead of real tensors can lead to an arbitrarily large reduction in the number of parameters of the network. This helps elucidate a topical open question on the potential role of complex numbers in machine learning models of real functions [297].

Furthermore, we introduce a new tensor-network model of discrete multivariate probability distributions with provably better expressive power than the previously introduced models. This tensor network, which retains an efficient learning algorithm, is referred to as a locally purified state (LPS) due to its origin in the classical simulation of quantum systems [298–301]. We demonstrate through numerical experiments on both random probability distributions as well as realistic data sets that our theoretical findings are relevant in practice - i.e. that LPS should be preferred over HMM and Born machines for probabilistic modeling.

6.1 Tensor-network models of probability distributions

Consider a multivariate probability mass function $P(X_1, \ldots, X_N)$ over N discrete random variables $\{X_i\}$ taking values in $\{1, \ldots, d\}$. This probability mass function is naturally represented as a multi-dimensional array, or tensor, with N indices, each of which can take d values. As such, we use the notation P to refer simultaneously to both the probability mass function and the equivalent tensor representation. More specifically, for each configuration $X_1, \ldots, X_N$ the tensor element $P_{X_1,\ldots,X_N}$ stores the probability $P(X_1, \ldots, X_N)$. Note that as P is a representation of a probability mass function, it is a tensor with non-negative entries summing to one.

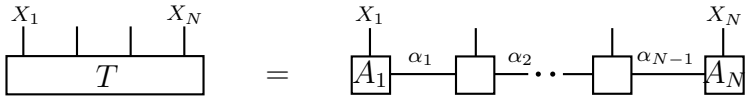
Here we are interested in the case where N is large. Since the number of elements of this tensor scales exponentially with N, it is quickly impossible to store. In cases where there is some structure to the variables, one may use a compact representation of P which exploits this structure, such as Bayesian networks or Markov random fields defined on a graph. In the following we consider models, known as tensor networks, in which a tensor T is factorized into the contraction of multiple smaller tensors. As long as T is non-negative, one can model P as $P = T/Z_T$, where $Z_T = \sum_{X_1,\ldots,X_N} T_{X_1,\ldots,X_N}$ is a normalization factor. For all tensor networks considered in this work, this normalization factor can be evaluated efficiently, as explained in Section 6.4.

In particular, we define the following tensor networks, in both algebraic and graphical notation. In the diagrams each box represents a tensor and lines emanating from these boxes represent tensor indices. Connecting two lines implies a contraction, which is a summation over the connected index.

1. **Tensor-train/matrix product state ($\mathbf{MPS}_{\mathbb{F}}$)**: A tensor T, with N d-dimensional indices, admits an $\mathrm{MPS}_{\mathbb{F}}$ representation of TT-rank$_{\mathbb{F}}$ r when the

entries of T can be written as

$$T_{X_1,\ldots,X_N} = \sum_{\{\alpha_i=1\}}^{r} A^{\alpha_1}_{1,X_1} A^{\alpha_1,\alpha_2}_{2,X_2} \cdots A^{\alpha_{N-2},\alpha_{N-1}}_{N-1,X_{N-1}} A^{\alpha_{N-1}}_{N,X_N} , \tag{6.1}$$

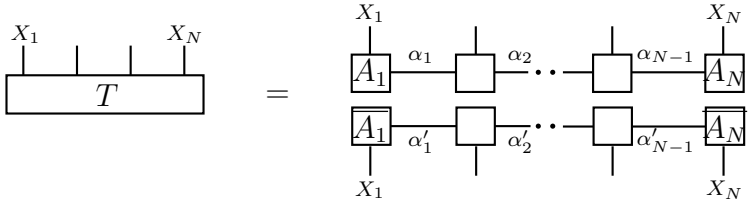


(6.2)

where A_1 and A_N are $d \times r$ matrices, and A_i are order-3 tensors of dimension $d \times r \times r$, with elements in $\mathbb{F} \in \{\mathbb{R}_{\geq 0}, \mathbb{R}, \mathbb{C}\}$. The indices α_i of these constituent tensors run from 1 to r and are contracted (summed over) to construct T.

2. **Born machine ($\mathbf{BM}_{\mathbb{F}}$)**: A tensor T, with N d-dimensional indices, admits a $\mathrm{BM}_{\mathbb{F}}$ representation of Born-rank$_{\mathbb{F}}$ r when the entries of T can be written as

$$T_{X_1,\ldots,X_N} = \left| \sum_{\{\alpha_i=1\}}^{r} A^{\alpha_1}_{1,X_1} A^{\alpha_1,\alpha_2}_{2,X_2} \cdots A^{\alpha_{N-2},\alpha_{N-1}}_{N-1,X_{N-1}} A^{\alpha_{N-1}}_{N,X_N} \right|^2 , \tag{6.3}$$

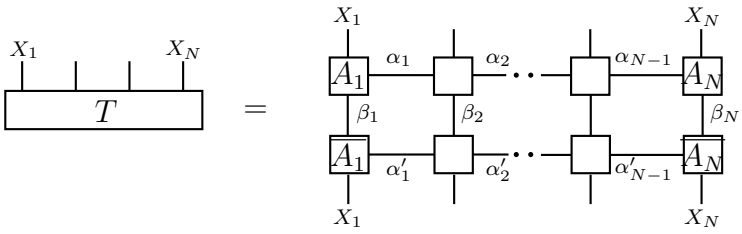


(6.4)

with elements of the constituent tensors A_i in $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, i.e., when T admits a representation as the absolute-value squared (element-wise) of an $\mathrm{MPS}_{\mathbb{F}}$ of TT-rank$_{\mathbb{F}}$ r.

3. **Locally purified state ($\mathbf{LPS}_{\mathbb{F}}$)**: A tensor T, with N d-dimensional indices, admits an $\mathrm{LPS}_{\mathbb{F}}$ representation of TT-rank$_{\mathbb{F}}$ r and purification dimension μ when the entries of T can be written as

$$T_{X_1,\ldots,X_N} = \sum_{\{\alpha_i,\alpha'_i=1\}}^{r} \sum_{\{\beta_i=1\}}^{\mu} A^{\beta_1,\alpha_1}_{1,X_1} \overline{A^{\beta_1,\alpha'_1}_{1,X_1}} A^{\beta_2,\alpha_1,\alpha_2}_{2,X_2} \overline{A^{\beta_2,\alpha'_1,\alpha'_2}_{2,X_2}} \cdots A^{\beta_N,\alpha_{N-1}}_{N,X_N} \overline{A^{\beta_N,\alpha'_{N-1}}_{N,X_N}}, \tag{6.5}$$



(6.6)

where A_1 and A_N are order-3 tensors of dimension $d \times \mu \times r$ and A_i are order-4 tensors of dimension $d \times \mu \times r \times r$. The indices α_i run from 1 to r, the indices β_i run from 1 to μ, and both are contracted to construct T. Without loss of generality we can consider only $\mu \leq rd^2$.

Note that all the representations defined above yield non-negative tensors by construction, except for $\mathrm{MPS}_{\mathbb{R}/\mathbb{C}}$. In this chapter, we consider only the subset of $\mathrm{MPS}_{\mathbb{R}/\mathbb{C}}$ which represent non-negative tensors.

Given a non-negative tensor T we define the TT-rank$_{\mathbb{F}}$ (Born-rank$_{\mathbb{F}}$) of T as the minimal r such that T admits an $\mathrm{MPS}_{\mathbb{F}}$ ($\mathrm{BM}_{\mathbb{F}}$) representation of TT-rank$_{\mathbb{F}}$ (Born-rank$_{\mathbb{F}}$) r. We define the puri-rank$_{\mathbb{F}}$ of T as the minimal r such that T admits an $\mathrm{LPS}_{\mathbb{F}}$ representation of puri-rank$_{\mathbb{F}}$ r, for some purification dimension μ. We note that if we consider tensors T with 2 d-dimensional indices (i.e., matrices) then the TT-rank$_{\mathbb{R}_{\geq 0}}$ is the non-negative rank, i.e., the smallest k such that T can be written as $T = AB$ with A being $d \times k$ and B being $k \times d$ matrices with real non-negative entries. The TT-rank$_{\mathbb{R}/\mathbb{C}}$ is the conventional matrix rank, the Born-rank$_{\mathbb{R}}$ (Born-rank$_{\mathbb{C}}$) is the real (complex) Hadamard square-root rank, i.e., the minimal rank of a real (complex) entry-wise square root of T, and finally the puri-rank$_{\mathbb{R}}$ (puri-rank$_{\mathbb{C}}$) is the real (complex) positive semidefinite rank [302]. These abbreviations, definitions and relations are summarized in Table 6.1 below, where we use the notations of ref. [302] for the different matrix ranks.

Table 6.1: Summary of notations for the different tensor-network representations and their ranks.

Tensor representation	$\mathrm{MPS}_{\mathbb{R}\geq 0}$	$\mathrm{MPS}_{\mathbb{R}/\mathbb{C}}$	$\mathrm{BM}_{\mathbb{R}/\mathbb{C}}$	$\mathrm{LPS}_{\mathbb{R}/\mathbb{C}}$
Tensor rank	TT-rank$_{\mathbb{R}\geq 0}$	TT-rank$_{\mathbb{R}/\mathbb{C}}$	Born-rank$_{\mathbb{R}/\mathbb{C}}$	puri-rank$_{\mathbb{R}/\mathbb{C}}$
Matrix rank [302]	rank_+	rank	$\mathrm{rank}_{\mathbb{R}/\mathbb{C}_\surd}$	$\mathrm{rank}_{\mathbb{R}/\mathbb{C},psd}$

For a given rank and a given tensor network, there is a set of non-negative tensors that can be exactly represented, and as the rank is increased, this set grows. In the limit of arbitrarily large rank, all tensor networks we consider can represent any non-negative tensor. This work is concerned with the relative expressive power of these different tensor-network representations, i.e. how do these representable sets compare for different tensor networks. This will be characterized in Section 6.3 in terms of the different ranks needed by different tensor networks to represent a non-negative tensor.

6.2 Relationship to hidden Markov models and quantum circuits

In order to provide context for the factorizations introduced in Section 6.1, we show here how they are related to other representations of probability distributions based on probabilistic graphical models and quantum circuits. In particular, we show that there is a mapping between hidden Markov models with constant number

of hidden units per variable and $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ with constant TT-$\mathrm{rank}_{\mathbb{R}_{\geq 0}}$, as well as between local quantum circuits of fixed depth and Born machines of constant Born-$\mathrm{rank}_{\mathbb{C}}$. These relations imply that results on the expressive power of the former directly provide results on the expressive power of the latter.

6.2.1 Hidden Markov models are non-negative matrix product states

Consider a hidden Markov model (HMM) with observed variables $\{X_i\}$ taking values in $\{1, \ldots, d\}$ and hidden variables $\{H_i\}$ taking values in $\{1, \ldots, r\}$. Let us show that such an HMM can be mapped to an $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ with TT-$\mathrm{rank}_{\mathbb{R}_{\geq 0}}$ equal to r, as depicted in Fig. 6.1a.

The probability of the observed variables in an HMM may be expressed as

$$P(X_1, \ldots, X_N) = \sum_{H_1, \ldots, H_N} P(X_1|H_1) \prod_{i=2}^{N} P(H_i|H_{i-1}) P(X_i|H_i). \tag{6.7}$$

Notice that $P(H_i|H_{i-1})$ and $P(X_i|H_i)$ are matrices with non-negative elements. Now define the tensors $A^j_{1,l} = P(X_i = l|H_1 = j)$, and $A^{j,k}_{i,l} = P(H_i = k|H_{i-1} = j)P(X_i = l|H_i = k)$. Then the MPS with TT-$\mathrm{rank}_{\mathbb{R}_{\geq 0}} = r$ defined with tensors A_i defines the same probability distribution on the observed variables as the HMM.

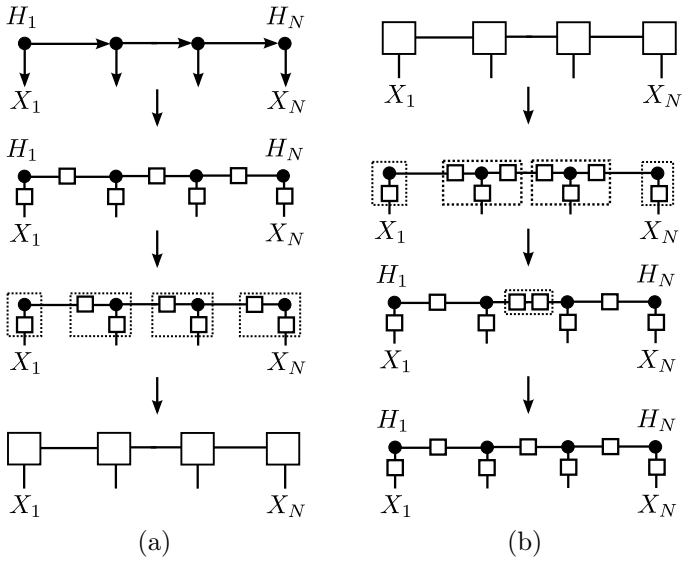


Figure 6.1: (a) Mapping of an HMM to an MPS with non-negative tensor elements. (b) Mapping of an MPS with non-negative tensor elements to an HMM.

Conversely, given an $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ with TT-$\mathrm{rank}_{\mathbb{R}_{\geq 0}} = r$, there exists an HMM, with hidden variables of dimension $r' \leq \min(dr, r^2)$, defining the same probability mass function, as shown in Fig. 6.1b. To construct this HMM, we split the tensors using an exact non-negative canonical polyadic decomposition such that $A^{j,k}_{i,l} = \sum_{s=1}^{r'} B^{j,s}_i C^{l,s}_i D^{s,k}_i$, where $r' \leq \min(dr, r^2)$. We can now set $P(X_i = l|H_i = s) = C^{l,s}_i$ and $P(H_i = s|H_{i-1} = j) = \sum_u D^{j,u}_{i-1} B^{u,s}_i$, where the probabilities must

be normalized properly, which can be done by first constructing the unnormalized factor graph and then normalizing the probabilities on every edge. We have then defined a HMM with hidden variables of dimension r' with the same probability of the observed units as the one arising from the MPS. We note also that by using a different graph for the HMM, it is possible to construct an equivalent HMM with hidden variables of dimension r [261, 289]. As such, any results on expressivity derived for $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ hold also for HMM.

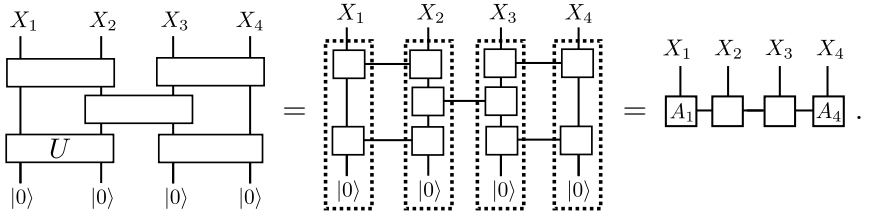
6.2.2 Quantum circuits are Born machines or locally purified states

In this section we provide a brief description of the connection between Born machines, locally purified states and quantum circuits, assuming some prior knowledge of the formalism of quantum computing. A more introductory presentation of this connection, assuming no knowledge of quantum mechanics and quantum computing, is contained in the supplementary material in Ref. [278].

Consider a 2-local quantum circuit of depth D acting on N d-dimensional qudits, with fixed orthonormal basis $\{|X_i\rangle\}_{X=1}^{d}$ for each local Hilbert space. The output state of the quantum circuit can be written as

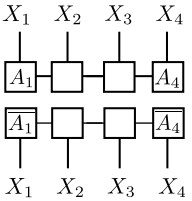
$$|\psi\rangle = \sum_{X_1=1}^{d} \cdots \sum_{X_N=1}^{d} \psi(X_1, \ldots, X_N)|X_1\rangle \otimes \ldots \otimes |X_N\rangle. \tag{6.8}$$

Furthermore, the amplitudes of this state are given by the entries of an MPS with TT-rank$_{\mathbb{C}}$ less than d^{D+1}. More specifically, as shown in Equation (6.9) below (explicitly for the case $N = 4$), by starting from the circuit diagram, reshaping and splitting each gate via a singular value decomposition, and then contracting the resulting diagram as indicated, one finds that

$\psi(X_1, \ldots, X_4) =$ 

(6.9)

The probability of outcome $X_1, \ldots, X_4$ when performing a measurement in the specified basis is obtained from the Born rule. It is thus given by the Born machine defined from the MPS representation of the quantum circuit, i.e.,

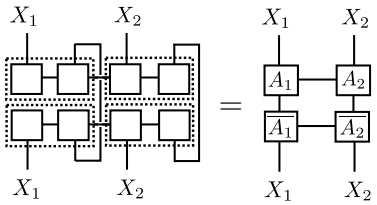
$P(X_1, \ldots, X_4) = |\psi(X_1, \ldots, X_4)|^2 =$ 

. (6.10)

By performing measurements on the output state of a 2-local quantum circuit, we are therefore effectively sampling from the probability mass function of N discrete d-dimensional random variables $\{X_i\}$ which is given by the Born machine defined from the MPS representation of the quantum circuit, as shown in Equation (6.10). Given this correspondence, any results on the expressive power of Born machines hold also for local quantum circuits, when considered as probabilistic models via the Born rule.

In order to understand the correspondence between local quantum circuits and locally purified states, note that if we consider each second qudit as an ancilla, or "hidden" qudit, then the probability of outcome X_1, X_2 when measuring only the visible qudits is given by the marginal

$$P(X_1, X_2) = \sum_{H_1, H_2} P(X_1, H_1, X_2, H_2) =$$



. (6.11)

In particular, we see that by performing such measurements we are sampling from the probability mass function given by the locally purified state which is obtained from the MPS representation of the circuit via the contractions indicated in Equation (6.11). Once again, this correspondence implies that any results on the expressive power of locally purified states hold also for local quantum circuits with alternating visible and hidden qudits.

6.3 Expressive power of tensor-network representations

In this section we present various relationships between the expressive power of all representations, which constitute the primary results of this work. The proofs of the propositions in this section can be found in the supplementary material in Ref. [278].

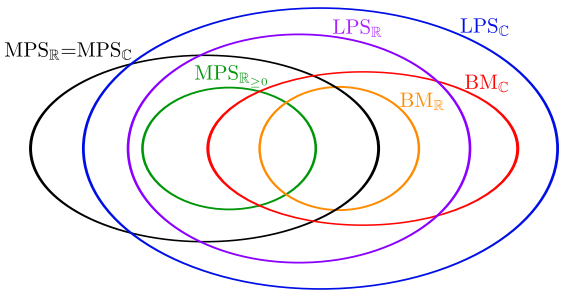


Figure 6.2: Representation of the sets of non-negative tensors that admit a given tensor-network factorization. In this figure we fix the different ranks of the different tensor networks to be equal.

For a given rank, there is a set of non-negative tensors that can be exactly represented by a given tensor network. These sets are represented in Fig. 6.2 for the case in which the ranks of the tensor networks are equal. The inclusion relationships between these sets can be characterized in terms of inequalities between the ranks, as detailed in Proposition 1.

Proposition 1. *For all non-negative tensors TT-rank$_{\mathbb{R}_{\geq 0}} \geq$ TT-rank$_{\mathbb{R}}$, Born-rank$_{\mathbb{R}} \geq$ Born-rank$_{\mathbb{C}}$, Born-rank$_{\mathbb{R}} \geq$ puri-rank$_{\mathbb{R}}$, Born-rank$_{\mathbb{C}} \geq$ puri-rank$_{\mathbb{C}}$, puri-rank$_{\mathbb{R}} \geq$ puri-rank$_{\mathbb{C}}$, TT-rank$_{\mathbb{R}_{\geq 0}} \geq$ puri-rank$_{\mathbb{R}}$, TT-rank$_{\mathbb{R}} =$ TT-rank$_{\mathbb{C}}$.*

Next, as detailed in Proposition 2, and summarized in Table 6.2, we continue by showing that all the inequalities of Proposition 1 can in fact be strict, and that for all other pairs of representations there exist probability distributions showing that neither rank can always be lower than the other. This shows that neither of the two corresponding sets of tensors can be included in the other. The main new result is the introduction of a matrix with non-negative rank strictly smaller than its complex Hadamard square-root rank, i.e. TT-rank$_{\mathbb{R}_{\geq 0}} <$ Born-rank$_{\mathbb{C}}$.

Proposition 2. *The ranks of all introduced tensor-network representations satisfy the properties contained in Table 6.2. Specifically, denoting by r_{row} (r_{column}) the rank appearing in the row (column), $<$ indicates that there exists a tensor satisfying $r_{row} < r_{column}$ and $<,>$ indicates that there exists both a tensor satisfying $r_{row} < r_{column}$ and another tensor satisfying $r_{column} > r_{row}$.*

Table 6.2: Results of Proposition 2

	TT-rank$_{\mathbb{R}}$	TT-rank$_{\mathbb{R}_{\geq 0}}$	Born-rank$_{\mathbb{R}}$	Born-rank$_{\mathbb{C}}$	puri-rank$_{\mathbb{R}}$	puri-rank$_{\mathbb{C}}$
TT-rank$_{\mathbb{R}}$	$=$	$<$	$<,>$	$<,>$	$<,>$	$<,>$
TT-rank$_{\mathbb{R}_{\geq 0}}$	$>$	$=$	$<,>$	$<,>$	$>$	$>$
Born-rank$_{\mathbb{R}}$	$<,>$	$<,>$	$=$	$>$	$>$	$>$
Born-rank$_{\mathbb{C}}$	$<,>$	$<,>$	$<$	$=$	$<,>$	$>$
puri-rank$_{\mathbb{R}}$	$<,>$	$<$	$<$	$<,>$	$=$	$>$
puri-rank$_{\mathbb{C}}$	$<,>$	$<$	$<$	$<$	$<$	$=$

We now answer the question: By how much do we need to increase the rank of a tensor network such that the set of tensors it can represent includes the set of tensors that can be represented by a different tensor network of a different rank? More specifically, consider a tensor that has rank r according to one representation and rank r' according to another. Can we bound the rank r as a function of the rank r' only? The results of Proposition 3, presented via Table 6.3, indicate that in many cases there is no such function - i.e. there exists a family of non-negative tensors, describing a family of probability distributions over N binary variables, with the property that as N goes to infinity r' remains constant, while r also goes to infinity.

Proposition 3. *The ranks of all introduced tensor-network representations satisfy the relationships without asterisk contained in Table 6.3. A function $g(x)$ denotes*

that for all non-negative tensors $r_{row} \leq g(r_{column})$. "No" indicates that there exists a family of probability distributions of increasing N with $d = 2$ and r_{column} constant, but such that r_{row} goes to infinity, i.e. that no such function can exist.

Table 6.3: Results of Proposition 3.

	TT-rank$_{\mathbb{R}}$	TT-rank$_{\mathbb{R}_{\geq 0}}$	Born-rank$_{\mathbb{R}}$	Born-rank$_{\mathbb{C}}$	puri-rank$_{\mathbb{R}}$	puri-rank$_{\mathbb{C}}$
TT-rank$_{\mathbb{R}}$	$=$	$\leq x$	$\leq x^2$	$\leq x^2$	$\leq x^2$	$\leq x^2$
TT-rank$_{\mathbb{R}_{\geq 0}}$	No	$=$	No	No	No	No
Born-rank$_{\mathbb{R}}$	No	No	$=$	No	No	No
Born-rank$_{\mathbb{C}}$	No	No*	$\leq x$	$=$	No*	No*
puri-rank$_{\mathbb{R}}$	No	$\leq x$	$\leq x$	$\leq 2x$	$=$	$\leq 2x$
puri-rank$_{\mathbb{C}}$	No	$\leq x$	$\leq x$	$\leq x$	$\leq x$	$=$

We conjecture that the relationships with an asterisk in Table 6.3 also hold. The existence of a family of matrices with constant non-negative rank but unbounded complex Hadamard square-root rank, together with the techniques introduced in the supplementary material in Ref. [278], would provide a proof of these conjectured results. It is also worth noting that lower-bounding this rank can be cast into the form of polynomial optimization problems amenable to hierarchies of convex relaxations [303], which may be useful to solve these conjectures. Proposition 3 indicates the existence of various families of non-negative tensors for which the rank of one representation remains constant, while the rank of another representation grows with the number of binary variables, however, the rate of this growth is not given. The following propositions provide details of the asymptotic growth of these ranks.

Proposition 4 ([299])**.** *There exists a family of non-negative tensors over $2N$ binary variables and constant TT-rank$_{\mathbb{R}}$=3 that have puri-rank$_{\mathbb{C}} = \Omega(N)$, and hence also puri-rank$_{\mathbb{C}}$, Born-rank$_{\mathbb{R}/\mathbb{C}}$ and TT-rank$_{\mathbb{R}_{\geq 0}} \geq \Omega(N)$.*

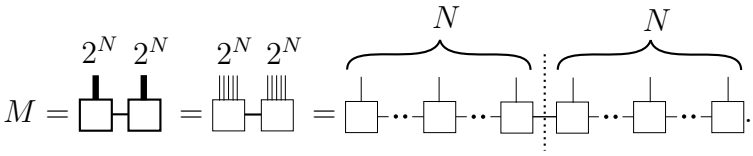
Proposition 5. *There exists a family of non-negative tensors over $2N$ binary variables and constant TT-rank$_{\mathbb{R}_{\geq 0}}$=2 (and hence also puri-rank$_{\mathbb{R}/\mathbb{C}} = 2$) that have Born-rank$_{\mathbb{R}} \geq \pi(2^{N+1})$, where $\pi(x)$ is the number of prime numbers up to x, which asymptotically satisfies $\pi(x) \sim x/\log(x)$.*

Proposition 6. *There exists a family of non-negative tensors over $2N$ binary variables and constant Born-rank$_{\mathbb{R}}$=2 (and hence also constant Born-rank$_{\mathbb{C}}$ and puri-rank$_{\mathbb{R}/\mathbb{C}}$) that have TT-rank$_{\mathbb{R}_{\geq 0}} \geq N$.*

Proposition 7. *There exists a family of non-negative tensors over $2N$ binary variables and constant Born-rank$_{\mathbb{C}}$=2 that have Born-rank$_{\mathbb{R}} \geq N$.*

As the techniques via which the results of Proposition 3 have been obtained are of interest, we provide a sketch of the proof for all "No" entries here (the full proofs can be found in the supplementary material in Ref. [278]). Assume that for a given pair of representations there exists a family of non-negative matrices

with the property that the rank r_{column} of one representation remains constant as a function of matrix dimension, while the rank r_{row} of the other representation grows. Now, consider such a matrix M of dimension $2^N \times 2^N$. The first step is to show that M can be unfolded into a tensor network of constant rank r_{column}, for $2N$ binary variables, such that M is a reshaping of the central bipartition of this tensor as



(6.12)

If the rank r_{row} of matrix M is large, the rank r_{row} of the corresponding tensor-network representation of the unfolded tensor will also be large. While above unfolding requires a particular matrix dimension, it is in fact possible to write any $N \times N$ matrix M as a submatrix of a $2^N \times 2^N$ matrix, to which the above unfolding strategy can then be used as a tool for leveraging matrix rank separations [302, 304–306] into tensor rank separations.

Finally, in order to discuss the significance of these results, note firstly that the TT-rank$_{\mathbb{R}}$ can be arbitrarily smaller than all other ranks, however, optimizing a real MPS to represent a probability distribution presents a problem since it is not clear how to impose positivity of the contracted tensor network [259, 301]. All other separations are relevant in practice since, as discussed in the following section, they apply to tensor networks that can be trained to represent probability distributions over many variables. Taken together, these results then show that LPS should be preferred over MPS$_{\mathbb{R}_{\geq 0}}$ or BM, since the puri-ranks will always be lower bounded compared to the other ranks. Additionally, complex BM should also be preferred to real BM as they can lead to an arbitrarily large reduction in the number of parameters of the tensor network. Note that because of the structure of the tensor networks we consider, these results also apply to more general tensor factorizations relying on a tree structure of the tensor network. How these results are affected if one considers approximate as opposed to exact representations remains an interesting problem that we leave for future work.

6.4 Learning algorithms

While the primary results of this work concern the expressive power of different tensor-network representations of probability distributions, these results are relevant in practice since MPS$_{\mathbb{R}_{\geq 0}}$, BM$_{\mathbb{R}/\mathbb{C}}$ and LPS$_{\mathbb{R}/\mathbb{C}}$ admit efficient learning algorithms, as shown in this section. In particular, given samples from a probability distribution, they can be trained to approximate this distribution through maximum likelihood estimation. Alternatively, these representations can also be used to compress a given non-negative tensor.

6.4.1 Maximum likelihood estimation

Consider first the setting in which one is given samples $\{\mathbf{x_i} = (X_1^i, \ldots, X_N^i)\}$ from a discrete multivariate distribution and would like to obtain an efficient approximation of this distribution. This can be done by minimizing the negative log-likelihood,

$$L = -\sum_i \log \frac{T_{\mathbf{x_i}}}{Z_T}, \tag{6.13}$$

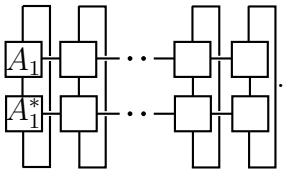
where i indexes training samples and $T_{\mathbf{x_i}}$ is given by the contraction of one of the tensor-network models we have introduced. The derivative of the log-likelihood with respect to a parameter w in the tensor network is given by

$$\partial_w L = -\sum_i \frac{\partial_w T_{\mathbf{x_i}}}{T_{\mathbf{x_i}}} - \frac{\partial_w Z_T}{Z_T}. \tag{6.14}$$

The negative log-likelihood can be minimized using a mini-batch gradient-descent algorithm. At each step of the optimization, the sum is computed over a batch of training instances. The parameters in the tensor network are then updated by a small step in the inverse direction of the gradient. Note that when using complex tensors, the derivatives are replaced by Wirtinger derivatives with respect to the conjugated tensor elements. This algorithm requires the computation of $T_{\mathbf{x_i}}$ and $\partial_w T_{\mathbf{x_i}}$ for a training instance, as well as of Z_T and $\partial_w Z_T$.

We first focus on the computation of these quantities for LPS. Since Born machines are LPS of purification dimension $\mu = 1$, they can directly use the same algorithm [269]. For an $\mathrm{LPS}_{\mathbb{C}}$ of puri-rank r (Equation (6.5)), the normalization Z_T can be computed by contracting the tensor network

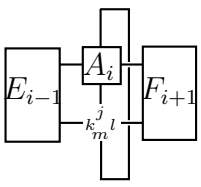
$$Z_T = \sum_{X_1,\ldots,X_N} T_{X_1,\ldots,X_N} =$$



. (6.15)

This contraction is performed in $\mathcal{O}(d\mu r^3 N)$ operations from left to right by contracting at each step the two vertical indices and then each of the two horizontal indices. During this contraction, intermediate results from the contraction of the first i tensors are stored in E_i, and the same procedure is repeated from the right with intermediate results of the contraction of the last $N - i$ tensors stored in F_{i+1}. The derivatives of the normalization for each tensor are then computed as

$$\frac{\partial Z_T}{\partial \bar{A}_{i,m}^{j,k,l}} =$$



, (6.16)

which also costs $\mathcal{O}(d\mu r^3 N)$ operations. Computing $T_{\mathbf{x_i}}$ for a training example and its derivative is done in the same way, except that the contracted index corresponding to an observed variable is now fixed to its observed value.

Note that here the training is done by computing the gradients of the log-likelihood over all tensors for each batch of training example and then updating all tensors at once in a gradient-descent optimization scheme. A different approach would be a DMRG-like algorithm where only a few tensors are updated at a time. The computation of Z_T and its derivative may be greatly simplified by using canonical forms [188].

The algorithm we use for training $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ is a variation of the one given above for LPS and is detailed in the supplementary material in Ref. [278]. $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ could also be trained using the expectation-maximization (EM) algorithm, but as BM and LPS use real or complex tensors, different algorithms are required. In Section 6.5 we will compare the algorithms described here with the EM algorithm for HMM. Note that in all these models not only the likelihood can be evaluated efficiently: marginals and correlation functions can be computed in a time linear in the number of variables, while exact samples from the distribution can also be generated efficiently [269, 307].

6.4.2 Approximate non-negative tensor factorization

Instead of approximating a distribution from samples, it might also be useful to compress a probability mass function P given in the form of a non-negative tensor. Since the original probability mass function has a number of parameters that is exponential in N, this is only possible for a small number of variables. It can be done by minimizing the Kullback–Leibler (KL) divergence

$$D(P||T/Z_T) = \sum_{X_1,\ldots,X_N} P_{X_1,\ldots,X_N} \log\left(\frac{P_{X_1,\ldots,X_N}}{T_{X_1,\ldots,X_N}/Z_T}\right), \tag{6.17}$$

where T is represented by a tensor-network model. The gradient of the KL-divergence can be obtained in the same way as the gradient of the log-likelihood and gradient-based optimization algorithms can then be used to solve this non-linear optimization problem. Note that for the case of matrices and $\mathrm{MPS}_{\mathbb{R}_{\geq 0}}$ more specific algorithms have been developed [308], and finding more efficient algorithms for factorizing a given tensor in the form of a BM or LPS represents an interesting problem that we leave for future work.

6.5 Numerical experiments

The results of Section 6.3 show that there exist probability distributions for which there are separations between the required ranks of the different tensor-network factorizations. Using the algorithms discussed in Section 6.4 we numerically investigate the extent to which these separations apply in both the setting of approximating a distribution from samples, and the setting of compressing given probability mass functions. All code, data sets and choice of hyperparameters are available in the provided repository [309].

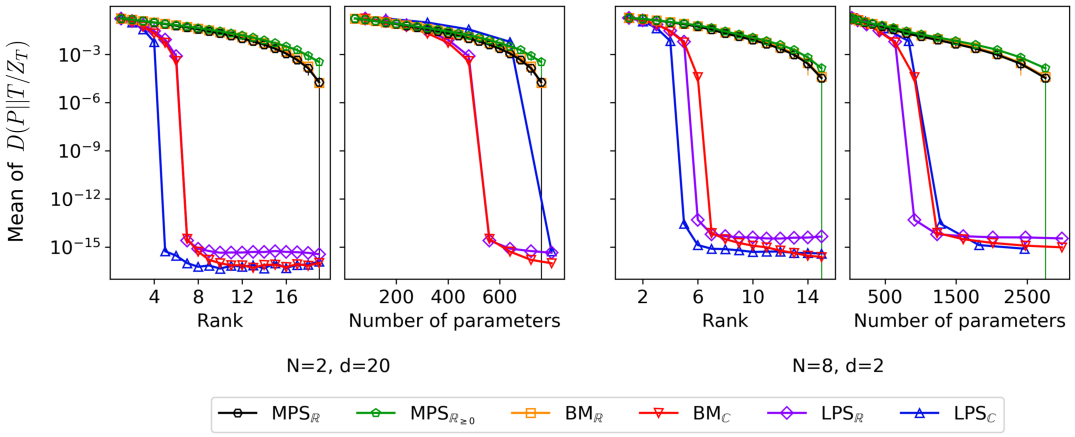


Figure 6.3: Mean of the minimum error of the approximation of 50 random tensors P with tensor networks of fixed rank, as a function of the rank or the number of (real) parameters. Left: 20×20 matrix. Right: tensor over 8 variables of dimension 2. The errors bars represent one standard deviation, and are omitted below 10^{-12}.

6.5.1 Random tensor factorizations

We first generate random probability mass functions P by generating a tensor with elements chosen uniformly in $[0, 1]$ and normalizing it. We then minimize the KL-divergence $D(P||T/Z_T)$, where T is the tensor defined by an MPS, BM or LPS with given rank r. We choose LPS to have a purification dimension of 2. Details of the optimization are available in the supplementary material in Ref. [278].

Results are presented in Fig. 6.3 for a 20×20 matrix and a tensor with 8 binary variables. They show that complex BM as well as real and complex LPS generically provide a better approximation to a tensor than an MPS or real BM, for fixed rank as well as for fixed number of real parameters.

6.5.2 Maximum likelihood estimation on realistic data sets

We now investigate how well the different tensor-network representations are able to learn from realistic data sets. We train $\text{MPS}_{\mathbb{R}_{\geq 0}}$, $\text{BM}_{\mathbb{R}}$, $\text{BM}_{\mathbb{C}}$, $\text{LPS}_{\mathbb{R}}$ and $\text{LPS}_{\mathbb{C}}$ (of purification dimension 2) using the algorithm of Section 6.4.1 on different data sets of categorical variables. Since we are interested in the expressive power of the different representations we use only training sets and no regularization.

The results in Fig. 6.4 show the best negative log-likelihood per sample obtained for each tensor network of fixed rank. As a comparison we also include the best negative log-likelihood obtained from an HMM trained using the Baum-Welch algorithm [313] on the same data. We observe that despite the different algorithm choice, the performance of the HMM and $\text{MPS}_{\mathbb{R}_{\geq 0}}$ are similar, as we could expect from their relationship. On all data sets, BM and LPS lead to significant improvements for the same rank over $\text{MPS}_{\mathbb{R}_{\geq 0}}$. Complex BM also outperform real BM, again as expected from the theoretical results.

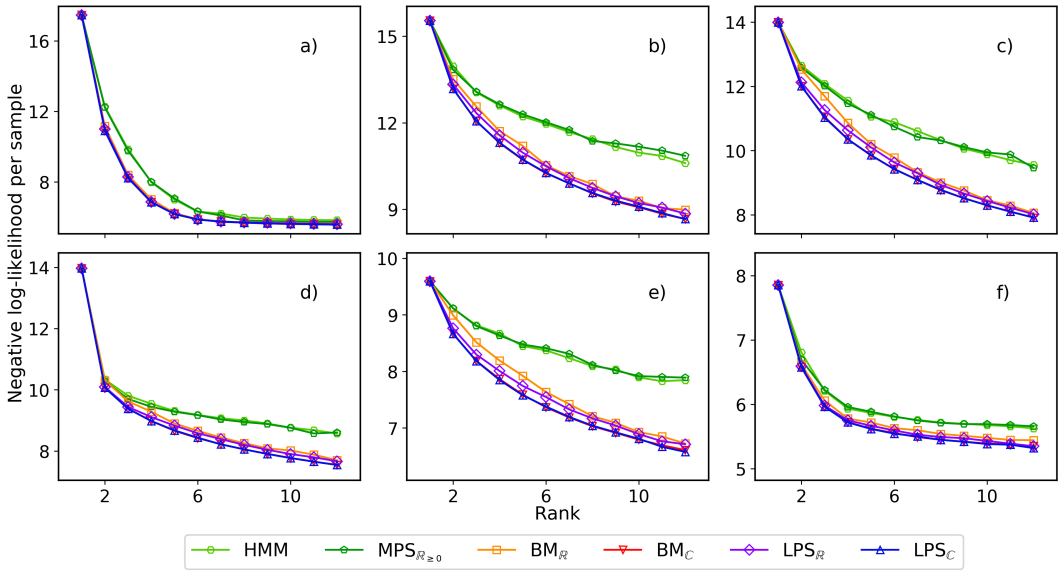


Figure 6.4: Maximum likelihood estimation with tensor networks and HMM on different data sets: a) biofam data set of family life states from the Swiss Household Panel biographical survey [310]; data sets from the UCI Machine Learning Repository [311]: b) Lymphography [312], c) SPECT Heart, d) Congressional Voting Records, e) Primary Tumor [312], f) Solar Flare.

6.6 Conclusion

We have characterized the expressive power of various tensor-network models of probability distributions, in the process enhancing the scope and applicability of the tensor-network toolbox within the broader context of learning algorithms. In particular, our analysis has concrete implications for model selection, suggesting that in generic settings LPS should be preferred over both hidden Markov models and Born machines. In the context of parameterized quantum circuits as probabilistic models, this implies that ancilla qubits can provide definitive advantages. Furthermore, our results prove that unexpectedly the use of complex tensors over real tensors can lead to an unbounded expressive advantage. This result provides impetus for further study of the potential benefits of complex numbers in other model classes, such as deep neural networks, where this question remains open. Additionally, this work contributes to the growing body of rigorous results concerning the expressive power of learning models, which have been obtained via tensor-network techniques. A formal understanding of the expressive power of state-of-the-art learning models is often elusive; it is hoped that both the techniques and spirit of this work can be used to add momentum to this program. Finally, through the formal relationship of LPS and Born machines to quantum circuits, our work provides a concrete foundation for both the development and analysis of quantum machine learning algorithms for near-term quantum devices.

Part V

Conclusions

In this thesis we contributed to the effort of developing new methods for the study of condensed matter systems, as well as to the adoption of these methods to tackle problems outside the physical real.

In Part II we focused on the characterization of quantum localization. First, we studied a many-body localized system where localization is induced by a disordered potential. We aimed at the description of Griffith effects in the vicinity of the phase transition and the mechanism for which localization emerges from disorder. We showed that the slow-operator technique can be employed not to only to find localized conserved charges, but also to compute statistical properties related to Griffith effects. The probability distributions associated to the eigenvalues corresponding to the slowest operators can be understood in the framework of Extreme Value Theory, we showed that the generalized extreme value distribution — the distribution which encompass all possible extreme value distributions — embeds information about the underlying physical system, its phase and dynamical properties. These findings give solid motivations for the use of the slow operator technique for the study of slow quantum dynamics and, at the same time, they provided a constructive way of building quasi-conserved local quantities (if they exist). The slow operator technique gives also an efficient way to explore the structure of the conserved charges and information about their dynamical properties. Extreme Value Theory has already proved itself as a powerful theory for the study of classical and quantum disordered systems. In this thesis, we showed how it can be used along with tensor-network techniques for the study of dynamical properties of many-body localized systems.

Many body localization has attracted considerable attention in the last fifteen years both in theory and experiments. This big effort has resulted into a large number of interesting open questions and, in particular, whether localization is possible in absence of disorder. In chapter 2 we dealt with one possible approach. Indeed, we showed that the quantum East model — a model inspired from classical structural glasses — exhibits a fast-to-slow transition throughout the spectrum. On the slow side of the transition we showed that, at least for small sizes, several eigenstates are exponentially localized. We analyzed in detail the localization of the ground state, and we used it to construct non-thermal excited states at arbitrary energy densities, some of which fulfill the area law of entanglement. We proved that an even larger class of weakly entangled states retain long memory of initial conditions under time evolution and, consequently, that the time required in order to entangle a region is exponential in the size of the region. These findings have interesting implications both for theoretical and practical purposes. They suggest the presence of a novel phase transition which dramatically affects the eigenstates, and thus the dynamics, of the system. Simulating quantum dynamical systems is an extremely challenging task, here we showed that the quantum East model is an exception. In particular, there exist a large class of states for which the simulation is efficient — yet non trivial — up to exponentially long time. On the other hand, the simplicity of the Hamiltonian sparks hopes for the experimental exploration of quantum localization and, at the same time, for the storage of quantum information. We extended our study to two-dimensional systems for which quantum localization is undeniably an important open question. Our DMRG simulations of the quantum

North-or-East model showed strong localization of the ground state. Extending the results from the one-dimensional model to the 2D case is an exciting research direction which may uncover new mechanisms for ergodicity breaking in higher dimensional lattices.

In Part III we investigated an opposite scenario. Our goal was to accelerate quantum thermodynamic protocols whilst keeping their dissipated work constant. To do so, we assumed we had perfect control over the interaction between the system and its surrounding environment, and of the internal energy of the system. Although, we considered a unitary evolution of the combined system and bath, we assumed weak interaction between them. This allowed us to perform a Markovian approximation for the thermalization properties of the system and, in particular, to assume that the thermalization time is decreased as the system-bath coupling is increased. We showed that by appropriately tuning the coupling between system and bath, it is possible to achieve faster or, equivalently, more efficient protocols. Our results were a combination of heuristics, rigorous analytics, and numerical simulations. In particular, we tested the "speed-ups to isothermality" for different systems, including ensembles of free bosons and fermions, and locally interacting spin models. The efficiency of our protocols interpolates between two established regimes: the Curzon-Ahlborn efficiency and the Carnot efficiency. We performed exact numerical simulations of the Caldeira-Legget model for bosonic degree of freedom coupled to a large but finite bath and the resonant-level model for a fermionic degree of freedom coupled to an infinite bath. The numerics are in good agreement with our results.

In Part IV, we investigated the interplay and relationships between machine learning and methods for condensed matter and quantum information. In chapter 4, we performed an in-depth comparison of Neural Network Quantum states and Tensor Network states. We combined analytical results and numerical simulations in order to characterize the descriptive capacity of different ansätze. We showed that a popular class of neural network states, the Restricted Boltzmann machine, can be exactly mapped to a subclass of tensor networks called String Bond states. This mapping highlights all the mathematical strengths and weaknesses of the Restricted Boltzmann machine and, moreover, it allows for simple generalizations. We showed that these states possess strong descriptive power by approximating the ground state of a chiral spin liquid via Variational Monte Carlo, including their correlations and topological order.

Restricted Boltzmann Machines are particular examples of probabilistic graphical models, popular techniques widely used in machine learning. Our connection from chapter 4 allowed us in chapter 5 to generalize the mapping between String Bond States and Restricted Boltzmann Machine to the broader classes of tensor networks and probabilistic graphical model. This relationship opens a wide variety of possibilities for applying tensor networks in machine learning. In particular, in chapter 5, we employed tensor networks to tackle supervised learning tasks such as image and sound recognition. We showed that they can perform extremely well and, if combined with state-of-the-art methods, they can reach performances comparable to modern deep learning techniques.

The exact correspondence between tensor networks and probabilistic graphical

models also allows for straightforward implementations of successful models from condensed matter to unsupervised learning tasks. In chapter 6 we employed tensor factorizations inspired from the simulation of one dimensional quantum systems to the approximation of probability distributions coming from data. First, we showed that Hidden Markov models can be mapped exactly to Matrix Product states with real and positive tensors. These constrains can be easily relax by imitating the probabilistic interpretation of quantum mechanics. In particular, we compared the expressive power of several models with tractable likelihood, we showed that by using complex or real tensors can induce an unbounded separation between different factorizations. And that there exists a particular choice inspired from the simulation of open quantum systems — the Locally Purified states — that possesses provably better expressive power than all the other models. In order to asses the practical usefulness of our results, we numerically approximated random probability distributions as well as distributions coming from real data. The simulations showed that these models often provide also practical advantages in approximation performances at comparable costs.

Our work contributes to the ever growing quest of connecting machine learning with ideas coming from quantum mechanics. Tensor networks are deeply rooted in our understanding of how a quantum mechanical wave functions behave. Their mathematical formulation is broad and robust. These connections might provide ground for developing more rigorous results in the field of machine learning which often advances by heuristics. At the same time, the intimate relationship between tensor networks and quantum circuits might provide possible solutions for testing machine learning algorithms on Noisy Intermediate-Scale Quantum computers.

Bibliography

[1] Kerson Huang. *Statistical Mechanics*. John Wiley & Sons, 2 edition, 1987.

[2] Paul Adrien Maurice Dirac and Ralph Howard Fowler. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 117(778):610–624, 1928.

[3] P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[4] Philip W Anderson. *More and Different*. WORLD SCIENTIFIC, 2011.

[5] Wikipedia contributors. Boltzmann distribution — Wikipedia, the free encyclopedia, 2019. [Online; accessed 9-January-2020].

[6] Wikipedia contributors. Monte carlo method — Wikipedia, the free encyclopedia, 2019. [Online; accessed 9-January-2020].

[7] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117 – 158, 2014.

[8] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96 – 192, 2011. January 2011 Special Issue.

[9] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008.

[10] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix Product State Representations. *arXiv e-prints*, pages quant–ph/0608197, Aug 2006.

[11] F. Verstraete and J. I. Cirac. Matrix product states represent ground states faithfully. *Phys. Rev. B*, 73:094423, Mar 2006.

[12] M B Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08024–P08024, aug 2007.

[13] B Pirvu, V Murg, J I Cirac, and F Verstraete. Matrix product operator representations. *New Journal of Physics*, 12(2):025012, feb 2010.

[14] Steven R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863–2866, Nov 1992.

[15] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91:147902, Oct 2003.

[16] Steven R. White and Adrian E. Feiguin. Real-time evolution using the density matrix renormalization group. *Phys. Rev. Lett.*, 93:076401, Aug 2004.

[17] 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, Nov 2004.

[18] Jutho Haegeman, J. Ignacio Cirac, Tobias J. Osborne, Iztok Pižorn, Henri Verschelde, and Frank Verstraete. Time-dependent variational principle for quantum lattices. *Phys. Rev. Lett.*, 107:070601, Aug 2011.

[19] N Schuch, M M Wolf, K G H Vollbrecht, and J I Cirac. On entropy growth and the hardness of simulating time evolution. *New J. Phys.*, 10(3):033032, mar 2008.

[20] Tobias J Osborne. Hamiltonian complexity. *Rep. Prog. Phys.*, 75(2):022001, jan 2012.

[21] Pasquale Calabrese and John Cardy. Evolution of entanglement entropy in one-dimensional systems. *J. Stat. Mech.*, 2005(04):P04010, apr 2005.

[22] Lev Vidmar, Lucas Hackl, Eugenio Bianchi, and Marcos Rigol. Entanglement entropy of eigenstates of quadratic fermionic hamiltonians. *Phys. Rev. Lett.*, 119:020601, Jul 2017.

[23] DJ Ashton, LO Hedges, and JP Garrahan. Fast simulation of facilitated spin models. *J. Stat. Mech.*, 2005.

[24] Jens Eisert, M Friesdorf, and C Gogolin. Quantum many-body systems out of equilibrium. *Nature Physics*, 11(February):7, 2014.

[25] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Adv. Phys.*, 65(3):239–362, 2016.

[26] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958.

[27] D.M. Basko, I.L. Aleiner, and B.L. Altshuler. Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states. *Ann. of Phys.*, 321(5):1126–1205, may 2006.

[28] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, Apr 2007.

[29] Rahul Nandkishore and David A. Huse. Many-body localization and thermalization in quantum statistical mechanics. *Annual Review of Condensed Matter Physics*, 6(1):15–38, 2015.

[30] Dmitry A. Abanin and Zlatko Papić. Recent progress in many-body localization. *Annalen der Physik*, 529(7):1700169, 2017.

[31] Rahul Nandkishore and David A. Huse. Many-body localization and thermalization in quantum statistical mechanics. *Annual Review of Condensed Matter Physics*, 6(1):15–38, 2015.

[32] Herbert B Callen. Thermodynamics and an introduction to thermostatistics, 1998.

[33] David Newman, Florian Mintert, and Ahsan Nazir. Performance of a quantum heat engine at strong reservoir coupling. *Phys. Rev. E*, 95:032139, Mar 2017.

[34] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert. Strong coupling corrections in quantum thermodynamics. *Phys. Rev. Lett.*, 120:120602, Mar 2018.

[35] Michael Wiedmann, Jürgen T Stockburger, and Joachim Ankerhold. Out-of-equilibrium operation of a quantum heat engine. *arXiv preprint arXiv:1903.11368*, 2019.

[36] David Newman, Florian Mintert, and Ahsan Nazir. A quantum limit to non-equilibrium heat engine performance imposed by strong system-reservoir coupling. *arXiv preprint arXiv:1906.09167*, 2019.

[37] Tim Schmiedl and Udo Seifert. Optimal finite-time processes in stochastic thermodynamics. *Phys. Rev. Lett.*, 98:108301, Mar 2007.

[38] David A. Sivak and Gavin E. Crooks. Thermodynamic metrics and optimal paths. *Phys. Rev. Lett.*, 108:190602, May 2012.

[39] Armen E. Allahverdyan, Karen V. Hovhannisyan, Alexey V. Melkikh, and Sasun G. Gevorkian. Carnot cycle at finite power: Attainability of maximal efficiency. *Phys. Rev. Lett.*, 111:050601, Aug 2013.

[40] Erik Torrontegui, Sara Ibáñez, Sofia Martínez-Garaot, Michele Modugno, Adolfo del Campo, David Guéry-Odelin, Andreas Ruschhaupt, Xi Chen, and Juan Gonzalo Muga. Shortcuts to adiabaticity. In *Advances In Atomic, Molecular, and Optical Physics*, pages 117–169. Elsevier, 2013.

[41] Jiawen Deng, Qing-hai Wang, Zhihao Liu, Peter Hänggi, and Jiangbin Gong. Boosting work characteristics and overall heat-engine performance via shortcuts to adiabaticity: Quantum and classical systems. *Phys. Rev. E*, 88:062122, Dec 2013.

[42] A. del Campo, J. Goold, and M. Paternostro. More bang for your buck: Super-adiabatic quantum engines. *Scientific Reports*, 4(1), August 2014.

[43] Mathieu Beau, Juan Jaramillo, and Adolfo del Campo. Scaling-up quantum heat engines efficiently via shortcuts to adiabaticity. *Entropy*, 18(5):168, April 2016.

[44] Jun Jing, Lian-Ao Wu, Marcelo S. Sarandy, and J. Gonzalo Muga. Inverse engineering control in open quantum systems. *Phys. Rev. A*, 88:053422, Nov 2013.

[45] Jie Song, Zi-Jing Zhang, Yan Xia, Xiu-Dong Sun, and Yong-Yuan Jiang. Fast coherent manipulation of quantum states in open systems. *Optics Express*, 24(19):21674, September 2016.

[46] Jun Jing, Marcelo S. Sarandy, Daniel A. Lidar, Da-Wei Luo, and Lian-Ao Wu. Eigenstate tracking in open quantum systems. *Phys. Rev. A*, 94:042131, Oct 2016.

[47] G Vacanti, R Fazio, S Montangero, G M Palma, M Paternostro, and V Vedral. Transitionless quantum driving in open quantum systems. *New Journal of Physics*, 16(5):053017, May 2014.

[48] Victor Mukherjee, Alberto Carlini, Andrea Mari, Tommaso Caneva, Simone Montangero, Tommaso Calarco, Rosario Fazio, and Vittorio Giovannetti. Speeding up and slowing down the relaxation of a qubit by optimal control. *Phys. Rev. A*, 88:062326, Dec 2013.

[49] Nishchay Suri, Felix C. Binder, Bhaskaran Muralidharan, and Sai Vinjanampathy. Speeding up thermalisation via open quantum system variational optimisation. *The European Physical Journal Special Topics*, 227(3-4):203–216, September 2018.

[50] Roie Dann, Ander Tobalina, and Ronnie Kosloff. Shortcut to equilibration of an open quantum system. *arXiv preprint arXiv:1812.08821*, 2018.

[51] L Dupays, IL Egusquiza, A del Campo, and A Chenu. Shortcuts in open quantum systems: Superadiabatic control of an open quantum oscillator. *arXiv preprint arXiv:1910.12088*, 2019.

[52] S Alipour, A Chenu, AT Rezakhani, and A del Campo. Shortcuts to adiabaticity in driven open quantum systems: Balanced gain and loss and non-markovian evolution. *arXiv preprint arXiv:1907.07460*, 2019.

[53] Ignacio A. Martínez, Artyom Petrosyan, David Guéry-Odelin, Emmanuel Trizac, and Sergio Ciliberto. Engineered swift equilibration of a brownian particle. *Nature Physics*, 12(9):843–846, May 2016.

[54] Anne Le Cunuder, Ignacio A. Martínez, Artyom Petrosyan, David Guéry-Odelin, Emmanuel Trizac, and Sergio Ciliberto. Fast equilibrium switch of a micro mechanical oscillator. *Applied Physics Letters*, 109(11):113502, September 2016.

[55] Marie Chupeau, Sergio Ciliberto, David Guéry-Odelin, and Emmanuel Trizac. Engineered swift equilibration for brownian objects: from underdamped to overdamped dynamics. *New Journal of Physics*, 20(7):075003, July 2018.

[56] Suriyanarayanan Vaikuntanathan and Christopher Jarzynski. Escorted free energy simulations: Improving convergence by reducing dissipation. *Phys. Rev. Lett.*, 100:190601, May 2008.

[57] Geng Li, H. T. Quan, and Z. C. Tu. Shortcuts to isothermality and nonequilibrium work relations. *Phys. Rev. E*, 96:012144, Jul 2017.

[58] Ayoti Patra and Christopher Jarzynski. Shortcuts to adiabaticity using flow fields. *New Journal of Physics*, 19(12):125009, December 2017.

[59] Roie Dann and Ronnie Kosloff. Quantum signatures in the quantum carnot cycle. *arXiv preprint arXiv:1906.06946*, 2019.

[60] Tamiro Villazon, Anatoli Polkovnikov, and Anushya Chandran. Swift heat transfer by fast-forward driving in open quantum systems. *arXiv preprint arXiv:1902.05964*, 2019.

[61] Ander Tobalina, Ion Lizuain, and J Gonzalo Muga. Vanishing efficiency of speeded-up quantum otto engines. *arXiv preprint arXiv:1906.07473*, 2019.

[62] Ken Funo, Neill Lambert, Bayan Karimi, Jukka Pekola, Yuta Masuyama, and Franco Nori. Speeding-up a quantum refrigerator via counter-diabatic driving. *arXiv preprint arXiv:1905.03480*, 2019.

[63] Alexander B Boyd, Ayoti Patra, Christopher Jarzynski, and James P Crutchfield. Shortcuts to thermodynamic computing: The cost of fast and faithful erasure. *arXiv preprint arXiv:1812.11241*, 2018.

[64] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[65] Ehsan Zahedinejad, Joydip Ghosh, and Barry C. Sanders. Designing high-fidelity single-shot three-qubit gates: A machine-learning approach. *Phys. Rev. Applied*, 6:054005, Nov 2016.

[66] Moritz August and Xiaotong Ni. Using recurrent neural networks to optimize dynamical decoupling for quantum memory. *Phys. Rev. A*, 95:012335, Jan 2017.

[67] Leonardo Banchi, Nicola Pancotti, and Sougato Bose. Quantum gate learning in qubit networks: Toffoli gate without time-dependent control. *npj Quant. Inf.*, **2**:16019, 2016.

[68] Giacomo Torlai and Roger G. Melko. Neural decoder for topological codes. *Phys. Rev. Lett.*, 119:030501, Jul 2017.

[69] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Many-body quantum state tomography with neural networks. *arxiv:1703.05334*, 2017.

[70] Evert P. L. van Nieuwenburg, Ye-Hua Liu, and Sebastian D. Huber. Learning phase transitions by confusion. *Nat. Phys.*, 13(5):435–439, May 2017.

[71] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nat. Phys.*, 13(5):431–434, May 2017.

[72] Peter Broecker, Juan Carrasquilla, Roger G. Melko, and Simon Trebst. Machine learning quantum phases of matter beyond the fermion sign problem. *Scientific Reports*, 7(1):8823, 2017.

[73] Lei Wang. Discovering phase transitions with unsupervised learning. *Phys. Rev. B*, 94:195105, Nov 2016.

[74] Louis-Francois Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, and Andrew J. Millis. Machine learning for many-body physics: The case of the Anderson impurity model. *Phys. Rev. B*, 90:155136, Oct 2014.

[75] Junwei Liu, Yang Qi, Zi Yang Meng, and Liang Fu. Self-learning Monte Carlo method. *Phys. Rev. B*, 95:041101, Jan 2017.

[76] Li Huang and Lei Wang. Accelerated Monte Carlo simulations with restricted Boltzmann machines. *Phys. Rev. B*, 95:035105, Jan 2017.

[77] Giacomo Torlai and Roger G. Melko. Learning thermodynamics with Boltzmann machines. *Phys. Rev. B*, 94:165134, Oct 2016.

[78] Hyungwon Kim, Mari Carmen Bañuls, J. Ignacio Cirac, Matthew B. Hastings, and David A. Huse. Slowest local operators in quantum spin chains. *Phys. Rev. E*, 92:012128, Jul 2015.

[79] David A. Huse, Rahul Nandkishore, and Vadim Oganesyan. Phenomenology of fully many-body-localized systems. *Phys. Rev. B*, 90(17):174202, 2014.

[80] Maksym Serbyn, Z. Papić, and Dmitry A. Abanin. Local conservation laws and the structure of the many-body localized states. *Phys. Rev. Lett.*, 111(12):127201, 2013.

[81] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008.

[82] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96 – 192, 2011. January 2011 Special Issue.

[83] Kartiek Agarwal, Sarang Gopalakrishnan, Michael Knap, Markus Müller, and Eugene Demler. Anomalous diffusion and griffiths effects near the many-body localization transition. *Phys. Rev. Lett.*, 114:160401, Apr 2015.

[84] Kartiek Agarwal, Ehud Altman, Eugene Demler, Sarang Gopalakrishnan, David A. Huse, and Michael Knap. Rare region effects and dynamics near the many-body localization transition. *Ann. Phys.*, 2017.

[85] Yevgeny Bar Lev, Guy Cohen, and David R. Reichman. Absence of diffusion in an interacting system of spinless fermions on a one-dimensional disordered lattice. *Phys. Rev. Lett.*, 114:100601, 2015.

[86] Marko Žnidarič, Antonello Scardicchio, and Vipin Kerala Varma. Diffusive and Subdiffusive Spin Transport in the Ergodic Phase of a Many-Body Localizable System. *Phys. Rev. Lett.*, 117(4):040601, 2016.

[87] Laurens de Haan and Ana Ferreira. Extreme value theory: An introduction. *Springer-Verlag New York*, 2006.

[88] David J. Luitz, Nicolas Laflorencie, and Fabien Alet. Many-body localization edge in the random-field heisenberg chain. *Phys. Rev. B*, 91:081103, Feb 2015.

[89] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, Apr 2007.

[90] Marcin Mierzejewski, Peter Prelovšek, and Toma ž Prosen. Identifying local and quasilocal conserved quantities in integrable systems. *Phys. Rev. Lett.*, 114:140601, Apr 2015.

[91] M. Mierzejewski, M. Kozarzewski, and P. Prelovsek. Counting local integrals of motion in disordered spinless-fermion and Hubbard chains. *ArXiv e-prints*, August 2017.

[92] C.-J. Lin and O. I. Motrunich. Explicit construction of quasi-conserved local operator of translationally invariant non-integrable quantum spin chain in prethermalization. *ArXiv e-prints*, September 2017.

[93] T. E. O'Brien, Dmitry A. Abanin, Guifre Vidal, and Z. Papić. Explicit construction of local conserved operators in disordered many-body systems. *Phys. Rev. B*, 94:144208, Oct 2016.

[94] Jean-Philippe Bouchaud and Marc Mézard. Universality classes for extreme-value statistics. *Journal of Physics A: Mathematical and General*, 30(23):7997, 1997.

[95] Giulio Biroli, Jean-Philippe Bouchaud, and Marc Potters. Extreme value problems in random matrix theory and other disordered systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07):P07019, 2007.

[96] Róbert Juhász, Yu-Cheng Lin, and Ferenc Iglói. Strong griffiths singularities in random systems and their relation to extreme value statistics. *Phys. Rev. B*, 73:224206, Jun 2006.

[97] Arvydas Astrauskas. From extreme values of i.i.d. random fields to extreme eigenvalues of finite-volume anderson hamiltonian. *Probab. Surveys*, 13:156–244, 2016.

[98] A. K. Kulshreshtha, A. Pal, T. B. Wahl, and S. H. Simon. Behaviour of l-bits near the many-body localization transition. *ArXiv e-prints*, July 2017.

[99] M. Goihl, M. Gluza, C. Krumnow, and J. Eisert. Construction of exact constants of motion and effective models for many-body localized systems. *ArXiv e-prints*, July 2017.

[100] Ronen Vosk, David A. Huse, and Ehud Altman. Theory of the many-body localization transition in one-dimensional systems. *Phys. Rev. X*, 5:031032, Sep 2015.

[101] Andrew C. Potter, Romain Vasseur, and S. A. Parameswaran. Universal properties of many-body delocalization transitions. *Phys. Rev. X*, 5:031033, Sep 2015.

[102] Sarang Gopalakrishnan, Markus Müller, Vedika Khemani, Michael Knap, Eugene Demler, and David A. Huse. Low-frequency conductivity in many-body localized systems. *Phys. Rev. B*, 92:104202, Sep 2015.

[103] David Pekker, Bryan K. Clark, Vadim Oganesyan, and Gil Refael. Fixed points of wegner-wilson flows and many-body localization. *Phys. Rev. Lett.*, 119(7):075701, 2017.

[104] Rahul Nandkishore, Sarang Gopalakrishnan, and David A. Huse. Spectral features of a many-body-localized system weakly coupled to a bath. *Phys. Rev. B*, 90:064203, Aug 2014.

[105] Mark H Fischer, Mykola Maksymenko, and Ehud Altman. Dynamics of a many-body-localized system coupled to a bath. *Phys. Rev. Lett.*, 116(16):160401, 2016.

[106] Emanuele Levi, Markus Heyl, Igor Lesanovsky, and Juan P Garrahan. Robustness of many-body localization in the presence of dissipation. *Phys. Rev. Lett.*, 116(23):237203, 2016.

[107] Mariya V Medvedyeva, Tomaž Prosen, and Marko Žnidarič. Influence of dephasing on many-body localization. *Phys. Rev. B*, 93(9):094205, 2016.

[108] Sarang Gopalakrishnan, K. Ranjibul Islam, and Michael Knap. Noise-induced subdiffusion in strongly localized quantum systems. *Phys. Rev. Lett.*, 119(4):046601, 2017.

[109] Merlijn van Horssen, Emanuele Levi, and Juan P. Garrahan. Dynamics of many-body localization in a translation-invariant quantum glass model. *Phys. Rev. B*, 92:100305, Sep 2015.

[110] Philip Crowley. PhD thesis, University College London, 2017.

[111] Philip Crowley, Andrew Green, and Vadim Oganesyan. *In preparation*, 2019.

[112] J. Jäckle and S. Z. Eisinger. A hierarchically constrained kinetic ising model. *Z. fur Phys. B*, 85, 1991.

[113] F. Ritort and P. Sollich. Glassy dynamics of kinetically constrained models. *Adv. Phys.*, 52(4):219–342, 2003.

[114] Juan P. Garrahan. Aspects of non-equilibrium in classical and quantum systems: Slow relaxation and glasses, dynamical large deviations, quantum non-ergodicity, and open quantum dynamics. *Physica A*, 504:130–154, 2018.

[115] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić. Weak ergodicity breaking from quantum many-body scars. *Nature Physics*, 14(7):745–749, 2018.

[116] Zhihao Lan, Merlijn van Horssen, Stephen Powell, and Juan P. Garrahan. Quantum slow relaxation and metastability due to dynamical constraints. *Phys. Rev. Lett.*, 121:040603, Jul 2018.

[117] Johannes Feldmeier, Frank Pollmann, and Michael Knap. Emergent glassy dynamics in a quantum dimer model. *Phys. Rev. Lett.*, 123:040601, Jul 2019.

[118] Daniel S. Rokhsar and Steven A. Kivelson. Superconductivity and the quantum hard-core dimer gas. *Phys. Rev. Lett.*, 61:2376–2379, Nov 1988.

[119] Claudio Castelnovo, Claudio Chamon, Christopher Mudry, and Pierre Pujol. From quantum mechanics to classical statistical physics: Generalized rokhsar–kivelson hamiltonians and the "stochastic matrix form" decomposition. *Ann. of Phys.*, 318(2):316 – 344, 2005.

[120] P. Sollich and M. R. Evans. Glassy time-scale divergence and anomalous coarsening in a kinetically constrained spin chain. *Phys. Rev. Lett.*, 83:3238, 1999.

[121] Juan P. Garrahan and David Chandler. Geometrical explanation and scaling of dynamical heterogeneities in glass forming systems. *Phys. Rev. Lett.*, 89(3):035704, Jul 2002.

[122] Alessandra Faggionato, Fabio Martinelli, Cyril Roberto, and Cristina Toninelli. The East model: recent results and new progresses. 2012.

[123] P Chleboun, A Faggionato, and F Martinelli. Time scale separation in the low temperature east model: rigorous results. *J. Stat. Mech.*, 2013(04):L04001, apr 2013.

[124] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. Dynamical first-order phase transition in kinetically constrained models of glasses. *Phys. Rev. Lett.*, 98:195702, May 2007.

[125] Mari Carmen Bañuls and Juan P. Garrahan. Using matrix product states to study the dynamical large deviations of kinetically constrained models. 2019.

[126] Hyungwon Kim, Mari Carmen Bañuls, J. Ignacio Cirac, Matthew B. Hastings, and David A. Huse. Slowest local operators in quantum spin chains. *Phys. Rev. E*, 92:012128, Jul 2015.

[127] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[128] Xiongjie Yu, David Pekker, and Bryan K. Clark. Finding matrix product state representations of highly excited eigenstates of many-body localized hamiltonians. *Phys. Rev. Lett.*, 118:017201, Jan 2017.

[129] Jun John Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, Reading, MA, 1994.

[130] Alexander L Fetter and John Dirk Walecka. *Quantum Theory of Many-Particle Systems*. McGraw-Hill, New York, 1971.

[131] A maximal bond dimension of 1000 allows the truncation error to be below 10^{-14}.

[132] Paul Fendley, K. Sengupta, and Subir Sachdev. Competing density-wave orders in a one-dimensional hard-boson model. *Phys. Rev. B*, 69:075106, Feb 2004.

[133] Igor Lesanovsky. Many-body spin interactions and the ground state of a dense rydberg lattice gas. *Phys. Rev. Lett.*, 106:025301, Jan 2011.

[134] C. Ates, J. P. Garrahan, and I. Lesanovsky. Thermalization of a strongly interacting closed spin system: From coherent many-body dynamics to a fokker-planck equation. *Phys. Rev. Lett.*, 108:110603, Mar 2012.

[135] Abhinav Prem, Jeongwan Haah, and Rahul Nandkishore. Glassy quantum dynamics in translation invariant fracton models. *Phys. Rev. B*, 95:155133, Apr 2017.

[136] Rahul M. Nandkishore and Michael Hermele. Fractons. *Annu. Rev. Condens. Matter Phys.*, 10:295–313, Mar 2019.

[137] Vedika Khemani and Rahul Nandkishore. Local constraints can globally shatter Hilbert space: a new route to quantum information protection. Apr 2019.

[138] Vedika Khemani, Michael Hermele, and Rahul M. Nandkishore. Localization from shattering: higher dimensions and physical realizations. 2019.

[139] Tibor Rakovszky, Pablo Sala, Ruben Verresen, Michael Knap, and Frank Pollmann. Statistical localization: from strong fragmentation to strong edge modes. Oct 2019.

[140] Pablo Sala, Tibor Rakovszky, Ruben Verresen, Michael Knap, and Frank Pollmann. Ergodicity-breaking arising from Hilbert space fragmentation in dipole-conserving Hamiltonians. Apr 2019.

[141] Rubem Mondaini, Krishnanand Mallayya, Lea F. Santos, and Marcos Rigol. Comment on "systematic construction of counterexamples to the eigenstate thermalization hypothesis". *Phys. Rev. Lett.*, 121:038901, Jul 2018.

[142] Naoto Shiraishi and Takashi Mori. Shiraishi and mori reply. *Phys. Rev. Lett.*, 121:038902, Jul 2018.

[143] James M Hickey, Sam Genway, and Juan P Garrahan. Signatures of many-body localisation in a system without disorder and the relation to a glass transition. *J. Stat. Mech.*, 2016(5):054047, may 2016.

[144] The Hamiltoanian of the quantum North-or-East model is defined as

$$H_{2D} = -\frac{1}{2}\sum_{i,j} n_{i,j}\left[e^{-s}\left(\sigma^x_{i+1,j} + \sigma^x_{i,j+1}\right) - \mathbb{1}\right].$$

The boundary conditions are chosen such that the $(1,1)$-site is pinned at 1, and an additional field n is placed on the opposite corner.

[145] Massimiliano Esposito, Ryoichi Kawai, Katja Lindenberg, and Christian Van den Broeck. Efficiency at maximum power of low-dissipation carnot engines. *Phys. Rev. Lett.*, 105:150603, Oct 2010.

[146] Paolo Abiuso and Martí Perarnau-Llobet. Optimal cycles for low-dissipation heat engines. *arXiv preprint arXiv:1907.02939*, 2019.

[147] Tim Schmiedl and Udo Seifert. Optimal finite-time processes in stochastic thermodynamics. *Phys. Rev. Lett.*, 98:108301, Mar 2007.

[148] Dibyendu Mandal and Christopher Jarzynski. Analysis of slow transitions between nonequilibrium steady states. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(6):063204, June 2016.

[149] Vasco Cavina, Andrea Mari, and Vittorio Giovannetti. Slow dynamics and thermodynamics of open quantum systems. *Phys. Rev. Lett.*, 119:050601, Aug 2017.

[150] J. Eisert, M. Friesdorf, and C. Gogolin. Quantum many-body systems out of equilibrium. *Nature Physics*, 11(2):124–130, February 2015.

[151] A. O. Caldeira and A. J. Leggett. Influence of dissipation on quantum tunneling in macroscopic systems. *Phys. Rev. Lett.*, 46:211–214, Jan 1981.

[152] A. O. Caldeira and A. J. Leggett. Path integral approach to quantum brownian motion. *Physica A*, 121(3):587–616, sep 1983.

[153] María Florencia Ludovico, Jong Soo Lim, Michael Moskalets, Liliana Arrachea, and David Sánchez. Dynamical energy transfer in ac-driven quantum systems. *Phys. Rev. B*, 89:161306, Apr 2014.

[154] Massimiliano Esposito, Maicol A. Ochoa, and Michael Galperin. Quantum thermodynamics: A nonequilibrium green's function approach. *Phys. Rev. Lett.*, 114:080602, Feb 2015.

[155] Massimiliano Esposito, Maicol A. Ochoa, and Michael Galperin. Nature of heat in strongly coupled open quantum systems. *Phys. Rev. B*, 92:235440, Dec 2015.

[156] Anton Bruch, Mark Thomas, Silvia Viola Kusminskiy, Felix von Oppen, and Abraham Nitzan. Quantum thermodynamics of the driven resonant level model. *Phys. Rev. B*, 93:115318, Mar 2016.

[157] Patrick Haughian, Massimiliano Esposito, and Thomas L. Schmidt. Quantum thermodynamics of the resonant-level model with driven system-bath coupling. *Phys. Rev. B*, 97:085435, Feb 2018.

[158] Michele Campisi, Sergey Denisov, and Peter Hänggi. Geometric magnetism in open quantum systems. *Phys. Rev. A*, 86:032114, Sep 2012.

[159] Thiago V Acconcia, Marcus VS Bonança, and Sebastian Deffner. Shortcuts to adiabaticity from linear response theory. *Physical Review E*, 92(4):042148, 2015.

[160] María Florencia Ludovico, Francesca Battista, Felix von Oppen, and Liliana Arrachea. Adiabatic response and quantum thermoelectrics for ac-driven quantum systems. *Phys. Rev. B*, 93:075136, Feb 2016.

[161] Patrick R. Zulkowski and Michael R. DeWeese. Optimal control of overdamped systems. *Phys. Rev. E*, 92:032117, Sep 2015.

[162] Matteo Scandi and Martí Perarnau-Llobet. Thermodynamic length in open quantum systems. *Quantum*, 3:197, October 2019.

[163] Harry JD Miller, Matteo Scandi, Janet Anders, and Martí Perarnau-Llobet. Work fluctuations in slow processes: quantum signatures and optimal control. *arXiv preprint arXiv:1905.07328*, 2019.

[164] J. Nulton, P. Salamon, B. Andresen, and Qi Anmin. Quasistatic processes as step equilibrations. *The Journal of Chemical Physics*, 83:334, 1985.

[165] G. E. Crooks. Measuring thermodynamic length. *Phys. Rev. Lett.*, 99:100602, Sep 2007.

[166] H. P. Breuer and F. Petruccione. *The theory of open quantum systems*. Oxford University Press, Great Clarendon Street, 2002.

[167] V. Cavina, A. Mari, and V. Giovannetti. Slow dynamics and thermodynamics of open quantum systems. *Phys. Rev. Lett.*, 119:050601, Aug 2017.

[168] Ángel Rivas, A Douglas K Plato, Susana F Huelga, and Martin B Plenio. Markovian master equations: a critical study. *New Journal of Physics*, 12(11):113032, November 2010.

[169] Ahsan Nazir and Gernot Schaller. The reaction coordinate mapping in quantum thermodynamics. In F. Binder, L. Correa, C. Gogolin, J. Anders, and Adesso G., editors, *Thermodynamics in the Quantum Regime*, pages 551–577. Springer International Publishing, 2018.

[170] Albert W. Marshall and Ingram Olkin. Inequalities for the trace function. *aequationes mathematicae*, 29(1):36–39, Dec 1985.

[171] Philipp Strasberg, Gernot Schaller, Neill Lambert, and Tobias Brandes. Nonequilibrium thermodynamics in the strong coupling and non-markovian regime based on a reaction coordinate mapping. *New Journal of Physics*, 18(7):073007, July 2016.

[172] Philipp Strasberg, Gernot Schaller, Thomas L. Schmidt, and Massimiliano Esposito. Fermionic reaction coordinates and their application to an autonomous maxwell demon in the strong-coupling regime. *Physical Review B*, 97(20), May 2018.

[173] Sebastian Restrepo, Javier Cerrillo, Philipp Strasberg, and Gernot Schaller. From quantum heat engines to laser cooling: Floquet theory beyond the born–markov approximation. *New Journal of Physics*, 20(5):053063, May 2018.

[174] Javier Prior, Alex W. Chin, Susana F. Huelga, and Martin B. Plenio. Efficient simulation of strong system-environment interactions. *Phys. Rev. Lett.*, 105:050404, Jul 2010.

[175] A. Strathearn, P. Kirton, D. Kilda, J. Keeling, and B. W. Lovett. Efficient non-Markovian quantum dynamics using time-evolving matrix product operators. *Nat. Commun.*, 9(1), aug 2018.

[176] Alejandro D. Somoza, Oliver Marty, James Lim, Susana F. Huelga, and Martin B. Plenio. Dissipation-assisted matrix product factorization. *Phys. Rev. Lett.*, 123:100502, Sep 2019.

[177] Ken Funo, Jing-Ning Zhang, Cyril Chatou, Kihwan Kim, Masahito Ueda, and Adolfo del Campo. Universal work fluctuations during shortcuts to adiabaticity by counterdiabatic driving. *Physical Review Letters*, 118(10), March 2017.

[178] Steve Campbell and Sebastian Deffner. Trade-off between speed and cost in shortcuts to adiabaticity. *Physical Review Letters*, 118(10), March 2017.

[179] Obinna Abah, Ricardo Puebla, Anthony Kiely, Gabriele De Chiara, Mauro Paternostro, and Steve Campbell. Energetic cost of quantum control protocols. *New Journal of Physics*, 21(10):103048, October 2019.

[180] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[181] Zi Cai. Approximating quantum many-body wave-functions using artificial neural networks. *arxiv:1704.05148*, 2017.

[182] Hiroki Saito. Solving the Bose-Hubbard model with machine learning. *J. Phys. Soc. Japan*, 86(9):093001, 2017.

[183] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985.

[184] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nat. Commun.*, **8**(662):662, 2017.

[185] Yichen Huang and Joel E. Moore. Neural network representation of tensor network and chiral states. *arxiv:1701.06246*, 2017.

[186] M. B. Hastings. An area law for one-dimensional quantum systems. *J. Stat. Mech.*, **2007**(01):P08024, 2007.

[187] Frank Verstraete, Valentin Murg, and J. Ignacio Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Adv. Phys.*, 57(2):143–224, 2008.

[188] Ulrich Schollwöck. The density-matrix renormalization group in the age of matrix product states. *Ann. Phys.*, 326(1):96 – 192, 2011.

[189] Steven R. White. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B*, 48:10345–10356, Oct 1993.

[190] Frank Verstraete and J Ignacio Cirac. Renormalization algorithms for quantum-many body systems in two and higher dimensions. *arxiv:0407066*, 2004.

[191] Norbert Schuch, Michael M. Wolf, Frank Verstraete, and J. Ignacio Cirac. Computational complexity of projected entangled pair states. *Phys. Rev. Lett.*, 98:140506, Apr 2007.

[192] Andrej Gendiar and Tomotoshi Nishino. Latent heat calculation of the three-dimensional $q = 3$, 4, and 5 potts models by the tensor product variational approach. *Phys. Rev. E*, 65:046702, 2002.

[193] Fabio Mezzacapo, Norbert Schuch, Massimo Boninsegni, and J. Ignacio Cirac. Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo. *New J. Phys.*, 11(8):083026, 2009.

[194] Hitesh J. Changlani, Jesse M. Kinder, C. J. Umrigar, and Garnet Kin-Lic Chan. Approximating strongly correlated wave functions with correlator product states. *Phys. Rev. B*, 80:245116, 2009.

[195] Norbert Schuch, Michael M. Wolf, Frank Verstraete, and J. Ignacio Cirac. Simulation of quantum many-body systems with strings of operators and Monte Carlo tensor contractions. *Phys. Rev. Lett.*, 100:040501, 2008.

[196] Alessandro Sfondrini, Javier Cerrillo, Norbert Schuch, and J. Ignacio Cirac. Simulating two- and three-dimensional frustrated quantum systems with string-bond states. *Phys. Rev. B*, 81:214426, 2010.

[197] A. W. Sandvik and G. Vidal. Variational quantum Monte Carlo simulations with tensor-network states. *Phys. Rev. Lett.*, 99:220602, Nov 2007.

[198] A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Ann. Phys.*, 303(1):2 – 30, 2003.

[199] 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, Jun 2006.

[200] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Exact machine learning topological states. *arxiv:1609.09060*, 2016.

[201] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted Boltzmann machines and tensor network states. *Phys. Rev. B*, 97:085104, 2018.

[202] D. C. Tsui, H. L. Stormer, and A. C. Gossard. Two-dimensional magneto-transport in the extreme quantum limit. *Phys. Rev. Lett.*, 48:1559–1562, May 1982.

[203] Michael Levin and Ady Stern. Fractional topological insulators. *Phys. Rev. Lett.*, 103:196803, Nov 2009.

[204] DN Sheng, Zheng-Cheng Gu, Kai Sun, and L Sheng. Fractional quantum Hall effect in the absence of landau levels. *Nat. Commun.*, 2:389, 2011.

[205] Titus Neupert, Luiz Santos, Claudio Chamon, and Christopher Mudry. Fractional quantum Hall states at zero magnetic field. *Phys. Rev. Lett.*, 106:236804, Jun 2011.

[206] Yi-Fei Wang, Zheng-Cheng Gu, Chang-De Gong, and D. N. Sheng. Fractional quantum Hall effect of hard-core bosons in topological flat bands. *Phys. Rev. Lett.*, 107:146803, Sep 2011.

[207] Kai Sun, Zhengcheng Gu, Hosho Katsura, and S. Das Sarma. Nearly flatbands with nontrivial topology. *Phys. Rev. Lett.*, 106:236803, Jun 2011.

[208] N. Regnault and B. Andrei Bernevig. Fractional Chern insulator. *Phys. Rev. X*, 1:021014, Dec 2011.

[209] Anne E B Nielsen, Germán Sierra, and J Ignacio Cirac. Local models of fractional quantum Hall states in lattices and physical implementation. *Nat. Commun.*, **4**(2864):2864, 2013.

[210] B. Bauer, L. Cincio, B. P. Keller, M. Dolfi, G. Vidal, S. Trebst, and A. W. W. Ludwig. Chiral spin liquid and emergent anyons in a kagome lattice Mott insulator. *Nat. Commun.*, **5**(5137):5137, 2014.

[211] Yin-Chen He, D. N. Sheng, and Yan Chen. Chiral spin liquid in a frustrated anisotropic kagome Heisenberg model. *Phys. Rev. Lett.*, 112:137202, Apr 2014.

[212] Shou-Shu Gong, Whei Zhu, and D. N. Sheng. Emergent chiral spin liquid: Fractional quantum Hall effect in a kagome Heisenberg model. *Scientific Reports*, 4:6317, Apr 2014.

[213] G. Baskaran, Z. Zou, and P.W. Anderson. The resonating valence bond state and high-Tc superconductivity — a mean field theory. *Solid State Communications*, 63(11):973 – 976, 1987.

[214] Ian Affleck, Z. Zou, T. Hsu, and P. W. Anderson. $SU(2)$ gauge symmetry of the large-U limit of the Hubbard model. *Phys. Rev. B*, 38:745–747, Jul 1988.

[215] Xiao-Gang Wen. Projective construction of non-Abelian quantum Hall liquids. *Phys. Rev. B*, 60:8827–8838, Sep 1999.

[216] Wen-Jun Hu, Wei Zhu, Yi Zhang, Shoushu Gong, Federico Becca, and D. N. Sheng. Variational Monte Carlo study of a chiral spin liquid in the extended Heisenberg model on the kagome lattice. *Phys. Rev. B*, 91:041124, Jan 2015.

[217] V. Kalmeyer and R. B. Laughlin. Equivalence of the resonating-valence-bond and fractional quantum Hall states. *Phys. Rev. Lett.*, 59:2095–2098, Nov 1987.

[218] T. B. Wahl, H.-H. Tu, N. Schuch, and J. I. Cirac. Projected entangled-pair states can describe chiral topological states. *Phys. Rev. Lett.*, 111:236805, Dec 2013.

[219] J. Dubail and N. Read. Tensor network trial states for chiral topological phases in two dimensions and a no-go theorem in any dimension. *Phys. Rev. B*, 92:205307, Nov 2015.

[220] Shuo Yang, Thorsten B. Wahl, Hong-Hao Tu, Norbert Schuch, and J. Ignacio Cirac. Chiral projected entangled-pair state with topological order. *Phys. Rev. Lett.*, 114:106803, Mar 2015.

[221] Didier Poilblanc, J. Ignacio Cirac, and Norbert Schuch. Chiral topological spin liquids with projected entangled pair states. *Phys. Rev. B*, 91:224431, Jun 2015.

[222] Shuo Yang, Thorsten B. Wahl, Hong-Hao Tu, Norbert Schuch, and J. Ignacio Cirac. Chiral projected entangled-pair state with topological order. *Phys. Rev. Lett.*, 114:106803, Mar 2015.

[223] W. L. McMillan. Ground state of liquid He^4. *Phys. Rev.*, 138:A442–A451, Apr 1965.

[224] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, Jan 2001.

[225] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**(6):1087, 1953.

[226] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1):97–109, 1970.

[227] Sandro Sorella. Generalized Lanczos algorithm for variational quantum Monte Carlo. *Phys. Rev. B*, 64:024512, Jun 2001.

[228] Sandro Sorella. Wave function optimization in the variational Monte Carlo method. *Phys. Rev. B*, 71:241103, Jun 2005.

[229] C. J. Umrigar and Claudia Filippi. Energy and variance optimization of many-body wave functions. *Phys. Rev. Lett.*, 94:150201, Apr 2005.

[230] M. P. Nightingale and Vilen Melik-Alaverdian. Optimization of ground- and excited-state wave functions and van der Waals clusters. *Phys. Rev. Lett.*, 87:043401, Jul 2001.

[231] Julien Toulouse and C. J. Umrigar. Optimization of quantum Monte Carlo wave functions by energy minimization. *J. Chem. Phys.*, 126(8):084102, 2007.

[232] C. J. Umrigar, Julien Toulouse, Claudia Filippi, S. Sorella, and R. G. Hennig. Alleviation of the fermion-sign problem by optimization of many-body wave functions. *Phys. Rev. Lett.*, 98:110201, Mar 2007.

[233] Robert Jastrow. Many-body problem with strong forces. *Phys. Rev.*, 98:1479–1484, Jun 1955.

[234] Steven R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863–2866, Nov 1992.

[235] Shoudan Liang and Hanbin Pang. Approximate diagonalization using the density matrix renormalization-group method: A two-dimensional-systems perspective. *Phys. Rev. B*, 49:9214–9217, Apr 1994.

[236] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, May 2017.

[237] Chanchal K. Majumdar and Dipan K. Ghosh. On next nearest neighbor interaction in linear chain. i. *J. Math. Phys.*, 10(8):1388–1398, 1969.

[238] Ian Affleck, Tom Kennedy, Elliott H. Lieb, and Hal Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.*, 59:799–802, Aug 1987.

[239] Xun Gao. private communication.

[240] Olga Sikora, Hsueh-Wen Chang, Chung-Pin Chou, Frank Pollmann, and Ying-Jer Kao. Variational Monte Carlo simulations using tensor-product projected states. *Phys. Rev. B*, 91:165113, Apr 2015.

[241] Anne E. B. Nielsen, J. Ignacio Cirac, and Germán Sierra. Laughlin spin-liquid states on lattices obtained from conformal field theory. *Phys. Rev. Lett.*, 108:257206, Jun 2012.

[242] Z. C. Tu. Stochastic heat engine with the consideration of inertial effects and shortcuts to adiabaticity. *Phys. Rev. E*, 89:052148, May 2014.

[243] Ivan Glasser, J. Ignacio Cirac, Germán Sierra, and Anne E. B. Nielsen. Lattice effects on Laughlin wave functions and parent Hamiltonians. *Phys. Rev. B*, 94:245104, Dec 2016.

[244] Anne E. B. Nielsen, J. Ignacio Cirac, and Germán Sierra. Quantum spin Hamiltonians for the $SU(2)_k$ WZW model. *J. Stat. Mech.*, **2011**(11):P11014, 2011.

[245] Eric Neuscamman, C. J. Umrigar, and Garnet Kin-Lic Chan. Optimizing large parameter sets in variational quantum Monte Carlo. *Phys. Rev. B*, 85:045103, Jan 2012.

[246] Shunichi Amari. Natural gradient works efficiently in learning. *Neural Comput.*, 10(2):251–276, 1998.

[247] Matthew B. Hastings, Iván González, Ann B. Kallin, and Roger G. Melko. Measuring Renyi entanglement entropy in quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 104:157201, Apr 2010.

[248] Julia Wildeboer and N. E. Bonesteel. Spin correlations and topological entanglement entropy in a non-Abelian spin-one spin liquid. *Phys. Rev. B*, 94:045125, Jul 2016.

[249] Alexei Kitaev and John Preskill. Topological entanglement entropy. *Phys. Rev. Lett.*, 96:110404, Mar 2006.

[250] Michael Levin and Xiao-Gang Wen. Detecting topological order in a ground state wave function. *Phys. Rev. Lett.*, 96:110405, Mar 2006.

[251] O. S. Zozulya, M. Haque, K. Schoutens, and E. H. Rezayi. Bipartite entanglement entropy in fractional quantum Hall states. *Phys. Rev. B*, 76:125310, Sep 2007.

[252] Didier Poilblanc. Investigation of the chiral antiferromagnetic Heisenberg model using peps. *arxiv:1707.07844*, 2017.

[253] Daisuke Tahara and Masatoshi Imada. Variational Monte Carlo method combined with quantum-number projection and multi-variable optimization. *J. Phys. Soc. Japan*, 77(11):114701, 2008.

[254] Martin Greiter and Ronny Thomale. Non-Abelian statistics in a quantum antiferromagnet. *Phys. Rev. Lett.*, 102:207203, May 2009.

[255] Ivan Glasser, J Ignacio Cirac, Germán Sierra, and Anne E B Nielsen. Exact parent Hamiltonians of bosonic and fermionic Moore-Read states on lattices and local models. *New J. Phys.*, 17(8):082001, 2015.

[256] Alexander Novikov, Mikhail Trofimov, and Ivan Oseledets. Exponential machines. In *International Conference on Learning Representations*, 2017.

[257] Edwin Miles Stoudenmire and David J. Schwab. Supervised learning with quantum-inspired tensor networks. *Advances in Neural Information Processing Systems*, 29:4799, 2016.

[258] Andrew Critch and Jason Morton. Algebraic geometry of matrix product states. *SIGMA Symmetry Integrability Geom. Methods Appl.*, (10), 2014.

[259] Martin Kliesch, David Gross, and Jens Eisert. Matrix-product operators and states: NP-hardness and undecidability. *Phys. Rev. Lett.*, 113:160503, 2014.

[260] Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D. Rodriguez, and J. Ignacio Cirac. Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X*, 8:011006, 2018.

[261] Elina Robeva and Anna Seigal. Duality of graphical models and tensor networks. *arXiv:1710.01437*, 2017.

[262] Kunihiko Fukushima. Neural network model for a mechanism of pattern recognition unaffected by shift in position - neocognitron. *Trans. IECE*, J62-A(10):658–665, 1979.

[263] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. Back-propagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[264] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. In *Parallel distributed processing*, volume 1. The MIT Press, 1986.

[265] Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, 2002.

[266] Ivan V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[267] Nadav Cohen, Or Sharir, and Amnon Shashua. Convolutional rectifier networks as generalized tensor decompositions. In *International Conference on Machine Learning*, pages 955–963, 2016.

[268] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Bridging many-body quantum physics and deep learning via tensor networks. *arxiv:1803.09780*, 2018.

[269] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. Unsupervised generative modeling using matrix product states. *Phys. Rev. X*, 8:031012, 2018.

[270] Yuhan Liu, Xiao Zhang, Maciej Lewenstein, and Shi-Ju Ran. Learning architectures based on quantum entanglement: a simple matrix product state algorithm for image recognition. *arXiv:1803.09111*, 2018.

[271] Ding Liu, Shi-Ju Ran, Peter Wittek, Cheng Peng, Raul Blázquez García, Gang Su, and Maciej Lewenstein. Machine learning by two-dimensional hierarchical tensor networks: A quantum information theoretic perspective on deep architectures. *arXiv:1710.04833*, 2017.

[272] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv:1708.07747*, 2017.

[273] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/, 1998.

[274] Karol J. Piczak. Environmental sound classification with convolutional neural networks. In *IEEE 25th International Workshop on Machine Learning for Signal Processing*, pages 1–6, 2015.

[275] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[276] Brendan J. Frey. Extending factor graphs so as to unify directed and undirected graphical models. In *Conference on Uncertainty in Artificial Intelligence*, pages 257–264, 2003.

[277] Qibin Zhao, Masashi Sugiyama, and Andrzej Cichocki. Learning efficient tensor representations with ring structure networks. *arXiv:1705.08286*, 2018.

[278] Ivan Glasser, Ryan Sweke, Nicola Pancotti, Jens Eisert, and J. Ignacio Cirac. Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning. *arXiv e-prints*, page arXiv:1907.03741, Jul 2019.

[279] Jacob Biamonte, Stephen R. Clark, and Dieter Jaksch. Categorical tensor network states. *AIP Advances*, 1(4):042172, 2011.

[280] Stephen R. Clark. Unifying neural-network quantum states and correlator product states via tensor networks. *arXiv:1710.03545*, 2017.

[281] Hugo Larochelle and Yoshua Bengio. Classification using discriminative restricted Boltzmann machines. In *International Conference on Machine learning*, pages 536–543, 2008.

[282] Hugo Larochelle, Michael Mandel, Razvan Pascanu, and Yoshua Bengio. Learning algorithms for the classification restricted Boltzmann machine. *Journal of Machine Learning Research*, 13:643–669, 2012.

[283] Song Cheng, Jing Chen, and Lei Wang. Information perspective to probabilistic modeling: Boltzmann machines versus born machines. *arXiv:1712.04144*, 2017.

[284] Jin-Guo Liu and Lei Wang. Differentiable learning of quantum circuit born machine. *arXiv:1804.04168*, 2018.

[285] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. A dataset and taxonomy for urban sound research. In *International Conference on Multimedia*, pages 1041–1044, Orlando, USA, 2014.

[286] Vasily Pestun and Yiannis Vlassopoulos. Tensor network language model. *arXiv:1710.10248*, 2017.

[287] Vasily Pestun, John Terilla, and Yiannis Vlassopoulos. Language as a matrix product state. *arXiv:1711.01416*, 2017.

[288] William Huggins, Piyush Patel, K. Birgitta Whaley, and E. Miles Stoudenmire. Towards quantum machine learning with tensor networks. *arxiv:1803.11537*, 2018.

[289] Ivan Glasser, Nicola Pancotti, and J. Ignacio Cirac. Supervised learning with generalized tensor networks. *arXiv:1806.05964*, 2018.

[290] Amnon Shashua and Tamir Hazan. Non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 792–799, New York, NY, USA, 2005. ACM.

[291] Chengran Yang, Felix C. Binder, Varun Narasimhachar, and Mile Gu. Matrix product states for quantum stochastic modeling. *Physical Review Letters*, 121:260602, 2018.

[292] Vasily Pestun, John Terilla, and Yiannis Vlassopoulos. Language as a matrix product state. *arXiv:1711.01416*, 2017.

[293] E. Miles Stoudenmire. Learning relevant features of data with multi-scale tensor networks. *Quantum Science and Technology*, 3(3):034003, 2018.

[294] James Stokes and John Terilla. Probabilistic modeling with matrix product states. *arXiv:1902.06888*, abs/1902.06888, 2019.

[295] Marcello Benedetti, Delfina Garcia-Pintos, Yunseong Nam, and Alejandro Perdomo-Ortiz. A generative modeling approach for benchmarking and training shallow quantum circuits. *arxiv:1801.07686*, 2018.

[296] Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, and Simone Severini. Hierarchical quantum classifiers. *arXiv:1804.03680*, 2018.

[297] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, João Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J. Pal. Deep complex networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[298] Frank Verstraete, Juan José García-Ripoll, and J. Ignacio Cirac. Matrix product density operators: Simulation of finite-temperature and dissipative systems. *Physical Review Letters*, 93:207204, 2004.

[299] Gemma De las Cuevas, Norbert Schuch, David Pérez-García, and J. Ignacio Cirac. Purifications of multipartite states: limitations and constructive methods. *New Journal of Physics*, 15(12):123021, 2013.

[300] Thomas Barthel. Precise evaluation of thermal response functions by optimized density matrix renormalization group schemes. *New Journal of Physics*, 15(7):073010, 2013.

[301] Albert H. Werner, Daniel Jaschke, Pietro Silvi, Martin Kliesch, Tommaso Calarco, Jens Eisert, and Simone Montangero. Positive tensor network approach for simulating open quantum many-body systems. *Physical Review Letters*, 116(23):237201, 2016.

[302] Hamza Fawzi, João Gouveia, Pablo A. Parrilo, Richard Z. Robinson, and Rekha R. Thomas. Positive semidefinite rank. *Mathematical Programming*, 153(1):133–177, 2015.

[303] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796, 2001.

[304] Joel E. Cohen and Uriel G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149 – 168, 1993.

[305] António Pedro Goucha, Jo ao Gouveia, and Pedro M. Silva. On ranks of regular polygons. *SIAM Journal on Discrete Mathematics*, 31(4):2612–2625, 2017.

[306] João Gouveia, Pablo A. Parrilo, and Rekha R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013.

[307] Andrew J. Ferris and Guifre Vidal. Perfect sampling with unitary tensor networks. *Physical Review B*, 85:165146, 2012.

[308] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS) 2000, Denver, CO, USA*, pages 556–562, 2000.

[309] `https://github.com/glivan/tensor_networks_for_probabilistic_modeling`.

[310] Nicolas S. Müller, Matthias Studer, and Gilbert Ritschard. Classification de parcours de vie à l'aide de l'optimal matching. In *XIVe Rencontre de la Société francophone de classification, Paris (SFC 2007)*, page 157–160, 2007.

[311] Dheeru Dua and Casey Graff. UCI machine learning repository, 2019.

[312] This lymphography and tumor domains were obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia. Thanks go to M. Zwitter and M. Soklic for providing the data.

[313] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.