Probing universality with entanglement entropy via quantum Monte Carlo

by

Bohdan Kulchytskyy

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following members served on the examination committee for this thesis. The decision of the examination committee is by majority vote.

Supervisor:	Roger Melko Associate Professor, Dept. of Physics & Astronomy, University of Waterloo Associate Faculty, Perimeter Institute for Theoretical Physics
Committee Member:	Anton Burkov Associate Professor, Dept. of Physics & Astronomy, University of Waterloo
Committee Member:	Juan Carrasquilla Adjunct Professor, Dept. of Physics & Astronomy, University of Waterloo Faculty Member, Vector Institute
Internal-External Examiner:	Pierre-Nicholas Roy Professor, Dept. of Chemistry, University of Waterloo
External Examiner:	Fabien Alet Researcher, Laboratoire de Physique Théorique de Toulouse

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Our understanding of physical phenomena hinges on finding universal core mechanisms that unite them. The concept of universality is deeply ingrained in the study of quantum many-body systems. At zero temperature, microscopically different systems with long-range order collapse into their universal state described by a handful of universal parameters. Establishing those parameters implies identification of the universal theory that is effectively describing the system and ultimately providing the desired understanding. We take up this task with the help of Rényi entanglement entropy. Defined with respect to a system bipartition, this measure quantifies information shared between the subsystems. Understanding what insights into universality are encoded within this information-theoretic quantity as well as developing numerical tools to efficiently estimate the Rényi entanglement entropy are the subjects of this thesis.

On the computational end of this far-reaching goal, we develop a novel theoretical framework for constructing improved Rényi entanglement entropy estimators in the context of $d+1$ quantum Monte Carlo methods. The discovery of a connection of this methodology to the well-established Kandel-Domany formalism provides a clear path towards generalization. Additionally, we embrace a data-driven approach towards learning the ground state wavefunction. We demonstrate how a restricted Boltzmann machine can be used to reconstruct the Rényi entanglement entropy from projective measurements of a quantum ground state. Furthermore, we extend this classical generative architecture to a quantum analogue that we call the quantum Boltzmann machine.

On the theoretical side of this endeavour, we study the Rényi entanglement entropy scaling terms for two quantum lattice models embedded in two dimensional space via extensive quantum Monte Carlo simulations. For the ground state of the XY model, we provide conclusive numerical evidence for a logarithmic contribution that uniquely characterizes the continuous symmetry of the emerging order parameter. Moreover, we confirm the form of the subleading universal geometric contribution arising due to the bosonic nature of low-energy degrees of freedom in this model. For the critical ground state of the transverse field Ising model, we develop a novel scaling procedure to extract a universal number κ_2 revealed via a cylindrical entangling bipartition in the thin-slice limit. The combined product of our work sheds new light on the entanglement-based classification of universality and brings a suite of new powerful numerical tools to continue illuminating this theoretical program in the future.

Acknowledgements

Concluding my PhD, I find myself on a path I wanted to be on. With the hindsight, the trajectory that has led up to this point was decided during my middle school years. Therefore, I open my acknowledgements by expressing my sincere gratitude to my physics teacher, Yevhenij Petrovych, and my teacher of mathematics, Maryna Valentynivna, for inflicting on me an insatiable curiosity for natural processes. I feel lucky for having had the opportunity to closely collaborate with Adrian Del Maestro who guided my first steps in research during the two summers we spent working together, including my choice of the graduate program. My supervisor, Roger Melko, has never made me regret my decision. I am thankful to Roger for creating so many unique opportunities, for all the enlightening advice and contagious upbeat disposition. Every time I left your office, my worries would remain behind. Also, I would like to take this opportunity to thank my committee members, Pierre-Nicholas Roy and Anton Burkov, with whom it has been a great pleasure to discuss my work over the years.

My academic experience would have never been the same without the curious minds I have met along the way. I would like to thank Chris, Stephen and Juan for all the passed knowledge. I thank Florian, Jaan, Perry, Alex, Andres, Matt, Behnam, Michael, Dan and Majid for embodying the process of science in an exciting face. I cannot phantom the PhD experience without the friends I have made during the six years - Lauren, David, Giac and Hina. I am very grateful for all the adventures we have had together. A special thanks goes to Lauren for politely scrutinizing every hand-wavy argument I have made and, of course, for the γ that made κ possible.

Among my Montreal friends, I would like to single out Martin whose unhealthy criticism never ceased over the years despite the distance. In retrospect, I cherish the memory of our collaborative last-minute composition of a scholarship application on Google Docs more than the money I never ended up receiving. On this note, I would like to thank the government of Canada and Ontario for investing in science and supporting my research throughout the years. Finally, I would like to thank my family for all the support and a proud glance in the eyes.

Dedication

I dedicate this thesis to my parents, Alla and Rost, whose love and support I have never questioned.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Quantum ground states conspire into an intricate superposition of product states. At zero temperature, all correlations between parts of a system are mediated by the entanglement. Therefore, it is only natural to study quantum ground states from the point of view of entanglement. A vast majority of quantum states residing in a Hilbert space are highly entangled. A random state generated via a selection of its amplitudes from a uniform distribution in a given basis is almost surely maximally entangled with respect to any bipartition [1]. This statement can be put on a quantitative ground via a measure of entanglement known as entanglement entropy (EE). Defined with respect to a subsystem A, EE quantifies the quantum information contained within the subsystem about its complement. With respect to the EE, the highly-entangled structure of a random quantum state takes the specific form of a volume law, such that

$$S(A) \sim \text{Volume}(A). \tag{1.1}$$

In other words, the EE is expected to be proportional to the volume of any subsystem A. This feature of a general quantum state is in striking contrast to the EE structure obeyed by physical ground states. It transpires that a physical Hamiltonian imprints a stringent structure on the corresponding ground state via the so-called *area law*,

$$S(A) \sim \text{Area}(A). \tag{1.2}$$

Comparison of the above equations reveals a fundamental truth about our universe - ground states are very different from a typical quantum state. In the thermodynamic limit, ground states effectively reside in a subspace of measure zero. Understanding the structure of this subspace via its universal EE characteristics is the goal of this thesis.

1.1 Universality in the entanglement

EE is a highly non-local quantity that summarizes the underlying structure of a physical state embedded in a vast Hilbert space via a single number. In order to extract information from such a compressed description, we study how the EE changes in response to the modifications in the region A such as its size and shape. Entrancingly, the resulting scaling of the EE is found to be universal among a vast array of condensed matter systems. In fact, those systems can be classified by the structure of the entanglement as revealed by the EE scaling.

In quantum many-body systems, the area law can be traced back to the locality of interactions and, therefore, directly encodes the microscopic details of the underlying Hamiltonian. Non-local features of ground states transpire in the EE as well. On the extreme side of this spectrum, one finds topologically-ordered states such as spin-liquids. Endowed with patterns of long-range entanglement, those systems correspond to an EE that scales independently on the extent of the subsystem [2, 3]. In fact, EE is often the only probe that can detect topological order [4].

For systems characterized by long-range order, the EE encodes universal information about the effective low-energy Hamiltonian. In this case, universal numbers extracted from EE are directly related to the geometry of the region A. For instance, the geometry considered in this thesis together with the physics uncovered by studying the corresponding EE scaling are summarized in Fig. 1.1. This flexibility in the range of information extracted from EE motivates a EE-based "geometrical" paradigm of critical systems. Traditionally within the Landau paradigm, universal fixed points are classified via a set of critical exponents extracted from n-point functions. It is an open question whether universal features contained in the EE represent a complementary or distinct view on the relationship between critical theories.

A strong motivation for a research program aiming to develop an EE-based geometrical classification originates from the well-known analytic result in one dimension. Namely, for a conformally invariant system, the EE scaling provides direct access to the central charge, c [5, 6]. This universal number is arguably the most important quantity for identifying the relevant conformal field theoretical description of the system. Its further importance is underlined in Zamolodchikov's c-theorem establishing a strict relationship between different fixed points in the renormalization group (RG) plane based on the central charge value [7, 8]. The manifest relationship between the EE and the central charge can be taken as the definition of latter - that is, a coefficient characterizing the logarithmic divergence of the EE. While in one dimension, this redefinition of the central charge is mostly syntactic in

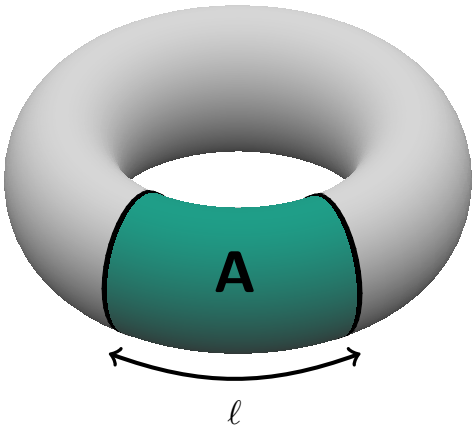


Figure 1.1: The bipartition employed in this thesis to probe universality of underlying microscopic Hamiltonians. The system has periodic boundary conditions such that it can be visualized on the surface of a torus. Region A is a cylindrical subregion that wraps around the y-direction and has length ℓ along the x-direction. In ground states with spontaneously broken continuous symmetry, such a region A geometry directly reveals the symmetry of the order parameter (see Sec. 4.1). In scale-invariant systems, the thin-slice limit, $\ell \to 0$, unravels a universal contribution to the entanglement entropy characterized by a universal number κ (see Sec. 4.2).

nature, in higher dimensions, such interpretation generalizes to the F-theorem in $2+1$ and the a-theorem in $3+1$, both relying on the geometry of a hyper-spherical region A to extract an RG-monotone [9, 10, 11]. However, the study of such regions on a lattice is challenging due to an unavoidable pixelation. Therefore, in order to employ the powerful RG theory as a practical tool for the study of condensed matter systems, the search for universal numbers acting as monotones under RG-flow must be extended to other entangling surfaces.

1.2 Entanglement entropy scaling theory

In order to gain intuition about the origin of various terms contributing to the total EE corresponding to a region A, we consider the following phenomenological argument in the context of a renormalization group (RG) flow for a d-dimensional system [12, 13]. Specifically, we consider the real-space RG where at each step n of the renormalization procedure the lattice constant δ is renormalized as $r_n = b^n\delta$. Our scaling argument then

relies on two postulates. First, we assume that the total EE is composed of independent contributions from every length scale r_n along the (RG) flow such that

$$S(A) = \sum_n S(r_n, A). \tag{1.3}$$

In integral form, this expression takes the form[1]

$$S(A) = \int_\delta^{r_{\max}} S(r, A)\, \mathrm{d}\Big(\ln \frac{r}{\delta}\Big)\,. \tag{1.4}$$

Here, $r_{\max}$ determines the infra-red (IR) cutoff in the RG procedure. This bound is set by the scale at which region A is no longer entangled with its complement. In general, $r_{\max} = \min(\xi, \ell)$ where ξ is the correlation length and ℓ is the characteristic length of region A. In a scale-invariant system, the correlation length diverges, i.e. $\xi \to \infty$, so that the only IR length scale is provided by the entangling region, $r_{\max} = \ell$. Therefore, by simply changing the size ℓ, the EE is able to probe the system at different scales.

Second, we assume that at each scale r, the contribution to the EE is local to the entangling boundary. As such, it can be represented as a surface integral of a local geometric function $g(A, r)$,

$$S(r, A) = \int_{\partial A} \frac{\mathrm{d}s_{d-1}}{r^{d-1}} g(A, r). \tag{1.5}$$

Here, $\mathrm{d}s_{d-1}$ represents a dimensionful differential surface element. As the EE is a dimensionless quantity, we normalize the surface element by the inverse scale $1/r^{d-1}$. Combining both assumptions into one equation, we get

$$S(A) = \int_\delta^{r_{\max}} \frac{\mathrm{d}r}{r^d} \int_{\partial A} \mathrm{d}s_{d-1} g(A, r). \tag{1.6}$$

In a flat one dimensional system, the inner integral can only be a constant. In this case,

$$S(A) \propto \ln r_{\max} + c_\delta, \tag{1.7}$$

where c_δ is a constant that has absorbed the microscopic length scale δ. In a gapped system with sufficiently large region A, $r_{\max} = \xi$, indicating that the EE saturates once the entangling region exceeds the correlation length. As the boundary size is also constant, this

[1]The logarithmic measure appears due to the change of variables $r_n = b^n\delta \to n = e^b \ln(r/\delta) \to \mathrm{d}n = \mathrm{d}\ln(r/\delta)$.

constant contribution to the EE represents the area law. However, in a one-dimensional critical system, the area law is violated. Instead, we recover the well known-result [5]

$$S(A) \propto \ln \ell + c_\delta. \tag{1.8}$$

As the non-universal constant δ does not appear in the logarithm, the coefficient of this term must be universal. We note that so far our argument has ignored the effect of the boundary conditions, which is equivalent to considering region A being embedded in an infinite chain. For a conformally invariant system, the full analytic calculation on a finite system L reveals a signature of a universal geometric function that depends the aspect ratio ℓ/L [6]. It also reveals that the universal number characterizing the logarithmic divergence above is the central charge.

In two spatial dimensions, various options for the geometric integral Eq. (1.5) due various choice for the geometry of region A lead to various EE scaling forms [12]. As a rule of thumb, region A geometries that produce a logarithmic contribution allow one to extract universal numbers, since the microscopic effects parameterized by the ultraviolet cut-toff, δ, separate out into an additive constant proportional to $\ln \delta$. However, as the following argument shows, non-logarithmic contributions can lead to universal numbers as well.

Consider a region A with flat boundaries embedded in a critical system of size L, such as in Fig. 1.1. In this case the inner geometric integral in Eq. (1.6) simply gives $2Lc_g$, where $c_g = g(A, r)$ is the constant local geometric feature. Ignoring the effects due to the boundary conditions and assuming that the entangling region's height is ℓ, Eq. (1.6) takes the form

$$\begin{aligned} S(A) &= \int_\delta^\ell \frac{2c_g L}{r^2} \mathrm{d}r \\ &\propto \frac{L}{\delta} - \frac{L}{\ell}. \end{aligned} \tag{1.9}$$

Unlike in one dimension, we recover the leading area law even for a critical system. Moreover, the "subleading" contribution is free of the UV cut-off parameter δ, thus hinting at its universality. As in the one-dimensional case considered above, the effects of the boundary conditions modify this result to

$$S(A) = a\frac{L}{\delta} + \chi(u), \tag{1.10}$$

where $\chi(u)$ is a universal function specified by the aspect ratio $u = \frac{\ell}{L}$. More generally, this geometric function depends on all macroscopic geometric characteristics of the system.

For instance, in a torus of unequal dimensions, it acquires a dependence on the system aspect ratio, L_x/L_y, as well. For this reason, unless this function is the targeted source of universal information (as in Sec. 4.2), it is advisable to keep all the aspect ratios constant during the EE data fitting analysis.

1.3 Entanglement-inspired numerical approaches

While brute force approaches that directly tackle the exponential enormity of the Hilbert space are equally good, or rather equally bad, at describing any of the highly-entangled states, a compact and scalable representation of a ground state must take into an account the lesson learned from the entanglement structure - not only it must be good at capturing the area law, but it also must be bad at capturing the volume law. This lesson has been an ongoing area of research for the last two decades.

The evolution of numerical methods targeting the estimation of entanglement entropy paralleled our understanding of many-body systems through its entanglement structure. Arguably the most fruitful interaction between the development of theory and numerical tools happened in one dimensional systems. In this low-dimensional realm, tensor networks methods have flourished taking their routes from the matrix product state (MPS) [14] and the density matrix renormalization group (DMRG) method [15]. The success of this variational ansatz (MPS) and the optimization procedure behind it (DMRG) turned out to be ultimately tied to each other and to the entanglement structure of physical states they were designed to represent. Indeed, on one hand, in its iterative procedure, DMRG heavily relies on the entanglement between two complementary subsystems in order to sort and prune out the exponential large Hilbert space in terms of its significance to the ground state wavefunction. On the other hand, in one-dimension, the area law bounds the amount of entanglement entropy to a constant and, therefore, represents a physical phenomenon that makes it possible for this procedure to succeed. More importantly, the revelation that the entanglement structure of a quantum state needs to be incorporated as an inductive bias into the structure of the variational ansatz has driven the advancement of tensor networks ever since.

An important development in tensor networks came from the study of one-dimensional critical systems. In those scale-invariant systems, it was noticed that the entanglement entropy seemed to slowly grow without bounds with the system size [16]. As a result, the computational costs associated with the MPS ansatz optimization also grow polynomially in the system size, thus requiring a brute-force approach for the largest system sizes. The multi-scale renormalization ansatz (MERA)[17] provides an elegant solution to this

problem. Its design reflects the multi-scale structure of entanglement present in systems it was designed to imitate [17]. In such systems, the entanglement is equally redistributed among all the scales. In other words, the contribution of each renormalized scale of a system remains constant. MERA is able to capture this entanglement structure through a deep layered architecture mimicking the scale transformation between the progressive levels of the network while preserving its invariance from one level to the next. Therefore, MERA provides another example of a classical method that allowed to harness the quantum complexity by exploiting its intrinsic structure.

Unlike in one dimension, in two dimensions, the boundary of a subsystem is directly related to its size. For this reason, the area law provides a formidable challenge to tensor networks, even to the specialized architectures such as projected entangled pair states (PEPS) [18]. In order to overcome this limitation, in this thesis, we rely on the realization that the exact knowledge of the ground state sought by tensor networks is not mandatory for the extraction of accurate information about it. Stochastic methods such as quantum Monte-Carlo (QMC) introduce a controlled *bias-variance trade-off* to the equation. Those methods provide an unbiased tool to estimate quantities of interest at a price of uncertainty in their values. The computational scaling of QMC sampling is not affected by the entanglement structure of the system and, therefore, is not paralyzed by the area law in dimensions larger then one in the same way as tensor networks. For this reason, QMC methods have proved to be a vital tool for the study of EE in large classes of quantum models beyond the reach of tensor networks.

1.4 Numerical approaches for entanglement entropy

The sources of complexity in QMC methods targeting the extraction of EE concentrated on two fronts. The first challenge is to reduce the variance of the estimated EE. While the EE does not impact the sampling performance of a QMC, its magnitude directly affects the *signal-to-noise* ratio of the extracted estimate. A conceptually simple but an efficient method called *ratio trick* allows to control the strength of the *signal* targeted by an EE estimator. However often in practice, this method alone is not able to generate sufficiently accurate estimates required for probing the informative subleading contributions to the EE. As a tool to mitigate this problem, we develop a *projected ensemble* framework that allows to dramatically reduce the *noise* by constructing improved estimators based on a reformulated classical representation of a quantum system in terms of highly non-local loops.

The second source of complexity inherent to QMC methods is intrinsically linked to the

quantum nature QMC methods are designed to emulate. In a path-integral representation, d-dimensional quantum states are $d+1$ objects with complicated interference patterns in the imaginary-time dimension. Contrarily, QMC algorithms are inherently classical algorithms that rely on the probabilistic interpretation of a quantum state and ignore its complex phase structure. Fundamentally, the success of this representation hinges on the absence of destructive interference within a targeted quantum system. When this condition is not satisfied, the classically biased representation leads to the so called *sign problem* within QMC methods [19]. Namely, a QMC ends up exploring the classical state space that would self-average to zero by means of destructive interference in a true quantum state. Ultimately, this unguided exploration leads to the loss of ability to generate physically relevant samples and, hence, a noise-dominated EE estimate. Unfortunately, most $d+1$ QMC methods targeting spin systems are plagued by the sign problem.

An inspiration towards a possible solution once again can be gained by considering the problem from the point-of-view of the bias-variance framework. Applied to a quantum system with a large EE and a strong sign structure, tensor network and $d+1$ QMC are both expected to fail as they land on the opposite extreme ends on the bias-variance axis. Fortunately, variational QMC (VMC) [20] allows to trade between those extremes. Working directly in d-dimensional physical space of the underlying problem, VMC is devoid of the sign problem. At the same time, being a QMC method, its sampling complexity is not dependent on the entanglement. The bias and variance of this method are ultimately tied to the expressiveness of the variational ansatz. Traditionally, the form of this trial wavefunction has been physically motivated. A remarkable story of success is the BCS wavefunction that elegantly captures the phenomenon of superconductivity in a Cooper pair break-up of the wavefunction. Numerically, the Jastrow wavefunction with inbuilt locality is often used to study lattice systems. The paradox of VMC is that a good trial wavefunction requires a good understanding of the relevant physics of the problem in the first place.[2] In the language of machine learning, a good trial wavefunction requires good features.

Compelled by recent successes of deep learning in fields like vision and reinforcement learning relying on automatic processes of feature-engineering, the core ideas of self-learning algorithms have propelled the VMC into the new era [21]. Framed as a generative learning problem, the search for the ground state now relies on powerful neural-networks with multi-

[2]In fact, the same statement applies to tensor networks and QMC. For instance, MERA was inspired from the insight of the scale-invariant structure of entanglement. Similarly, the success of QMC relies on the efficiency of cluster updates whose design is intrinsically linked to the understanding of relevant physics. The difference with the VMC is that unlike in the cases of the aforementioned frameworks, there is no asymptotic guarantees for the exactness of this method.

scale architectures and asymptotic guarantees of being universal function approximators. Moreover, the variational optimization of those networks can be flexibly adjusted between energy and data-based loss functions, making them directly applicable for an experimental setup [22]. From the data-driven point of view, an experimentally-realized natural process can be regarded as a physical implementation of the sampling component of a QMC. Naturally produced datasets can be exploited to encourage ergodic exploration of the state-space, thus, eliminating mode-seeking biases inherent to the energy-based optimization that VMC relies on.[3] In return, the learned variational wavefunction can be used to provide access to physical quantities not easily accessible in a laboratory such as EE. This symbiotic relationship is explored in this thesis via a classically and quantumly inspired architectures for the trial wavefunction.

1.5 Outline

Below, we provide a brief overview of the chapters composing this manuscript.

Chapter 2 starts with an introduction to the concept of Monte Carlo-based calculations and various strategies for implementation such as importance sampling and Markov Chain Monte Carlo. It then proceeds with the development of two flavours of $d+1$ QMC. First, it discusses the framework of path-integral Monte-Carlo that can be used to map a quantum model into its higher-dimensional classical analog. As a particular example, the discussion revolves around the transverse field Ising model (TFIM) mapping to a classical Ising model. As part of the corresponding section, the Fortuin-Kasteleyn decomposition of an Ising model is shown to arise in a particular limit of a stochastic mapping based on a restricted Boltzmann machine (RBM). This view allows a natural interpretation for the Swensden-Wang cluster update as a blocked Gibbs update of the RBM. As part of the RBM consideration, we also derive an effective model for the marginal distribution over the hidden units, known as the random-cluster model. This model is later used in Ch. 3 in order to derive an improved EE estimator. Second, Ch. 2 discusses the stochastic series expansion (SSE). The details of this QMC framework necessarily for implementation are discussed in the context of XXZ model. Lastly, SSE estimators required to implement an unbiased likelihood-based training of a quantum Boltzmann machine (introduced in Ch. 5) are derived.

Chapter 3 builds up different strategies for estimating EE within d and $d+1$ QMC methods. In particular, it introduces a novel *projected ensemble* approach that provides a

[3] For further discussion of the trade-offs inherent to a particular choice of a cost-function, see Sec.5.1

theoretical model for deriving new improved EE estimators in $d+1$ QMC methods. Its application within the path-integral and SSE frameworks are discussed in detail.

Chapter 4 showcases two EE scaling analyses. First, the quantum XY model is studied. An intuitive state-counting argument for the origin of a universal logarithmic EE contribution is provided. The numerical analysis is shown to accurately confirm the theoretical prediction of Ref. [23]. Second, an EE study is applied to the critical TFIM. A universal number κ_2 characterizing the geometric contribution to EE is extracted in the limit where the entangling region is a thin cylinder. For this task, a novel scaling analysis procedure is developed and benchmarked on a data set from a non-interacting theory with known results.

Chapter 5 introduces the concept of generative learning and, in particular, focuses on the RBM. The full likelihood-based training algorithm is provided. As an application, an RBM is trained on projective spins measurements of the TFIM ground state. In this setting, the training is equivalent to a wavefunction reconstruction. As a proof-of-concept, we demonstrate that entanglement entropy is correctly recovered from the reconstructed state. Additionally, we introduce a novel quantum extension of the RBM dubbed the quantum Boltzmann machine (QBM). A scalable likelihood bound-based training algorithm is developed and successfully tested on an artificial classical dataset.

Chapter 6 summarizes the results of this thesis and provides thoughts on the future of EE detection for the purpose of characterizing quantum many-body systems.

Chapter 2

Quantum Monte Carlo methods

Evaluation of expectation values is of interest to many subfields of science. In general this problem can be cast in the form

$$\langle O \rangle = \frac{1}{Z} \sum_c O\left(c\right) W\left(c\right), \tag{2.1}$$

where $W\left(c\right) \geq 0$ is the unnormalized probability of a configuration c and the sum is taken over all configuration space. The normalizing constant Z can be written as

$$Z = \sum_c W\left(c\right), \tag{2.2}$$

which turns the weight $W(c)$ into a normalized probability $P(c) = W(c)/Z$. In the context of statistical mechanics, those quantities adopt their domain specific terminology. The normalizing constant Z is known as the partition function, the fundamental object of interest in the field. The weight takes a particular form, $W\left(c\right) = e^{-\beta E(c)}$, known as the Boltzmann weight. Here, $E(c)$ is the energy and $\beta = 1/T$ is the inverse temperature. Hence, Eq. (2.1) takes the following domain specific form

$$\langle O \rangle = \frac{1}{Z} \sum_s O\left(c\right) e^{-\beta E(c)} \tag{2.3}$$

The notation is once again readjusted when talking about statistical quantum mechanics, such that

$$\left\langle \hat{O} \right\rangle = \frac{1}{Z} \mathrm{Tr}\left[e^{-\beta \hat{H}}\right] = \frac{1}{Z} \sum_s \langle s|\hat{O} e^{-\beta \hat{H}}|s\rangle. \tag{2.4}$$

Now, the observable $\hat{O}$ and Hamiltonian $\hat{H}$ are quantum operators, mathematically described by matrices. The probability of a state is weighted by a quantum generalization of the Boltzmann weight, $e^{-\beta\hat{H}}$. Similarly to the classical case, the partition function is computed by tracing out all physical degrees of freedom, $Z = \mathrm{Tr}\left[e^{-\beta\hat{H}}\right]$. However, in contrast to the previous interpretation of s as a configuration state with a well-defined probability, in this context, s represents a quantum state living in the Hilbert space over which the Hamiltonian $\hat{H}$ is defined. With some work and for some quantum models, it is possible to find a configuration space c (not necessarily directly related to s) such that Eq. (2.4) can be cast in the form of Eq. (2.1). This thesis' focus is on quantum systems for which such a transformation is possible. The underlying approaches towards this transformation are commonly referred to as quantum Monte Carlo (QMC) algorithms. Later in this chapter, we will explore the details of two flavours of QMC.

The unifying challenge that all Monte Carlo methods [24] are designed to tackle is concealed in the summation operation in Eq. (2.1). The difficulty lies in the exponential growth of the number of terms required to evaluate this expression exactly when the system size grows linearly. Alternatives to the brute-force solution are made possible due an underlying structure often present in problems of interest. In those problems, most of terms in the sum have a negligible contribution to the expectation. The challenge of locating the relevant terms, which are vastly outnumbered among other irrelevant terms, is known as the curse of dimensionality. Monte Carlo methods aim to circumvent this problem via a statistical approach. Namely, they approximate the true expectation value with an unbiased estimate

$$\langle O \rangle \approx \frac{1}{N_s} \sum_s O\left(c\right), \tag{2.5}$$

where the sum is over N_s samples generated according to $W(c)$. As long as those samples are i.i.d (identically and independently distributed), the variance of the estimator's mean can be controlled by enlarging the sample set as, according to the Central Limit theorem, it will decay as $1/N_s$. In practice, generation of i.i.d. samples from an arbitrary $p\left(s\right)$ is a very challenging problem. For this reason, the curse of dimensionality also plagues Monte Carlo methods. However, its reformulation as a sampling problem provides a constructive handle to tackle the exponential.

In this thesis, we mostly focus on problems with a particular realization of the curse of dimensionality known as *critical slowing down*. This physical phenomenon is marked by a relative slowing down of the thermal dynamics generated by local updates with respect to the diverging correlation length near a continuous phase transition. This phenomenon is the strongest at a critical point where the physical system becomes scale-invariant,

implying that excitations on all scales are equally important. At such a point, local spin updates that work on the scale of the lattice spacing fail to effectively update the system. As a result, the relaxation time diverges with the system size, thus prohibitively requiring ever-growing simulation times. The solution to this problem comes in the form of cluster algorithms, which are a class of Monte Carlo updates capable of efficiently sampling the system on all length scales. Such algorithms will be described in detail in this chapter.

The ergodicity loss in Monte Carlo simulations can also be caused by other effects such as rough energy landscapes in glasses or first order phase transitions. In those problems, the Markov chain is incapable of jumping over local energy barriers and tends to localize in a particular configuration space. The techniques to tackle this challenge rely on the notion of extended ensembles. The general idea of those methods is to extend the configuration space of the problem in order to facilitate the mixing of the Markov chain, and, effectively, create global updates in the original configuration space. Some examples of such methods are parallel tempering, simulated annealing, umbrella sampling and Wang-Landau methods.

Quantum systems are described by exponentially large Hilbert spaces and, therefore, numerical approaches to studying those systems must provide an efficient way to deal with exponentially many degrees of freedom. Luckily, Monte Carlo methods such as the Metropolis-Hastings technique provide a powerful tool for efficiently computating multi-dimensional integrals. For this reason, Monte Carlo methods constitute the natural choice for tackling the exponential complexity associated with quantum systems. In order to apply this numerical approach in practice, a translation from the language of quantum mechanics to the language of Monte Carlo, expressed with the help of classical probability theory, is required.

In this chapter, we employ two constructions to achieve the mapping between quantum degrees of freedom and equivalent classical ones. The first mapping is obtained via the application of a Suzuki-Trotter decomposition provided by the imaginary path-integral view of the quantum system. For certain d-dimensional quantum system, this approach allows one to construct an equivalent $d+1$-dimensional classical systems for the price of a controllable systematic bias. We will apply this mapping to study the universal properties of the $2d$ transverse field Ising model (TFIM). In this context, the universality arguments can be used in order to reestablish the accuracy of the mapping. The second approach relies on a series expansion of the quantum density matrix. Known as the Stochastic Series Expansion (SSE), this mapping also presents a classical system with an extra dimension compared to its quantum analog. For SSE, the emergent extra dimension is no longer directly related to the imaginary time. Instead, it adopts an algorithmic interpretation. In this case, the physical observables must be expressed in terms of the algorithmic variables. For instance, the expected energy of the quantum system is directly related to the length

of the extra dimension. The advantage of SSE is that it is a bias free method. We will utilize it to study the quantum XY model at effectively zero temperature as well as in the form of a tool for training the quantum Boltzmann machine.

In addition to the choice of the numerical representation of a quantum system, one must also be careful about the choice of an estimator. It is not always clear on how to derive an estimator for the physical observable of interest within a given Monte Carlo framework. In the sections that follow, we derive estimators for the entanglement entropy in the path-integral as well as SSE formulations. We also derive all the estimators required for training the quantum Boltzmann machine both in clamped and unclamped ensembles. Additionally, different estimators for the same physical observables can have very different statistical properties. In order to optimize the use of available computational resources, one is interested in an estimator with the smallest variance. We introduce novel cluster and loop versions of entanglement estimators that provide an exponential efficiency speed up, proving to be fundamental in order to push our calculations to the cutting edge.

2.1 Preliminaries

2.1.1 Importance sampling

As an illustration of the Monte Carlo approach, we consider importance sampling. This flavour of Monte Carlo is based on the derivation

$$\begin{aligned} \langle O \rangle_p &= \langle O(c) \rangle_{c \sim p(c)} \\ &= \sum_c O(c)\, p(c) \\ &= \sum_c O(c)\, p(c)\, \frac{q(c)}{q(c)} \\ &= \left\langle \frac{O(c)\, p(c)}{q(c)} \right\rangle_{c \sim q(c)} . \end{aligned}$$

The only working assumption in the derivation is implicit in the third line. Namely, the distribution $q(s)$ must be non-zero whenever the product $O(s)p(s)$ is non-zero. Via these simple steps, we have managed to reformulate the problem of sampling from the target probability $p(s)$ to that of sampling from a distribution $q(s)$, which we are free to choose. This reformulation can be useful when $q(s)$ is much simpler to generate samples from than

it is for $p(s)$. However, the computational gain of this approach can be limited by the quality of the distribution $q(s)$. Intuitively, this estimator performs well if it is able to place a large probability density mass on the modes of the product $O(s)p(s)$. Oppositely, its performance quickly deteriorates when $q(s)$ spreads out its distribution to regions where $O(s)p(s)$ is negligible.

A classical example of an application for importance sampling is a sampling-based calculation of the mathematical constant π. In order to set this problem up as an importance sampling problem, consider two uniform distributions $p_{\bigcirc}(x,y), p_{\square}(x,y)$ which uniformly distribute their probability mass over a unit circle and a unit square centered at the origin, correspondingly. More explicitly, the unnormalized probability density for the circle is

$$W_{\bigcirc}(x,y) = \begin{cases} 1, & \text{if } \sqrt{x^2+y^2} < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}, \tag{2.6}$$

while for the square is

$$W_{\square}(x,y) = \begin{cases} 1, & \text{if } |x|, |y| < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}. \tag{2.7}$$

To proceed, we make a few observations. First of all, a sample from $p_{\square}$ can be easily produced by generating two random numbers from a uniform distribution $[-\frac{1}{2}, \frac{1}{2}]$ corresponding to the x and y coordinates of the sample. Secondly, the normalization constants for the two distributions are just the areas of the underlying geometric shapes, such that $Z_{\bigcirc} = \pi/4$ and $Z_{\square} = 1$. Therefore, if we can estimate the partition function of the circle based on the samples generated from the square distribution, we can extract an estimate for the value of π. For this purpose, we take advantage of the following identity

$$\begin{aligned} \left\langle \frac{W_{\bigcirc}(c)}{W_{\square}(c)} \right\rangle_{c \sim p_{\square}} &= \int_{\square} \mathrm{d}x\mathrm{d}y p_{\square}(x,y) \frac{W_{\bigcirc}(x,y)}{W_{\square}(x,y)} \\ &= \frac{1}{Z_{\square}} \int_{\square} \mathrm{d}x\mathrm{d}y W_{\bigcirc}(x,y) \\ &= \frac{Z_{\bigcirc}}{Z_{\square}} \\ &= Z_{\bigcirc}. \end{aligned} \tag{2.8}$$

As a matter of fact, this identity lies at the heart of numerical measurement of entanglement entropy we will demonstrate in Ch. 3. For now, we apply it for the evaluation of π. Basing our calculation on $N_s = 200$ samples from $p_{\square}$ as shown in Fig. 2.1(a), we achieve an

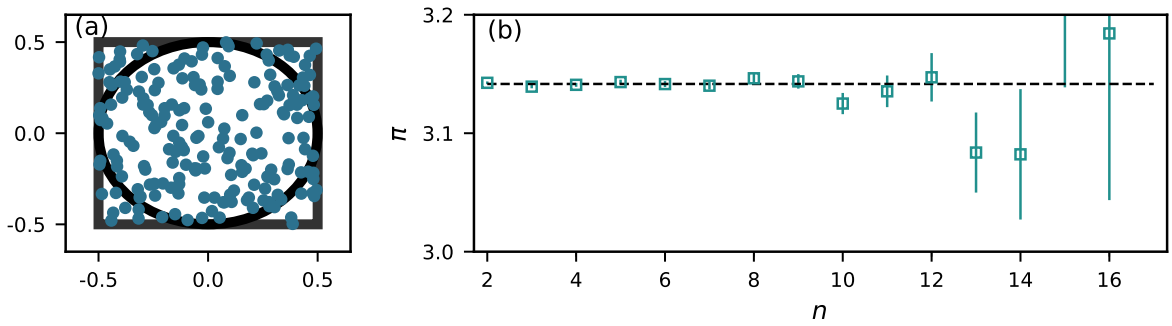


Figure 2.1: (a) A circle of radius $\frac{1}{2}$ embedded in a square of side length 1. The dots represent I.I.D. samples from the square probability distribution $p_{\square}$ defined in Eq. (2.7). These samples are used to compute the estimate $\pi = 3.00(12)$ based on Eq. (2.8). (b) A similar procedure is used to extract an estimate for π from an n-dimensional generalization. The dashed line represents the true value $\pi = 3.14159$

estimate for π of 3.00(12). Since the uncertainty scales favourably as $N_s^{-\frac{1}{2}}$, the accuracy can be greatly improved for a reasonable computational cost.

However, let's consider conducting the same procedure in higher dimensions by generalizing Eq. (2.6) and Eq. (2.7) to a hyper-sphere and a hyper-cube, correspondingly. In dimension n, the volume of a ball of radius $\frac{1}{2}$ is related to a power of π through $Z_{S^n} = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}\frac{1}{2}^n$, where Γ is the gamma function. The value of the partition function can be estimated via a trivial generalization of Eq. (2.8) with the help of samples generated from a hyper-cube. This time we increase the number of samples to $N_s = 200 \times 10^4$ and plot the estimate for π as function of the dimension n in Fig. 2.1(b). We observe that while the true value remains in statistical proximity from the numerical estimates, the uncertainty in the estimates increases exponentially. In fact, the uncertainty in π extracted from the calculation in dimension $n = 16$ is the same as in our previous calculation in $n = 2$ which was based on 10^4 times fewer samples. This perhaps-counter-intuitive result reflects the fact that, as n becomes large, the ball effectively shrinks to a point with respect to its embedding space. The curse of dimensionality is directly related to this phenomenon since the physical distribution of interest (for example, the ground state of a local Hamiltonian) occupies an infinitesimal fraction of the space it lives in. In order to gain tractability despite the curse of dimensionality, methods that exploit the structure of a problem must be developed. The framework of Markov Chain Monte Carlo provides a clear path towards this goal.

2.1.2 Markov chain Monte Carlo

Unlike the previous section where a full configuration is generated in one shot, one step of *Markov Chain Monte Carlo* (MCMC) [25] only partially updates a given state. The independence between the samples is achieved after a series of those steps take place. Effectively, MCMC represents a stochastic dynamical process under the evolution of which a particle described by its coordinates, c, explores all the configuration space according to the probability distribution of interest, $p(c)$. Mathematically, this process is described via a transition distribution at the MCMC time t

$$T(c_t \to c_{t+1}) = Pr(c_{t+1}|c_t), \tag{2.9}$$

that specifies the transition probability of accessing the state c_{t+1} from the current state c_t. In its definition, the transition probability distribution embeds the Markovian property of MCMC. Namely, an independence of T on the chain of events $c_0, \ldots, c_{t-1}$ leading to c_t. For stationary distributions, the transition dynamics T are assumed to be independent of the time variable t. With those simplifications in place, the MCMC dynamics appropriate for a particular distribution $p(c)$ can be derived by requiring this distribution to be invariant under T. This requirement implies that the flow of probability under the action of T from a particular configuration c needs to be balanced out by the flow of probability into c from the rest of state space, such that

$$\sum_{c'} p(c)T(c \to c') = \sum_{c'} p(c')T(c' \to c). \tag{2.10}$$

This equation is known as the *balance* equation. One way to satisfy it is by requiring the flow of probability to be matched at the level of each pair c and c', so that

$$p(c)T(c \to c') = p(c')T(c' \to c). \tag{2.11}$$

This particular solution to the balance equation is known as *detailed balance*. As a side remark, we note that it might be possible to satisfy Eq. (2.10) without requiring detailed balance. Such generalizations might provide better mixing of the Markov Chain in certain cases [26]. Proceeding further with our derivation based on the detailed balance equation, we deconstruct the transition dynamics into two stochastic sub-processes as

$$T(c \to c') = g(c \to c')A(c \to c'), \tag{2.12}$$

where $g(c \to c')$ is known as the *proposal distribution* and $A(c \to c')$ is the *acceptance distribution*. The task of the proposal distribution is to generate proposals for the next

state c'. The task of the acceptance distribution is to guarantee that Eq. (2.11) is satisfied for any proposal distribution. Such $A(c \to c')$ can always be found. For example, by plugging Eq. (2.12) into Eq. (2.11) and setting $A(c \to c')$ on the side of a larger pre-factor to one, we obtain a solution

$$A(c \to c') = \min\left[1, \frac{p(c')}{p(c)} \frac{g(c' \to c)}{g(c \to c')}\right]. \tag{2.13}$$

An MCMC algorithm based on this acceptance probability is known as a *Metropolis-Hastings algorithm*. This algorithm satisfies the detailed balance condition by construction. Moreover, via a careful design of the generation process defining $g(c' \to c)$ tailored for a particular $p(c)$ of interest, it is sometimes possible to achieve the perfect acceptance rate of 1. In Sec. 2.2.2, we provide an example of such a generative process that updates the Ising model on all length scales.

In addition to preserving the distribution of interest via the detailed balance condition, a valid $T(c \to c')$ is required to be *ergodic*. A process is said to be ergodic if there is a finite probability of transitioning between any pair of states in a finite number of steps. This additional requirement is needed to guarantee that the Markov chain does not get stuck in a particular region of the configuration space but instead explores the entire support of $p(c)$. Finally, as long as the requirements of ergodicity and detailed balanced are satisfied, it can be shown that the Markov chain produced by $T(c \to c')$ converges to $p(c)$ as the simulation time $t \to \infty$.

The asymptotic convergence of MCMC methods provides a solid theoretical ground for practical applications of MCMC. However, MCMC methods are far from a black-box tool when constrained by finite computational resources. A part of the challenge is to determine how many MCMC steps are required to produce two independent samples. Relatively recently, there has been a development in this direction known as the coupling from the past (also known by the name of its authors as Propp-Wilson) [27, 28]. This framework provides a rigorous way of generating provably I.I.D. samples. However, a tractable algorithm can be developed only when the underlying physical configuration space assumes a *partial order* that is left unaltered by the MCMC dynamics. For example, this condition is satisfied for the Ising model under the heat-bath dynamics of local pairs of spins [29]. However, such analysis is not always possible and, we argue, is not necessary. That is because in practice, producing I.I.D. configuration samples $\{c\}$ is just a means to an end. Indeed, the MCMC practitioner is only interested in independent samples of an observable $\{O(c)\}$ that can decorrelate on a scale much shorter than c does. As witnessed by such an observable, it is common to see a transient period during which the value $O(c_t)$ sets into a stationary state starting from its initial value at $O(c_0)$ for a randomly

selected c_0. This "burn-in" period is usually discarded and the auto-correlation analysis, such as the binning analysis [30], is applied to the rest of the time-series. The result of such analysis is an implicit or explicit extraction of the *integrated auto-correlation time* τ [30, 31]. This constant provides an effective time-scale (as measured in Monte-Carlo update iterations) required to decorrelate an observable. In other words, a simulation run producing N_s samples, effectively contains only N_s/τ independent samples.

2.2 Imaginary time path-integral

2.2.1 Quantum-classical correspondence

Fundamentally, the difficulty of studying a quantum Hamiltonian H_q lies in the non-commutativity of its elements. Different strategies exist to tackle this problem numerically. In this section, in order to make progress, we decompose the Hamiltonian into conglomerate sub-parts as

$$H_q = H_a + H_b,$$

such that all terms making up $H_a(H_b)$ commute between themselves. Further, we can always rewrite the corresponding density matrix as a product of M terms as

$$\begin{aligned} Z_q &= \mathrm{Tr}\left[e^{-\beta H_q}\right] \\ &= \mathrm{Tr}\left[\left(e^{-\Delta\tau H_q}\right)^M\right], \end{aligned} \tag{2.14}$$

with $\Delta\tau = \beta/M$ defining a discretized step in the imaginary time direction of length β. Keeping $\Delta\tau$ small allows one to separate the non-commuting Hamiltonian terms via a controlled approximation as

$$e^{-\Delta\tau H} = e^{-\Delta\tau(H_a+H_b)} = \begin{cases} e^{-\Delta\tau H_a}e^{-\Delta\tau H_b} + O(\Delta\tau^2) \\ e^{-\Delta\tau H_b/2}e^{-\Delta\tau H_a}e^{-\Delta\tau H_b/2} + O(\Delta\tau^3). \end{cases} \tag{2.15}$$

where the error conceals a proportionality factor that is equal to the commutator $[H_a, H_b]$ and, therefore, is extensive in the system size. Plugging in the second approximation into

Eq. (2.14) yields

$$\begin{aligned} Z_q &= \mathrm{Tr}\left[\left(e^{-\Delta\tau H_b/2}e^{-\Delta\tau H_a}e^{-\Delta\tau H_b/2}+O(\Delta\tau^3)\right)^M\right] \\ &= \mathrm{Tr}\left[\left(e^{-\Delta\tau H_b/2}e^{-\Delta\tau H_a}e^{-\Delta\tau H_b/2}\right)^M\right]+O(\Delta\tau^2) \\ &= \mathrm{Tr}\left[\left(e^{-\Delta\tau H_a}e^{-\Delta\tau H_b}\right)^M\right]+O(\Delta\tau^2) \\ &\equiv Z_c+O(\Delta\tau^2). \end{aligned} \tag{2.16}$$

Here, in the third line, we have used the cyclic invariance of the trace in order to move the first term to the very end of the expression and, then, merged together adjacent exponentials of H_b. The second-to-last line defines Z_c. This partition function represents the part of Z_q that can be captured by the classical model up to $O(\Delta\tau^2)$. We proceed by focusing solely on this part and insert $M-1$ resolutions of the identity $\sum_s |s\rangle\langle s|$, such that

$$\begin{aligned} Z_c &= \sum_{s_0}\langle s_0|\left(e^{-\Delta\tau H_a}e^{-\Delta\tau H_b}\right)^M|s_0\rangle \\ &= \sum_{s_0,s_1,\ldots,s_{M-1}}\langle s_0|\,e^{-\Delta\tau H_a}e^{-\Delta\tau H_b}\,|s_1\rangle\langle s_1|\ldots|s_{M-1}\rangle\langle s_{M-1}|\,e^{-\Delta\tau H_a}e^{-\Delta\tau H_b}\,|s_0\rangle \\ &= \sum_{s_0,s_1,\ldots,s_{M-1}}\prod_{k=1}^{M}\langle s_{k-1}|\underbrace{e^{-\Delta\tau H_a}e^{-\Delta\tau H_b}}_{\equiv T}|s_k\rangle\,, \end{aligned} \tag{2.17}$$

with the last set of spins identified with the first one such that $s_M = s_0$. This expression is known as the Suzuki-Trotter decomposition, and it provides a path-integral representation of a quantum system in equilibrium. Within this picture, spin particles s evolve under the action of Hamiltonians H_a and H_b from one slice s_k to the next s_{k+1} and, as a result, trace out world lines in imaginary time. As such, this picture constitutes the foundation for a class of algorithms known as world line quantum Monte Carlo. However, we take a different route and proceed with the mapping to a classical Hamiltonian. For this, we consider each term in the product as a matrix element of a transfer-matrix T. To be more specific, we specialize the next steps to the model of interest, namely, the transverse field Ising model (TFIM) on a 2 dimensional bipartite rectangular grid with periodic boundary conditions. The Hamiltonian for this model is given by

$$H_{\mathrm{TFIM}} = -\overbrace{J_q\sum_{\langle i,j\rangle}\sigma_i^z\sigma_j^z}^{H_a} - \overbrace{\Gamma\sum_i\sigma_i^x}^{H_b}, \tag{2.18}$$

where σ^z, σ^x are Pauli matrices and the couplings are over the nearest neighbours. Taking $|s\rangle$ to be the basis set of the σ^z operator, we obtain

$$\begin{aligned} T_{s_k,s_{k+1}} &\equiv \langle s_k| \, e^{\Delta\tau J_q \sum_{\langle i,j\rangle} \sigma_i^z \sigma_j^z} e^{\Delta\tau\Gamma \sum_i \sigma_i^x} \, |s_{k+1}\rangle \\ &= e^{\Delta\tau J_q \sum_{\langle i,j\rangle} s_{k,i} s_{k,j}} \, \langle s_k| \, e^{\Delta\tau\Gamma \sum_i \sigma_i^x} \, |s_{k+1}\rangle \\ &= e^{\Delta\tau J_q \sum_{\langle i,j\rangle} s_{k,i} s_{k,j}} \prod_i \langle s_{k,i}| \underbrace{\mathbf{1}\cosh(\Delta\tau\Gamma) + \sigma^x \sinh(\Delta\tau\Gamma)}_{\equiv T^{\mathrm{loc,x}}} |s_{k+1,i}\rangle \end{aligned} \tag{2.19}$$

where the last term represents an element of a 2×2 matrix $T^{\mathrm{loc,x}}$,

$$T^{\mathrm{loc,x}} = \cosh(\Delta\tau\Gamma) \begin{pmatrix} 1 & \tanh(\Delta\tau\Gamma) \\ \tanh(\Delta\tau\Gamma) & 1 \end{pmatrix} = e^{J_\tau} \begin{pmatrix} 1 & e^{-2J_\tau} \\ e^{-2J_\tau} & 1 \end{pmatrix}.$$

On the right hand side, we have written an explicit expression for a local transfer-matrix $e^{J_\tau s_{k,i} s_{k+1,i}}$. Both sides can be made equal up to a proportionality factor by setting $J_\tau = -\frac{1}{2} \ln \tanh(\Delta\tau\Gamma)$. The proportionality factor is not important as it is independent of the spin configuration and serves only to match the ground state energies between the quantum and classical Hamiltonians. Therefore, we can substitute the classical representation $T^{\mathrm{loc,x}} = e^{J_\tau s_{k,i} s_{k+1,i}}$ into Eq. (2.19) and, then, plug the resulting expression back in Eq. (2.17) to arrive at the expression

$$Z_c = \sum_{s_0, s_1, \ldots, s_{M-1}} \prod_{k=1}^{M} e^{\Delta\tau J_q \sum_{\langle i,j\rangle} s_{k,i} s_{k,j} + J_\tau \sum_i s_{k,i} s_{k+1,i}} \tag{2.20}$$

This expression is nothing else but the partition function of a classical Ising model with the reduced Hamiltonian

$$H_c/T_c = -J_c \sum_{k,\langle i,j\rangle} s_{k,i} s_{k,j} - J_\tau \sum_{k,i} s_{k,i} s_{k+1,i} \tag{2.21}$$

where T_c is the classical temperature not related to β^{-1} and the reduced couplings J_c, J_τ are highly anisotropic since

$$\begin{aligned} J_c &= \Delta\tau J_q \\ J_\tau &= -\frac{1}{2} \ln \tanh(\Delta\tau\Gamma). \end{aligned} \tag{2.22}$$

Since the original Hamiltonian was two-dimensional, its classical analog is three-dimensional, where the third dimension is periodic and composed of M sites (see Fig. 2.2). As is clear

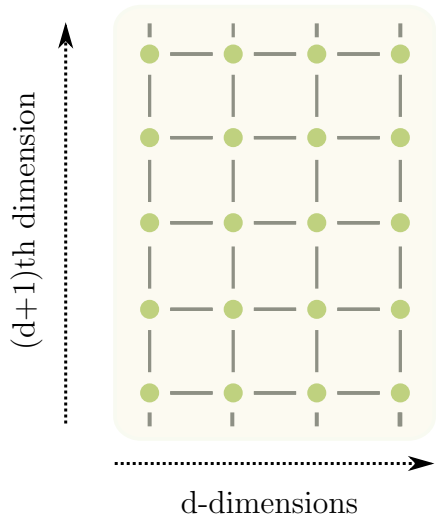


Figure 2.2: A $d+1$ classical representation of the transverse field Ising model in d dimensions. The original model is a chain of four quantum spins. The circles represent classical Ising spins. The lines represent ferromagnetic couplings the spins. The half-lines couple the boundary spins in the $+1$ direction. Those boundary interactions are due to the trace in the quantum expectation value, Eq. (2.14).

from Eq. (2.16), the equivalence between the models is obtained in the continuum limit $\Delta\tau \to 0$ of the quantum model and in the weak inter-slice coupling limit of the Ising model $J_c \to 0$. Going beyond this limit implies that Eq. (2.15), which validates the equivalence between the quantum and classical transfer matrices, breaks down. Specifically, a higher order expansion of the quantum imaginary-time propagator $e^{-\Delta\tau H_q}$ generates higher powers of H_q. Upon a mapping similar to Eq. (2.2.1), those powers of the quantum Hamiltonian effectively translate to non-local classical interactions, which can no longer be captured with a physical Hamiltonian.

However, while the microscopic correspondence between the quantum and classical models can only be established in the particular limit discussed above, the same mapping can be used to argue for the equivalence of the critical quantum and classical systems as far as their universal properties are concerned. This equivalence is in particular true for critical systems where the dynamic scaling exponent is equal to one, i.e., $z = 1$, such as TFIM. In those systems, the exponents that govern the divergences of the correlation lengths along the spatial and temporal directions are the same, implying that the couplings in both directions must appear on the same footing (up to a scalar prefactor) in the effective low-momentum action of the Hamiltonian. The full equivalence between the couplings in

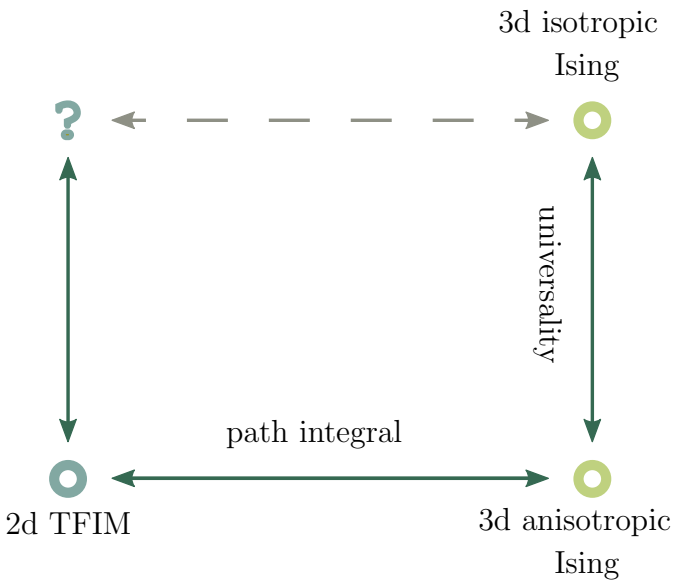


Figure 2.3: Microscopic critical theories that share the same universality class of the Wilson-Fisher fixed point and the relations between them. The critical 2d TFIM can be mapped onto the 3d anisotropic Ising model via a path-integral. The universality argument shows that isotropic and anisotropic Ising model are equivalent at criticality. However, it is not clear how to map the transfer matrix of the 3d isotropic Ising model to its 2d quantum microscopic analogue (hence the dashed line). Although the exact form of the later is unknown, it must be universally equivalent to the 2d TFIM.

the imaginary time and space directions can be thus fully regained by a simple rescaling one of the corresponding coordinate axis. In the language of the classical Ising model, this equivalence corresponds to mapping to a model with isotropic interactions.

The exact correspondence between the critical quantum and classical models can also be argued from the renormalization point of view. Indeed, any scaling function dependent on $\Delta\tau$ can only depend on it through a dimensionless ratio $\Delta\tau/\beta$. Naturally, the thermodynamic limit corresponds to $\Delta\tau/\beta \to 0$ since this quantity, which effectively corresponds to a classical dimension, represents a relevant operator that determines the dimensionality of the system. Therefore, under a renormalization step, the ratio must transform as $\Delta\tau/\beta \to b\Delta\tau/\beta$ where b is the rescaling factor. For a fixed β, this implies that $\Delta\tau \to b^{-1}\tau$ and, therefore, the discretization in the imaginary time represents an irrelevant variable, in spirit similar to the lattice spacing constant δ [32].

The quantum-classical correspondence provides a powerful tool to characterize critical theories. We use this mapping to study universal properties of the Wilson-Fisher fixed

point as reflected in the entanglement structure of the underlying critical system. In order to quantify the amount of entanglement, we use the notion of the entanglement entropy which is defined with respect to a subregion A of a many-body system. As we will see in the Sec. 3.4, the stochastic evaluation of the entanglement entropy requires working on a topologically modified configuration space where the imaginary time is first sliced midway and, then, is restitched according to the geometry of region A. This procedure translates into modified boundary conditions along the imaginary time direction. All other details of the quantum-classical mapping remain unchanged. Therefore, the universal features of the entanglement can be accessed via a Monte Carlo simulation of the anisotropic critical Ising model on a modified 3d lattice. Combined with the universal argument provided in the previous paragraphs, we have license to forget about the underlying microscopic quantum model and instead simply study the isotropic Ising model tuned to its critical point. The relationship between these models is visualized in Fig. 2.3. The motivation behind this choice of numerical strategy is a practical one. Constrained by the available computational resources, we desire to scale our simulations to the largest system sizes possible. For comparison, a natural alternative approach would be to use the framework of the Stochastic Series Expansion in order to directly simulate the critical 2d TFIM. We have seen how this microscopic model is equivalent to a highly anisotropic Ising model. However, the ultimate goal is to access the universal critical theory captured by the ϕ^4 theory, which treats the space and time directions on equal footing. Therefore, it is more natural to study a microscopic theory with inherent isotropic interactions - the isotropic Ising model. Although we haven't conducted direct numerical comparisons, we note that the systems sizes reached in our simulations (effectively up to $L = 102$) exceed the ones obtained in Ref. [33] which employed the SSE approach and only reached $L = 36$.

2.2.2 Cluster updates with restricted Boltzmann machine

In this section, we take a non-traditional approach to derive a rejection-free cluster update required for an efficient MCMC update of the critical Ising model. The widely successful Wolf [34] and Swendsen-Wang [35] cluster updates are largely based on the Fortuin-Kasteleyn cluster decomposition [36] of the Ising model. This transformation stochastically maps an Ising configuration into the one of a random cluster model. We show how a restricted Boltzmann machine (RBM, introduced in Sec. 5.2) with a special architecture can be seen as a bridge between those two thermodynamic models. Such a representation makes explicit the equivalence between the cluster update and the two-step blocked Gibbs update of the RBM that takes advantage of the inherited $\mathbf{Z}_2$ symmetry. The Fortuin-Kasteleyn transformation is recovered in a particular limit of the RBM parameters. The derivation

that follows is largely based on Ref. [37].

Motivated by the quantum-to-classical mapping derived in the previous section, we are interested in simulating the three dimensional isotropic Ising model which we write in the form

$$E(\boldsymbol{s}) = -J \sum_{l=(l_1,l_2)} s_{l_1} s_{l_2}, \tag{2.23}$$

where the sum is over the bond/link variable l defined as a tuple $l = (l_1, l_2)$ listing the spin indices. This energy function induces a Boltzmann distribution over the spin variables, given by

$$p_{\text{Ising}}(\boldsymbol{s}) = \frac{1}{Z_{\text{Ising}}} e^{-\beta E(\boldsymbol{s})} = e^{-\beta(E(\boldsymbol{s}) - F_{\text{Ising}})}, \tag{2.24}$$

where the partition function is $Z_{\text{Ising}} = \sum_{\boldsymbol{s}} e^{\beta E(\boldsymbol{s})}$ and the free energy is $F_{\text{Ising}} = -1/\beta \ln Z_{\text{Ising}}$. Using machine learning vocabulary, we next treat the physical Ising spins, $\mathbf{s}$, as visible units and couple them to auxiliary/hidden units, $\boldsymbol{h}$. Each new hidden unit is coupled to a bond of the original Ising model resulting in the energy function

$$E(\boldsymbol{s}, \boldsymbol{h}) = -\sum_{l} \left(W s_{l_1} s_{l_2} + b\right) h_l. \tag{2.25}$$

Unlike the spin variable s taking values ± 1, the hidden variable $h_l \in \{0, 1\}$ with the two states referred to as open and closed correspondingly. Here, W and b are free parameters. In principle these parameters can depend on the index l but in order to respect the symmetries of the original Ising Hamiltonian they are assumed to be independent of the index. This energy function also induces a probability distribution over its variables via the Boltzmann distribution, for which we set the temperature to one for convenience such that

$$p_{\text{BM}}(\boldsymbol{s}, \boldsymbol{h}) = \frac{1}{Z_{\text{BM}}} e^{-E(\boldsymbol{s}, \boldsymbol{h})} = e^{-(E(\boldsymbol{s}, \boldsymbol{h}) - F_{\text{BM}})}. \tag{2.26}$$

This model falls within the family of higher-order RBMs [38]. Unlike the standard RBM, this model is based on three-body interactions. Such a generalization allows us to "train" the RBM to perfectly capture the Ising model. For this, we require the RBM marginal probability over spin variables to exactly reproduce the Ising distribution such that

$$p_{\text{Ising}}(\boldsymbol{s}) = p_{\text{BM}}(\boldsymbol{s}) \equiv \sum_{h} p_{\text{BM}}(\boldsymbol{s}, \boldsymbol{h}). \tag{2.27}$$

Instead of the gradient-based numerical methods adopted in Ch. 5, here we rely on theoretical manipulations in order to find the RBM parameters for which this equality is satisfied.

As the first step, we bring both distributions into a product form. The Ising model reads

$$\begin{aligned} p_{\text{Ising}}(\boldsymbol{s}) &= e^{F_{\text{Ising}}} \prod_{l=1}^{N_l} e^{\beta J s_{l_1} s_{l_2}} \\ &= \prod_{l=1}^{N_l} e^{f_{\text{Ising}} + \beta J s_{l_1} s_{l_2}}, \end{aligned} \tag{2.28}$$

where $f = F/N_l$ is the free energy per link. Similarly, for the RBM, we get

$$\begin{aligned} p_{\text{BM}}(\boldsymbol{s}) &= \sum_{h} p_{\text{BM}}(\boldsymbol{s}, \boldsymbol{h}) \\ &= e^{F_{\text{BM}}} \sum_{h} \prod_{l=1}^{N_l} e^{\left(W s_{l_1} s_{l_2} + b\right) h_l} \\ &= \prod_{l=1}^{N_l} e^{f_{\text{BM}}} \left(e^{W s_{l_1} s_{l_2} + b} + 1\right). \end{aligned} \tag{2.29}$$

As the next step, we fulfill Eq. (2.27) by requiring the equality to be satisfied at the level of each individual link such that

$$e^{\Delta f + \beta J s_{l_1} s_{l_2}} = e^{W s_{l_1} s_{l_2} + b} + 1, \tag{2.30}$$

with $\Delta f \equiv f_{\text{Ising}} - f_{\text{BM}}$. We expand this equation into the system of equations

$$\begin{cases} e^{b+W} + 1 = e^{\Delta f + \beta J} \\ e^{b-W} + 1 = e^{\Delta f - \beta J} \end{cases} \Rightarrow \begin{cases} b + W = \ln\left(e^{\Delta f + \beta J} - 1\right) & \text{for } s_{l_1} s_{l_2} = +1 \\ b - W = \ln\left(e^{\Delta f - \beta J} - 1\right) & \text{for } s_{l_1} s_{l_2} = -1 \end{cases} \tag{2.31}$$

Note that Δf can be made to take any value by offsetting the ground state energy in Eqs. (2.23) and (2.25). However, the system of equations is well-defined only for a positive arguments of the logarithm, such that we require $\Delta f \geq \beta |J|$. As we are interested in the ferromagnetic Ising model with $J \geq 0$, we set $\Delta f = \beta J + \epsilon$ where ϵ is a small non-negative parameter. In the limit $\epsilon \to 0$, to which we refer as the Fortuin-Kasteleyn (FK) limit for reasons that will become apparent in the next section, we get

$$\begin{cases} b + W = \ln\left(e^{2\beta J} - 1\right) \\ b - W = -\infty \end{cases} \tag{2.32}$$

It is interesting to consider the dynamics of the blocked Gibbs update (see Sec.5.2.2) in this limit. We start with the conditional probability over the hidden units. We note that, conditioned on a spin state, the conditional is factorized as

$$p(\boldsymbol{h} \mid \boldsymbol{s}) = \prod_{l=1}^{N_l} p(h_l \mid \boldsymbol{s}), \tag{2.33}$$

where the activation probability of an individual hidden unit is given by

$$\begin{aligned} p(h_l = 1 \mid \boldsymbol{s}) &= \sigma\left(W s_{l_1} s_{l_2} + b\right) \\ &= \begin{cases} \sigma\left(b + W\right) & \text{for } s_{l_1} s_{l_2} = +1 \\ \sigma\left(b - W\right) & \text{for } s_{l_1} s_{l_2} = -1 \end{cases}, \end{aligned} \tag{2.34}$$

and $\sigma(x) = 1/\left(1 + e^{-x}\right)$ is a sigmoid function. In the FK limit, this activation probability converges to

$$p(h_l = 1 \mid \boldsymbol{s}) \overset{\Delta f = \beta J}{=} (1 - e^{-2\beta J}) \delta_{s_{l_1}, s_{l_2}} \equiv p \delta_{s_{l_1}, s_{l_2}}, \tag{2.35}$$

such that the hidden unit is deterministically set to zero when the associated spins variables are misaligned. When they are aligned, the hidden unit is activated with probability $p \equiv 1 - e^{-2\beta J}$.

The inverse conditional $p(\boldsymbol{s} \mid \boldsymbol{h})$ does not factorize as easily and, therefore, it is not seemingly amenable for a blocked Gibbs update. However, it has an important symmetry that can be exploited. To make this symmetry apparent, we note that according to Bayes' law,

$$p(\boldsymbol{s} \mid \boldsymbol{h}) = \frac{p(\boldsymbol{s}, \boldsymbol{h})}{p(\boldsymbol{h})} \propto e^{-E(\boldsymbol{s} \mid \boldsymbol{h})}, \tag{2.36}$$

where the constant of proportionality is not dependent on the spin state $\mathbf{s}$. For convenience, we introduce the notation for the clamped RBM energy, $E(\boldsymbol{s} \mid \boldsymbol{h})$; it is the same as the original RBM energy, $E(\boldsymbol{s}, \boldsymbol{h})$, but the hidden units are clamped to the conditioning vector $\boldsymbol{h}$. Upon a close examination of $E(\boldsymbol{s} \mid \boldsymbol{h})$, we note that a non-activated hidden unit h_l removes the original spin-spin interaction on the link l. It follows that bonds that have been disactivated in this manner partition the spin state into a set of non-interacting clusters. Moreover, independent Hamiltonians describing each of those clusters inherit the $\mathbb{Z}_2$ invariance of their parental Ising model. Lastly, the clusters are necessarily composed of aligned spins (due to the activation of $\mathbf{h}$ according to Eq. (2.35)). Combining those

observations, we conclude that Eq. (2.36) factorizes as

$$p(\boldsymbol{s} \mid \boldsymbol{h}) = \prod_{c=1}^{N_c} p(\boldsymbol{s}_c \mid \mathbf{h}) \tag{2.37}$$

where the index c iterates over N_c clusters and each cluster includes its own set of aligned spin variables $\boldsymbol{s}_c$. Additionally, due to the previously mentioned $\mathbb{Z}_2$ invariance, both collective states of a cluster $\boldsymbol{s}_c$ are equally likely implying that

$$p(\boldsymbol{s}_c = \mathbf{1} \mid \mathbf{h}) = p(\boldsymbol{s}_c = -\mathbf{1} \mid \mathbf{h}) = \frac{1}{2}, \tag{2.38}$$

which concludes the derivation. Indeed, the two-steps blocked Gibbs update defined by Eqs. (2.33), (2.35) and (2.38) is equivalent to the Swendsen-Wang cluster update [35]. The condition of detailed balance, Eq. (2.11), with respect to the spin state can be verified since

$$\begin{aligned} p(s)T(\mathbf{s} \to \mathbf{s}') &= \sum_{\mathbf{h}} p(\mathbf{s}' \mid \mathbf{h}) p(\mathbf{h} \mid \mathbf{s}) p(\mathbf{s}) \\ &= \sum_{\mathbf{h}} p(\mathbf{s}', \mathbf{h}) \frac{1}{p(\mathbf{h})} p(\mathbf{h}, \mathbf{s}) \\ &= \sum_{\mathbf{h}} p(\mathbf{h} \mid \mathbf{s}') p(\mathbf{s}') p(\mathbf{s} \mid \mathbf{h}) \\ &= p(\mathbf{s}') T(\mathbf{s}' \to \mathbf{s}), \end{aligned}$$

thus validating the RBM update within the MCMC framework. The inefficiency overhead of this update is related to the fact that the new state, $\boldsymbol{s}'$, is only indirectly dependent on the previous state, $\boldsymbol{s}$, through an intermediary hidden state, $\boldsymbol{h}$. This indirect dependence leads to the possibility that some clusters are not flipped despite the invested computational cost of building them. Building just a single cluster that is always flipped provides an elegant solution to this problem. Such a cluster update is known as the Wolff update [34].

2.2.3 Fortuin-Kasteleyn limit

In this section, the origin of the name given to the limit leading to Eq. (2.32) finally becomes clear. Namely, we show how, in the FK limit, the marginal probability distribution over the hidden units is effectively the random cluster model [39]. The latter model lies at the heart of the Fortuin-Kasteleyn cluster decomposition of the Ising model [36]. This dual model provides the basis of an improved entanglement entropy estimator that we derive

in Sec. 3.4.2. With this goal in mind, we use Bayes' rule to rewrite the RBM probability distribution, Eq. (2.25), in a form that lets us reuse the previously derived results, so that

$$\begin{aligned} p_{\mathrm{RBM}}(\boldsymbol{s},\boldsymbol{h}) &= p(\boldsymbol{h} \mid \boldsymbol{s}) p_{\mathrm{RBM}}(\boldsymbol{s}) \\ &\sim p(\boldsymbol{h} \mid \boldsymbol{s}) p_{\mathrm{Ising}}(\boldsymbol{s}), \end{aligned}$$

where the constant of proportionality is $e^{-N_l \Delta f} = e^{-N_l \beta J}$. We proceed by substituting the expressions for the conditional, Eqs. (2.33) and (2.35), as well as the marginal, Eq. (2.28), into the above, yielding

$$\begin{aligned} p_{\mathrm{RBM}}(\boldsymbol{s},\boldsymbol{h}) &= e^{-N_l \beta J} \prod_{l=1}^{N_l} p\left(h_l \mid \boldsymbol{s}\right) \prod_{l=1}^{N_l} e^{\beta J s_{l_1} s_{l_2}} \\ &= \prod_{l=1}^{N_l} \left[p\left(h_l = 1 \mid \boldsymbol{s}\right) \delta_{h_l,1} + p(h_l = 0 \mid \boldsymbol{s}) \delta_{h_l,0}\right] \prod_{l=1}^{N_l} e^{\beta J s_{l_1} s_{l_2} - 1} \\ &= \prod_{l=1}^{N_l} \left[p \delta_{s_{l_1},s_{l_2}} \delta_{h_l,1} + (1 - p\delta_{s_{l_1},s_{l_2}}) \delta_{h_l,0}\right] \prod_{l=1}^{N_l} \left[(1-p)\delta_{s_{l_1},-s_{l_2}} + \delta_{s_{l_1},s_{l_2}}\right] \\ &= \prod_{l=1}^{N_l} \left[p \delta_{s_{l_1},s_{l_2}} \delta_{h_l,1} + (1-p)\delta_{h_l,0}\right], \end{aligned} \tag{2.39}$$

with $p = 1 - e^{-2\beta J}$ previously defined in Eq. (2.35). Here, in going from the third to the last line we have used the identities

$$\begin{aligned} \delta_{s_{l_1},s_{l_2}} \delta_{s_{l_1},-s_{l_2}} &= 0, \\ \delta_{s_{l_1},s_{l_2}} + \delta_{s_{l_1},-s_{l_2}} &= 1, \\ \delta_{s_{l_1},s_{l_2}} \delta_{s_{l_1},s_{l_2}} &= \delta_{s_{l_1},s_{l_2}}. \end{aligned}$$

Eq. (2.39) corresponds to Eq. (2.25) in the FK limit. By construction, the marginal of this joint distribution over spins, $p_{\mathrm{RBM}}(\boldsymbol{s})$, is the Ising model. However, we are interested to explore $p_{\mathrm{RBM}}(\boldsymbol{h})$ which is given by

$$\begin{aligned} p_{\mathrm{RBM}}(\boldsymbol{h}) &= \sum_{\boldsymbol{s}} p_{\mathrm{BM}}(\boldsymbol{s},\boldsymbol{h}) \\ &= \sum_{\boldsymbol{s}} \prod_{l=1}^{N_l} \left[p \delta_{s_{l_1},s_{l_2}} \delta_{h_l,1} + (1-p)\delta_{h_l,0}\right]. \end{aligned}$$

We proceed by reindexing the links based on the activation of the hidden units. The active (open) links are reindexed to the range $[1, n(\mathbf{h})]$ where $n(\mathbf{h})$ is the number of active links. All other $N_l - n(\mathbf{h})$ links must be non-activated (closed). Using the fact that the cross product terms cancel out, $\delta_{h_l,1}\delta_{h_l,0} = 0$, we expand the product in the above expression as

$$\begin{aligned} p_{\mathrm{RBM}}(\boldsymbol{h}) &= \sum_{\boldsymbol{s}} \prod_{\substack{l=1 \\ h_l=1}}^{n(\boldsymbol{h})} p\delta_{s_{l_1},s_{l_2}} \prod_{\substack{l=n(\boldsymbol{h})+1 \\ h_l=0}}^{N_l} (1-p) \\ &= \sum_{\boldsymbol{s}} p^{n(\boldsymbol{h})}(1-p)^{N_l-n(\boldsymbol{h})} \prod_{\substack{l=1 \\ h_l=1}}^{n(\boldsymbol{h})} \delta_{s_{l_1},s_{l_2}}, \end{aligned}$$

where the second line follows from the fact that the product over closed links is independent of the spin state, which is not the case for the product over active links. In fact, this product can be extended to include all the links, thus generalizing to an indicator function defined as

$$\mathbb{1}(\boldsymbol{s}, \boldsymbol{h}) \equiv \prod_{l=1}^{N_l} \delta_{s_{l_1},s_{l_2}} \delta_{h_l,1}, \tag{2.40}$$

where spins-link configurations that are not compatible with this rule are automatically set to zero by this function. Upon substitution into the previous equation, we get

$$\begin{aligned} p_{\mathrm{RBM}}(\boldsymbol{h}) &= \sum_{\boldsymbol{s}} p^{n(\boldsymbol{h})}(1-p)^{N_l-n(\boldsymbol{h})} \mathbb{1}(\boldsymbol{s}, \boldsymbol{h}) \\ &= p^{n(\boldsymbol{h})}(1-p)^{N_l-n(\boldsymbol{h})} 2^{N_c(\boldsymbol{h})}. \end{aligned} \tag{2.41}$$

To derive the last line, we note that the active links define a set of connected but independent clusters. The indicator function nullifies all the spin configurations that contain misaligned spins within those link clusters. The non-zero contributions of equal weight, $p^{n(\boldsymbol{h})}(1-p)^{N_l-n(\boldsymbol{h})}$, originate from spin configurations that are aligned within the clusters. Those configurations can be easily counted. Indeed, there are two sets of spin configurations to satisfy the alignment condition within each cluster. Since the clusters are independent, in total, there must be 2^{N_c} valid spin configurations.

Concluding the Fortuin-Kasteleyn decomposition, we sum up the resulting marginal probability over all hidden states, thus, obtaining a partition function for the random cluster model,

$$Z_{RC} \equiv \sum_{\mathbf{h}} p^{n(\boldsymbol{h})}(1-p)^{N_l-n(\boldsymbol{h})} 2^{N_c(\boldsymbol{h})}. \tag{2.42}$$

We also note the relationship between the value of partition functions between the joint and marginal distributions, given by

$$Z_{RC} = Z_{RBM} = Z_{Ising} e^{\beta J N_l}. \tag{2.43}$$

At criticality, the clusters come in all possible sizes and their distribution follows a power law [40]. This critical behaviour, together with the clusters' non-local geometrical nature often make it useful to rewrite the MCMC estimators in terms of clusters statistics [31]. The cluster-based estimators usually have a dramatically smaller variance and, therefore, are known as *improved estimators*. In Sec. (3.4.2) we derive an improved estimator for the Rényi entanglement entropy with an exponentially improved performance.

2.3 Stochastic series expansion

2.3.1 Operator list representation

Contrary to the quantum-classical correspondence derived in the previous section based on the path-integral interpretation of the partition function, the starting point for the development of a quantum Monte Carlo framework known as the Stochastic Series Expansion (SSE) is the Taylor expansion of the partition function,

$$Z = \mathrm{Tr}\left[e^{-\beta H}\right] = \sum_{s}\sum_{n=0}^{\infty} \langle s| \frac{(-\beta H)^n}{n!} |s\rangle \,. \tag{2.44}$$

Here, $|s\rangle$ represents a state from a complete basis set. For concreteness, we consider the standard spin-$\frac{1}{2}$ S_z basis,

$$|s\rangle = |S_1^z, S_2^z, \ldots, S_N^z\rangle \,,$$

where N is the number of spins. In order to build-in the physical locality in the method, we write the Hamitlonian as a sum of local interactions such that

$$H = -\sum_{b=1}^{N_b} H_b,$$

where N_b is the number of interacting bonds between spins. For example, the reduced Hamiltonian of a quantum spin $\frac{1}{2}$ XXZ-model assumes the form

$$\begin{aligned} H_{XXZ}/J &= \sum_{b=1}^{N_b} \left[S_{i(b)}^x S_{j(b)}^x + S_{i(b)}^y S_{j(b)}^y + \Delta S_{i(b)}^z S_{j(b)}^z \right] \\ &= \sum_{b=1}^{N_b} \left[\frac{1}{2}\left(S_{i(b)}^+ S_{j(b)}^- + S_{i(b)}^- S_{j(b)}^+ \right) + \Delta S_{i(b)}^z S_{j(b)}^z \right]. \end{aligned} \tag{2.45}$$

Specifically, in this expression, the variable b indexes interaction bonds between spins $i(b)$ and $j(b)$. The parameter Δ tunes the degree of the anisotropy between the couplings. At two special symmetry points, $\Delta = 0$ and $\Delta = 1$, the resulting models are known as the XY and Heisenberg models correspondingly.

Based on this basis choice, we further subdivide the bond into diagonal and off-diagonal ones as

$$H_b = H_{D,b} - H_{O,b},$$

where

$$H_{D,b} = C - \Delta S^z_{i(b)} S^z_{j(b)} \tag{2.46}$$
$$H_{O,b} = \frac{1}{2}\left(S^+_{i(b)} S^-_{j(b)} + S^-_{i(b)} S^+_{j(b)}\right).$$

Here, the constant C has been added to each bond in order to make the diagonal bond positive, which is achieved for

$$C = C_0 + \epsilon, \quad C_0 = \frac{\Delta}{4},$$

where $\epsilon \geq 0$ is a hyper-parameter which can be used to optimize the dynamics of the simulation. The shift in the ground state energy due to C has no effect on the physical observables except for the expectation of energy. In order to recover the energy of the original model, it is sufficient to subtract $N_b C$ from the estimated value.

Sequences of local bond Hamiltonians that result from the expansion in powers of H in Eq. (2.44) can be indexed by the so-called operator-list,

$$S_n = [a_1, b_1][a_2, b_2] \ldots [a_n, b_n]. \tag{2.47}$$

Here, the index i iterates over the elements in the operator list; the label $a_i \in \{D, O\}$ indicates the type of an operator (diagonal or off-diagonal as per Eq. (2.46)); the index $b_i \in \{1, \ldots, N_b\}$ labels the bond index; the list length, n, is directly related to the expansion power, H^n. We now rewrite the Taylor expansion Eq. (2.44) in the new notation as

$$Z = \sum_s \sum_{n=0}^{\infty} \sum_{S_n} (-1)^{n_2} \frac{\beta^n}{n!} \langle s| \prod_{i=1}^{n} H_{a_i,b_i} |s\rangle, \tag{2.48}$$

where we have redefined $\beta \leftarrow \beta/J$ and introduced an integer variable n_2 that counts the number of off-diagonal operators in the operator-list. For some Hamiltonians referred to as *sign-free* Hamiltonians, it is possible to get rid of the oscillating negative sign. With this goal in mind, we define the propagated spin state,

$$|s(p)\rangle = \prod_{i=1}^{p} H_{a_i,b_i} |s\rangle, \tag{2.49}$$

which is the state obtained by application of the first p operators from the operator-list onto the state $|s\rangle$. Due to the specific bond decomposition in Eq. (2.46), there is no branching-out of states. Therefore, the propagated state can always be represented by a single basis

state. The trace in Eq. (2.48) imposes periodic boundary conditions for the propagated state in the expansion direction such that $\langle s(0)|s(n)\rangle = 1$. Only operator sequences S_n, that respect this condition have a positive contribution to the partition function. For a bipartite lattice, this condition is only satisfied by operator sequences with an even number of off-diagonal operators and, therefore, $(-1)^{n_2} = 1$ in Eq. (2.48). Hence, each term in the partition function of Eq. (2.48), given by

$$W(S_n, s) = \frac{\beta^n}{n!} \langle s| \prod_{i=1}^{n} H_{a_i,b_i} |s\rangle \,, \tag{2.50}$$

is positive. Therefore, a Markov chain Monte Carlo procedure can be implemented based on this weight. The targeted configuration space consists of all possible operator sequence and basis state pairs, $c = (S_n, s)$, that satisfy the periodic boundary conditions in the expansion direction. An example of such configuration is depicted in Fig. 2.4 (a).

The operator sequence length n is a random quantity fluctuating around its natural mean value, $\langle n\rangle$. In order to simplify a further algorithmic implementation, it is useful to keep the operator length a constant size $M > \langle n\rangle$. Therefore, we truncate the sum iterating over the index n in Eq. (2.48) to M terms. Operator sequences with more then M operators are removed. As for the sequences that are smaller than M, they are extended to size M by inserting $M - n$ identity operators $H_{\mathbf{1},0} = \mathbf{1}$ in all the possible ways. To compensate for this added redundancy, we divide the weight of each string by its degeneracy $\binom{M}{M-n}$. The new expression for the partition function reads

$$Z = \sum_{s} \sum_{S_M} \frac{\beta^n (M-n)!}{M!} \langle s| \prod_{i=1}^{n} H_{a_i,b_i} |s\rangle \,, \tag{2.51}$$

where now n counts the number of non-identity operators in the sequence S_M, $[a_i, b_i] \neq [0, 0]$. As will be shown in Sec. 2.3.5, the average operator list length is directly related to the expected energy $\langle n\rangle = -\beta \langle E\rangle$. Therefore, fluctuations in n should be of order $\sqrt{\langle n\rangle}$. M should be set to a value that can accomodate any statistically significant fluctuations in n. An adjustable procedure is adopted in order to choose M in practice. During the burn-in stage of a simulation, M is repeatedly readjusted to a value rn_{max} where r is a small coefficient (we take values between 1.2 and 1.5) and n_{max} is the largest encountered operator length. Typically the value of M stabilizes very fast and, although truncated, Eq. (2.51) provides an unbiased representation for the full partition function. In order to efficiently sample the space of configurations contributing to this partition function, two types of a MCMC update are required, as we discuss in the following two subsections.

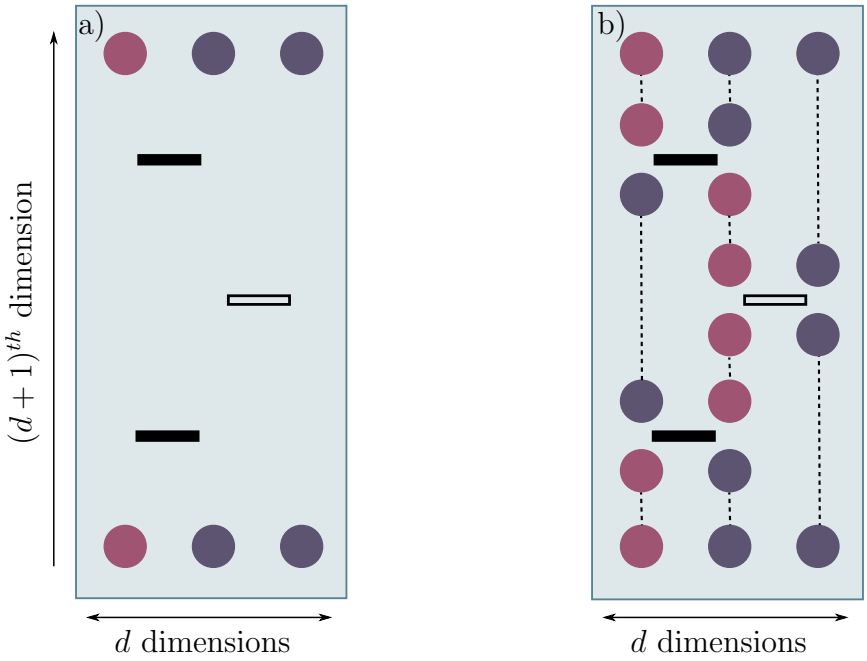


Figure 2.4: An example of a SSE configuration as expressed in terms of a) a binary state s (spin up and down) and an operator list pair. The off-diagonal/diagonal operators are denoted by solid/hollow rectangles. b) linked vertex list. In this representation, the type of operator can be directly determined from its action on the spins; the rectangles are kept only to provide a visual clue. We also choose to redundantly display the spin state for the spins state which, otherwise, can be inferred from the vertices. In both representations, aside from the d physical dimensions, there is an additional $(d+1)^{th}$ dimension corresponding to the Taylor series expansion.

2.3.2 Diagonal update

As the name suggests, the diagonal update only modifies the diagonal operators present in the operator list. The update consequently iterates over the elements in S_M. When it encounters a diagonal element (including the identity operator $H_{\mathbf{1},0}$), it proposes a bidirectional replacement $[0_p, 0_p] \leftrightarrow [D_p, b_p]$ that is accepted with the probabilities

$$P\left([0_p, 0_p] \to [D_p, b_p]\right) = \frac{N_b \beta \left\langle s(p)|H_{D,b}|s(p)\right\rangle}{M-n} \tag{2.52}$$

$$P\left([0_p, 0_p] \leftarrow [D_p, b_p]\right) = \frac{M-n+1}{N_b \beta \left\langle s(p)|H_{D,b}|s(p)\right\rangle} \tag{2.53}$$

where the bond b_p is chosen with a uniform probability out of all possible N_b bonds. The factor N_b appears because the proposal dynamics connects a unique identity state, $[0_p, 0_p]$, with N_b diagonal bond states, $[D_p, b_p]$. We also note that this update changes the operator list length, n, by ± 1 which explains the corresponding factors in the acceptance probabilities above. The state $|s(p)\rangle$ is constantly updated in accordance with Eq. (2.49) as the off-diagonal operators are encountered during the traversal of the operator list.

2.3.3 Off-diagonal update

The off-diagonal update is responsible for generating processes of type $[D_p, b_p] \leftrightarrow [O_p, b_p]$. The auxiliary identity operator, $[\mathbf{1}, 0]$, is left untouched. For this reason, during this update, it is useful to work with the original operator-list S_n instead of S_M and the associated configuration weight defined in Eq. (2.50).

Contrary to the diagonal-update, the off-diagonal update cannot be performed on the level of a single operator because such an operation requires a spin-flip, which can easily lead to a forbidden configuration. In order to make sure that such inconsistencies do not arise, it is useful to change the representation of the configuration space from the (S_M, s) tuple to the one that explicitly tracks the spin state. Such a representation is known as the *linked vertex list*.

The vertex is the fundamental object in the vertex list representation. It is defined with respect to a bond operator and the spins it acts on. More specifically, the spins affected by the operator acting on the bond b_p are

$$\left|s_{b_p}(p)\right\rangle = \left|S^z_{i(b_p)}, S^z_{j(b_p)}\right\rangle, \tag{2.54}$$

while all other spins in the propagated state $|s(p)\rangle$, as defined in Eq. (2.49), remain unchanged. The vertex is a representation for one of the matrix elements of a local bond operator $\left\langle s_{b_p}(p)\middle|H_{b_p}\middle|s_{b_p}(p-1)\right\rangle$. For the XXZ model, Eq. (2.46), there are 6 possible

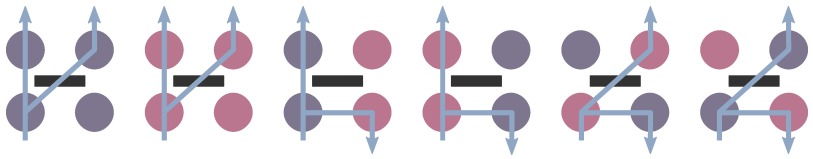


Figure 2.5: The six valid vertices that arise in the simulation of the XXZ model. They represent matrix elements of a local bond operator denoted by a filled bar. The operator acts on a pair of entrance spins ("legs") below and produce two exit legs as the output above. The 4 diagonal operators lead to the same state of the exit legs while the 2 off-diagonal operators flip both legs. On top of the vertices, the arrows indicate the possible loop construction actions that can be taken at a given vertex starting from the bottom left entrance leg. Those processes go under self-explanatory names and are known as "continue-straight", "switch-and-continue" and "switch-and-reverse".

vertices,

$$\begin{aligned} \langle\uparrow\uparrow|H_b|\uparrow\uparrow\rangle &= \langle\downarrow\downarrow|H_b|\downarrow\downarrow\rangle = \epsilon \\ \langle\downarrow\uparrow|H_b|\downarrow\uparrow\rangle &= \langle\uparrow\downarrow|H_b|\uparrow\downarrow\rangle = \frac{1}{2} \\ \langle\downarrow\uparrow|H_b|\uparrow\downarrow\rangle &= \langle\uparrow\downarrow|H_b|\downarrow\uparrow\rangle = \Delta/2 + \epsilon. \end{aligned} \tag{2.55}$$

Those vertices are visualized in Fig. 2.5. The spins that are part of a vertex are referred to as *vertex legs*. Note that setting the hyper-parameter ϵ to 0 decreases the number of allowed vertices from 6 to 4, thus reducing the total number of possible configurations. However, in practice, having $\epsilon > 0$ is helpful for the mixing of a Markov chain. The ϵ-weighted vertices provide intermediate configurations that help the Markov chain to tunnel from one configuration to another. At the same time, it is discouraged setting ϵ to a large value as the operator list is expected to grow as $\epsilon\beta N_b$ due an additional energy contribution. The weight of a configuration defined by a list of vertices directly follows from Eq. (2.50) and is given by

$$W(S_n, s) = \frac{\beta^n}{n!} \prod_{p=1}^{n} \left\langle s_{b_p}(p) \right| H_{b_p} \left| s_{b_p}(p-1) \right\rangle . \tag{2.56}$$

A SSE configuration in the vertex-list representation is depicted besides the equivalent operator-list one in Fig. 2.4. The vertices are *linked* between themselves as two consecutive vertices acting on the same spin are not independent. Indeed, the output leg of the preceding vertex in the operator sequence must be set to the same state as the input leg of the vertex that follows. Caution should be taken due to the periodicity in the expansion dimension. This constraint implies that the link between the output leg of the last vertex

with respect to a spin is connected to the input leg of the first vertex over the same spin. In the case when there is only one vertex associated with a spin, the link is created between the vertex' own legs.

The linked list of vertices is the essential object required for the execution of the off-diagonal update. It provides all of the necessary information for the implementation of the so called *loop algorithm*. In this algorithm, the "loop" refers to a type of the update proposal. Specifically, a loop is a chain of spin flips in the vertex-list representation that is closed, i.e. the chain loops back to its own start. The topology of a loop guarantees that no defects arise in the updated configuration.

Construction of a loop is done incrementally via a chain of small steps. As the first step, the initial leg is randomly selected (chosen out of $4n$ possibilities). This leg is referred to as the "entrance" leg. The loop is then extended by an addition of an "exit" that shares a vertex with the entrance leg. As the spins along the path of the loop are eventually flipped, not every leg is considered in the choice. For example, Fig. 2.5 shows the possible moves for the XXZ vertices. Choosing the entrance leg as the exit leg is a valid choice. Such a move is known as the *bounce move* and, if it happens as the first move, the loop construction is terminated leading to no change in the configuration. However, if one of the alternatives is selected, the exit leg is added to the loop and the loop construction proceeds to the leg linked to the exit leg. The new leg becomes the entrance leg and the process repeats itself. As the loop expands, there are two links with mismatched legs that exist at the head and tail of the loop. The loop construction terminates when the mismatched links cancel each other out, that is, when the head and tail meet. This event can arise in two ways: either through a vertex move or through a move following a link. Only in the second case does the initial spin change its state when the spins along the loop are flipped. As a result of a loop update, the vertices along the loop change their type. Vertices that have been subjected to the *switch*-type of moves (see Fig. 2.5) lead to the change of the associated operator type. We also note that the network of the linked vertex list is left unaltered during this update.

The previous paragraph describes the general idea of a loop construction. However, in order to provide a valid MCMC update, it must respect the detailed balance equation in Eq. (2.11). This criterion locally conditions the vertex-moves probabilities and relates them by the so called *directed loops equations* [41]. As detailed balance is satisfied on the level of each move, there is no rejection step at the end of the loop construction, leading to an efficient cluster update. Ref. [41] provides an excellent resource deriving these equations directly from the detailed balance condition. Additionally, this reference gives examples of optimized solutions to those equations based on the minimization of the bounce moves. Such optimization in the loop construction process can lead to a dramatic

increase in the performance of the algorithm [41]. In special cases of certain Hamiltonians, the back-tracking caused by the bounce move can be avoided without the need for the directed loops equations. Such examples include the high-symmetry points of the XXZ model, Eq.(2.45), such as the Heisenberg ($\Delta = 1$) and XY ($\Delta = 0$) models. For those models, there is a choice of the parameter ϵ, Eq. (2.3.1), that leads to a set of vertices that all share the same weight. As the loop update only changes the vertex type, the SSE configuration weight is preserved after the loop flip. This simplification allows one to partition the vertices into a set of independent loops, analogously to the Swendsen-Wang clusters (see Sec.2.2.2), that can be independently flipped with probability $\frac{1}{2}$. In the case of the Heisenberg model, the loop break-up is deterministic as each vertex allows only one kind of move. In the case of the XY model, the break-up is stochastically sampled by choosing one of two possible complementary moves with equal probability at each vertex. For instance, if a vertex allows for switch-and-continue and switch-and-reverse moves (see Fig. 2.5), then only one of them is randomly selected for each vertex irrespective of which leg happens to be the entrance leg. This restriction assures that loops do not overlap and are independent from each other.

2.3.4 Estimation of observables

In order to estimate an observable in the SSE framework, it first must be converted to its SSE representation. This necessity is due to the fact that while an observable of interest usually acts on a Hilbert space, its corresponding estimator needs to operate on the space of SSE configurations. At the core of such a transformation is the idea of importance sampling (see Sec. 2.1.1), that is, the idea of reweighting samples generated from the base distribution in order to evaluate properties of the target distribution. The methodology for this conversion is the following. First, a procedure similar to the one outlined in the previous sections is required in order to map the mathematical expression $\left\langle Oe^{-\beta H}\right\rangle$ to its SSE equivalent. Alternatively, this technical derivation can be sped up if the observable can be expressed as a derivative of the partition function. In this case, the derivative is directly applied to the SSE representation for the partition function. In either case, as a result, one obtains an estimator, $O(c)$, that is now defined with respect to a SSE configuration c. In parallel, the same conversion is applied to the partition function as before. Consequently,

one obtains

$$\langle O \rangle = \frac{\mathrm{Tr}\left[O e^{-\beta H_\theta}\right]}{\mathrm{Tr}\left[e^{-\beta H_\theta}\right]} \tag{2.57}$$

$$= \frac{\sum_{c' \in \{c'\}} O(c')}{\sum_{c \in \{c\}} W(c)}. \tag{2.58}$$

As long as the operator, O, is based on operators appearing in the Hamiltonian, H_θ, the configuration space attached to the sum in the numerator is included in the configuration space of the denominator sum, $\{c'\} \subseteq \{c\}$. Such an assumption is required for the next step. Otherwise, extended ensemble techniques such as the worm algorithm might be required [41]. We consider an example of this assumption breakdown in the case of an entanglement entropy estimator in Sec. 3.5. In that case, we develop a loop-based mapping between the configuration spaces that allows us to circumvent the problem. As for now, we consider this assumption to hold. Following the standard importance sampling steps, we get

$$\begin{aligned} \langle O \rangle &= \frac{\sum_c \frac{O(c)}{W(c)} W(c)}{\sum_c W(c)} \\ &= \sum_c \frac{O(c)}{W(c)} P(c) \\ &= \left\langle \frac{O(c)}{W(c)} \right\rangle_{c \sim P}, \end{aligned} \tag{2.59}$$

where the c are sampled according to the probability $P(c)$, which is based on the unnormalized probability $W(c)$. Such a procedure is described in detail in the preceding sections.

2.3.5 Log-likelihood gradient estimators

As an illustration of the procedure described in the previous section, we apply it to derive SSE estimators required for the training of the quantum Boltzmann machine (QBM). As discussed in Sec. 5.3.1, the QBM training is based on the gradient

$$\frac{\partial \ln p_\theta(\mathbf{v})}{\partial \theta} = -\frac{\partial \ln Z_\theta(\mathbf{v})}{\partial \theta} + \frac{\partial \ln Z_\theta}{\partial \theta} \tag{2.60}$$

$$= -\frac{\partial_\theta Z_\theta(\mathbf{v})}{Z_\theta(\mathbf{v})} + \frac{\partial_\theta Z_\theta}{Z_\theta}, \tag{2.61}$$

where $Z_\theta \equiv \mathrm{Tr}\left[e^{-\beta H_\theta}\right]$ and $Z_\theta(\mathbf{v}) \equiv \mathrm{Tr}\left[e^{-\beta H_\theta} M_{\mathbf{v}}\right]$ are the partition functions of the unclamped and clamped ensembles correspondingly. The mathematical expressions for those gradients will be derived in Sec. 5.3.1 and are given in Eqs. (5.31) and (5.31). The direct approach that converts those expressions into the SSE language requires extensive mathematical manipulations. A more elegant way is to take advantage of the gradient form of Eq. (2.60) and directly work with the operator list representation of the partition function, Eq. (2.48). For convenience, we duplicate this equation here as well as provide the SSE representation for $Z_\theta(\mathbf{v})$, which are

$$\begin{aligned} Z_\theta &= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \langle \mathbf{s} | \prod_{i=1}^{n} H_{a_i,b_i} | \mathbf{s} \rangle \\ Z_\theta(\mathbf{v}) &= \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \langle \mathbf{v} | \prod_{i=1}^{n} H_{a_i,b_i} | \mathbf{v} \rangle . \end{aligned}$$

Here, we note that the lack of a trace over the spins in the clamped partition function is what distinguishes the two ensembles. As the QBM does not impose any symmetries on the Hamiltonian, each bond operator, $H_{a,b}$, has its own adjustable parameter $\theta_{a,d}$. Similarly, the derivation that follows is ensemble independent. Therefore, we restrain our attention to the unclamped ensemble Z.

In order to derive an importance estimator according to Eq.(2.59), $O(c)$ and $W(c)$ need to be matched for every configuration c. For the unclamped ensemble, we adopt definition for c and $W(c)$ given by

$$c = \langle \mathbf{v} | \prod_{i=1}^{n} H_{a_i,b_i} | \mathbf{v} \rangle , \tag{2.62}$$

$$W(c) = \frac{\beta^n}{n!}. \tag{2.63}$$

To derive an expression for $O(c)$, we apply the gradient to the partition function and use the product rule, thus arriving at

$$\begin{aligned} \frac{\partial Z_\theta}{\partial \theta_{a,d}} &= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \langle \mathbf{s} | \sum_{k=1}^{n} \left(\prod_{i=k+1}^{n} H_{a_i,b_i} \frac{\partial H_{a_i,b_i}}{\partial \theta_{a,d}} \prod_{i=1}^{k-1} H_{a_i,b_i} \right) | \mathbf{s} \rangle \\ &= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \langle \mathbf{s} | \sum_{k=1}^{n} \left(\prod_{i=k+1}^{n} H_{a_i,b_i} \frac{\partial H_{a_k,b_k}}{\partial \theta_{a,d}} \prod_{i=1}^{k-1} H_{a_i,b_i} \right) | \mathbf{s} \rangle . \end{aligned} \tag{2.64}$$

For a diagonal operator $a = D$, the derivative $\frac{\partial H_{D_k,b_k}}{\partial \theta_{D,d}} \equiv H_{D,d}$ represents a bare diagonal operator associated with the coupling $\theta_{D,d}$. Denoting the eigenvalues of this operator with respect to a state $|\mathbf{s}\rangle$ as $\lambda(\mathbf{s}, D, d)$, we have

$$H_{D,d} |\mathbf{s}\rangle = \lambda_{D,d}(\mathbf{s}) |\mathbf{s}\rangle . \tag{2.65}$$

Moreover, according to the definition in Eq. (2.49), the operator product $\prod_{i=1}^{k-1} H_{a_i,b_i} |\mathbf{s}\rangle$ represents the propagated state $|\mathbf{s}(k-1)\rangle$. This state remains unaltered under the action of a diagonal operator. Therefore, Eq. (2.64) reduces to

$$\begin{aligned}
\frac{\partial Z_\theta}{\partial \theta_{D,d}} &= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \left(\sum_{k=0}^{n-1} \lambda_{D,d}(k-1) \right) \langle \mathbf{s}| \prod_{i=1}^{n-1} H_{a_i,b_i} |\mathbf{s}\rangle \\
&= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \underbrace{\frac{\beta^{n+1}}{(n+1)!} \left(\sum_{k=0}^{n} \lambda_{D,d}(k-1) \right)}_{O(c)} \underbrace{\langle \mathbf{s}| \prod_{i=1}^{n} H_{a_i,b_i} |\mathbf{s}\rangle}_{c},
\end{aligned}$$

where in the second line we have relabeled $n \to n-1$ to extract the weight $O(c)$ of a configuration c as per Eq. (2.62). Concluding the derivation by applying the importance ratio Eq.(2.59), we obtain the estimator

$$\frac{\partial \ln Z_\theta}{\partial \theta_{D,d}} = \beta \left\langle \bar{\lambda}_{D,d} \right\rangle , \tag{2.66}$$

where we have defined the estimator $\bar{\lambda}_{D,d} \equiv \frac{1}{n+1} \sum_{k=0}^{n} \lambda_{D,d}(k-1)$. An equivalent estimator is valid for the clamped distribution. Therefore, the full log-likelihood derivative with respect to a parameter coupled to a diagonal operator is

$$\frac{\partial p_\theta(\mathbf{v})}{\partial \theta_{D,d}} = -\beta \left\langle \bar{\lambda}_{D,d} \right\rangle_{\mathbf{v}} + \beta \left\langle \bar{\lambda}_{D,d} \right\rangle . \tag{2.67}$$

Proceeding with the derivative with respect to a parameter coupled to an off-diagonal operator $\theta_{O,d}$, we consider again the expression Eq. (2.64) as the starting point. This time we note that an off-diagonal operator changes the state $|s\rangle$ it acts upon, thus invalidating the previous approach. Instead, for an off-diagonal operator, we have

$$\frac{\partial H_{D_i,b_i}}{\partial \theta_{O,d}} = \frac{1}{\theta_{O,d}} H_{a_i,b_i} \delta_{a_i,O} \delta_{b_i,d}, \tag{2.68}$$

where δ is the Kronecker delta function. Substituting this identity into Eq. (2.64) leads to

$$\begin{aligned}
\frac{\partial Z_\theta}{\partial \theta_{O,d}} &= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \langle \mathbf{s} | \sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} H_{a_i,b_i} \left(\frac{1}{\theta_{O,d}} H_{a_k,b_k} \delta_{a_k,O} \delta_{b_k,d} \right) \prod_{i=k+1}^{n} H_{a_i,b_i,\theta_i} \right) |\mathbf{s}\rangle \\
&= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \frac{\beta^n}{n!} \frac{1}{\theta_{O,d}} \sum_{k=1}^{n} \left(\delta_{a_k,O} \delta_{b_k,d} \right) \langle \mathbf{s} | \prod_{i=1}^{n} H_{a_i,b_i} |\mathbf{s}\rangle \\
&= \sum_{\mathbf{s}} \sum_{n=0}^{\infty} \sum_{S_n} \underbrace{\frac{\beta^n}{n!} \frac{1}{\theta_{O,d}} n_{O,d}}_{O(c)} \underbrace{\langle \mathbf{s} | \prod_{i=1}^{n} H_{a_i,b_i} |\mathbf{s}\rangle}_{c},
\end{aligned} \tag{2.69}$$

where $n_{O,d} = \sum_{k=1}^{n} \delta_{a_k,O} \delta_{b_k,d}$ is the number of occurrences of an off-diagonal operator at the bond d in the operator sequence S_n. As the configuration c directly matches Eq. (2.62), the importance estimator is easily read off as

$$\frac{\partial \ln Z_\theta}{\partial \theta_{O,d}} = \frac{1}{\theta_{O,d}} \langle n_{O,d} \rangle . \tag{2.70}$$

Therefore, the full log-likelihood derivative with respect to a parameter coupled to an off-diagonal operator is

$$\frac{\partial p_\theta(\mathbf{v})}{\partial \theta_{O,d}} = -\frac{1}{\theta_{O,d}} \langle n_{O,d} \rangle_{\mathbf{v}} + \frac{1}{\theta_{O,d}} \langle n_{O,d} \rangle . \tag{2.71}$$

It is interesting to compare the SSE expression for the gradient in Eqs. (2.67) and (2.71) against their mathematical counterparts in the Eqs. (5.31) and (5.32). The SSE representation is fully equivalent between the clamped and unclamped ensemble while the mathematical expressions take very different forms in the two ensembles.

Chapter 3

Entanglement entropy estimation in quantum Monte Carlo

3.1 Introduction

Looking back at the history of the innovation of entanglement entropy (EE) methods in MC methods, there seems to be three quasi-independent branches of development. Those are zero temperature quantum Monte Carlo (QMC) [42], finite temperature QMC [43] and $d+1$ classical [44] MCMC aiming to study critical phenomena. These parallel developments have led to the birth of similar ideas. However, the connection between them has been masked by the contextualities of a particular application. This lack of common language has been detrimental to the mixing of important developments and cross-fertilization between the subfields. In order to restore the ergodicity, we frame the problem of EE measurement in a common language and accentuate on the main ideas that have led to the improvements.

The point of contact between the MCMC frameworks mentioned in the previous paragraph is the density matrix, ρ, describing the statistical ensemble under study (be it a mixed state or a pure state). Irrespective of the framework, we can always represent the matrix elements of ρ in a bipartite basis. Indeed, the Schmidt decomposition allows us to build a complete orthogonal basis set out of states like

$$|\mathbf{s}_A, \mathbf{s}_{\bar{A}}\rangle = |\mathbf{s}_A\rangle \otimes |\mathbf{s}_{\bar{A}}\rangle \,. \tag{3.1}$$

Here are A and $\bar{A}$ refer to two regions that partition the system into two complementary subsystems; $\mathbf{s}_A$ and $\mathbf{s}_{\bar{A}}$ are assumed to be orthogonal basis states spanning their corre-

sponding Hilbert spaces. In this basis, the density matrix elements are

$$\rho(\mathbf{s}_A, \mathbf{s}_{\bar{A}}, \mathbf{s}'_A, \mathbf{s}'_{\bar{A}}) = \langle \mathbf{s}_A, \mathbf{s}_{\bar{A}} | \rho | \mathbf{s}'_A, \mathbf{s}'_{\bar{A}} \rangle \tag{3.2}$$

In particular, we are interested in an object from the density matrix known as the reduced density matrix ρ_A, which satisfies

$$\rho_A \propto \sum_{\mathbf{s}_{\bar{A}}} \langle \mathbf{s}_{\bar{A}} | \rho | \mathbf{s}_{\bar{A}} \rangle \,. \tag{3.3}$$

In this expression, the proportionality factor is the partition function,

$$Z = \mathrm{Tr}\left[\rho_A\right] = \mathrm{Tr}\left[\rho\right], \tag{3.4}$$

which normalizes the sum of diagonal elements to add up to one, thus allowing one to interpret them as a probability distribution. The reduced density matrix fulfills the same role as ρ but for a particular subsystem A and is the central object for the EE computation. Its matrix elements are

$$\rho_A(\mathbf{s}_A, \mathbf{s}'_A) = \frac{1}{Z} \sum_{\mathbf{s}_{\bar{A}}} \rho(\mathbf{s}_A, \mathbf{s}_{\bar{A}}, \mathbf{s}'_A, \mathbf{s}_{\bar{A}}). \tag{3.5}$$

Depending on the QMC flavour, the representation for the density matrices changes. However, the basic mathematical formalism that follows is independent of those details.

3.1.1 Replica trick and Rényi entropy

Unlike a typical physical observable, the EE is not an average property of sampled configurations. In order to access this quantity in QMC simulations, a non-trivial modification to the simulation cell is required. This modification constitutes the common ground that all QMC methods rely on in order to access the EE and is known as the *replica trick* [6]. Originally, the replica trick was developed as an analytic tool to compute the *von Neumann EE* in conformal field theories, given by

$$S_1(A) = -\,\mathrm{Tr}\left[\rho_A \ln \rho_A\right]. \tag{3.6}$$

The authors in Ref. [6] use an analytic continuation in order to compute S_1 through

$$S_1(A) = -\lim_{n \to 1} \frac{\partial}{\partial n} \mathrm{Tr}\left[\rho_A^n\right], \tag{3.7}$$

where n is a real number. When n is a positive integer larger then 1, we can write

$$\begin{aligned}\mathrm{Tr}\left[\rho_A^n\right] &= \frac{\sum_{\mathbf{s}_{1,A},\ldots,\mathbf{s}_{n,A}} \prod_{k=1}^{n} \rho_A(\mathbf{s}_{k,A}, \mathbf{s}_{k+1,A})}{\sum_{\mathbf{s}_{1,A},\ldots,\mathbf{s}_{n,A}} \prod_{k=1}^{n} \rho_A(\mathbf{s}_{k,A}, \mathbf{s}_{k,A})} \\ &= \frac{\sum_{\mathbf{s}_{1,A},\ldots,\mathbf{s}_{n,A}} \prod_{k=1}^{n} \rho_A(\mathbf{s}_{k,A}, \mathbf{s}_{k+1,A})}{\prod_{k=1}^{n} \sum_{\mathbf{s}_{k,A}} \rho_A(\mathbf{s}_{k,A}, \mathbf{s}_{k,A})} \qquad (3.8) \\ &= \frac{Z\left[n, A\right]}{Z\left[n, \varnothing\right]}. \qquad (3.9)\end{aligned}$$

In the first line, the sum and the product in the denominator commute, which leads to the expression in the second line. The last line introduces the notation for a *replicated partition function*, $Z[n, A]$. This notation specifies the number, n of *replicas* of the reduced density matrix in the composition as well as the region A. With an empty region $A = \varnothing$, the replicas are fully independent of each other, so that $Z[n, \varnothing] = Z^n$. With a non-empty region A, the replicas become interconnected. For instance, in the $d + 1$ representation of ρ, specific boundary conditions are imposed between replicas in the $+1$ direction (see Fig.3.4). Such notation shows that the replica trick allows us to express the trace of the density matrix as a the ratio of partition functions.

While the generalization of the definition above to real n is possible for analytic calculations, in QMC methods only integer n can be probed. With this limitation, it is constructive to consider a different entropy measure known as the *Rényi EE*, given by

$$S_n = \frac{1}{1-n} \ln \mathrm{Tr}\left[\rho_A^n\right]. \qquad (3.10)$$

In this context, n is known as the *Rényi index*. Contrasted with the von Neumann entropy, Rényi entropies are lacking the property of sub-additivity. However, despite this missing feature, all Rényi entropies collectively carry at least as much information as the von Neumann entropy, which is evident since the latter can be expanded in powers of ρ_A. Moreover, Rényi entropy can be seen as the generalization of the von Neumann entropy due to the limit

$$\lim_{n \to 1} S_n = S_1. \qquad (3.11)$$

More importantly from the practical viewpoint, it has been found that for a given n, S_n often carries a similar informational content about a physical system as the Von Neumann entropy. For this reason and due to practical considerations of the linear scaling of the computational costs in n, we solely focus on the second Rényi entropy S_2 as a numerical probe for the EE. We note that the techniques required to estimate S_2 can be trivially extended to S_n.

3.1.2 Swap operator

It is insightful to extract a physical interpretation of the replicated partition function. For this, we consider the expression it takes in an augmented state space spanned by a complete basis set $|\mathbf{s},\mathbf{s}'\rangle = |\mathbf{s}\rangle_1 \otimes |\mathbf{s}'\rangle_2$. In this tensor product, each factor spans over the Hilbert space associated with the non-trivial part of one of the replicated density matrices embedded in this larger space as $\rho_1 \equiv \rho \otimes I$ and $\rho_2 \equiv I \otimes \rho$ correspondingly. As we also want to identify the spins associated with region A, we adopt the notation $\left|\mathbf{s}_A,\mathbf{s}_{\bar{A}},\mathbf{s}'_A,\mathbf{s}'_{\bar{A}}\right\rangle = |\mathbf{s}_A,\mathbf{s}_{\bar{A}}\rangle_1 \otimes \left|\mathbf{s}'_A,\mathbf{s}'_{\bar{A}}\right\rangle_2$. In light of these definitions, we rewrite Eq. (3.9) for $n=2$ as

$$\begin{aligned}
Z[2,A] &= \sum_{\substack{\mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\\ \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}}} \langle \mathbf{s}_{1,A}|\rho_A|\mathbf{s}_{2,A}\rangle \, \langle \mathbf{s}_{2,A}|\rho_A|\mathbf{s}_{1,A}\rangle \\
&= \sum_{\substack{\mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\\ \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}}} \left\langle \mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\right| \rho \left|\mathbf{s}_{2,A},\mathbf{s}_{1,\bar{A}}\right\rangle \left\langle \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}\right| \rho \left|\mathbf{s}_{1,A},\mathbf{s}_{2,\bar{A}}\right\rangle \\
&= \sum_{\substack{\mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\\ \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}}} \left\langle \mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\right|_1 \left\langle \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}\right|_2 \rho_1\rho_2 \left|\mathbf{s}_{2,A},\mathbf{s}_{1,\bar{A}}\right\rangle_1 \left|\mathbf{s}_{1,A},\mathbf{s}_{2,\bar{A}}\right\rangle_2 \\
&= \rho_1 \left[\sum_{\substack{\mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}}\\ \mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}}} \left|\mathbf{s}_{2,A},\mathbf{s}_{1,\bar{A}},\mathbf{s}_{1,A},\mathbf{s}_{2,\bar{A}}\right\rangle \left\langle \mathbf{s}_{1,A},\mathbf{s}_{1,\bar{A}},\mathbf{s}_{2,A},\mathbf{s}_{2,\bar{A}}\right| \right] \rho_2 \\
&= \rho_1 \mathrm{Swap}_A \rho_2 .
\end{aligned} \tag{3.12}$$

The last line defines the operator Swap_A, whose matrix elements are iterated over in the preceding line. This operator acts on the joint space of the replicated system. As suggestive by its name, its action is to swap the spins that fall within the region A between the replicas. It is easy to verify that this operator is Hermitian and, therefore, represents a physical observable. To see this more explicitly, we note that in the special case when the density matrix is the one of a pure state $|\rho\rangle = |\psi\rangle\langle\psi|$, the full system is decoupled, and, therefore, is also in a pure state $|\Psi\rangle\langle\Psi| \equiv \rho_1\rho_2 = |\psi\rangle\langle\psi|_1 \otimes |\psi\rangle\langle\psi|_2$. In this case, the above equation is equivalent to the expectation

$$Z[2,A] = \langle\Psi|\mathrm{Swap}_A|\Psi\rangle \, . \tag{3.13}$$

Additionally, for $A = \varnothing$, this expression simplifies to

$$Z[2,\varnothing] = \langle\Psi|\mathrm{Swap}_\varnothing|\Psi\rangle = \langle\Psi|\Psi\rangle \, . \tag{3.14}$$

Combining the last two equations, we conclude that the ratio

$$\frac{Z[2,A]}{Z[2,\varnothing]} = \frac{\langle\Psi|\mathrm{Swap}_A|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{3.15}$$

is the normalized expectation value of the Swap operator. As discussed in the previous section, this quantity is directly related to S_2. Therefore, the swap operator provides a practical route for an experimental detection of the EE [45]. In fact, this interpretation of the Swap operator extends to mixed states, where we have

$$\frac{Z[2,A]}{Z[2,\varnothing]} = \frac{\mathrm{Tr}\left[\mathrm{Swap}_A\rho_1\rho_2\right]}{\mathrm{Tr}\left[\rho_1\rho_2\right]}, \tag{3.16}$$

which can be checked upon a substitution into Eq. (3.12).

Historically, the Swap operator opened the door to EE measurement in QMC in Ref. [42]. In this seminal work, the authors employed a projector SSE to study the area law contribution to the EE of a spin system at zero temperature. Later on, Ref. [43] introduced an EE estimator within a finite-temperature SSE framework (see Sec. 3.2.3) without relying on the Swap operator terminology. Unlike this historic development suggests, the reason for the different approaches to the EE evaluation undertaken in the zero- and finite-temperature methods is not due to a particular nature of the density matrix but, rather, due to the choice of the underlying state space representation within a QMC framework. Fundamentally, the Swap operator is just a reformulation of the replica trick. For this reason, on a conceptual level we find it more fruitful to think of the EE estimation within QMC methods not as an evaluation of the Swap operator but, directly, as an estimation of the partition function ratio in Eq. (3.9).

3.1.3 Free energy based methods

The replica trick has allowed us to turn the problem of EE estimation into the problem of evaluating

$$S_2(A) = -\ln\frac{Z\left[2,A\right]}{Z\left[2,\varnothing\right]}. \tag{3.17}$$

The first approach to solving this problem came from a realization that this expression can be written as a free energy difference as

$$S_2(A) = -\ln Z\left[2,A\right] + \ln Z\left[2,\varnothing\right]. \tag{3.18}$$

As such, methods designed to evaluate the free energies can be used, such as Wang-Landau [46] or, at finite temperature, thermodynamic integration [47, 48]. In the latter case, the partition function adopts an additional temperature label, $Z[2, A, \beta]$. It can be found by integrating the well-known thermodynamic relationship between the free energy and the expected energy $-\frac{\partial \ln Z}{\partial \beta} = \langle E \rangle$ resulting in the working expression

$$\ln Z[2, A, \beta] - \ln Z[2, A, \beta = 0] = -\int_0^\beta \mathrm{d}\beta' \, \langle E \rangle_{Z[2,A,\beta']}, \tag{3.19}$$

where the subscript in the expectation value identifies the underlying ensemble. The free energy at infinite temperature i.e., the second term on the left-hand side, is easily related to the number of degrees of freedom. Computation of the integral on the right is done numerically and requires MC estimation of the expectation value $\langle E \rangle_{2,A,\beta'}$ at regular discrete intervals filling up the entire window $\beta' \in [0, \beta]$. One of the fruitful results that came out from this approach was a demonstration that the EE can be used as a tool for detecting classical phase transitions [47] (for the example of the 2d quantum XY model, see Fig. 4.3). However, the numerical integration introduces a systematic bias [49] and an increasing uncertainty as $\beta \to \infty$, rendering this approach impractical for studying systems at zero temperature. More generally, the methods that rely on the estimation of the free energy are poorly matched for the measurement of EE because the free energy is expected to scale extensively, thus introducing an uncertainty from the volume law to the extracted EE signal.

3.2 Fundamentals of partition functions ratio estimation

Historically, in applicable domains, the free energy based approaches were the first numerical procedures to tackle the measurement of EE. However, they were quickly replaced by *ratio based* methods, which directly aim at constructing estimators for the ratio

$$r_{\varnothing}^{A} \equiv \frac{Z[2, A]}{Z[2, \varnothing]}, \tag{3.20}$$

where the subscript (superscript) in $r_{\varnothing}^{A}$ identifies the boundary conditions between replicated partition functions in the denominator (numerator). This definition naturally leads to a generalization that proves to be of the utmost importance and is given by

$$r_{A'}^{A} \equiv \frac{Z[2, A]}{Z[2, A']}. \tag{3.21}$$

The next several sections examine in detail strategies towards an efficient evaluation of this quantity in various contexts arising within QMC simulations.

3.2.1 Exponential speed-up through the ratio trick

It is educative to consider the expected behaviour of the ratio $r^A_{A'}$ from the physical point of view and contrast it with $r^A_\varnothing$. As a first step, we turn our attention to the latter and we invert Eq. (3.17) to write

$$r^A_\varnothing = e^{-S_2(A)} \approx e^{-\partial A}, \tag{3.22}$$

where the approximation is based on the assumption that the EE of the underlying system is captured by the area law and ∂A is the boundary of region A. Therefore, we expect an exponential decay in $r^A_{A'}$ as the system's boundary increases. For an accurate estimation, the uncertainty, $\delta r^A_\varnothing$, must be on the order of the value itself. In QMC simulations, the uncertainty decreases as $\sqrt{N_s}$ as the number of samples. Therefore, the scaling in Eq. (3.22) implies an exponential increase in the computational cost, which severely restricts the largest possible regions A that can be effectively studied with QMC.

On the other hand for Eq. (3.21), via small manipulations we arrive at

$$\begin{aligned} r^A_{A'} &= \frac{r^A_\varnothing}{r^{A'}_\varnothing} \\ &= e^{-(S_2(A)-S_2(A'))} \\ &\approx e^{-\partial A-\partial A'}. \end{aligned} \tag{3.23}$$

It becomes directly evident that for a judicious choice of the reference region $A' = A_0$ with a boundary size equal to that of region A, we have $\partial A_0 = \partial A$ such that $r^A_{A_0}$ becomes insusceptible to the area law, thus giving direct access to the subleading quantities of interest. Therefore, using $r^A_{A_0}$ instead of $r^A_\varnothing$ provides an efficiency leap, in spirit similar to using $r^A_\varnothing$ instead of the free-energy estimator such as Eq. (3.18). This intriguing property of the ratio $r^A_{A_0}$ has somehow been ignored in the literature. Instead, the general adopted technique for the elimination of the area law has been based on fitting. In fact, we make the same inefficiency mistake in our first EE study described in Sec. 4.1. However in the subsequent study detailed in Sec. 4.2, we correct ourselves mainly because such mistake was no longer an option within the budgeted computational time.

In cases where the area law contribution cannot be straightforwardly removed or when even subleading contributions to the EE lead to an exponential decrease in $r^A_{A'}$ as the gap

between the two regions, $|A - A'|$, increases, [1] a simple but powerful technique known as the *ratio trick* can be employed [42] (sometimes also referred to as the *incremental trick*). The idea is to subdivide the region between A and A' into N subregions A_i such that A_{i+1} contains A_i, and $|A_{i+1} - A_i| \leq \Delta A$, where ΔA is an integer hyper-parameter. With the help of these intermediate regions, we derive

$$r^A_{A'} = \prod_{i=0}^{N-1} r^{A_{i+1}}_{A_i}, \tag{3.24}$$

where $A_0 = A'$ and $A_N = A$. This expression can be verified by substituting it into Eq. (3.21) and noticing that the denominator and numerator in consecutive terms cancel each other. The ratio trick requires N separate processes, with each evaluating an individual ratio $r^{A_{i+1}}_{A_i}$. Despite this increased number of processes, it provides a dramatic improvement.

In order to understand the advantage of the ratio trick, we can imagine a partition $\{A_i\}_{i=0}^N$ such that each intermediate ratio is approximately the same, $r^{A_{i+1}}_{A_i} \approx r$ and, therefore, similar computation resources are required to achieve the accuracy δr in their estimate. To see that such N always exists, we write

$$r = R^{1/N} = e^{-(S_2(A) - S_2(A'))/N}$$

where we have defined $R \equiv r^A_{A'}$ to simplify the notation and used Eq. (3.23) in the second equality. We note that the EE difference, $S_2(A) - S_2(A')$, scales at most as the area law, say, $\sim \partial A$. At the same time, N has no such restriction and, in principle, can scale as the volume law. However, it is sufficient to set $N \sim \partial A$ to make r constant independent of A, A' and the system size L.

Concurrently, since each ratio evaluation is implemented in an independent simulation, the cumulative uncertainty in the product can be expressed in terms of the individual fractional uncertainty via a simple propagation of error as

$$\frac{\delta R}{R} = \sqrt{N}\frac{\delta r}{r}. \tag{3.25}$$

To make the fractional uncertainty on the left constant, one therefore needs to get more accurate estimates in r as N increases such that $\delta r \sim 1/\sqrt{N}$. The need for a more accurate evaluation in turn requires producing more samples from each simulation, scaling as $N_s \sim N$. Taking a total of all the combined resources required, we conclude that with the ratio trick the cost to evaluate the EE difference transform from an exponential scaling $\mathcal{O}\left(e^{|S(A) - S(A')|}\right)$ to a polynomial one $\mathcal{O}\left([S(A) - S(A')]^2\right)$.

[1] we assume A' is contained in A; if not, the same reasoning applies after the relabelling $A' \Leftrightarrow A$

3.2.2 Two basic estimators

The ratio trick provides an opportunity for a scalable computation of the EE by breaking up the task of evaluating $r_{\varnothing}^{A}$ into a set of manageable chunks that each focus on evaluating intermediate $r_{A'}^{A}$. As discussed prior in this chapter, the replica trick lets us express $r_{A'}^{A}$ as a ratio of partition functions as

$$r_{A'}^{A} = \frac{Z[2, A']}{Z[2, A]}. \tag{3.26}$$

Our goal is to find a MC estimator for this quantity. According to our previous discussion, the general expression for the partition function with entangling region A is

$$Z[2, A] = \sum_{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}} \rho(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}}) \rho(\mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}, \mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}}). \tag{3.27}$$

We will fall back on this definition whenever we need to derive an estimator specific to a particular application. However, for now, in the spirit of generalization, we define the weight $W(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}})$ to be the product of density matrix elements appearing in the sum and rewrite the partition function as

$$Z[2, A] = \sum_{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}} W(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}). \tag{3.28}$$

In $d + 1$ QMC simulations, the weight $W(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}})$ is not directly accessible. Instead, it is indirectly sampled via updates on a configuration space $\{c\}$ via a framework-dependent weight $W(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}, c)$. In this induced weight, the region A acts to impose certain restrictions that must be obeyed by the configurations c. This observation motivates us to define a more succinct and more general expression for the partition function as

$$Z[2, A] = \sum_{c \in \{c\}_A} W_A(c). \tag{3.29}$$

Here, the set $\{c\}_A$ represents the set of allowed configurations, c, as determined by region A and $W_A(c)$ is the representation-dependent weight of a configuration. This general form captures all the subtleties that lead to different estimators within different MCMC frameworks. Plugging this expression in Eq. (3.26) gives

$$r_{A'}^{A} = \frac{\sum_{c \in \{c\}_A} W_A(c)}{\sum_{c \in \{c\}_{A'}} W_{A'}(c).} \tag{3.30}$$

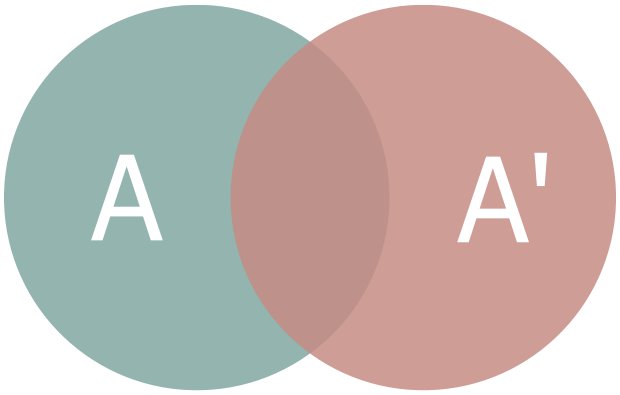


Figure 3.1: A Venn diagram of the relationship between the configuration spaces $\{c\}_A$ and $\{c\}_{A'}$ attached to the partition functions $Z[2, A]$, $Z[2, A']$ correspondingly.

This equation is reminiscent of the approach we took to estimate the value of the mathematical constant π in the Sec. 2.1.1. In that example, the path to the solution relied on a reformulation of the problem as an estimation of the ratio of specifically-chosen partition functions. An importance estimator was then derived to evaluate the ratio. Here, we encounter a similar problem with an important distinction – possibly non-overlapping configuration spaces (see Fig. 3.1 and contrast it with the Fig. 2.1(a)). To keep track of this feature, we define auxiliary configurations spaces based on the original ones.

$$\{c\}_{A\cap A'} \equiv \{c\}_A \cap \{c\}_{A'} \quad \text{intersection} \tag{3.31}$$
$$\{c\}_{A\cup A'} \equiv \{c\}_A \cup \{c\}_{A'} \quad \text{union} \tag{3.32}$$
$$\{c\}_{A'\setminus A} \equiv \{c\}_{A'} \setminus \{c\}_A \quad \text{compliment of } A \tag{3.33}$$
$$\{c\}_{A\setminus A'} \equiv \{c\}_A \setminus \{c\}_{A'} \quad \text{compliment of } A'. \tag{3.34}$$

Using those new objects, we split the sum in the numerator of Eq. (3.30) in two as

$$\begin{aligned} r^A_{A'} &= \frac{\sum_{c\in\{c\}_{A\cap A'}} W_A(c) + \sum_{c\in\{c\}_{A\setminus A'}} W_A(c)}{\sum_{c\in\{c\}_{A'}} W_{A'}(c)} \\ &= \frac{\sum_{c\in\{c\}_{A\cap A'}} W_A(c)}{\sum_{c\in\{c\}_{A'}} W_{A'}(c)} + \frac{\sum_{c\in\{c\}_{A\setminus A'}} W_A(c)}{\sum_{c\in\{c\}_{A'}} W_{A'}(c)} \\ &= \frac{Z_{A\cap A'}[2, A]}{Z[2, A']} + \frac{Z_{A\setminus A'}[2, A]}{Z[2, A']}, \end{aligned} \tag{3.35}$$

where the last line introduces a new notation for the partition functions restricted to the configuration space indicated in the subscript. To derive an estimator for the first term, we

first extend the definition of the weight W_A to configurations beyond its original domain, $\{c\}_{A'\setminus A}$, by explicitly setting it to zero outside of its original domain, such that

$$W_A(c) = \begin{cases} W_A(c) & \text{if } c \in \{c\}_A \\ 0 & \text{if } c \notin \{c\}_A. \end{cases} \tag{3.36}$$

With this version of $W_A(c)$, we extend the range of the first sum in the numerator of Eq. (3.35), which leads to

$$\begin{aligned} \frac{Z_{A\cap A'}[2,A]}{Z[2,A']} &= \frac{\sum_{c\in\{c\}_{A'}} W_A(c)}{\sum_{c\in\{c\}_{A'}} W_{A'}(c)} \\ &= \frac{\sum_{c\in\{c\}_{A'}} \frac{W_A(c)}{W_{A'}(c)} W_{A'}(c)}{\sum_{c\in\{c\}_{A'}} W_{A'}(c)} \\ &= \sum_{c\in\{c\}_{A'}} \frac{W_A(c)}{W_{A'}(c)} P_{A'}(c) \\ &= \left\langle \frac{W_A(c)}{W_{A'}(c)} \right\rangle_{A'}. \end{aligned} \tag{3.37}$$

When the configuration spaces fully overlap such that $\{c\}_{A'} = \{c\}_{A'\cap A} = \{c\}_A$, the second term in Eq. (3.35) is zero and the last expression yields the desired estimator

$$r^A_{A'} = \left\langle \frac{W_A(c)}{W_{A'}(c)} \right\rangle_{A'} \quad \textbf{when } \{c\}_A = \{c\}_{A'}. \tag{3.38}$$

When the configurations spaces differ, the second term in Eq. (3.35) needs to be estimated. Unfortunately, this is not achievable with samples generated from $P_{A'}$ as they provide no information outside of their domain $\{c\}_{A\setminus A'}$. Instead, we decompose

$$r^A_{A'} = \frac{Z_{A\cap A'}[2,A]}{Z[2,A']} \frac{Z_{A'\cap A}[2,A']}{Z_{A\cap A'}[2,A]} \frac{Z[2,A]}{Z_{A'\cap A}[2,A']}. \tag{3.39}$$

The estimator for the first term is given in Eq. (3.37). The inverse of the third term can be calculated via the same estimator with the replacement $A \Leftrightarrow A'$. As for the middle term, the configuration spaces composing its partition functions fully overlap by construction. In other words, those configurations are simultaneously compatible with regions A and A', implying that the extent of the entangling region plays no role in the weight, and

$$W_A(c) = W_{A'}(c) \quad \textbf{for } c \in \{c\}_{A'\cap A}. \tag{3.40}$$

Therefore, the middle term reduces to one. It follows that

$$r^{A}_{A'} = \frac{Z_{A'\cap A}[2,A]}{Z[2,A']} \Big/ \frac{Z_{A'\cap A}[2,A']}{Z[2,A]} = \left\langle \frac{W_A(c)}{W_{A'}(c)} \right\rangle_{A'} \Big/ \left\langle \frac{W_{A'}(c)}{W_A(c)} \right\rangle_{A} . \tag{3.41}$$

Additionally, we note that Eqs. (3.36) and (3.40) imply the equality

$$\frac{W_{A'}(c)}{W_A(c)} = \delta_{A\cap A'}(c), \tag{3.42}$$

where $\delta_{A'\cap A}(c)$ is an indicator function verifying the compatibility of a configuration c with the restrictions imposed by both regions A and A', i.e.,

$$\delta_{A\cap A'}(c) = \begin{cases} 1 & \text{if } c \in \{c\}_{A'\cap A} \\ 0 & \text{if } c \notin \{c\}_{A'\cap A} \end{cases} . \tag{3.43}$$

Therefore, we can rewrite Eq. (3.41) in its final form as

$$r^{A}_{A'} = \frac{\langle \delta_{A\cap A'}(c) \rangle_{A'}}{\langle \delta_{A\cap A'}(c) \rangle_{A}}. \tag{3.44}$$

3.2.3 Extended ensemble approach

The last equation derived in the previous section, Eq. (3.44), requires an independent simulation for each ensemble associated with regions A and A'. Those two simulations can be combined in one where the Markov chain explores the joint space $\{c\}_{A\cup A'}$. This *extended-ensemble* can be simulated via an introduction of a new MCMC move that switches between the underlying sampled distributions as indicated by A and A' according to the transition T that must satisfy the detailed balance,

$$W(A)T(A \to A') = W(A')T(A' \to A). \tag{3.45}$$

The simplest way to satisfy this condition is to propose to switch whenever the Markov chain finds itself in $\{c\}_{A\cap A'}$. In this configuration-space, the weight of a configuration is independent on the entangling region as per Eq. (3.40), so any symmetric choice of $T(A' \to A)$ satisfies detailed balance. A schematic depiction of a Markov chain with a deterministic transition (always switch) in the extended-space is shown in Fig. 3.2. The

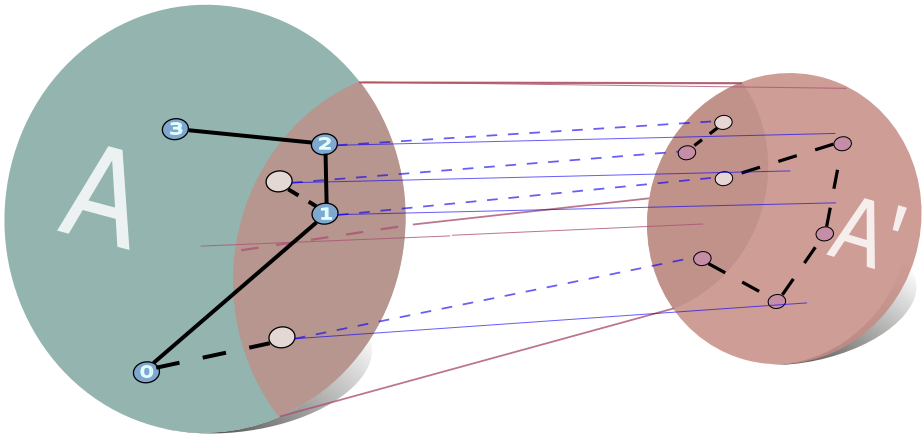


Figure 3.2: Three-dimensional visualization of a MCMC walk exploring the joint configuration space, $\{c\}_{A\cup A'}$ (see Eq.(3.32)). This figure is based on the Venn diagram visualization of the configuration space in Fig. 3.1. The trajectory of the consecutive moves taken by the Markov chain is traced out with dashed lines. The blue and dark dashed lines indicate inter-ensemble (displayed in the 3rd dimension) and intra-ensemble (displayed in-plane) moves correspondingly. In this case, an inter-ensemble move is made every time the Markov chain finds itself in the configuration space compatible with both regions. Configurations of the Markov chain sampled in the configuration space $\{c\}_A$ are numbered. The effective trajectory with respect to this space is traced out with a solid line. Note how the configurations labelled as 0 and 1 are connected via a multiple-step chain in the A' configuration space. Such effective tunneling via inter-ensemble moves helps with the Markov chain mixing.

same Eq. (3.44) applies in order to estimate $r^A_{A'}$. However, one must be careful how to interpret the two expectations in the extended-ensemble. In fact, the authors who introduce this method make a mistake in deriving their estimator [43]. Unlike those authors imply, those expectations need to be computed separately during the simulation run and only combined as sample averages once the simulation terminates. For instance, the estimate based on the example depicted in Fig. 3.2 is

$$r^A_{A'} = \frac{\frac{2}{4}}{\frac{2}{5}} = \frac{5}{4}. \tag{3.46}$$

From Fig. 3.2, the source of the advantage in simulating the extended-ensemble as opposed to performing two separated simulations also becomes transparent. Indeed, by introducing the inter-ensemble move, the Markov chain enjoys favourable mixing dynamics that act as a multi-step update with respect to each configuration space.

3.2.4 Projected ensemble approach

The extended ensemble method described in the previous section presents a sampling strategy that allows one to estimate the EE within a single simulation even for non-fully-overlapping configurations spaces, $\{c\}_A$ and $\{c\}_{A'}$. However, fundamentally, this estimation still relies on the estimator based on Eq. (3.41). In contrast to an estimator based on Eq. (3.38), this type of partition functions ratio estimator is disadvantageous as its statistics are confined to the two extreme values[2], thus intuitively leading to a high variance estimate. At first sight, the two estimators seem to be orthogonal to each other as far as their domains of application are concerned. Indeed, their areas of application seem to be strictly set by the relationship between $\{c\}_A$ and $\{c\}_{A'}$. However, the relationship between the configuration spaces is representation specific and, sometimes, it is possible to project the original representation into a new configuration space labelled via a new variable l that erases the distinction between the two configuration spaces, see Fig. 3.3.

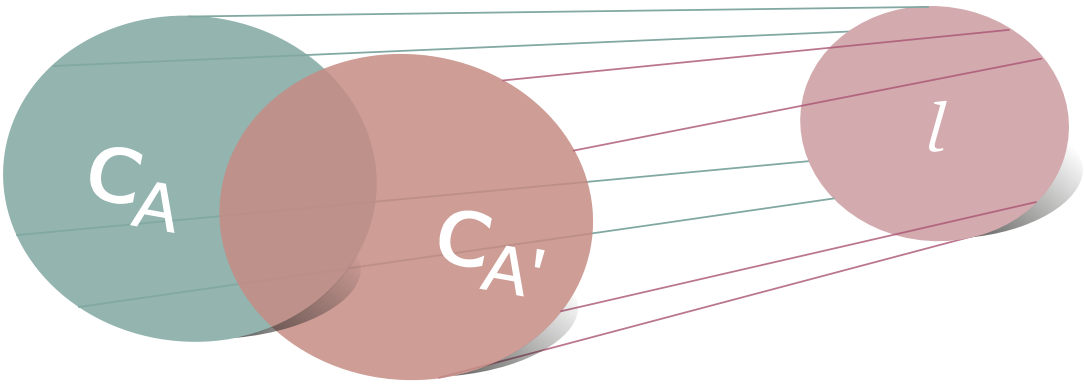


Figure 3.3: Three-dimensional visualization of the projected configuration space, $\{l\}$ with respect to the original spaces $\{c\}_A$ and $\{c\}_{A'}$. This figure is based on the Venn diagram in Fig. 3.1. Notice the change of variable $c \to l$ between the original and projected spaces. The idea is to find a new label l that is conditionally related to the original label c such that the projections of both $\{c\}_A$ and $\{c\}_{A'}$ overlap, i.e., $\{l\}_A = \{l\}_{A'}$.

Theoretically, the relationship between the configuration spaces is fully captured via a joint probability distribution, $p(c, l)$, defined over the original, $\{c\}$, and projected, $\{l\}$, configuration spaces. Such a distribution is implicitly defined by the conditional $p(l|c)$. In

[2]A single sample estimate for this estimator is restricted to be either 0 or 1. Therefore, the derived single-sample estimates for S_2 take the extreme values $(0,\infty)$.

fact, this conditional implements the stochastic map executing the conversion $c \to l$. The projection into the new space is achieved via the simple series of steps

$$\begin{aligned} Z_c[2, A] &= \sum_{c \in \{c\}_A} W_A(c) \\ &= \sum_{c \in \{c\}_A} W_A(c) \sum_l p(l|c) \\ &= \sum_l \sum_{c \in \{c\}_A} p(l|c) W_A(c) \\ &= \sum_l W_A(l) \\ &= Z_l[2, A]. \end{aligned}$$

Here,

$$W_A(l) \equiv \sum_c p(l|c) W_A(c) \tag{3.47}$$

is the unnormalized weight of a configuration l in the projected ensemble. In the second line, we explicitly assume that the conditional is normalized and, therefore, sums up to one. Such assumption directly leads to the resulting equality of the partition functions. Combined with the assumption that both $\{c\}_A$ and $\{c\}_{A'}$ are projected into the same configuration space $\{l\}$, we are justified to apply Eq. (3.38), which gives

$$r^A_{A'} = \left\langle \frac{W_A(l)}{W_{A'}(l)} \right\rangle_{A'}. \tag{3.48}$$

Here, a sample configuration l is to be generated with respect to the A' ensemble. Such sampling can be implemented as a two-step procedure: $c \sim p_{A'}(c)$ followed by $l \sim p(l|c)$. That is, first, c is generated via a QMC algorithm working in the original representation; then, an instance of the new label is generated by sampling the conditional $p(l|c)$ based on the c obtained in the previous step.

In practice, the success of the above approach to partition function ratio estimation hinges on a few ingredients. More specifically, given a stochastic map $p(l|c)$, the following conditions must be met:

- Equivalence of the original configuration spaces in the projected space must be demonstrated such that $\{l\}_A = \{l\}_{A'}$.

- Generation of i.i.d. samples from the conditional $p(l|c)$ must be possible, for instance, via a blocked Gibbs sampling.
- $p(l|c)$ must be normalized.
- Evaluation of the ratio $\frac{W_A(l)}{W_{A'}(l)}$ from Eq. (3.47) must be efficient.

In Secs. 3.4.2 and 3.5, we provide two examples of $p(l|c)$ satisfying all of those requirements. In both cases, the applied label-conversion map is directly related to the cluster/loop algorithm employed in the corresponding MCMC scheme. Such estimators are generally referred to as *improved estimators* [31]. The new estimators show a dramatically reduced variance as compared to alternative estimators working in the original representation. Such increase in performance can be attributed to the summation in Eq. (3.47). Conditioned on a particular configuration l, the sum iterates over all original configurations c that are mapped to l while reweighting their original weight, $W(c)$, by their likelihood to be mapped to l. In practice, such a sum can include thousands of terms. Therefore, while an estimator in the original representation is limited to base its estimate on a chain of sampled configurations, c, the new estimator effectively searches for large contributions to the EE in the configuration space all around the chain. This increased exploration of the configuration space stabilizes the variance of the estimator.

3.3 Estimator in d dimensions: variational Monte Carlo

As the first application of the theory developed in the preceding sections, we consider the problem of extracting the EE from a trial wavefunction $\psi(\mathbf{s})$. This parameterized wavefunction is usually optimized to represent a ground state of a Hamiltonian. Variational QMC [50] and projector QMC [51] are two flavours of QMC that build a direct representation of a pure state. However, those methods adopt very different internal representations for the wavefunction. The configuration space of the projector QMC is very similar to the standard $d+1$ SSE simulation cell (see Sec. 2.3.2). We consider EE estimators applicable within the $d+1$ setting in the next sections. In this section, we work directly in the physical d dimensions of the physical system. Such setting is natural to the framework of the variational QMC.

As a first step, we build a representation for the density matrix. For a pure state $|\psi\rangle$, the density matrix takes the form

$$\rho = |\psi\rangle\langle\psi| \,. \tag{3.49}$$

Therefore, the density matrix elements, Eq. (3.2), adopt the structure

$$\rho(\mathbf{s}_A, \mathbf{s}_{\bar{A}}, \mathbf{s}'_A, \mathbf{s}'_{\bar{A}}) = \psi(\mathbf{s}_A, \mathbf{s}_{\bar{A}})\psi^*(\mathbf{s}'_A, \mathbf{s}'_{\bar{A}}). \tag{3.50}$$

Consequently, the replicated partition function, Eq. (3.9), takes the form

$$\begin{aligned} Z[2, A] &= \sum_{\substack{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}} \\ \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}}} \rho_1(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}})\rho_2(\mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}, \mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}}) \\ &= \sum_{\substack{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}} \\ \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}}} \psi_1(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}})\psi_1^*(\mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}})\psi_2(\mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}})\psi_2^*(\mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}}) \\ &= \sum_{\substack{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}} \\ \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}}} \Psi^*(\mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}})\Psi(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}). \end{aligned} \tag{3.51}$$

In the second line, the wavefunction index indicates its corresponding replica index. In the next line, this index disappears as it is implicitly encoded in the position within the arguments of the combined wavefunction $|\Psi\rangle \equiv |\psi\rangle \otimes |\psi\rangle$. Its amplitudes are specified by four indices,

$$\Psi(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}) = \psi_1(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}})\psi_2(\mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}). \tag{3.52}$$

A similar expression holds for $Z[2, A']$. Therefore, the partition function ratio, Eq.(3.26), is

$$r^A_{A'} = \frac{\sum_{\substack{\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}} \\ \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}}} \Psi^*(\mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}})\Psi(\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}})}{\sum_{\substack{\mathbf{s}_{1,A'}, \mathbf{s}_{1,\bar{A}'} \\ \mathbf{s}_{2,A'}, \mathbf{s}_{2,\bar{A}'}}} \Psi^*(\mathbf{s}_{2,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{1,A'}, \mathbf{s}_{2,\bar{A}'})\Psi(\mathbf{s}_{1,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{2,A'}, \mathbf{s}_{2,\bar{A}'})}. \tag{3.53}$$

We note that both sums iterate over the same state space. Moreover, the states in each space can be bijectively mapped into each other simply by setting $\mathbf{s} \equiv (\mathbf{s}_{1,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{2,A}, \mathbf{s}_{2,\bar{A}}) = (\mathbf{s}_{1,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{2,A'}, \mathbf{s}_{2,\bar{A}'})$. As variational QMC can directly sample this configuration space, the EE estimator can be derived from Eq. (3.38). Treating the product $\Psi^*(\mathbf{s})\Psi(\mathbf{s})$ as the weight of a configuration, we derive the estimator

$$r^A_{A'} = \left\langle \frac{\Psi^*(\mathbf{s}_{2,A}, \mathbf{s}_{1,\bar{A}}, \mathbf{s}_{1,A}, \mathbf{s}_{2,\bar{A}})}{\Psi^*(\mathbf{s}_{2,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{1,A'}, \mathbf{s}_{2,\bar{A}'})} \right\rangle_{A'}. \tag{3.54}$$

Here, configurations $\mathbf{s}$ are assumed to be sampled based on the probability distribution

$$P_{A'}(\mathbf{s}) = \Psi^*(\mathbf{s}_{2,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{1,A'}, \mathbf{s}_{2,\bar{A}'})\Psi(\mathbf{s}_{1,A'}, \mathbf{s}_{1,\bar{A}'}, \mathbf{s}_{2,A'}, \mathbf{s}_{2,\bar{A}'}). \tag{3.55}$$

Such a probabilistic interpretation of the product above is possible when the wave-function is positive-definite. We study such a case in Sec. 5.2.4, where we employ a restricted Boltzmann machine as an ansatz to parameterize $\Psi(\mathbf{s})$ and sample $P_{A'}(\mathbf{s})$.

3.4 Estimator in $d+1$: classical path-integral representation

Transitioning to the application of the EE QMC estimation theory to a $d+1$ setting, we note that the extra dimension provides a new degree of complexity. However, such a representation encodes an extra structure about the underlying system that can be harvested for the development of an improved estimator. While the ideas present in this chapter were originally developed by the author, they were later discovered to have been risen in an early publication (Ref. [44]) that studied the $d+1$ classical representation of critical theories. This work predates the seminal work of the EE measurement in quantum many-body systems [42]. It is interesting to note that although Ref. [44] develops a very efficient improved estimator, the ratio trick was not developed at the time, thus impeding the final accuracy of the extracted EE. The first work combining the two ideas can be credited to Ref. [52].

3.4.1 Energy-based estimator

The goal is to study the scaling of the EE of the critical transverse field Ising model in 2 dimensions [53]. We employ the imaginary-time path-integral framework to convert this quantum model to a classical representation. As discussed at length in Sec. 2.2.1, the output of this mapping leads to a configuration space of an isotropic classical model in $2+1$ dimensions. The weight of each configuration c is given in terms of its Boltzmann weight.

The first step to derive an EE estimator is to obtain a representation for the replicated partition function, Eq. (3.27). Following a similar derivation as in Sec. 2.2.1, we obtain a classical representation for $\rho(\mathbf{s},\mathbf{s}')$, displayed in Fig. 3.4(a). The only difference from the previous procedure is that we keep open the possibility that $\mathbf{s}\neq\mathbf{s}'$. Next, we apply the replica trick. For this, the second simulation cell is introduced and an identification between the spin pairs on the inter-replica boundaries is carried out. In accordance with Eq. (3.27), inter-replica boundary spins that are part of region A are merged between the replicas; the spins that fall in the complement of A are merged within the replicas. This process is visualized in Fig. 3.4(b). The resulting simulation cell is displayed in Fig. 3.4(c). Based on this figure, we conclude that region A determines the bonds between inter-replica boundary spins and, therefore, influences the energy of an Ising configuration, $E_A(c)$. Importantly, region A does not introduce any restrictions on the configurations space. Therefore, the configuration spaces of $Z[2,A]$ and $Z[2,A']$ overlap. This fact validates an application of

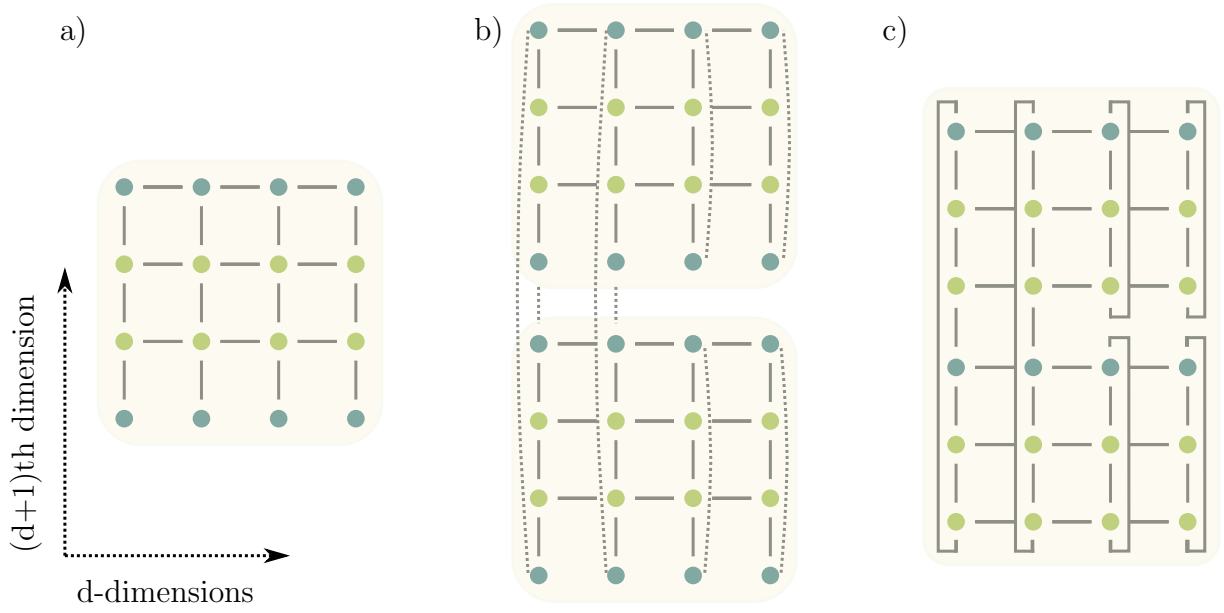


Figure 3.4: Classical $d+1$ representation of the transverse field Ising model (see Sec. 2.2.1). a) Depiction of a density matrix element, $\rho(s, s')$ in this representation. All circles represent classical spin degrees of freedom. The blue circles on the boundary are directly related to the density matrix element indices (s, s') . The Ising bonds between the spins are indicated with short lines. Note how the bottom layer is missing horizontal bonds. b) A replicated system of two independent densities $\rho(s_1, s_1')$ and $\rho(s_2, s_2')$. The dotted lines anticipate the pairs of spins that get identified with each other when the replicas are connected to form $Z[2, A]$ according to Eq. (3.27). In this case, the region A is comprised of 2 spins on the left. c) The resulting replicated system, $Z[2, A]$, after the matched pairs of blue spins are individually glued into a single entity. Note that only half of the blue spins remain as a result of this step. Additionally, some of the inter-replicas bonds are "stretched out". We remark that in a simulation, there is no distinction between the circles once the representation in c) is obtained.

the estimator from Eq. (3.38), which takes the form

$$r^A_{A'} = \left\langle \frac{e^{-\beta E_A(c)}}{e^{-\beta E_{A'}(c)}} \right\rangle_{A'} . \tag{3.56}$$

A further simplification is possible through the realization that the bulk contribution to the energy is exactly the same between E_A and $E_{A'}$. Therefore, the difference in energies of a configuration c is entirely due to the local difference in energies of the inter-replica

bonds in the region $\Delta A = A' - A$. Although computationally inexpensive to evaluate, the boundary-energy-based estimator relies on a fraction of the simulated configuration space. Indeed, in practice, the $(d+1)^{\text{th}}$ dimension is usually much larger than the rest of d dimensions. Therefore, the bulk of spins are left invisible to the estimator. Moreover, as the gap between the regions, ΔA, increases, the statistics of this estimator are progressively dominated by rare events. Those considerations motivate us to search for an improved estimator.

3.4.2 Improved estimator

As discussed in Sec. 2.2.3, the Ising model admits a Fortuin-Kasteleyn decomposition that casts the original spin model in terms of an equivalent random-cluster model (see Eq. (2.42)). In the context of the discussion in Sec. 3.2.4, we treat this decomposition as a stochastic map into a new projected space. As a result of this mapping, the binary degrees of freedom are mapped from spins to bonds, $s \to b$, via the normalized conditional $p(b|c)$ specified in Eq. (2.35). Sampling of this conditional is based on a Gibbs sampling leading to an easy generation of i.i.d. samples.

The bond variable b can be either in an open (active) or a closed (passive) state. Each configuration of bonds assumes a clustering decomposition. Clusters are formed by subgroups of open bonds split from each other by boundaries of closed bonds. The number of those clusters, N_c, is directly related to the number of spin configurations that are mapped to a bond configuration. Specifically, for each bond configuration, there are 2^{N_c} spin configurations mapping to it. The marginal of a bond configuration has an analytic expression, Eq. (2.42). For convenience, we restate it here as

$$W_{RC}(b) = p^{n(b)}(1-p)^{N_b - n(b)} 2^{N_c(b)}. \tag{3.57}$$

Here, $n(b)$ is the number of active bonds, N_b is the total number of bonds, $N_c(b)$ is the number of independent clusters formed within a bond configuration b, and $p \in [0,1]$ is a model parameter directly related to the parameters of the Ising model.

Based on this discussion, we conclude that the stochastic map, $p(b|c)$, satisfies all three efficiency requirements for an EE estimator in the projected space outlined in Sec. 3.2.4. The only condition left to demonstrate is that the projected configuration space is shared among all the entangling regions. To prove this point, we need to show that a given bond configuration is compatible with all regions A. For this we note that the spin configuration with all spins up, $c_\uparrow$, has a non-zero probability to be mapped to any bond configuration, such that $p(b|c_\uparrow) > 0 \,\forall\, b$. Therefore, this configuration alone is mapped to the whole bond

space. As such a spin configuration is compatible with all regions A, the range of the stochastic map is independent of the region A.

However, there still remains a subtlety of how to map a bond configuration compatible with region A into a bond configuration of another region A'. Its evidence transpires upon a closer examination of the replicated simulation cell from the point of view of bond variables, Fig. 3.4(c). The topology of the inter-replica bonds are directly influenced by the size of region A. However, their number is independent of this region. Therefore, there exists a one-to-one correspondence between the bond configuration spaces corresponding to two different regions A and A'. To specify this mapping unambiguously, consider region A' with three spins, that is one extra spin as compared to region A. To uniquely identify the inter-replica bonds, we refer to the boundary bonds via their coordinate $b_{r,l}^{r',l'}$, where r identifies the replica index and $l \in \{t, b\}$ specifies the location of the spin within the replica (top or bottom). With the topology of region A, bonds on the third spin are $b_{1,b}^{2,t}$ and $b_{2,b}^{1,t}$. The topology of region A' requires the inter-replica bonds over the third spin to be $b_{1,b}^{1,t}$ and $b_{2,b}^{2,t}$. It turns out that both possible mappings between those two pairs of bonds are valid. This is due to the fact that the direction of the arrow in the imaginary time axis is arbitrary. However once the convention is made it must be consistent and preserved across all the replicas for both ensembles A and A'. As the choice of the imaginary-time direction for each ensemble is independent, this leads to the two-fold possibility when the bond configurations are matched between the ensembles. This subtlety is important to keep in mind when simulating more than two replicas. At the same time, this time-reversal symmetry can be exploited to double the statistics by applying the EE estimator twice for each generated bond configuration b.

Having established a one-to-one correspondence between the configuration spaces of $Z_{WC}[2, A]$ and $Z_{WC}[2, A']$, we have satisfied all the requirements for the application of Eq. (3.48). Upon the substitution of the weight Eq. (3.57) into this equation, we derive an improved estimator,

$$\begin{aligned} r_{A'}^{A} &= \left\langle \frac{W_A(b)}{W_{A'}(b)} \right\rangle_{A'} \\ &= \left\langle \frac{2^{N_c^A(b)}}{2^{N_c^{A'}(b)}} \right\rangle_{A'} . \end{aligned} \tag{3.58}$$

In the resulting expression, a similar simplification occurs as that described for Eq. (3.56): the bulk clusters cancel out in the difference $N_c^A(\boldsymbol{b}) - N_c^{A'}(\boldsymbol{b})$, such that only the clusters that connect the replicas in region ΔA need to be built. Algorithmically, this estimator can be implemented by starting the generation of every new cluster from spins located at

the inter-replica bonds and inside of region ΔA until all of those spins are partitioned into corresponding clusters.

Due to the non-local extent of the clusters on which it is built, this estimator has an exponentially improved performance with respect to the estimator based on Eq. (3.56).

3.5 Estimator in $d+1$: stochastic series expansion representation

In this section, we encounter for the first time the case when the configuration spaces associated with the replicated ensembles of region A and A' do not fully overlap in the original representation, $\{c\}_A \neq \{c\}_{A'}$. We develop a stochastic map based on a segment partition of the original configuration that projects into a configuration space independent of the entangling region A. Our work is based on some of the ideas initially discussed in the thesis by Stephen Inglis [54]. This work developed a similar estimator applicable to the standard SSE representation of the transverse field Ising model. However, our theoretical formulation of the mapping as a stochastic projection allows us to generalize beyond the deterministic loop partition developed in Ref. [54]. In fact, the deterministic approach breaks down for the XY model considered here as an application. Our EE estimator shows an exponential efficiency improvement compared to the estimator working in the original representation. Our method is easily extensible to the Heisenberg model. We also derive a generalization to the XXZ family class and discuss the possibility for an efficient implementation.

3.5.1 Incompatible configuration spaces

For completeness, we briefly summarize the configuration space of a SSE framework. More details can be found in Sec. 2.3. A configuration in the SSE representation can be represented by a spin state and an operator list, where the operators are local terms in the Hamiltonian. Alternatively, we can represent the same configuration by a linked list of vertices. Such a representation is composed of two labels: the list of links l and the set of vertex types $v^A(l)$. The links connect consecutive vertices acting over the same spins. A given link structure can be compatible with many configurations of vertices. The superscript for the vertex configuration is needed since region A imposes hard boundary conditions at the replica boundaries, see Fig. 3.5. Just like the link structure, those boundary conditions

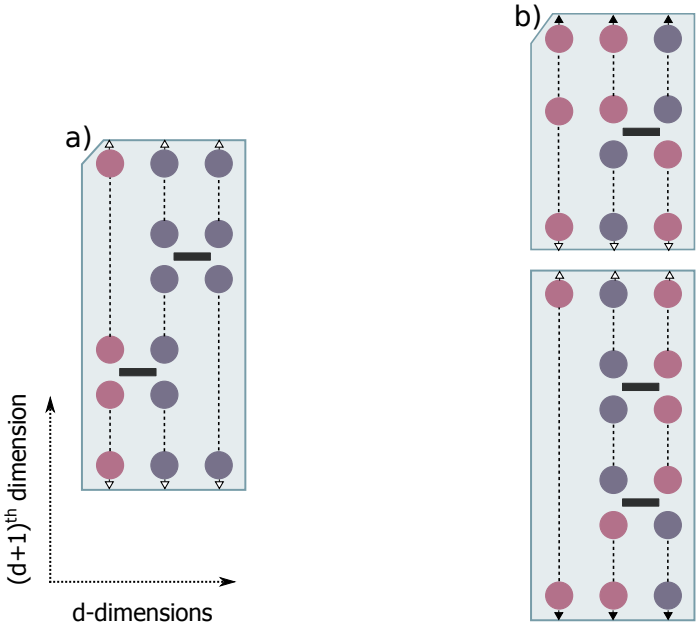


Figure 3.5: Stochastic series expansion representation of the XY model (see Sec. 2.3). The original d-dimensional quantum system is composed of three spins arranged in a chain. Here we adopt the linked-vertex representation. Links are dashed lines. The label l identifies different configurations of those lines. Links that connect pairs of vertices are internal links. Links that connect a vertex to a replica boundary spin are considered external. The two types of small arrows placed next to the replicas' boundary slices mark the boundary conditions along the dimension of the expansion: within a column, spins decorated with the same kind of arrows are connected. a) Depiction of a single replica. b) A configuration in the replicated ensemble, $Z[2, A]$, is shown. The region A extends over the whole system. Note the difference in the replica boundary conditions between a) (intra-replica) and b) (inter-replica).

determine the set of vertex configurations compatible with a given region A. For example, note how a valid configuration with respect to the $Z[2, A = 3]$ ensemble displayed in Fig. 3.5(b) is incompatible with region A' spanning the first two spins.

In order to keep the distinction between the vertex constraints imposed by the link structure and the inter-replica boundary conditions, we consider the link label l only to include *internal* links between vertices, i.e., those links that do not cross the inter-replica boundary. We leave the *exterior* links to be open. In the illustration of a SSE configuration in Fig. 3.5, those exterior links are connected to the boundary spins. Such a representation

deviates from the regular visualization of the SSE simulation cell [41] and gives us the flexibility to reconnect the external links on demand.

These two labels are sufficient to enumerate the configuration space of a quantum spin model. Thus, a replicated partition function can be expressed in terms of a double sum as

$$Z[2, A] = \sum_{l} \sum_{v^A(l)} W(v^A(l)), \tag{3.59}$$

where $W(v^A(l))$ is the weight of a configuration labelled by l and $v^A(l)$.

This general expression can be simplified for the XY model. The model's SSE vertices are explicitly displayed in Fig. 2.5 and also shown in Fig. 3.5. With an appropriate choice of adjustable SSE constants, the weight of each vertex becomes equal to $1/2J$ (see Sec. 2.3.3). Hence, the weight of a vertex configuration does not depend on a particular combination of vertices in the list and, therefore, is completely defined by the length of the corresponding operator list alone. We get that $W_{XY}(v^A(l)) = W_{XY}(l)$ and Eq. (3.59) is simplified to

$$Z_{XY}[2, A] = \sum_{l} W_{XY}(l) \sum_{v^A(l)} 1. \tag{3.60}$$

The second sum counts the degeneracy of vertex configurations compatible with the boundary conditions between replicas.

3.5.2 Segment partition

Motivated by the search for a variable independent of the inter-replica boundary conditions, we introduce a new label, $s(v^A(l))$, that enumerates all possible partitions of a vertex configuration labelled by $(l, v^A(l))$ into a set of non-overlapping segments. The following algorithm is used to trace a single instance of those segments.

1. Pick an unmarked leg located on a boundary slice. Mark it as visited.
2. By following the linked list, switch to a leg connected to it.
3. The new leg belongs to a vertex.
    - If this vertex is unmarked, pick with an equal probability one of two possible non-bounce moves for this vertex and switch to the corresponding leg. Mark this vertex as visited and store the move type.

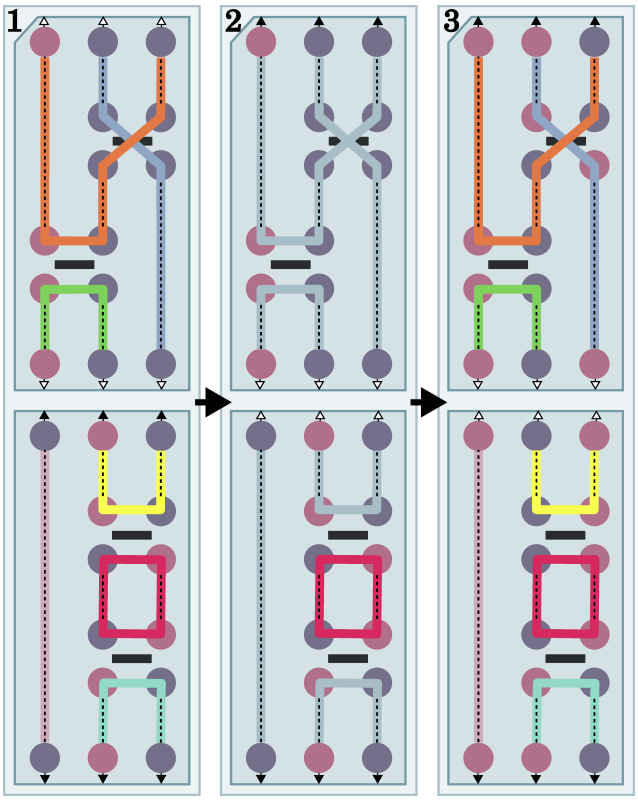


Figure 3.6: A three step conversion process of a vertex configuration $v^A(l)$ into $v^{A'}(l)$ that preserves its segment partition. The details of this process are illustrated in the text. At each step, a simulation cell composed of two replicas (top and bottom) is shown. In this way, the first simulation cell's region A is empty while the other two simulation cells' region A' contains all three spins. Open colored solid lines trace out a segment partition of the first and the third simulation cells (7 segments total). Note that within the same cells, there is also a single closed segment, an inner loop, composed of four legs. In the second simulation cell, the open segments are merged by boundary connections to form a single cross-replica loop identified by the same color. Mismatching boundary spins along this loop are flipped according to the algorithm presented in the text, resulting into a vertex configuration compatible with the new boundary conditions as displayed in the third simulation cell.

- If the vertex is marked, switch to the next leg by performing a move of the same type that was done before.

4. Repeat steps 2-4 until a leg on a boundary slice is reached.

By repeating this algorithm for all legs located on the boundary slices, all open segments are traced out. However, it is possible that some of the legs located on the inner slices have been left unmarked after this procedure. In order to partition those remaining legs too, the closed segments (loops) need to be traced. This is achieved by adjusting two steps of the algorithm. Now in the first step, the choice of legs to be picked is extended to all interior legs. Once the initial leg is picked, the algorithm proceeds in the same way until it reaches the same leg again. Hence, the condition to terminate the execution of the 4^{th} step has to be modified appropriately. By construction, any two segments built in such a manner can never pass through the same leg and, therefore, are non-intersecting.

The segment tracing continues until all legs are marked. By the end of this procedure every leg belongs to one single segment (closed or open). This constitutes a single instance of the partitioning of a vertex configuration into a set of non-overlapping segments. An example of such partition is shown in the first cell of Fig. 3.6.

3.5.3 Equivalence of segment partitioning

It can be also shown that for any vertex configuration in A, $v^A(l)$, partitioned as $s(v^A(l))$, there exists a vertex configuration in A', $v'^{A'}(l)$, with exactly the same partitioning, that is $s(v'^{A'}(l)) = s(v^A(l))$. The proof is by construction. If $v^A(l)$ and $v'^{A'}(l)$ were the same, the task is trivial. Otherwise, $v^A(l)$ has to be modified in order to satisfy the boundary conditions of A'. An example of such process is displayed in Fig. 3.6. Here, the first and third simulation cells represent $v^A(l)$ and $v'^{A'}(l)$ correspondingly. The second cell depicts an intermediate step of the correctional procedure. Here, the open segments are connected into a loop along which the boundary spins mismatches are fixed one-by-one. Further details of the algorithm are given below.

Proceeding column by column, consider each pair of boundary legs, (s_1^0, s_2^0), to be matched with respect to the new boundary conditions A'. The superscript refers to the pair index while the subscript refers to one of two spins included in the pair. If the legs align, $s_1^0 = s_2^0$, proceed to the next pair. Otherwise, randomly choose one of the two legs in the pair. Say it is s_1^0. Since this leg is located on a boundary slice, it belongs to an open segment. Flip all legs belonging to this segment. Now, the original pair of legs is properly

aligned, however there might be another mismatch at the other end of the segment. Call the new pair (s_1^1, s_2^1) where s_2^1 is the leg that has just been flipped as part of the open segment. By the same logic as before, s_1^1 must belong to an open segment whose other end is identified as another boundary leg s_2^2. If $s_1^1 \neq s_2^1$, flip this segment in order to align the (s_1^1, s_2^1) pair and move on to the next pair (s_2^2, s_1^2). Otherwise, proceed to the same pair without flipping the segment. In this way, one-by-one pairs of boundary legs are aligned with respect to A' boundary condition along a loop of open segments. An important subtlety occurs at the last step of this algorithm when the last pair (s_2^n, s_2^1) is considered. Unlike previously, s_2^1 cannot be flipped if those legs do not align. An attempt to do so would entail another iteration of corrections with the same result, thus, initiating the algorithm in an infinite loop.

However, this does not occur in the XY-model due to the special properties of its vertices directly linked to the preservation of particle number in this model. Notice that the only vertex move that connects two anti-aligned legs is the "switch-and-reverse" move (see Fig.2.5); this is the only move that reverses the vertical directionality of propagation of the segment's head. Consequently, once a segment tracing is initiated with the choice of a leg and its state, the spin state of the leg at the segment's head is determined by the vertical direction that the segment passes through the leg. On the last boundary connection, the direction of motion along the segment must be the same as the initial direction, and therefore the initial spin state at the head of the segment under construction is always the same its final state. We see then that for any segment partition of $v^A(l)$, it is always possible to construct a $v'^{A'}(l)$ with the same segment partition. This fact directly implies that all linked-vertex configurations are mapped into the same segment configuration space independently on the entangling region A.

3.5.4 Improved estimator

Now that we have shown the equivalence between any two configuration spaces constraint by boundary conditions A and A' in terms of the segment partitions, we have achieved our initial goal to find a label s independent of the entangling region A. This allows us to apply the procedure developed in Sec. 3.2.4 in order to derive a segments-based EE estimator. Adopting the language of that section, we view the segment partitioning as a stochastic map from the original vertex space into the projected segment configurations space, $p(s|l, v^A(l))$. We proceed with a careful examination of this conditional.

Note that at each vertex, there are two choices how to proceed with the construction

of a segment. Each of them leads to a different partition.[3] Therefore, a simulation cell that contains N_v vertices can be partitioned in 2^{N_v} distinct ways. Since N_v just counts the number of vertices without discerning their types, the number of partitions for a particular $v^A(l)$ is determined by the l label only $N_v(l)$. Therefore, the vertex dependence can be removed from the conditional which takes the following form

$$p(s|l, v^A(s(l))) = p(s|l) \equiv \frac{1}{2^{N_v(l)}}\delta_{s,l}, \tag{3.61}$$

where $\delta_{s,l}$ is an indicator function enforcing the compatibility of a segment partitioning s with the link structure l.

According to the procedure outlined in Sec. 3.2.4, we proceed with the marginalization over the vertex variable $v^A(l)$ to obtain the marginal $W^A(l, s(l))$. As the vertex weights are all the same, this step reduces to counting the number of vertices that are mapped to a segment partitioning $s(l)$ for a given linked structure l and region A. This vertex degeneracy is directly related to the number of closed loops formed by the segments. Indeed, under closure of open segments with respect to the boundary conditions A, $N_b^A(s(l))$ loops that cross inter-replica boundaries are formed. In addition to those loops, there are also $N_i(s(l))$ inner loops formed by the closed segments. Since all those loops do not intersect with each other by construction, the spins within them can be flipped independently. Each combination of the loops' flips leads to a valid vertex configuration. Equivalently, all of those vertex configurations generate the same segment partition $s(l)$. It follows, the marginal is

$$W^A(l, s(l)) = \frac{W_{XY}(l)}{2^{N_v(l)}} 2^{N_b^A(s(l))+N_i(s(l))}. \tag{3.62}$$

Since the inner-loops are unaffected by the boundary conditions, upon the substitution of this weight in Eq. (3.48) their number drops out and an elegant expression for the partition functions ratio is derived,

$$\begin{aligned} r^A_{A'} &= \left\langle \frac{W^A(l, s(l))}{W^{A'}(l, s(l))} \right\rangle_{A'} \\ &= \left\langle 2^{N_b^A(s(l))-N_b^{A'}(s(l))} \right\rangle_{A'}. \end{aligned} \tag{3.63}$$

In practice, this estimator can be implemented in fewer steps that were required to prove its validity. It requires two routines. One routine traces out a random single segment

[3] Two partitions are considered equal when all their segments are the same. In its turn, for two segments to be considered the same the order of legs in the construction of one segment must match exactly with the order of legs in the construction of another segment

partition $s(l)$ for a vertex configuration $v^A(l)$ as was outlined before. In order to speed up the execution, it is not necessary to identify the closed segments. The end product of this routine is to associate the pairs of boundary spins that are connected via open segments. Once this step is done, the second routine takes the set of those pairs together with replicas' boundary conditions as its inputs. Its task is to count N_b^A. This routine is executed for both A and A' with the same open segments. In the end, $N_b^{A'}(s(l))$ and $N_b^A(s(l))$ are known and the EE estimator can be evaluated according to Eq. (3.63).

3.5.5 Benchmarks

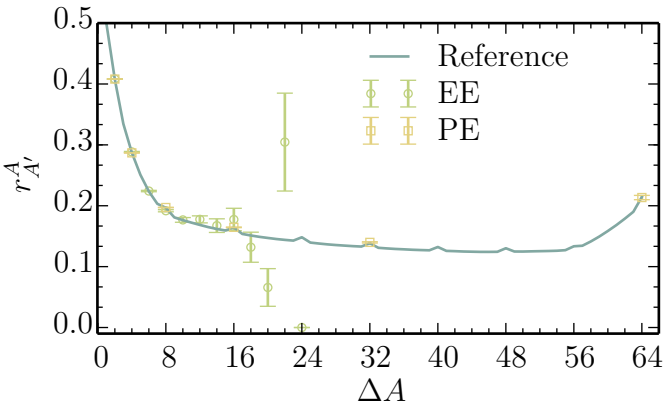


Figure 3.7: The comparison of ratio-based estimators efficiency in a 8×8 system with periodic boundary conditions at $\beta = 8$. The ratio of partition functions measurement is plotted against the size difference between their corresponding region, $\Delta A = A$. Our method is referred to as projected-ensemble (PE) estimator. The extended-ensemble (EE) method (see Sec. 3.2.4) completely fails for $A > 22$, and thus the data is not shown on the plot. The values obtained from the ratio trick serve as a reference. The statistical error in those values is contained within the width of the curve.

To illustrate the efficiency of the new estimator, its raw measurements are compared to the extended-ensemble method (see Sec. 3.2.3) working in the original vertex-list representation (see Sec. 3.2.4) in Fig. 3.7, on the 2D XY model of interest in the main text. Here, the deterioration of both estimators' statistics is seen as the difference between the region $\Delta A = A$ increases. As the reference values, we employ the results obtained from the ratio trick (see Sec. 3.2.1), which constitute a compilation of the extended ensemble (EE) ratio method results from many different Monte Carlo simulations, each executed

with ΔA=1. Note that the EE results are based on five times more Monte Carlo sweeps that were involved to produce the PE ratio method results. Even with this advantage, the ratio method statistics becomes increasingly poor towards ΔA=22. After this threshold, the estimator is no longer capable to capture any meaningful statistics within the running time of its simulations. The performance of the EE is strikingly better. Even when ΔA=64, the largest possible increment in the system, it produces an accurate result with a precision comparable to the ratio method precision at ΔA=10.

3.5.6 Extension to a larger class of models

So far, the discussion has focused specifically on the XY-model. However, the projected-ensemble estimator can be applied in a more general context. In order to understand the more generalized procedure, let us trace through the main steps in the derivation of Eq. (3.63) where the unique properties of the XY-model are used.

Most importantly, the discussion in the Sec. 3.5.3 showing the independence of the partition label $s(v^A(l))$ on boundary conditions A can be generalized. Let us stress that as long as this property holds for a model in question, the framework of the projected-ensemble estimator applies as discussed in Sec. 3.2.4. In general, the proof of this property is achieved through the demonstration that for any segment partition of a vertex configuration $v^A(l)$, there exists a configuration $v'^{A'}(l)$ with exactly the same partition. For the XY-model, the key point in the proof revolves around the properties of the vertex set displayed in Fig.2.5. Namely, it is the property that the only move that switches the leg color is the one that changes the vertical directionality of propagation as well. Let us note that due to this feature, the whole argument is oblivious to the type of lattice the Hamiltonian is defined on; thus, making the estimator Eq. (3.63) applicable to models on lattices beyond the bipartite one considered in the study. Furthermore, the above-mentioned set of vertices is not limited to the XY-model. Indeed, it is common to all XXZ-models (including the Heisenberg model) in a magnetic field or without it. Alternatively, for a different class of models defined on a different set of vertices, it is possible to construct an argument not relying on the aforementioned feature of the XXZ vertices. However, it is most likely to depend on the underlying lattice. For instance, on a bipartite lattice the fact that an even number of lateral moves is required to get back to the same spin can be used for vertex sets in which the only moves that change the leg color are the ones that move onto the neighbouring spins ("switch-" moves).

A a special property of the XY model is used in going from Eq. (3.59) to Eq. (3.60). Namely, this step assumes that all type of SSE vertices have equal weights. For other

models with different symmetries, we can identify two classes: those with equal weights, and those without. We further discuss the anticipated required modifications to the estimator in both cases.

For other models with equal-weight vertices, the extension of the algorithm is straightforward. Essentially, the relevant model weights simply replace the specific XY weights, W_{XY}. For models with imaginary-time loop updates, like the spin-1/2 Heisenberg model, Eq. (3.63) remains fully valid with the same procedure of counting the number of boundary loops, $N_b^A(l)$. One expects that all $SU(N)$-invariant models (with general N) will have this same form of highly-efficient estimator, which should facilitate the accurate estimation of Rényi entropies in these cases. More generally, in equal-weight models with other imaginary-time structures, such as branching clusters in the case of the transverse-field Ising model [55], these loop counters will simply be replaced by the numbers of analogous branching clusters (similarly to the clusters $d+1$ classical Ising model discussed in Sec. 3.4.2). Thus, we expect our method to straightforwardly produce an efficiency gain in a wide variety of important models.

The second class of models is the case of reduced symmetry, where the assumption of equal vertex weights is not longer valid. To incorporate this generalization in the derivation, we avoid the specialization to the XY-model by skipping Eq. (3.60). Assuming the existence of a common label $s(v^A(l))$ independent on the boundary conditions can be proven, the application of the same stochastic map as in Eq. (3.61) leads to the marginal,

$$\begin{aligned} W^A(l, s(l)) &= \sum_{v^A(s(l))} p(s|l, v^A(s(l)))W(v^A(s(l))) \\ &= \frac{1}{2^{N_v(l)}} \sum_{v^A(s(l))} W(v^A(s(l))). \end{aligned} \tag{3.64}$$

As before, this sum iterates over $2^{N_b^A(s(l))+N_i(s(l))}$ different vertex configurations compatible with boundary conditions defined by region A. However, it is not longer possible to factor out the weight as it can take different values for different vertex configurations.

Lastly, we substitute this weight into Eq. (3.48) to obtain a generalized version of the estimator Eq. (3.62):

$$r_{A'}^A = \left\langle \frac{\sum_{v^{A'}(s(l))} W^A(v^A(s(l)))}{\sum_{v^{A'}(s(l))} W^{A'}(v^{A'}(s(l)))} \right\rangle_{A'}. \tag{3.65}$$

Note that Eq. (3.63) can be recovered from this expression for any model with the vertex weights independent on a particular vertex configuration, that is $W(v^A(s(l))) = W(l)$. For instance, this is the case for the Heisenberg model.

In practice, the cost of a straightforward evaluation of this estimator is likely to overcome the gains associated with dramatically improved statistics observed for the XY-model (Fig. 3.7). This consideration is based on two complications that were not present in the simplified version Eq. (3.63). First of all, instead of just counting the number of closed loops compatible with the given boundary conditions, it is now necessary to iterate through all vertex configurations generated by flipping all combinations of those loops in order to calculate the sum in Eq. (3.65). Secondly, unlike in the case for models with equal vertex weights for which only open segments are required to be traced out, in a more general case the inner segments need to be identified as well. Furthermore, those segments can also be used to generate vertex configurations as mentioned in the previous point. Since the number of vertex configurations is exponential in the number of the inner segments, the evaluation of this estimator is prohibitive.

In order to regain the tractability of the estimator, we note that the sums in Eq. (3.65) are likely to be dominated by few vertex weights. Therefore, a sampling approach to their evaluation is expected to be very fruitful. Such importance sampling can be implemented by realizing that the current stochastic map, $p(s|l)$, from Eq. (3.61) ignores the vertex weights during the construction of the segments. If the stochastic construction procedure can be modified so that the resulting conditional, $p(s|l, v^A(l))$, cancels out the vertex weight in the first line of Eq. (3.64), this marginal would take exactly the same expression as in the XY case. As the result, the easy-to-evaluate estimator from Eq. (3.63) would replace the intractable expression Eq. (3.65). Such non-trivial stochastic process is actually at play during the stochastic map employed in Fortuin-Kasteleyn construction discussed in Sec.2.2.3. In the general case treated here, the relevant stochastic map is likely to be based on the generalized loop construction algorithm known as the Kandel-Domany framework [31]. It would be exciting to see the development of the EE estimator in this direction.

Chapter 4

Applications of entanglement entropy scaling analysis

In this chapter, we put to a heavy use the QMC EE estimation framework developed in the previous chapter in order to obtain the second Rényi entropy for two strongly-interacting quantum models. In both cases, we rely on a cylindrical geometry of region A shown in Fig. 4.1 an EE probe. As we show further, this entangling region geometry provides a deep insight into the universal nature of the underlying microscopic models.

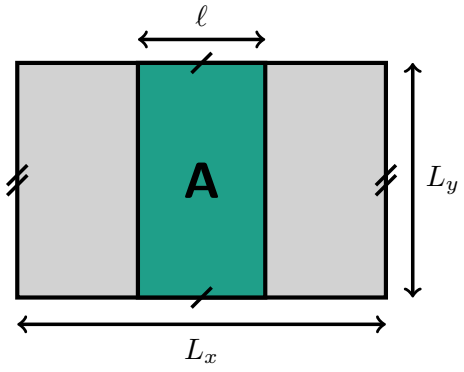


Figure 4.1: Entanglement bipartition used in our studies. The system has periodic boundary conditions such that it can be visualized on the two-dimensional plane. Region A is a cylindrical subregion that wraps around the y-direction and has height ℓ along the x-direction.

4.1 Characterizing continuously broken symmetry

In order to examine the effects of broken continuous symmetry on entanglement, we consider the 2D spin-1/2 XY model,

$$H = J \sum_{\langle ij \rangle} (S_i^x S_i^x + S_i^y S_i^y). \tag{4.1}$$

This model is known to realize a ground state where the $U(1)$ symmetry is spontaneously broken, resulting in one Nambu-Goldstone mode. We consider our system to be embedded in a torus partitioned into two cylindrical regions of linear dimension $L \times l$ and $L \times (L - l)$, see Fig. 4.1 (we take $L_x = L_y = L$). To study the EE of this ground state, we employ the finite-temperature SSE algorithm with specialized loop updates described in Sec. 2.3.3.

Historically, the numerical evidence for a subleading to the area law EE contribution of a smooth region A that scales logarithmically with the subsystem size was first observed in spin wave [56] and finite-size lattice numerics [57]. The apparently anomalous logarithm had no rigorous explanation until a comprehensive theory developed by Metlitski and Grover [23].

They argued that, for a finite-size subsystem with length scale L, the term is a manifestation of the two long-wavelength energy scales corresponding to the spin wave gap, and the *tower of states* arising from the restoration of symmetry in a finite volume [58, 59, 60]. Remarkably, their theory not only explains the subleading logarithm, but predicts that the coefficient is directly proportional to the number of Goldstone modes in the groundstate. Furthermore, describing a Goldstone mode with a free scalar field theory allows them to predict the value for an additional additive *geometric* constant, which is fully universal and should therefore be the same across a wide range of continuum theories and lattice models. We put their predictions to a rigorous test with a highly-efficient EE estimator described in Sec. 3.5.

4.1.1 Origin of the logarithmic contribution

To obtain a qualitative understanding of the origin of the logarithmic correction in Metlitski and Grover's theory, it is simplest to first envision decoupling the two spatial subsystems, A and $\bar{A}$, that define the entangled bipartition. The low-energy degrees of freedom in each subsystem can be described by an $O(N)$ rotor ($N = 2$ for the XY model), representing the direction of the order parameter. Here we are only allowing global fluctuations of the order

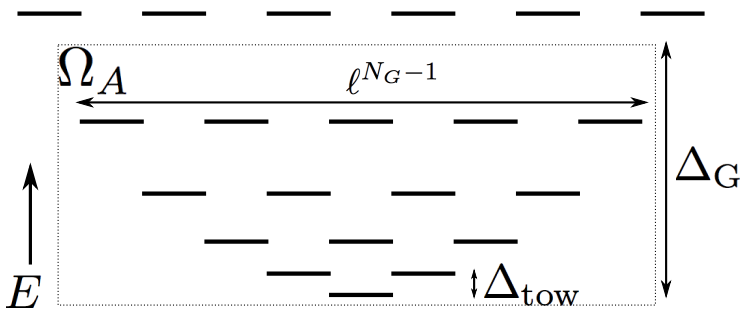


Figure 4.2: Schematic energy level structure of the low energy tower of states for finite-size systems with spontaneous breaking of a continuous symmetry. The correction to the entanglement entropy may be approximated by the log of the number of quantum rotor states below the Goldstone gap, Ω_A, which is represented by the states within the dotted box.

parameter within each subsystem, such that we may approximate the state of A and $\bar{A}$ each as a single independent quantum rotor. The effective Hamiltonian of each subsystem is $H = \mathbf{L}^2/2I$, where $\mathbf{L}^2$ is the total angular momentum operator with eigenvalues $\ell(\ell+1)$ and I is the effective moment of inertia which is extensive, proportional to the magnetic susceptibility χ: $I \sim \chi L^d$ in d spatial dimensions [61]. Thus the energy scale of the tower $\Delta_{\rm tow} = 1/\chi L^d$ vanishes with the system volume, faster than any other energy scale. The eigenstates of $\mathbf{L}^2$ result in the famous "tower of states" observed routinely in computational studies of systems with continuous symmetry breaking in a finite volume [59, 60].

The interaction between A and $\bar{A}$ which aligns the subsystem order parameters may be introduced via a Goldstone mode Hamiltonian H_G which couples the two rotors. The energy scale of H_G is the Goldstone mode gap Δ_G which is the scale of the lowest energy spin waves. Since $\Delta_G \sim c/L$ where c is the spin-wave velocity, $\Delta_G \gg \Delta_{\rm tow}$ in the thermodynamic limit for $d > 1$. In the limit $\Delta_G \to \infty$, there are no *relative* fluctuations in the order parameter between subsystems, and A and $\bar{A}$ act as a single rigid rotor. For finite Δ_G, there will be relative fluctuations between the subsystems order parameters due to the zero point fluctuations of H_G.

To estimate the entanglement entropy contribution from the tower of states, we can count the number of "accessible" states of subsystem A, Ω_A, when the total system is in the ground state, and use $S_{\rm tow} \sim \log \Omega_A$. In the limit $\Delta_G \to \infty$ and the rotors are rigidly coupled, the ground state is the ground state of the total system tower of states with zero total angular momentum: $\ell_{\rm A\bar{A}} = 0$. In this case all states in the A subsystem tower are

accessible to the ground state, as each state in A can be paired with an appropriate state in $\bar{A}$ to form a state with nonzero overlap with the $\ell_{A\bar{A}} = 0$ state. However, as discussed above, by including a *finite* Δ_G and thus allowing *relative* fluctuations of the subsystem order parameter between A and $\bar{A}$, the fluctuations in the subsystem angular momentum are finite and determined by the ratio of the energy scales: $\langle \mathbf{L}_A^2 \rangle \sim \Delta_G/\Delta_{\text{tow}}$ [23]. In fact, the reduced matrix of the subsystem takes the form of a thermal density matrix with an effective "entanglement Hamiltonian" given by H_{tow} and the "entanglement temperature" given by Δ_G [23]; the resulting tower of states structure in the entanglement spectrum has been seen in numerics [62, 63]. Thus the inclusion of Goldstone modes cuts off the accessible states of the subsystem to those with an energy below the spin wave gap, as illustrated in Fig. 4.2.

As an example of this mechanism, consider the case of $N = 2$ (valid for our XY model simulations below). Here the rotors have a single component of angular momentum ℓ^z and the orientation of the rotors is described by a single angle θ. For $\Delta_G \to \infty$ the ground state has $\ell^z_{A\bar{A}} = 0$, which has nonzero overlap with states of equal and opposite ℓ^z in each subsystem: $\left|\ell^z_A = \ell, \ell^z_{\bar{A}} = -\ell\right\rangle$; consequently all $|\ell^z_A\rangle$ states are accessible in this limit. We may include the effect of the lowest Goldstone mode by treating the dynamics of the relative angle between subsystems θ_δ as a single harmonic oscillator with frequency Δ_G and moment inertia $I_\delta \sim \Delta_{\text{tow}}^{-1}$, with an effective Hamiltonian

$$H_G = \frac{1}{2I_\delta} L_\delta^2 + \frac{1}{2} I_\delta \Delta_G^2 \theta_\delta^2. \tag{4.2}$$

Here, the fluctuations in the relative angular momentum L_δ are given by the ground state fluctuations of a harmonic oscillator: $\langle L_\delta^2 \rangle \sim I_\delta \Delta_{\text{G}}/2 \sim \Delta_{\text{G}}/\Delta_{\text{tow}}$. The key point here is that because the order parameter is canonically conjugate to the rotor angular momentum, *increasing* the relative fluctuations in the order parameter *reduces* the fluctuations in L^2. Thus, allowing relative fluctuations of the order parameter between subsystems effectively cuts off subsystem rotor states that are accessed in the ground state at order $\ell \sim (\Delta_G/\Delta_{\text{tow}})^{1/2}$ – a relationship that holds for all N [23].

We may therefore estimate Ω_A by counting the number of states (in A's tower of states) that lie below Δ_G. For systems with $O(N)$ symmetry, the tower of states is described by a rotor living on an $N_G = N - 1$ dimensional sphere, where N_G is the number of Goldstone modes. The degeneracy of each energy level is of order ℓ^{N_G-1}. We then may estimate the total number of states below Δ_G by integrating the degeneracy up to the cutoff $\ell_{\text{co}} = (\Delta_G/\Delta_{\text{tow}})^{1/2}$:

$$\Omega_A \sim \int_0^{\ell_{\text{co}}} d\ell \; \ell^{N_G-1} \sim \left(\frac{\Delta_G}{\Delta_{\text{tow}}}\right)^{N_G/2}. \tag{4.3}$$

Using the relation $\chi = \rho_s/c^2$ from hydrodynamic spin-wave theory where ρ_s is the stiffness [64], the entanglement entropy correction due to the tower of states becomes

$$S_{\text{tow}} \sim \frac{N_G}{2} \log\left(\frac{\rho_s}{c} L^{d-1}\right). \tag{4.4}$$

We see that the logarithmic correction to the area law arises due to the quasi-degeneracy of accessible bulk subsystem states, that scales as a power law in L for systems with spontaneously broken continuous symmetries. This contrasts with the leading-order area law, arising from the exponential scaling of the number of local boundary states with the boundary area. Clearly, the prefactor of the logarithmic correction is a universal number that simply counts the number of Goldstone modes.

4.1.2 Mutual information

On a finite lattice, the spectrum of a model that spontaneously breaks a continuous symmetry is endowed with a set of low-energy states known as tower-of-states. Those energy states scale as $1/L^2$. Therefore, a careful convergence to low temperature is required for an accurate probe of the ground state. Even after a careful temperature convergence, we witness very small thermal contributions to S_2. Unfortunately, those volume-like effects significantly affect our finite-size scaling analysis below. Fortunately, this thermal contribution can be essentially eliminated by employing the mutual information,

$$I_2(A) = S_2(A) + S_2(\bar{A}) - S_2(A \cup \bar{A}).$$

The bulk (volume-law) contributions from the two subsystems A and $\bar{A}$ approximately cancel the bulk term $S_2(A \cup \bar{A})$ in I_2. In the regard of this new quantity, the scaling form described by Metlitski and Grover becomes,

$$I_2 = aL + N_G \log(L\rho_s/c) + 2\gamma_{\text{ord}}. \tag{4.5}$$

Here, a is a non-universal constant, ρ_s and c are the spin-wave stiffness and velocity, and γ_{ord} is the geometric constant that depends on the aspect ratios of the cylinder A [23]. Note, since non-universal (cutoff) dependences are all contained within ρ_s and c, this geometric constant remains fully universal. For the spin-1/2 XY model on the square lattice, $\rho_s = 0.26974(5)J$ and $c = 1.1347(2)J$ were obtained from Ref. [65].

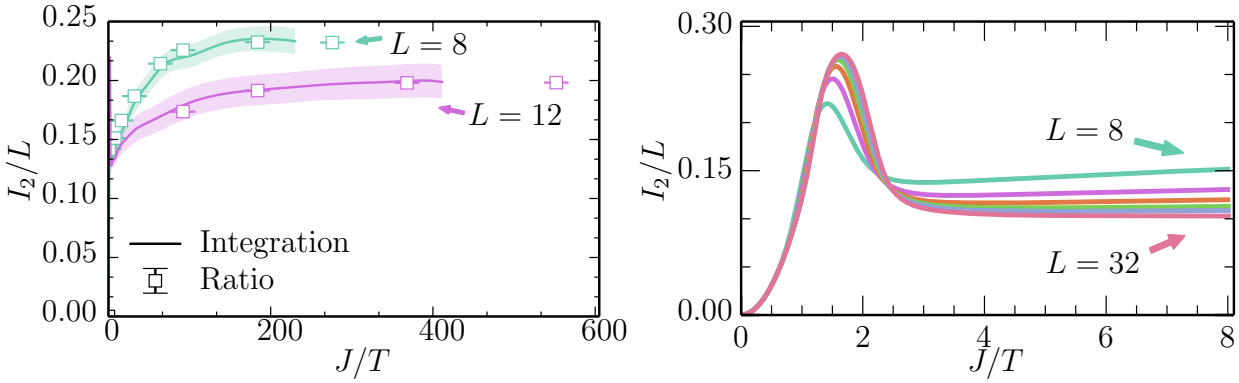


Figure 4.3: The mutual information of the $S = 1/2$ XY model as a function of $\beta \equiv J/T$. The full temperature range is displayed on the *left*) figure. A zoomed-in version focusing around the classical phase transition is shown on the *right*) figure. Solid lines are obtained through thermodynamic integration from $\beta = 0$, with statistical errors estimated by the shading. Square points with error bars are data obtained at a fixed β using the projected ensemble ratio method, described in Sec. 3.5.

4.1.3 Thermal convergence

For system sizes $L = 8, 12, 16$, separate finite-temperature tests are performed to explore the convergence of S_2. In agreement with the expected scaling of the tower-of-states gap, those results confirm that the convergence temperature scales approximately as $1/L^2$. We use this to estimate the convergence temperatures for $L = 20, 24, 28, 32$.

Figure 4.3 illustrates a representative convergence test for different system sizes. The continuous lines of EE measurements are obtained via the thermodynamic integration technique described in Sec. 3.1.3. As witnessed by the right plot, the mutual information peaks at temperatures above the Kosterlitz-Thouless transition of $(T/J)_{\mathrm{KT}} = 0.343$ (which can be detected by the crossing of the finite-size curves; see Ref. [66]). For $T/J < (T/J)_{\mathrm{KT}}$, the mutual information reaches a minimum (at $J/T \equiv \beta \approx 4$ in Fig. 4.3) before undergoing a slow rise. The left plot demonstrates that this rise continues until the approximate ground state is reached, for temperature below the finite-size scaling gap, which for system sizes larger than $L = 8$ occurs for $\beta > 100$. Thus, although the method of thermodynamic integration is useful to produce the general shape of the I_2 curve for a wide range of temperatures, it is difficult to control the systematic error introduced by numerical integration at low temperatures for $L > 12$. Therefore, data used in the below fits was converged at very low temperatures using the projected ensemble ratio method, described in Sec. 3.5.

Figure 4.4 illustrates the resulting temperature-converged mutual information for a va-

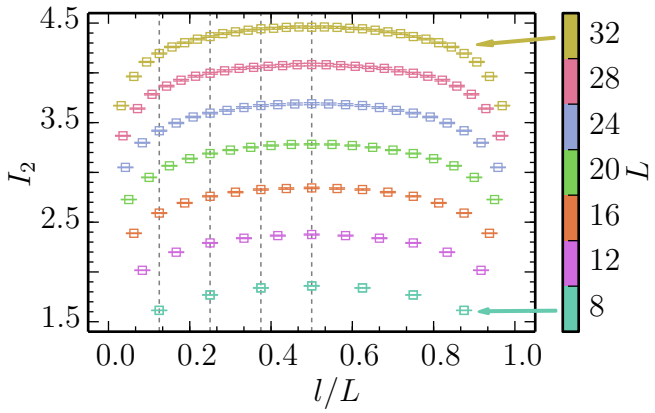


Figure 4.4: The mutual information as a function of torus aspect ratio, for the lowest temperatures examined for each system size. The corresponding β are 184, 368, 736, 1150, 1650, 2300, 3200 ordered from the smallest to the largest system size. Vertical dashed lines are the aspect ratio values employed in the fitting in Figs. 4.5 and 4.6. The measurements are obtained using the projected ensemble ratio method, described in Sec. 3.5.

riety of system sizes, as a function of the height of the cylindrical region, l. Since, for a subsystem A and its complement $\bar{A}$, $S_A = S_{\bar{A}}$ only at $T = 0$, the symmetry of the entanglement entropy about $l/L = 1/2$ provides a sensitive test of temperature convergence. Employment of the "bare" Rényi entropy results in a very slight asymmetry in the curve; use of I_2 restores this symmetry producing high-quality data that can be fitted using Eq. (4.5).

4.1.4 Fitting analysis

In order to extract the coefficients of interest, we subject our mutual information measurements to a thorough fitting analysis (see Table 4.1). The fits are performed independently for fixed aspect ratios l/L in Fig. 4.4. This choice allows us to control the geometric dependence of the mutual information that is absorbed into a constant term. We start by establishing the consistency of the data with the scaling behaviour of Eq. (4.5). This is achieved via fits 1 and 3 where the logarithmic coefficient as well as the geometric constant are set to the predicted values. By comparing the χ^2_k of those fits, we note the importance of the finite-size correction a_4/L. We then proceed with a four-parameter fit (fit 6) in an attempt to simultaneously extract both coefficients of interest. However, a dramatic increase in the error bars of the extracted values signals an overfitting, caused by a limited

Table 4.1: Fitting coefficients, labelled as a_i, extracted from the data points that fall on the dashed line in Fig. 4.4 with the help of various functional forms. The last column shows the χ^2 per degree of freedom that can be used as a measure of the goodness of fit. γ_{ord} refers to the theoretical value of the geometric constant which is aspect ratio dependent. The expected theoretical value[23] for a_2 is 1, while for a_3 they are $0.672, 0.851, 0.921, 0.941$ as arranged in the order from the smallest to the largest aspect ratio correspondingly. We set $c_{s,sw} = \rho_s/c_{sw}$.

#	Fitting function	L_x^A/L	a_1	a_2	a_3	a_4	χ_k^2
1	$a_1 L + \log(Lc) + 2\gamma_{ord}$	1/8	0.1561(4)	-	-	-	3.7
		2/8	0.155(1)	-	-	-	23.0
		3/8	0.155(2)	-	-	-	26.0
		4/8	0.155(1)	-	-	-	11.0
2	$a_1 L + a_2 \log(Lc) + 2a_3$	1/8	0.157(2)	1.01(3)	0.663(4)	-	0.85
		2/8	0.155(2)	1.06(3)	0.806(4)	-	0.67
		3/8	0.156(2)	1.05(3)	0.876(3)	-	0.27
		4/8	0.158(3)	1.02(4)	0.899(6)	-	0.73
3	$a_1 L + \log(Lc) + 2\gamma_{ord} + \frac{a_4}{L}$	1/8	0.1567(2)	-	-	-0.10(2)	0.42
		2/8	0.1569(2)	-	-	-0.51(4)	0.64
		3/8	0.1572(2)	-	-	-0.51(3)	0.2
		4/8	0.1570(4)	-	-	-0.52(6)	0.89
4	$a_1 L + \log(Lc) + 2a_3 + \frac{a_4}{L}$	1/8	0.157(1)	-	0.67(2)	-0.1(2)	0.82
		2/8	0.157(1)	-	0.84(2)	-0.4(2)	0.76
		3/8	0.158(1)	-	0.91(2)	-0.4(3)	0.32
		4/8	0.159(2)	-	0.91(2)	-0.1(3)	0.75
5	$a_1 L + a_2 \log(Lc) + 2\gamma_{ord} + \frac{a_4}{L}$	1/8	0.157(2)	1.00(3)	-	-0.10(4)	0.81
		2/8	0.158(2)	0.99(3)	-	-0.50(5)	0.78
		3/8	0.158(2)	0.99(3)	-	-0.49(4)	0.33
		4/8	0.161(3)	0.95(4)	-	-0.47(7)	0.82
6	$a_1 L + a_2 \log(Lc) + 2a_3 + \frac{a_4}{L}$	2/8	0.145(8)	1.4(3)	0.6(2)	3.0(2)	0.57
		4/8	0.140(9)	1.6(3)	0.5(2)	4.0(2)	0.44

range of available system sizes. Instead, three-parameter fits (4 and 5) with the a_4/L term included are performed by setting either a_2 or a_3 to their respective putative values; these fits provide values in an excellent agreement with the theory [23]. We also note from the

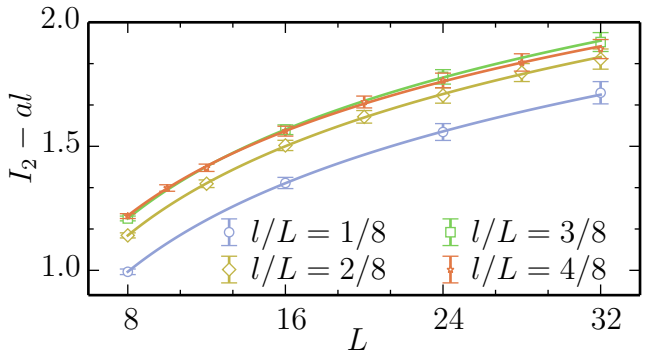


Figure 4.5: A three parameters fit to the functional form 2 in Table 4.1 for different torus aspect ratios with the goal to extract the Goldstone number $N_{\rm G}$ from the expected scaling based on Eq. (4.5). An additional three-parameters fit is used to extract the geometrical constant $\gamma_{\rm ord}$ (not plotted). This fit is based on the fit 4 from Table 4.1.

three-parameter fit 2 that disregarding the a_4/L term has a much larger negative effect on the extracted values of a_3 then on the values of a_2. Such effect is consistent with a stronger relative contribution from the logarithmic term to the total value of the mutual information, as compared to that from the constant term. Thus we conclude that with this data set, we can extract the log coefficient without any assumptions about the theoretical values of the coefficients, since fit 2 (with no theoretical assumptions or $1/L$ correction) results in a_2 values that are consistent with those extracted from fit 5 (which includes a $1/L$ correction at the cost of assuming the theoretical value of a_3). However, to accurately extract the geometric constant from this data, we must include a $1/L$ correction term and thus fix a_2 to its theoretical value.

The results of this analysis are illustrated in Fig 4.5. Here, I_2 is calculated at various aspect ratios (the vertical cuts in Fig. 4.4) and fit to the functional form Eq. (4.5). Specifically, to extract the coefficient of the subleading logarithm, the mutual information was fit via the functional form 2 from Table 4.1. As illustrated in Fig. 4.5, there is definitive evidence for the existence of a logarithm; furthermore, independent fits for the four aspect ratios studied each give $N_{\rm G} = 1$ to within error bars as seen in Fig. 4.6.

Even more striking, we are able to extract the universal shape-dependence of the geometric constant $\gamma_{\rm ord}$. To do so, fits were performed to the functional form 4 in Table 4.1 where $N_{\rm G}$ is fixed at unity in order to remove one parameter from the analysis. Thus calculated, $\gamma_{\rm ord}$ for $N = 2$ in two dimensions can be compared via a zero-parameter fit to the subleading constant term $\gamma_{\rm free}$ calculated in a free scalar field theory [23] through the relation $\gamma_{\rm ord} = \gamma_{\rm free} + \frac{1}{2}\log(2\pi)$, valid for the second Rényi entropy. The free field result

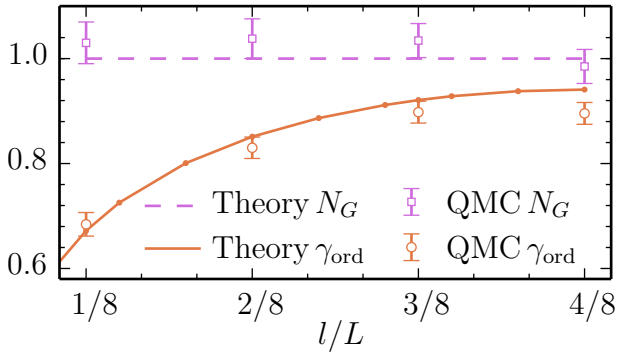


Figure 4.6: The extracted log-coefficient $N_{\rm G}$ together with the geometric function $\gamma_{\rm ord}$ are shown alongside the theoretically-predicted values. The corresponding fits are based on the 2nd and 4th functional forms in Table 4.1.

$\gamma_{\rm free}$, which depends on the aspect ratio l/L, can be calculated numerically for free bosons on the lattice using the correlation matrix technique (as in Ref. [67, 12]). As illustrated in Fig. 4.6, the resulting theoretical curve is in excellent agreement with our QMC results for $\gamma_{\rm ord}$.

4.1.5 Discussion

In addition to confirming the proportanality of the logarithmic scaling to the number of Goldstone modes, $N_{\rm G} = 1$, we are able to converge the value of an additional additive geometric constant $\gamma_{\rm ord}$, which is fully universal since all short-distance physics is confined to the (known) spin wave stiffness and velocity, contained within the argument of the logarithm. The resulting $\gamma_{\rm ord}$ has a functional dependence on the geometric aspect ratio of the entangled bipartition. This function matches, to within error bars, that calculated using a free scalar field theory regularized on a toroidal square lattice, with no adjustable parameters.

Armed with the *a posteriori* knowledge of Eq. (4.5), we revisit the dataset for the Heisenberg model that led to the detection of the anomalous EE scaling in QMC simulations [57]. Based on a similar analysis as that presented in the preceding section, we conclude that the Heisenberg model contains much weaker "signal-to-noise" ratio than the XY model. Indeed, for the Heisenberg model, the coefficient of the area law, as well as the coefficient of the $1/L$ correction, are more than an order of magnitude larger than the corresponding coefficients for the XY model. At the same time, the signal from the logarithmic and constant terms are comparable (or smaller) to those in the XY model. Hence, although

the Heisenberg data of Ref. [57] is consistent with the prediction of Ref. [23], it is difficult to obtain agreement with unconstrained fits.

In summary, our work mounts a strong evidence for the validity of the theoretical prediction of Metlitski and Grover [23]. In this case, with the full understanding of the universal structure of the entanglement entropy in the presence of a spontaneously broken continuous symmetry, the door is now open to the examination of Goldstone modes in a large variety of systems, through the calculation of Rényi entropies.

4.2 Probing universal geometric function

As an alternative to traditional critical exponents, universal numbers extracted from the EE can be used to characterize quantum critical points. In general, it is an open question whether the universal information about a fixed point contained in entanglement quantities differs from that contained in critical exponents. Recently, there has been considerable progress in relating the content of universal numbers from entanglement entropies to those obtained from two-point functions in $2+1$ dimensions. A particular success story is the contribution from a local corner in the entangling boundary. There, the universal coefficient of the corner contribution has been shown to be related to the central charge C_T defined by the two-point function of the stress tensor. This result was initiated by numerical studies of interacting quantum critical systems [33, 68], where the corner contribution was observed to scale with the number of degrees of freedom of the underlying field theory [69, 70]. Subsequent comparison to similar scaling in C_T, numerically calculated from the conformal bootstrap [71], motivated a conjecture relating the two quantities [72, 73, 74], which was eventually proven in general [75]. Further studies have examined the behaviour of universal corner coefficients for more general angles and Rényi indices in $2+1$ dimensions [76], while other work has examined universal features due to corners in higher dimensions [77, 78, 79, 80, 81, 82, 83, 84].

A corner or vertex is only one geometry that induces a universal contribution to the EE in $2+1$ dimensions. In this work, we once again we turn to a cylindrical entangling region on a torus (see Fig. 4.1). The EE of such bipartition for a critical system contains a universal scaling coefficient that we call κ. Contrary to the recent efforts described above to study a corner's universal contribution to the EE in both free and interacting theories, studies of κ have been relatively restricted, with results obtained only for free theories and theories with a gravitational dual. The cases where κ has been studied include the free scalar field theory with the dynamical exponent $z = 1$ [77, 73, 85, 86] and $z = 2$ (the

quantum Lifshitz model) [87, 88], as well as their fermionic analogues: Dirac fermions with $z = 1$ [77, 73, 89] and $z = 2$ (the quadratic band touching model) [89, 88].

In $(2 + 1)$-dimensional interacting theories, no exact results for κ have been obtained to date. Here, we examine the cylindrical-slice Rényi EE in a $(2 + 1)$-dimensional Ising model, an interacting theory for which the critical behaviour is governed by the scalar Wilson-Fisher fixed point.

4.2.1 Scaling in the thin-slice limit

As discussed in Sec. 1.2, the EE scaling of a cylinder embedded in a torus like in Fig. 4.1 is expected to take the following form,

$$S_n(\ell, L_x, L_y) = a_n \frac{L_y}{\delta} + \chi_n(u, b) + \cdots , \tag{4.6}$$

where $u = \ell/L_x$, $b = L_x/L_y$, and the ellipses denote further finite-size subleading corrections such as those that scale as δ/L. Compared to our previous discussion of Sec. 1.2, we have included a dependence on the Rényi index n. The behaviour of χ_n as a function of the aspect ratio u is known to obey certain restrictions. In particular, being an entropy measure of pure states, it must respect a symmetry around $u = \frac{1}{2}$ such that $\chi_n(u) = \chi_n(1 - u)$.

Alternatively, as per Sec. 1.2, the form of the geometric function can be analytically computed when the boundary conditions are ignored. Namely, in the limit where $u \ll 1$ (*i.e.*, $\ell \ll L_x$), the EE contributions at different length scales are expected to be indifferent to the infrared boundary conditions and we expect that we can estimate S_n by considering the EE of a strip-like region embedded in infinite space [77]. In this case, we expect that

$$S_n(\ell \ll L_x, L_y) = a_n \frac{L_y}{\delta} - \frac{\kappa_n}{bu} + \cdots , \tag{4.7}$$

where κ_n is a universal coefficient that can vary with the Rényi index n. By comparing with Eq. (4.6), we see that the second term in this equation reveals the behaviour of the universal scaling term for the cylinder entanglement in the thin slice limit such that

$$\chi_n(u \to 0, b) = -\frac{\kappa_n}{bu}. \tag{4.8}$$

Therefore, the thin-slice limit of the geometric function reveals a universal number that we make our mission to extract. The numerical value of κ_n has previously been calculated in a number of free theories, holographic duals, and phenomenological models [87, 89, 88, 85, 86, 88]. Specifically, values of κ for the free scalar field theory in $2 + 1$ are $\kappa_{1,\text{Gaussian}} = 0.0397$ for the von Neumann entropy [77] and $\kappa_{2,\text{Gaussian}} = 0.0227998$ for the second Rényi entropy $S_2(A)$ [73]. This second value $\kappa_{2,\text{Gaussian}}$ is of particular relevance to the present study.

4.2.2 Conical singularity

While the main focus of this study is to better understand the universal geometric function χ_n, it is important to control the behaviour of the other terms in order to properly extrapolate to the thermodynamic limit. In particular, the conical singularity term $g_n(\delta/\ell)$ is unavoidable in Monte Carlo calculations of the Rényi entropies. This anomalous term leads to a correction of the scaling ansatz in Eq. (4.6) to

$$S_n(\ell, L_x, L_y) = a_n \frac{L_y}{\delta} + \chi_n(u, b) + g_n\left(\frac{\delta}{\ell}\right) + \cdots , \tag{4.9}$$

Physically, this conical singularity is due to the restructuring of the lattice that occurs when one calculates Rényi entropies using the replica trick. While this modification does not change the coordination number of the lattice, it affects its topological structure, giving arise to a relevant operator that is locally confined. This effect transpires through the subleading anomaly correction term that we have called $g_n(\delta/\ell)$ [90]. As a recent study shows [91], when unaccounted for, the presence of this term can lead to erroneous extrapolation results. Unfortunately for us, while its scaling form is known for 1+1 systems, no such analytical result exists in $2+1$. However, the EE contribution from this term is expected to grow when ℓ becomes small, which is exactly the scaling regime in which we are interested. Therefore this conical singularity term cannot be safely neglected. For this reason, we develop an extraction procedure, designed to directly access the universal term χ_2 by isolating the effect of the conical singularity term.

4.2.3 Fitting approach

We now turn to a discussion of our numerical extraction of κ_2 for the second Rényi entropy. The raw data obtained from Monte Carlo simulations on the Ising model requires significant analysis due to a number of factors. In particular, the area law and the unknown conical singularity term pollute the universal contribution $\chi_2(u, b)$. Additionally, we are tasked with striking a fine balance when choosing the right cylinder height ℓ. On one hand, we are interested in the thin-slice limit $\ell \ll L_x$, while on the other hand the continuum EE scaling is only expected to apply when $\ell \gg \delta$. For this reason, we first perform the scaling analysis on the model of free bosons on a lattice to benchmark against previously known results [73]. The continuum low-energy theories of the free and interacting models are captured by the Gaussian and Wilson-Fisher fixed points correspondingly that are intimately related to each other in the RG space as shown in Fig. 4.7. From this point of view, it is interesting to see whether the story told by the numerical estimates of κ_2 between the two theories is

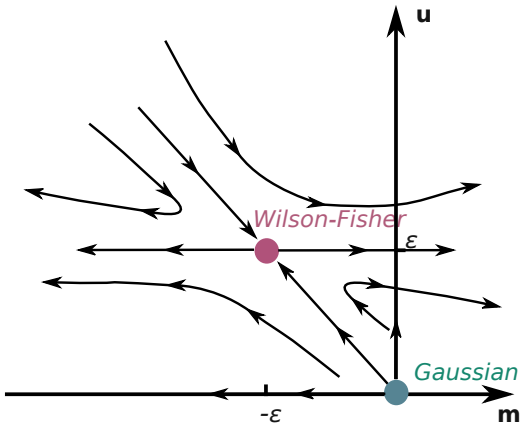


Figure 4.7: Visualization of the relationship between the Wilson-Fisher and Gaussian fixed points in the parameters space of the ϕ^4 field theory described by the action $S[\phi] = \int d^d x \left[(\nabla\phi)^2 + m\phi^2 + u\phi^4\right]$. The positive parameter ϵ encodes the dimensionality of the system, $d = 4 - \epsilon$. Arrows indicate the renormalization group flows.

different from the expectation of the ϵ-expansion procedure that links the fixed points via a perturbative expansion.

As we will show below this benchmark procedure illuminates the crucial importance of the fitting parameter γ that we introduce in order to extend the range of applicability of continuum results to the lattice. Surprisingly, this parameter enables us to take advantage of our whole dataset including the extreme thin-slice regime, $\ell = 1$.

We consider toroidal rectangular geometries and fix $L_y = L$ and $L_x = 4L$, where the factor of 4 allows us to better probe the $\ell \ll L_x$ limit. In this thin-cylinder limit the universal geometric function is expected to take the form given in Eq. (4.8) such that

$$\chi_2(u \to 0) \sim -\kappa_2 \frac{L}{\ell}, \tag{4.10}$$

where we have removed the implicit dependence on the aspect ratio b since we take it to be constant.

In order to remove the dominant area law contribution to the entropy scaling, we consider the difference in EE between two cylinders with different lengths ℓ and ℓ_0, with $\ell < \ell_0$. In the case where the first cylinder is in the thin-slice regime such that $\ell \ll L_x$, we

expect:

$$S_2(\ell, L) - S_2(\ell_0, L) = -\kappa_2 \frac{L}{\ell} + g_2\left(\frac{\delta}{\ell}\right) - \underbrace{\chi_2\left(\frac{\ell_0}{L}\right) - g_2\left(\frac{\delta}{\ell_0}\right)}_{\equiv B(\ell_0, L)}. \tag{4.11}$$

As discussed in Sec. 3.2.1, this difference is directly measured in QMC simulations, making our dataset completely free of the area law term and the associated statistical variance. We use the convention $\delta = 1$ for our lattice calculations.

4.2.4 Free theory

4.2.4.1 Gaussian scalar field theory

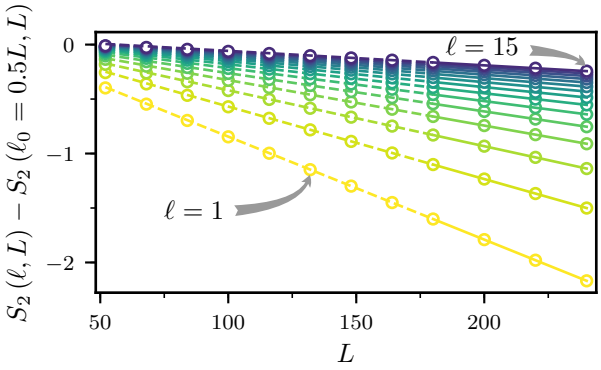


Figure 4.8: entanglement entropy with respect to a reference region for the free model as function of the system size. Different colours correspond to different sizes of region A between $\ell = 1$ and $\ell = 15$. The solid lines are linear fits to four largest system sizes. The dashed continuations of the same color are the extrapolation of those fits to lower system sizes. The linear fit seems to capture most of dependence on L. Only upon closer examination can one observe that the extrapolation quality deteriorates for larger ℓ, in line with the expectations discussed for the thin-slice limit in the main text.

We warm up with a free (Gaussian) scalar field theory on the toroidal system described above. In order to avoid zero modes that could lead to a logarithmic EE contribution [88], we impose anti-periodic boundary conditions along the y-axis while periodic boundary

conditions are kept for the x-direction. The Hamiltonian is given by

$$H_{\text{free}} = \frac{1}{2}\sum_i \left(\pi_i^2 + m^2\phi_i^2\right) + \frac{1}{2}\sum_{\langle i,j\rangle} \left(\phi_i - \phi_j\right)^2 , \tag{4.12}$$

where ϕ_i is a bosonic field with mass m and conjugate momentum π_i. Exact methods for calculating the Rényi entropies for such free scalar theories are described in detail in Refs. [67, 77, 92, 12].

4.2.4.2 Fitting analysis

We plot in Fig. 4.8 the left-hand side of Eq. (4.11) for this free theory versus L, grouping points by their value of ℓ. The behavior appears linear over a wide range of ℓ values, strongly hinting at a dominant contribution from the thin-cylinder form for $\chi_2(\ell/L)$ as in Eq. (4.10). Motivated by this observation, we perform fits of $S_2(\ell, L) - S_2(\ell_0, L)$ to a function f_1 that is linear in ℓ^{-1}. This is given by

$$f_1\left(\ell\right) = -\kappa_2^L \frac{L}{\ell} + C_1, \tag{4.13}$$

for each system size L, where there are two fitting parameters κ_2^L and C_1. The role of C_1 is to absorb the offset due to the ℓ-independent term $B\left(\ell_0, L\right)$. When performing these and the following fits, instead of fitting all points at once, we fit data over a sliding window. For this reason, the estimate for κ_2 has an explicit dependence on both ℓ and L. However, to reduce the notational clutter, we avoid showing the dependence on ℓ, since the x-axis in all the figures makes this dependence explicit.

The results for the extracted $\kappa^L_{2,\text{Gaussian}}$ are shown in the left panel of Fig. 4.9. This plot reveals non-linear dependencies on ℓ^{-1} and demonstrates the challenge in extracting an unbiased estimate for κ_2. Most notably, as judged by the proximity to the known exact result from the Gaussian theory, the best estimate for κ_2^L does not come from the thinnest cylinders with $\ell = \delta = 1$. This observation is not surprising since the EE scaling prediction is only expected to hold in the continuum, which, on a lattice, amounts to the requirement $\delta \ll \ell$. Both this condition and the thin-slice requirement $\ell \ll L$ constitute the challenge of tuning to a regime where both criteria are satisfied. One possible approach is to consider the largest extracted κ_2^L as the best estimate for each system size. However, these values converge quickly to an underestimate for the true value, implying that this extrapolation technique will yield a biased result.

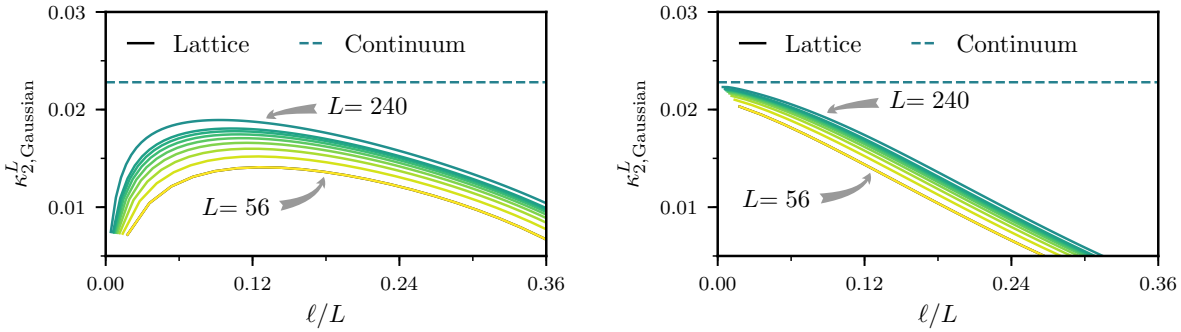


Figure 4.9: κ_2 values for the free theory extracted from Fig. 4.8 for different system sizes L *left*) via the linear fit in Eq. (4.13) *right*) via the non-linear fit in Eq. (4.14), with an additional parameter γ introduced to capture the short-distance behavior important in the regime of small ℓ. Each fit again uses six points, and the value of ℓ for the x-axis corresponds to the lowest value used in each fit. The density of the extracted values on the x-axis is so high that plotting each point results in too much clutter and thus a continuous line representation is used instead. The dashed line indicates the known value for the Gaussian theory in the thermodynamic limit [73].

An alternative approach originates from the following insight: any length on a lattice is only defined up to the lattice spacing. As a consequence, on a lattice it is impossible to distinguish between the class of continuum cylindrical regions A with lengths in the range $(\ell-\delta, \ell+\delta)$. In light of this realization, let us introduce a parameter $\gamma \in (-\delta, \delta)$ designed to capture the degree of freedom associated with mapping the lattice and continuum theories such that the left-hand side of Eq. (4.11) instead takes the form $-\kappa_2 \frac{L}{\ell+\gamma} + g_2\left(\frac{\delta}{\ell}\right) + B\left(\ell_0, L\right)$. Here we have included γ only within the term for which it contributes most significantly. Notice that with L and ℓ_0 kept constant, $B\left(\ell_0, L\right)$ is also a constant and, therefore, can be ignored in the following discussion. The expected first-order offset due to γ is then $-\frac{L}{\ell}\left(\kappa_2 - \kappa_2\frac{\gamma}{\ell} - \frac{g_2(\delta/\ell)\ell}{L}\right)$, and the expression in the parentheses is a good approximation for the value of κ_2^L extracted under a linear fit to Eq. (4.13). This expression thus explains well the biases observed on the left panel of Fig. 4.9. Namely, we observe that the estimate for κ_2 is polluted by two terms. While the term containing the conical singularity scaling function $g_2\left(\delta/\ell\right)$ can be eliminated by taking the thermodynamic limit $L \to \infty$, the systematic offset due to γ can only be suppressed by scaling ℓ to infinity as well. In addition, this term explains the observed increased negative offset in the regime $\ell \sim \delta$.

In order to take these considerations into account, we parameterize the ambiguity in

the definition of the cylinder's length by including the parameter γ in out fits. We perform new non-linear fits of $S_2(\ell, L) - S_2(\ell_0, L)$ to the function

$$f_2(\ell) = -\kappa_2^L \frac{L}{\ell + \gamma} + C_2, \tag{4.14}$$

for each L, now with three fitting parameters κ_2^L, γ and C_2. The outcome for the free theory is illustrated on the right panel of Fig. 4.9. The dependence of the extracted κ_2^L estimate on ℓ for this theory has now drastically changed. In particular, the introduction of the parameter γ has completely removed the downward drop observed at small ℓ in this plot. Moreover, the systematic offset, *i.e.*, the difference between the peak and the known value, which seemed to survive to the thermodynamic limit in left panel of Fig. 4.9, is also remedied.

With a firmly-supported understanding of the short-distance scaling of the Rényi entropy, we are ready to account for the conical singularity in our estimate for κ_2. For this we note that although the non-linear fit based on Eq. (4.14) cannot distinguish the contributions from the universal term and the conical singularity, the latter is independent of system size and therefore its relative magnitude decays as L^{-1}. To be more specific, $\kappa_2^L \approx \kappa_2 - \frac{g_2(\delta/\ell)\ell}{L}$. Consequently, we perform a second fit, extrapolating the previously extracted κ_2^L towards $L = \infty$ via a two-parameter fit linear in L^{-1} such that, for each ℓ, we fit the results from the plot of Fig. 4.9 to the function

$$f_{\text{extrap.}}(L) = -C_{\text{extrap.}}/L + \kappa_2^\infty, \tag{4.15}$$

where κ_2^∞ and $C_{\text{extrap.}}$ are fitting parameters. Here κ_2^∞ represents our final estimate for κ_2 extrapolated to the thermodynamic limit, with the results for the free theory illustrated in Fig. 4.10. Taking the value corresponding to the smallest region A on the largest system considered as our best numerical estimate, we find $\kappa_{2,\text{Gaussian}}^\infty = 0.0227558$, which is less than 0.2% below the known value of 0.0227998 [73]. To put this result into context, we can compare our estimate to that obtained in Ref. [85], which also aims to numerically extract $\kappa_{2,\text{Gaussian}}$ but does not take into account the proposed offset that scales as $\frac{\gamma}{\ell}$. Although the authors were able to collect a dataset for system sizes as large as $L/\delta = 2000$, their extrapolation to the thermodynamic limit yielded a value for $\kappa_{2,\text{Gaussian}}$ that is still 9% off from its theoretical value. Therefore, our fitting procedure produces an estimate which is two orders of magnitude closer to the theoretical value, despite considering system sizes an order of magnitude smaller.

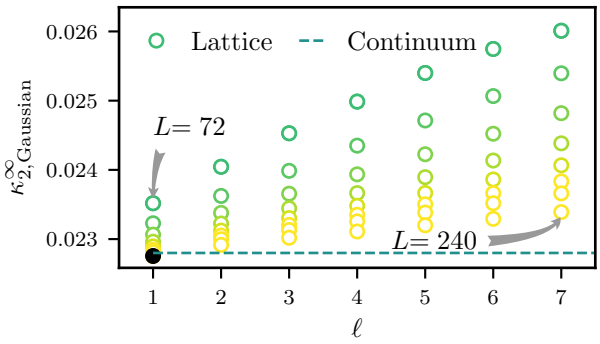


Figure 4.10: κ_2 values for the free theory extrapolated to the thermodynamic limit. These values are obtained via an additional six-point fit to the form in Eq. 4.15 to the results in the right plot of Fig. 4.9. The x-axis indicates the value of ℓ used in each fit. The dashed line represents the known continuum value for the Gaussian theory [73], while the solid black circle indicates our best numerical estimate, namely $\kappa^{\infty}_{2,\text{Gaussian}} = 0.0227558$, for the free theory in the thermodynamic limit.

4.2.5 Interacting theory

4.2.5.1 Wilson-Fisher fixed point

To extract universal characterizations of the Wilson-Fisher fixed point from the EE, we utilize a mapping from the quantum theory in $2+1$ dimensions to a 3-dimensional isotropic Ising model with classical spins $s \in \{-1, 1\}$ and reduced Hamiltonian given by

$$E(\mathbf{s}) = -J/T_c \sum_{\langle i,j \rangle} s_i s_j, \tag{4.16}$$

at its critical temperature $J/T_c = 0.2216544$ [93]. The validity of this approach is based on the path-integral mapping discussed in Sec. 2.2 and the construction of an improved EE estimator is described in Sec. 3.4.

Since the field theory of interest (the scalar ϕ^4 theory) is Lorentz invariant with a dynamic exponent $z = 1$, finite size scaling studies often scale the imaginary time dimension L_τ proportional to the linear spatial dimension L (with a proportionality constant being close to unity for some observables [94]). However, this dimension effectively represents the quantum inverse temperature, namely, $J^Q \beta^Q = L_\tau \Delta\tau$ where $\Delta\tau$ is the unitless discretization constant of the lattice in the imaginary time direction and J^Q is the interaction

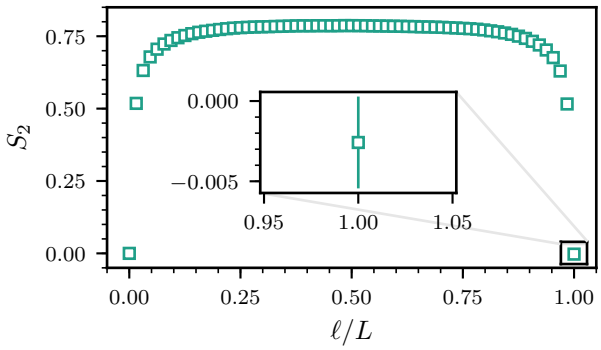


Figure 4.11: Test of the thermal convergence of S_2 for a system of linear size $L = 16$. The direction corresponding to the imaginary time is taken to be $L_\tau = 15L$. The reflection symmetry of the measured S_2 around $\ell/L = 0.5$ is in correspondence with the theoretical expectation that the entanglement entropy of region A is the same as the one of its complement in the ground state. The inset zooms in on a single datapoint corresponding to region A comprising the full system. Its value is zero within the errorbar further indicating the absence of thermal fluctuations.

parameter for the quantum model (for instance, the ferromagnetic coupling). For studies of the ground-state entanglement entropy, one must be careful to ensure that this new dimension is large enough to eliminate all thermal contributions. For this reason, we adopt the practical strategy of converging our simulations below the energy gap $\Delta(L)$ due to the finite system size L, such that $\beta^Q \gg \frac{1}{\Delta(L)}$. The convergence can be tested by studying the amount of thermal entropy contained in the entire system, which should be zero if we are indeed probing the (pure) quantum ground state. As shown on the inset of Fig. 4.11, the entropy of the whole state is zero within the statistical uncertainty for a proportionality constant $L_\tau/L = 15$.

4.2.5.2 Fitting analysis

We now use the insight gained in the consideration of the free theory in Sec. 4.2.4 to proceed with the extraction of $\kappa_{2,\mathrm{WF}}$ for the interacting Ising theory at the Wilson-Fisher fixed point. As before, we begin by plotting the full data set as a function of L (see Fig. 4.12), and we again note a dominant linear behaviour attributed to the universal term's contribution in the thin-slice limit. This observation suggests that the two-step fitting procedure developed above for the free theory can be carried over and applied for this

interacting data set.

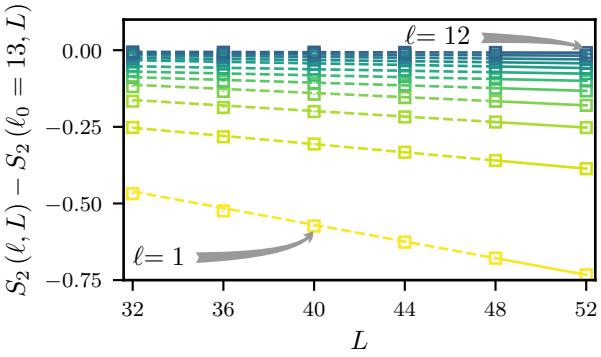


Figure 4.12: Entanglement entropy with respect to a reference region A of size $\ell_0 = 13$ for the 2d transverse field Ising model at criticality as a function of the system size. Different colours correspond to different sizes of region A from $\ell = 1$ to $\ell = 12$. The solid lines are linear fits to the two largest system sizes, and the dashed continuations of the same colour are the extrapolations of those fits to lower system sizes. As in Fig. 4.8, the linear fit seems to capture most of the dependence on L, and it is only upon closer examination that one can appreciate the need for the two-step fitting procedure developed in Sec. 4.2.4.2.

The effect of including the fit parameter γ for extracting an unbiased estimate for $\kappa_{2,\mathrm{WF}}$ can be seen by comparing fits with and without it as displayed on the left and right panels of Fig. **??** correspondingly. The estimates obtained via fits containing γ are all above the corresponding estimates without it. This situation is analogous to the systematic bias towards lower values of the extracted $\kappa_{2,\mathrm{Gaussian}}$ observed for the free theory when the γ parameter is not included in the fits (see Fig. 4.9).

Concentrating further on the fits including the parameter γ in the right plot of Fig. **??**, we note a systematic increase in error bar with increasing ℓ/L. This trend can be explained by considering the relative strength of the leading contribution to the universal term (proportional to L/ℓ), against the subleading terms. These terms originate from contributions due to the conical singularity and the next-order Taylor expansion in $\chi_2(u)$, and scale like $u = \ell/L$ relative to the leading term. Since our fitting form ignores these additional terms, the error bars can be seen as a qualitative indicator for the validity of the assumption of their relative insignificance. Indeed, the regime where the error bars are empirically small is also that in which we are interested ($\ell/\delta \to 1$), allowing us to proceed with confidence

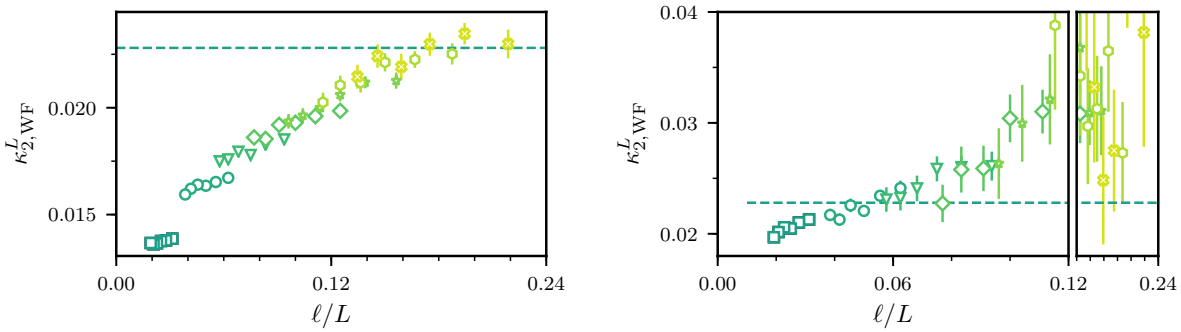


Figure 4.13: κ_2 values for the Ising theory extracted from Fig. 4.12 for different systems sizes L *right*) via the linear fit in Eq. 4.13 *right*) via the non-linear fit in Eq. 4.14 with an additional parameter γ introduced to capture the short-distance behavior important in the regime of small ℓ. On the *left*) the x-axis is split into two parts in order to focus on the values of the extracted $\kappa^L_{2,\mathrm{WF}}$ at low ℓ/L, since the values are noisy for $\ell/L > 0.12$. Markers with the same color and style share the same value of ℓ. The x-axis indicates the lowest value of ℓ used in each six-point fit divided by the system size. The dashed line represents the theoretical value for the Gaussian theory in the thermodynamic limit [73].

onto the second extrapolation that estimates $\kappa_{2,\mathrm{WF}}$. This step is completely analogous to that done in the free theory extrapolation (see Sec. 4.2.4.2, and our results for the Ising theory are shown in Fig. 4.13). As before, this step is based on the fact that the relative strength of the previously neglected terms decay as δ/L. In correspondence with our previous discussion, the error bars are significantly reduced for small ℓ. Furthermore, we note that for a decreasing ℓ the value for $\kappa^\infty_{2,\mathrm{WF}}$ seems to decrease at first, similar to the trend observed for the free theory in Fig. 4.10. However, our $\kappa^\infty_{2,\mathrm{WF}}$ estimates seem to stabilize within error-bars for $\ell \leq 3$. This fact can be interpreted as an indication that we have reached system sizes large enough to accurately probe the $u \to 0$ limit. Concluding our analysis, we take the extrapolated value $\kappa^\infty_{2,\mathrm{WF}} = 0.0174(5)$, which is extracted at $\ell = 1$, as our best estimate for κ_2 at the Wilson-Fischer fixed point.

4.2.6 Discussion

The coefficient κ_2 that we have extracted from cylindrical geometries in the thin-slice limit additionally serves to give insight about entangling geometries with corners that are difficult

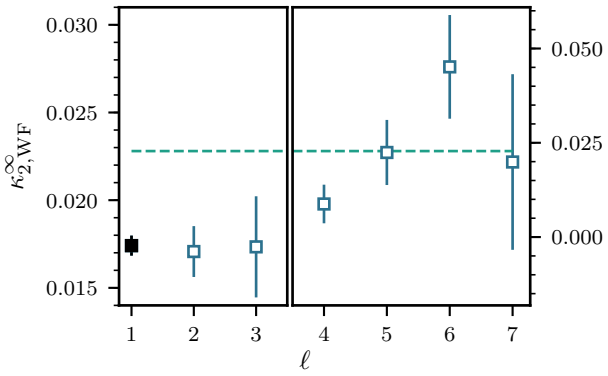


Figure 4.14: κ_2 values for the Ising theory extrapolated to the thermodynamic limit. These values are obtained by fitting the results from the right panel of Fig. **??** to form in Eq. 4.15. The x-axis indicates the ℓ value used in each six-point fit. The black marker for $\ell = 1$ indicates our best numerical estimate $\kappa^{\infty}_{2,\mathrm{WF}} = 0.0174(5)$, and the dashed line represents the known value for the Gaussian theory in the thermodynamic limit [73]. The plot is split into two halves, with the right half providing a wider range of y values. These values are not as accurate as the ones on the left half but show the trend of decreasing error-bars and a decreasing estimate for $\kappa^{\infty}_{2,\mathrm{WF}}$.

to access through direct means on a lattice. Specifically, κ_2 is related to the logarithmic coefficient $a_2(\theta)$ that arises in the scaling of the Rényi EE when the entangling geometry contains a corner such that, in the small-angle limit, $a_2(\theta \to 0) = \kappa_2/\theta$ [77, 95]. Further, combining our results for $\kappa_{2,\mathrm{WF}}$ with previous results for $a_{2,\mathrm{WF}}(\pi/2)$ in the same interacting theory, we expect that one can approximately reconstruct the behaviour of $a_{2,\mathrm{WF}}(\theta)$ for all angles θ using techniques similar to those proposed in Ref. [73].

Our numerical value of κ_2 for the interacting fixed point is relatively close to the value for the free theory, but does have a significant difference when statistical errors are taken into account. This finding is interesting in the context of a recent large-N calculation for the more general N-component $O(N)$ model [96], in which the κ_1 value, extracted from the von Neumann EE, of the Wilson-Fisher fixed point is expected to be directly related to the non-interacting Gaussian fixed point as $\kappa_{1,\mathrm{WF}}(N) \simeq N\kappa_{1,\mathrm{Gaussian}}$ to leading order in N^{-1}. This theoretical prediction is only applicable to κ_1, thus leaving an intriguing possibility that κ_2 captures non-trivial differences between the Gaussian and Wilson-Fisher fixed points that extend to the large-N limit. In addition, the observed proximity of $\kappa_{2,\mathrm{WF}}$ to $\kappa_{2,\mathrm{Gaussian}}$ for the Ising ($N = 1$) case is similar to the situation encountered for the universal coefficient of the logarithmic scaling term that arises due to a corner in an entangling boundary in

2+1 dimensions. There, the interacting value is numerically very close to the free Gaussian theory. This behaviour changes for $N > 1$, where extensive calculations show the universal coefficient increasing with N. An interesting avenue for future work would be to examine if κ_2 obeys a similar trend.

For the second Rényi entropy of the free field, we extract the value $\kappa_{2,\text{Gaussian}} = 0.0227558$, which is accurate to within 0.2% of the continuum value. After the extrapolation to the thermodynamic limit, our best estimate for the value of this universal coefficient at the interacting (Wilson-Fisher) fixed point is $\kappa_{2,\text{WF}} = 0.0174(5)$. In total, extracting this value used approximately 300 years of CPU time.

Chapter 5

Generative machine learning

Generative learning is a subfield of Machine Learning (ML). The task of generative modelling is to construct a probabilistic model for the "data". The later is a generic concept with the common premise of an underling probability distribution $p_{data}(\mathbf{v})$. In practice, we might interact with this distribution via various interfaces. For instance, the "data" might be a set of generated experimentally samples $V = \{\mathbf{v}^i\}_{i=1}^{N_s}$. In this case, the unknown $p_{data}(\mathbf{v})$ is approximated as the empirical distribution of the observed samples[1]

$$p_{data}(\mathbf{v}) \approx \frac{1}{N_s}\mathbb{1}_V(\mathbf{v}). \tag{5.1}$$

In the setting of statistical mechanics, the "data" might refer to the Gibbs state of a known but intractable Hamiltonian. In this case, the "data" refers to the unnormalized Gibbs probability $p^*_{data}(\mathbf{v}) = e^{-\beta E(\mathbf{v})}$ while the samples of this model might be not available.

In order to model the data distribution, we employ a parametric class of models $p_\theta(\mathbf{v})$ known as the restricted Boltzmann machine (RBM). The RBM played a crucial role in the early stages of *Deep* Machine Learning, leading to the first breakthroughs in the training of deep neural networks [97]. It was used both in supervised learning of feed-forward networks where it was employed in an unsupervised pre-training step, and as a fundamental constituent to the first successful methods in deep unsupervised learning as a building block of *deep belief networks* and initializer of deep Boltzmann machines. Enjoying the

[1] $\mathbb{1}_V(\mathbf{v})$ is the indicator function having the value 1 for elements v which belong to the set V and being zero otherwise. In adopting this notation we implicitly assume that there are no elements in V appearing more then once for the sake of simplicity. If this condition is not satisfied, the value of an element needs to be multiplied by its multiplicity.

initial success in the field of Machine Learning, its mainstream usage has been phased out due to its shallow architecture and the inability to scale up RBM-based deep architectures. Instead, new methods that are naturally compatible with deep neural networks have become prominent. However, being an instance of a Markov random field network, a more general ML concept, the RBM still has its niche in modern developments, for instance as a structural multimodal prior in variational autoencoders [98]. Furthermore, promising hardware approaches such as digital [99] and quantum adiabatic [100] annealers have the potential to do to the sampling efficiency, which is the stumbling stone of RBMs, what graphical processing units have done to the linear algebra computation.

Recently, RBMs have seen a newly-found interest in applications to physics problems [21, 22, 101]. This enthusiasm can be explained by several considerations. First of all, the language describing the RBM as an algorithm is the language of statistical physics and so it is easily internalized by a physicist. Secondly, the ability to control the expressiveness of the RBM-based variational ansatz through the number of hidden units is reminiscent of the *bond dimension* parameter in tensor networks which have proved to be a ground-breaking physics-inspired variational ansatz. This connection has been greatly explored and led to new insights [102, 103]. Lastly, its discrete nature is natural for the representation of the quantum wavefunction. However, while the RBM has facilitated the import of ML tools into condensed matter, the history repeats itself and the future of the RBM in physics is uncertain.

In this chapter, we take a data-driven approach to learning a wavefunction [50]. Towards this goal, we develop the classical framework of generative learning and apply it for the reconstruction of the ground state of a transverse field Ising chain. As a verification procedure, we extract the entanglement entropy from the RBM representation relying on the methods developed in Sec. 3.3. Successfully benchmarking the results against known values provides a validation of this process for experimental design. In addition, we consider a novel quantum generalization of the RBM known as the quantum Boltzmann machine (QBM) [104], for which we derive an exact training algorithm and its quantum Monte Carlo implementation in Sec. 2.3.5. The exact methodology allows us to explore the expressiveness of the quantum distribution with respect to its classical counterpart. However, this procedure is not tractable at larger scale. For this reason, we derive a scalable semi-classical algorithm that can be used in practical applications. We conduct numerical experiments attesting to the QBM's strengths and weaknesses in various applications. Lastly, we explore the RBM as a conceptual tool that can reformulate established methods in a new language with the potential to lead to untapped generalizations. As an example, we provide an RBM-based derivation of the Swendsen-Wang cluster update in Sec. 2.2.2.

5.1 The loss function

Once the class of probability models $p_\theta(\mathbf{v})$ is chosen, the modeling consists in the optimization procedure of the parameter vector θ with respect to some measure designed to capture the resemblance between the model and the "data", known as the *loss function*. The measure that is overwhelmingly used in Machine Learning applications is the Kullback-Leibler (KL) divergence which measures the similarity between two distributions $p(\mathbf{v})$ and $q(\mathbf{v})$,

$$D_{KL}\left(p||q\right) = \sum_{\mathbf{v}\in V} p(\mathbf{v}) \ln \frac{p(\mathbf{v})}{q(\mathbf{v})} \tag{5.2}$$

While this quantity is symmetric in the regime of infinitesimal deviations of p from q where it is equal to the Fisher information, in general, $D_{KL}(p \parallel q) \neq D_{KL}(q \parallel p)$ and, therefore, is not strictly a metric in the geometrical sense. Regardless, it possesses a range of desirable qualities among which are non-negativity and invariance under the reparameterization. Those imply that the optimization procedure with respect to θ is well defined.

Being an asymmetric quantity, the question arises in what mode should the KL-divergence be used to set up the optimization problem. Depending on the context, either formulation might be more suitable. However, when only the empirical distribution of the data as in Eq. (5.1) is known, the only choice falls onto the *forward* KL-divergence:

$$\begin{aligned} -\mathcal{L}_\theta \equiv D_{KL}(p_{data} \parallel p_\theta) + \ln N_s &= -\frac{1}{N_s} \sum_{\mathbf{v}\in V} \ln p_\theta(\mathbf{v}) \\ &= -\frac{1}{N_s} \ln \prod_{\mathbf{v}\in V} p_\theta(\mathbf{v}) \\ &= -\frac{1}{N_s} \ln p_\theta(\mathbf{v}^1, \ldots, \mathbf{v}^{N_s}) \end{aligned} \tag{5.3}$$

In this setup, the forward KL-divergence is equivalent to the negative log-likelihood $\mathcal{L}_\theta$ up to an irrelevant for the optimization purposes constant which corresponds to the data empirical entropy. The interpretation is that the best model is the one with the highest chance to produce the observed set of samples.

The second possibility is to have the knowledge of the data model up to a partition function Z_{data} but its sampling is unfeasible. This is a typical situation rising in statistical physics where the classical target Hamiltonian E_{data} and the inverse temperature $\beta = 1/T$

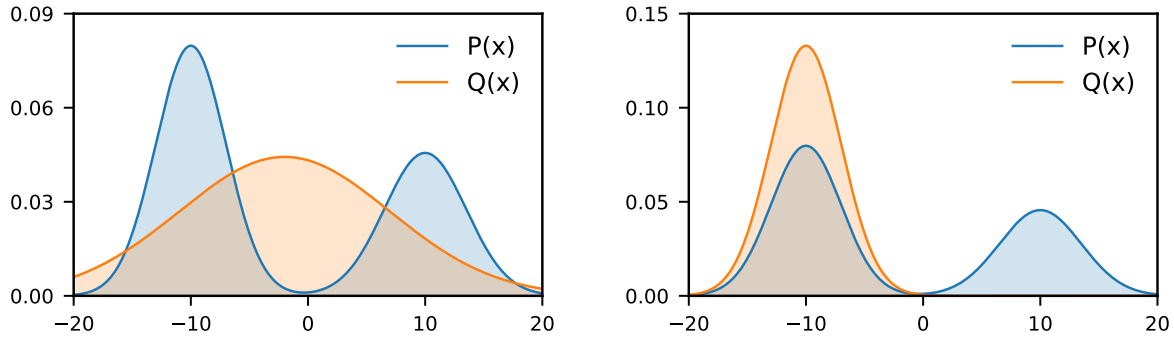


Figure 5.1: Training result of a uni-modal model on a bi-modal data set with the forward (left) and reverse (right) KL-divergence.

are known. This leads us to the *reverse* KL-divergence,

$$D_{KL}(p_\theta \parallel p_{data}) = \frac{1}{T}\left(\mathop{\mathbb{E}}_{\mathbf{v}\sim p_\theta}\left[E_{data}(\mathbf{v})\right] - TS_\theta + T\ln Z_{data}\right)$$
$$\propto F_\theta(E_{data}) - F_{data},$$

where S is the entropy, $F = -T\ln Z$ is the free energy and $F_\theta(E_{data}) \equiv \mathbb{E}_{\mathbf{v}\sim p_\theta}\left[E_{data}(\mathbf{v})\right] - TS_\theta$ is the variational free energy. Remarking that the KL-divergence is a non-negative quantity, we note that the variational free energy is lower bounded by the true free energy $F_\theta(E_{data}) \geq F_{data}$. So minimization of the reverse KL-divergence in this case is equivalent to the well-known *mean field* approach in physics. We also note that while this method provides a clear way how to setup the optimization problem with respect to θ, it cannot quantify the quality of the achieved optimum since the lower bound F_{data} is unknown. For instance, in the case of the 2-d classical Ising model, the mean-field approximation fails to detect its own failure at the critical point.

In order to develop some intuition for the qualitative behaviour for the KL-divergences, let's consider the following toy problem. Assume the data is described by a bi-model Gaussian distribution, and p_θ is a uni-model Gaussian with the mean μ and variance σ^2 being free parameters. The optimization of those parameters with respect to the forward and reverse KL-divergence leads to qualitatively different results as shown on Fig. (5.1). In the forward mode, the variational distribution spreads out in order to cover as much of the data as possible. This is consistent with the intuition of maximizing the likelihood to regenerate the data from the model. However, in the reverse mode, it is putting all its weight on the more pronounced mode while completely ignoring the other one. This *mode*

seeking behavior is consequence of the reverse KL-divergence not penalizing the model for ignoring part of the data as long as $p_\theta = 0$ there. Having limited representational capacity, the model is encouraged to specialize to represent well just part of the data. In the context of the mean-field, this implies that this approach will fail to faithfully represent the critical Ising model containing many distinct modes of variation.

The mode-capturing behaviour inherent to the reverse KL-based learning can be addressed in different ways. One option is to increase the temperature of the variational free energy to encourage the distribution to spread out under the dominance of the entropy term. Additionally, if the samples of the data are available, the forward and inverse KL-divergences can be combined into one loss function. Such a possibility could arise in the context of hardware assisted calculations, for example, in an experimental setup of cold atoms physically emulating a known target Hamiltonian. In this scenario, the samples provide a quick reference for the important state-space that the variational procedure needs to take into account. However, instead of combining the reverse and forward KL-divergence directly, it is better to use the *α-Jensen-Shannon (α-JS) divergence*, given by

$$D^{\alpha}_{JS}(p_\theta \parallel p_{data}) = \alpha D_{KL}(p_\theta \parallel p^{\alpha}_{mix}) + (1-\alpha) D_{KL}(p_{data} \parallel p^{\alpha}_{mix}), \tag{5.4}$$

where $p^{\alpha}_{mix} = \alpha p_\theta + (1-\alpha) p_{data}$ is a weighted mixture of two distributions controlled by the parameter $\alpha \in [0, 1]$. In the limits of $\alpha \to 0, 1$, this expression reduces to the forward and reverse KL-divergence correspondingly. For the equally-weighted mixture $\alpha = 0.5$, α-JS becomes symmetric and constitutes a true metric on the space of probability distributions. In this case, it is known simply as the *JS divergence*.

In summary, we have seen some of the available choices for the loss function in terms of generative learning. Under purely theoretical conditions of uncorrupted and infinite data supply and an infinitely-flexible model distribution, all loss functions considered would be equivalent. However, in practice the dataset is finite and often subject to various sources of noise. Additionally, imposing tractability on the model usually leads to a trade-off in its expressiveness. With those constraints, the optimization of otherwise equivalent loss loss-functions can lead to drastically different results. One must therefore be careful in setting up the optimization problem. In this thesis, we limit ourselves to the case where only the empirical distribution of the data is available, leaving no choice but to use the log-likelihood in Eq. (5.3) as the loss function. As for the family of parametric models p_θ, we focus on *Boltzmann Machines*, which we introduce in the following sections.

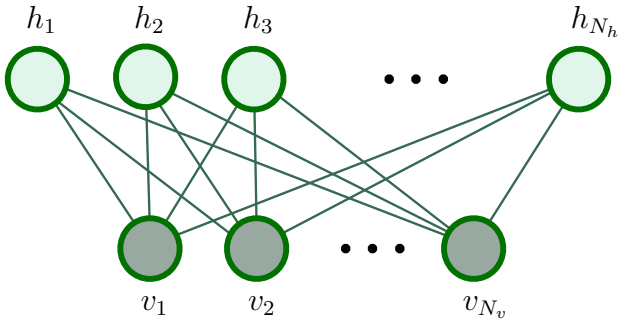


Figure 5.2: The bipartite structure of the Restricted Boltzmann Machine (RBM). The restriction imposes the connectivity to be limited only to the connections between visible, v, and hidden, h, nodes.

5.2 Classical Boltzmann machine

In the large scheme of machine learning models, Boltzmann Machine (BM) belongs to the class of energy-based models [105]. Those models assign, the familiar to a physicist, concept of energy to every configuration $\mathbf{v}$ via the reparameterization of the probability as

$$P_\theta(\mathbf{v}) = e^{-E_\theta(\mathbf{v}) - \ln Z_\theta}. \tag{5.5}$$

where $-\ln Z_\theta$ is added to make sure that the total probability sums up to one. As such, energy-based models constitute an alternative language for probabilistic notions. In practice, they provide a unifying framework for different algorithms designed for generative modelling, inference and discriminative learning [97]. Their usefulness stems from their ability to circumvent the computation of the normalizing constant/partition function Z_θ which is intractable in most applications.

BM is a type of energy-based model with binary degrees of freedom interacting via a two-body weight/interaction matrix W while subjected to an external bias/field represented by a vector b,

$$E_\theta(\mathbf{s}) = -\sum_{i,j} s_i W_{ij} s_j - \sum_i b_i s_i, \tag{5.6}$$

where $s_i \in \{-1, 1\}$ and θ is a collective label for model's parameters (W, b) [2]. Generalizations to higher discretization values [106], continuous variables [107, 38], complex numbers [108] and higher-order interactions exist [109, 38]. While Eq. (5.6) provides the most

[2]Note that the standard ML convention for a binary unit is to take values $\{0, 1\}$. Here, we deviate from the later in order to facilitate the transition to the quantum analog introduced in Sec. 5.3.

general two-body Hamiltonian, it is computationally advantageous to impose a certain structure on the couplings W. Restricted Boltzmann Machine architecture (RBM) results from splitting the degrees of freedom into two non-interacting subsets $\mathbf{s} = (\mathbf{v}, \mathbf{h})$,

$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_{i,j} v_i W_{ij} h_j - \sum_i b_i^v v_i - \sum_j b_j^h h_j. \tag{5.7}$$

Those subsets of variables are referred to as visible $\mathbf{v}$ and hidden $\mathbf{h}$ units correspondingly. Such grouping with the coupling matrix W *restricted* to a bipartite connectivity makes it possible to parallelize the sampling via the *blocked* Gibbs update as explained in Sec. 5.2.2. The visible units are usually associated with the physical degrees of freedom to be modeled over, while the hidden units' primary role is to increase the representation capacity of the marginal over visible units,

$$p_\theta(\mathbf{v}) = \sum_{\mathbf{h}} p_\theta(\mathbf{v}, \mathbf{h}) \tag{5.8}$$

$$= \frac{1}{Z_\theta} e^{\sum_{i=1}^{N_v} b_i^v v_i} \prod_{j=1}^{N_h} \left(1 + e^{b_j^h + \sum_{i=1}^{N_v} W_{ij} v_i}\right) \tag{5.9}$$

In the limit of infinite number of hidden units, RBM becomes a universal approximator with the marginal over visibles $P_\theta(\mathbf{v})$ capable to represent any discrete distribution: $P_\theta(\mathbf{v}) \overset{N_h \to \infty}{\to} P_{data}(\mathbf{v})$ [110].

The value of hidden units goes beyond their capacity to extend RBM's representational power. Useful insights can be gained by considering the emergent properties of the hidden state distribution. In fact, the marginal $P_\theta(\mathbf{h})$ defines an effective Hamiltonian over hidden variables via Eq. (5.5). Consequently, an RBM can be implicitly viewed both as a mapping between the visible and hidden spaces and, at the same time, as the transformation between the corresponding effective Hamiltonians. The richness of this transformation is underpinned by the flexibility of the RBM's architecture. For instance, limiting the number of hidden units to be strictly smaller then the number of visibles while imposing an information bottleneck, via, for example, the weight regularization, can lead to a coarse-graining interpretation of the RBM mapping analogous to the real-space renormalization group transformation [111, 112, 113]. Relaxing hidden units from discrete to continuous variables while imposing only a single connection per visible unit can be shown to exactly implement the Hubbard-Stratonovich transformation [114] whose physics textbook example is the mapping from the Ising model to an effective low-energy ϕ^4 field theory. Promoting the interaction term to a three-body coupling in conjunction with spatially localized hidden units can be shown to produce Fortuin-Kasteleyn dual of the Ising model in

the hidden space known as a Random-Cluster model [37, 39]. As an illustrative example, we exploit this mapping in order to provide a generalized derivation of the Swendsen-Wang cluster update for the Ising model in Sec. 2.2.2.

5.2.1 Log-likelihood gradient

The training consists in optimization of model's parameters with the goal to reproduce the "data" as quantified by the chosen loss function described in Sec. (5.1). Here we limit ourselves to an empirical data distribution Eq. (5.1) reducing the possible choices for the loss function to a single one, namely to the negative log-likelihood Eq. (5.3). The minimization is achieved with the help of the gradient descent or, alternatively, in order to reduce the number of occurrences of the treacherous minus sign, with the help of gradient ascent on the *positive* log-likelihood

$$\theta \to \theta + \eta \frac{\partial \mathcal{L}_\theta}{\partial \theta} \tag{5.10}$$

where η is the *learning rate* determining the length of the gradient step. Computation of the gradient for the specific case of RBM leads to

$$\begin{aligned} \frac{\partial \mathcal{L}_\theta}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(\frac{1}{N_s}\sum_{\mathbf{v}\in V} \ln p_\theta(\mathbf{v})\right) \\ &= \frac{1}{N_s}\sum_{\mathbf{v}\in V} \frac{1}{p_\theta(\mathbf{v})} \partial_\theta p_\theta(\mathbf{v})). \end{aligned} \tag{5.11}$$

We now compute separately the derivative of the marginal,

$$\begin{aligned} \frac{\partial p_\theta(\mathbf{v})}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(\frac{\sum_{\mathbf{h}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}{Z_\theta}\right) \\ &= -\frac{\sum_{\mathbf{h}} e^{-E_\theta(\mathbf{v},\mathbf{h})} \partial_\theta E_\theta(\mathbf{v},\mathbf{h})}{Z_\theta} + \frac{\sum_{\mathbf{h}} e^{-E_\theta(\mathbf{v},\mathbf{h})}}{Z_\theta} \frac{\sum_{\mathbf{h},\mathbf{v}} e^{-E_\theta(\mathbf{v},\mathbf{h})} \partial_\theta E_\theta(\mathbf{v},\mathbf{h})}{Z_\theta} \\ &= -p_\theta(\mathbf{v}) \left[\frac{\sum_{\mathbf{h}} e^{-E_\theta(\mathbf{v},\mathbf{h})} \partial_\theta E_\theta(\mathbf{v},\mathbf{h})}{\sum_{\mathbf{h}} e^{-E_\theta(\mathbf{v},\mathbf{h})}} - \frac{\sum_{\mathbf{h},\mathbf{v}} e^{-E_\theta(\mathbf{v},\mathbf{h})} \partial_\theta E_\theta(\mathbf{v},\mathbf{h})}{Z_\theta}\right] \\ &\equiv -p_\theta(\mathbf{v}) \left[\langle \partial_\theta E_\theta(\mathbf{v},\mathbf{h})\rangle_{\mathbf{v}} - \langle \partial_\theta E_\theta(\mathbf{v},\mathbf{h})\rangle\right], \end{aligned} \tag{5.12}$$

$$\tag{5.13}$$

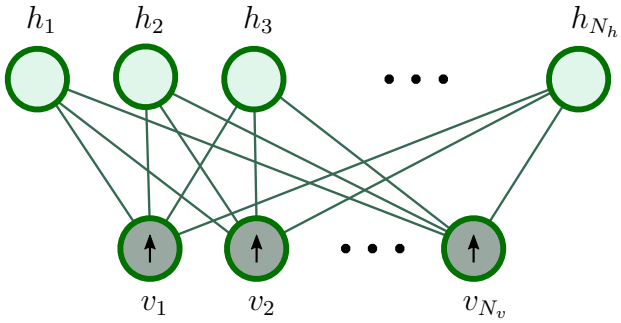


Figure 5.3: Restricted Boltzmann Machine in the clamped ensemble, $\langle\cdot\rangle_{\mathbf{v}}$, where the visible nodes are set to a particular configuration $\mathbf{v}$ and only hidden nodes are allowed to fluctuate stochastically.

where in the last equation we have defined $\langle\cdot\rangle_{\mathbf{v}}$ to be the expectation value in the so called *clamped* ensemble with the visible units strictly set to the vector $\mathbf{v}$ and only hidden units fluctuating according to the conditional $P(\mathbf{h} \mid \mathbf{v})$. This ensemble is visualized in Fig.5.3. The *unclamped* ensemble $\langle\cdot\rangle$ corresponds to the traditional expectation value over both $\mathbf{v}, \mathbf{h}$ which was previously depicted in Fig. 5.2.

Plugging the result of Eq. (5.12) in the expression for the derivative of the log-likelihood from Eq. (5.11) yields

$$\begin{aligned} \frac{\partial \mathcal{L}_\theta}{\partial \theta} &= -\frac{1}{N_s} \sum_{\mathbf{v} \in V} \langle \partial_\theta E_\theta(\mathbf{v}, \mathbf{h}) \rangle_{\mathbf{v}} + \langle \partial_\theta E_\theta(\mathbf{v}, \mathbf{h}) \rangle \\ &\equiv - \langle \partial_\theta E_\theta(\mathbf{v}, \mathbf{h}) \rangle_{\mathbf{v} \sim V} + \langle \partial_\theta E_\theta(\mathbf{v}, \mathbf{h}) \rangle \end{aligned} \tag{5.14}$$

where we have rendered the notation even more compact by defining the average of the clamped ensembles $\langle\cdot\rangle_{\mathbf{v}\sim V}$. Beyond compactness, this notation is consistent with the training procedure where, for the sake of efficiency, the full gradient over the whole batch V is approximated with an unbiased estimator which produces the gradient estimate with respect to a randomly sampled mini-batch of visible vectors. This expression can now be specialized to the parameters of the RBM as

$$\begin{aligned} \frac{\partial \mathcal{L}_\theta}{\partial W_{mn}} &= \langle v_m h_n \rangle_{\mathbf{v} \sim V} - \langle v_m h_n \rangle \\ \frac{\partial \mathcal{L}_\theta}{\partial b^v_m} &= \langle v_m \rangle_{\mathbf{v} \sim V} - \langle v_m \rangle \\ \frac{\partial \mathcal{L}_\theta}{\partial b^h_n} &= \langle h_n \rangle_{\mathbf{v} \sim V} - \langle h_n \rangle \end{aligned} \tag{5.15}$$

The the positive/negative expectations are also known as *positive/negative* phases.

5.2.2 Blocked Gibbs sampling

Blocked Gibbs sampling is a rejection-free Monte-Carlo method similar to the single-spin flip heat-bath algorithm with the advantage that, thanks to the restricted connectivity matrix, all units in a group can be updated at once. It works by iterative exact sampling of the conditionals $p(\mathbf{v} \mid \mathbf{h})$ and $p(\mathbf{h} \mid \mathbf{v})$. For the RBM, both of the conditionals adopt a factorized form,

$$\begin{aligned} p(\mathbf{v} \mid \mathbf{h}) &= \prod_{i=1}^{N_v} p(v_i|\mathbf{h}) \\ p(\mathbf{h} \mid \mathbf{v}) &= \prod_{j=1}^{N_h} p(h_j|\mathbf{v}), \end{aligned} \tag{5.16}$$

with the individual spin probability to be on is given by

$$\begin{aligned} p(v_i = 1 \mid \mathbf{h}) &= \sigma\left(b_i^v + \sum_{j=1}^{N_h} W_{ij} h_j\right) \\ p(h_j = 1 \mid \mathbf{v}) &= \sigma\left(b_j^h + \sum_{i=1}^{N_v} W_{ij} v_i\right), \end{aligned} \tag{5.17}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function.

5.2.3 Training

With the expression for the gradient Sec. (5.2.1) and the sampling procedure Sec. (5.2.2) in place, the training procedure is now defined. However, while the clamped expectation values can be evaluated exactly, the sampling of the unclamped ensemble presents a considerable challenge. This problem accentuates when the RBM's Hamiltonian has a rough potential landscape, a typical scenario in ML. In this case, the Markov chain generated by the blocked Gibbs sampling can get stuck in local minima. As a result, its mixing time can exceed the budgeted computational resources leading to a biased gradient estimate over the highly-correlated samples. Furthermore, the presence of this problem can go undetected since there is no foolproof indicators to test for the thermalization of the Markov chain.

One way to reduce the glassiness of the energy landscape is to impose a regularization on the weights. This is achieved by adding an additional term to the loss function penalizing the model for the parameters with large magnitude. The typical choices are l^1 and l^2 regularization which equate to the sum over all the parameters $\sum_i |\theta_i|$ and $\sum_i \theta_i^2$ correspondingly. l^1 regularization encourages sparse weights due to a constant gradient while l^2 regularization produces non-sparse weights due to the vanishing gradient around 0.

While a regularization can mitigate the poor mixing problem in the unclamped ensemble to a certain degree, it does not provide a complete solution. In practice, a range of methods designed to facilitate inter-modal Markov chain hopping have been introduced. Those include *parallel tempering* [115] and *fast-weights* RBM [116]. However, as noted in [117], even the unbiased sampling is a not a panacea. Indeed, an inherent variance of a statistical estimator can produce an unwanted effect on the stochastic gradient. By looking at Eq. (5.15), we note that the gradient flattens out when the clamped and unclamped estimates cancel out exactly. Thus, the behaviour of a gradient estimator with a high-variance close to a convergence would be dominated by a low signal-to-noise ratio. The net effect is an entropic force that repels the gradient from high to low variance regions in parameter space. As the RBM is designed to capture highly-modal distributions with a high variability between modes, the variance of the gradient estimator poses a big problem.

A partial solution to both, the biased sampling and high variance gradient, is conveniently endowed with additional benefits. Indeed, the *n-step contrastive divergence* (CD_n) [117] and, its derivative, *persistent contrastive divergence* [118] provide a computational advantage and an ease of implementation as well. CD_n operates on a modified loss-function which builds upon the Markov process by which the gradient optimization is achieved. In order to define it, we introduce a sequence of probability distributions $p^n(\mathbf{v})$ labelled by an integer number n. For $n = 0$, $p^0_\theta(\mathbf{v}) \equiv p_{data}(\mathbf{v})$ while the hidden variables are activated via one blocked Gibbs update. Probability distributions with a larger index n are induced by applying a full blocked Gibbs update Eq. (5.17) (composed of two blocked Gibbs updates: one updating hidden units and then one for the visibles), n times after that. Due to the convergence property of a Markov chain, the evolving distribution asymptotically approaches the thermal state of the model $p^\infty_\theta(\mathbf{v}) = p_\theta(\mathbf{v})$. The contrastive divergence is defined as

$$CD^n\left(p^0_\theta(\mathbf{v}) \mid p^n_\theta(\mathbf{v})\right) = D_{KL}(p^0_\theta(\mathbf{v}) \parallel p^\infty_\theta(\mathbf{v})) - D_{KL}(p^n_\theta(\mathbf{v}) \parallel p^\infty_\theta(\mathbf{v})). \tag{5.18}$$

The idea behind this loss function is based on two observations. First, the data distribution is the objective thermal distribution of the model. At the same time, the thermal state is invariant under a Markov chain operator by definition. Therefore, the data distribution

should be invariant under n blocked Gibbs updates upon the model's convergence. In this regard, the contrastive divergence loss-function is designed to counteract the opposite process. Namely, CD_n penalizes the tendency of the Markov chain to drift away from the data distribution. Additionally, it can be seen that this objective is well-defined. Indeed, $p^n_\theta(\mathbf{v})$ is necessarily closer to $p^\infty_\theta(\mathbf{v})$ then $p^0_\theta(\mathbf{v})$, and, therefore, CD_n cannot be negative. Only for the desired outcome, $p^0_\theta(\mathbf{v}) = p^n_\theta(\mathbf{v})$, it equals to zero.

The discussion above stipulates an obvious question. Why do not we minimize directly $D_{KL}\left(p^0_\theta(\mathbf{v}) \mid p^n_\theta(\mathbf{v})\right)$ to obtain the desired outcome? The answer lies in the intractability of $p^n_\theta(\mathbf{v})$ even for $n = 1$. CD_n alleviates this problem by approximating the true distribution with an empirical one obtained through the blocked Gibbs sampling. With this approximation, the gradient of Eq. (5.18) reduces to the difference of the gradients of two log-likelihood functions. Using the expression for the latter obtained in Eq. 5.15, we derive

$$\begin{aligned}
\frac{\partial \mathrm{CD}^n_\theta}{\partial W_{mn}} &= \langle v_m h_n \rangle_0 - \langle v_m h_n \rangle_n \\
\frac{\partial \mathrm{CD}^n_\theta}{\partial b^v_m} &= \langle v_m \rangle_0 - \langle v_m \rangle_n \\
\frac{\partial \mathrm{CD}^n_\theta}{\partial b^h_n} &= \langle h_n \rangle_0 - \langle h_n \rangle_n
\end{aligned} \tag{5.19}$$

where we have introduced a reduced notation $\langle \cdot \rangle_n = \langle \cdot \rangle_{\mathbf{v},\mathbf{h}\sim p^n_\theta}$ describing a sampled distribution with respect to n full blocked Gibbs updates as described previously. Note that the intractable expectation $\langle \cdot \rangle_\infty$ has been cancelled out between the two gradient terms. By changing n, one can adjust between the minimization of the contrastive divergence or log-likelihood. In practice $n = 1$ works well. However, the empirical approximation upon which CD_n is based upon can be shown to introduce a *slight* bias between the fixed points of the contrastive-divergence-based and the log-likelihood-based optimization. To remove the bias, one can increase n to a large number once CD_n has converged, resulting in a much faster, unbiased algorithm [119] outlined below. For practical tricks used to run it

refer to [120].

Algorithm 1: Stochastic gradient descent for the CD_n-based optimization of an RBM.

load data $\{\mathbf{d}\}_{data}$
random init $\theta = (\mathbf{W}, \mathbf{b^v}, \mathbf{v^h})$
set $k = 1$
repeat
  empty $\{\mathbf{v_0}\}$, $\{\mathbf{h_0}\}$, $\{\mathbf{v}\}$, $\{\mathbf{h}\}$
  $\{\mathbf{d}\}_{batch} = \{$sample N_m items from $\{\mathbf{d}\}_{data}\}$
  for $\mathbf{d}$ *in* $\{\mathbf{d}\}_{batch}$ **do**
    set $\mathbf{v_0} = \mathbf{d}$
    sample $\mathbf{h_0} \sim p_\theta(\mathbf{h} \mid \mathbf{v_0})$
    set $\mathbf{v} = \mathbf{v_0}$, $\mathbf{h} = \mathbf{h_0}$
    for *step=1 to* n **do**
      sample $\mathbf{v} \sim p_\theta(\mathbf{v} \mid \mathbf{h})$
      sample $\mathbf{h} \sim p_\theta(\mathbf{h} \mid \mathbf{v})$
    add $\mathbf{v_0}$, $\mathbf{h_0}$, $\mathbf{v}$, $\mathbf{h}$, to $\{\mathbf{v_0}\}$, $\{\mathbf{h_0}\}$, $\{\mathbf{v}\}$, $\{\mathbf{h}\}$
  compute $\frac{\partial \mathrm{CD}_\theta^n}{\partial \theta}$ according to Eq. 5.19
  normalize $\frac{\partial \mathrm{CD}_\theta^n}{\partial \theta} = \frac{\partial \mathrm{CD}_\theta^n}{\partial \theta} / \left\| \frac{\partial \mathrm{CD}_\theta^n}{\partial \theta} \right\|$
  set $\Delta\theta = \frac{1}{k} \frac{\partial \mathrm{CD}_\theta^n}{\partial \theta}$
  set $\theta = \theta - \Delta\theta$
  set $k = k + 1$
until $\Delta\theta \leq \epsilon$

5.2.4 Many-body wavefunction reconstruction

As an application of an RBM to many-body physics, we consider the problem of the ground state wavefunction reconstruction. Any such wavefunction $|\psi\rangle$ can be written in terms of its expansion in a complete basis set $\{|\mathbf{v}\rangle\}$

$$|\psi\rangle = \sum_{\mathbf{v}} \psi(\mathbf{v}) \, |\mathbf{v}\rangle \, . \tag{5.20}$$

The Born rule relates the amplitude $\psi(\mathbf{v})$ to the probability of the wavefunction collapse into a particular state $|\mathbf{v}\rangle$ upon a projective measurement,

$$p(\mathbf{v}) = |\psi(\mathbf{v})|^2. \tag{5.21}$$

A repeated application of those measurements can be used to generate a dataset $\{\mathbf{v}\}$. This dataset can be naturally used to evaluate physical observables which are diagonal in $|\mathbf{v}\rangle$. However, the evaluation of off-diagonal observables requires a repetition of the above procedure from scratch in the appropriate basis. Not only this approach is time-consuming and introduces complications to the experimental design, but some quantities of interest such as the EE requires a dedicated experimental setup [45]. This limitation motivates us to take an alternative route. Instead we attempt to reconstruct the underlying state $\psi(\mathbf{v})$ from $\{\mathbf{v}\}$ with a generative model. Once captured in a compact representation, its knowledge can be exploited to measure any observable.

Formally, we make the following connection between the targeted amplitude $\psi(\mathbf{v})$ and the probabilistic model $p_\theta(\mathbf{v})$

$$\psi(\mathbf{v}) = \sqrt{p_\theta(\mathbf{v})}. \tag{5.22}$$

In this representation, we implicitly assume that the wavefunction has no phase structure and is completely described by a positive number in the selected basis. Such assumption is well-founded for ground states of stoquastic Hamiltonians (the class of Hamiltonians amenable for quantum Monte-Carlo simulations). When no such assumption can be made, multiple basis measurements are required to capture the phase structure. Such generalization is an ongoing area of research [121].

As a practical demonstration, we adopt an RBM with 20 hidden units as the generative model distribution specified in Eq. (5.8). The dataset, $\{\mathbf{v}\}$, is composed of S_z measurements of the transverse field Ising model ground state with 10 sites arranged in a one-dimensional chain and tuned to criticality. This Hamiltonian is stoquastic. The training is no different from the standard machine learning procedure described in detail in

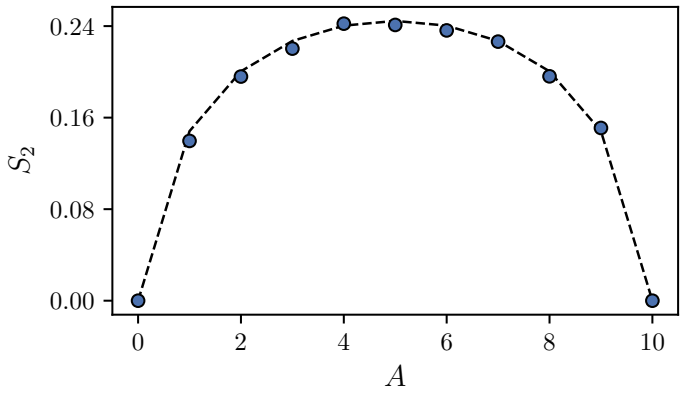


Figure 5.4: The second Rényi entropy for the transverse field Ising chain with $N = 10$ spins. The number of sites in the entangled bipartition A is indicated by the horizontal axis. The markers indicate values obtained through the "Swap" operator (see Sec. 3.3) applied to the samples from a trained RBM. The dashed line denotes the result from exact diagonalization.

Sec. 5.2.3. We employ the dedicated software package QUCUMBER (Quantum Calculator Used for Many-body Eigenstate Reconstruction) for this purpose [50].

Once RBM is trained, off-diagonal observables can be computed via standard variational Monte-Carlo techniques [50]. The samples required for this evolution can be either reused from $\{\mathbf{v}\}$ or generated by Gibbs sampling based on Eq. (5.17). We choose to benchmark the trained RBM on the second Rényi entropy, S_2, against values obtained with the exact diagonalization. Monte-Carlo estimation of the Rényi entropy require the usage of so called replica trick and is covered in Sec. 3.3. The results are shown in Fig. 5.4. The trained RBM correctly captures well the value of the second Rényi entropy for different subregion A sizes. Being a basis-independent observable, this constitutes a good test on the ability of the RBM to capture the full wave-function from the information contained in a single-basis dataset for transverse field Ising chain.

5.3 Quantum Boltzmann machine

Quantum Boltzmann Machine (QBM) [104] provides a generalization to the classical counterpart described in Sec. (5.2). The restricted versions of those two architectures are pictorially contrasted against each other in Fig. 5.5. In QBM, the configuration space is promoted from the state space of a classical binary vector to the state space of wave functions defined over spin one-half Hilbert space. Formally, this is done by replacing the spin variables v, h in the classical Hamiltonian Eq.(5.7) with σ^z Pauli matrices. Additionally, in order to create quantum fluctuations, a transverse field is introduced through σ^x term leading to the Hamiltonian,

$$-H_\theta = \sum_{v,h} \sigma_v^z W_{vh} \sigma_h^z + \sum_s \Gamma_s \sigma_s^x + \sum_s b_s \sigma_s^z, \tag{5.23}$$

which can be recognized as a transverse field Ising model (TFIM). Note how the notation differs from Eq.(5.7). From now on, $s = (v, h)$ is a collective *index* variable referring to both types of indices, the ones of visibles units, v, as well as of the hidden ones h.[3] In addition to the restricted architecture above, we will also be interested in a semi-restricted QBM

$$-H_\theta = \sum_{v,v} \sigma_v^z W_{vv} \sigma_v^z + \sum_{v,h} \sigma_v^z W_{vh} \sigma_h^z + \sum_s \Gamma_s \sigma_s^x + \sum_s b_s \sigma_s^z, \tag{5.24}$$

with supplementary all-to-all interactions between the visible qubits parameterized via matrix W_{vv}. Without hidden units, the semi-restricted QBM becomes a fully-connected fully-visible model. Those extensions to the restricted architecture of the classical BM come with a new training procedure which relies on a different sampling paradigm. Instead of the blocked Gibbs sampling introduced in the Sec. 5.2.2 for the RBM training, QBM relies on the sampling from a quantum hardware.

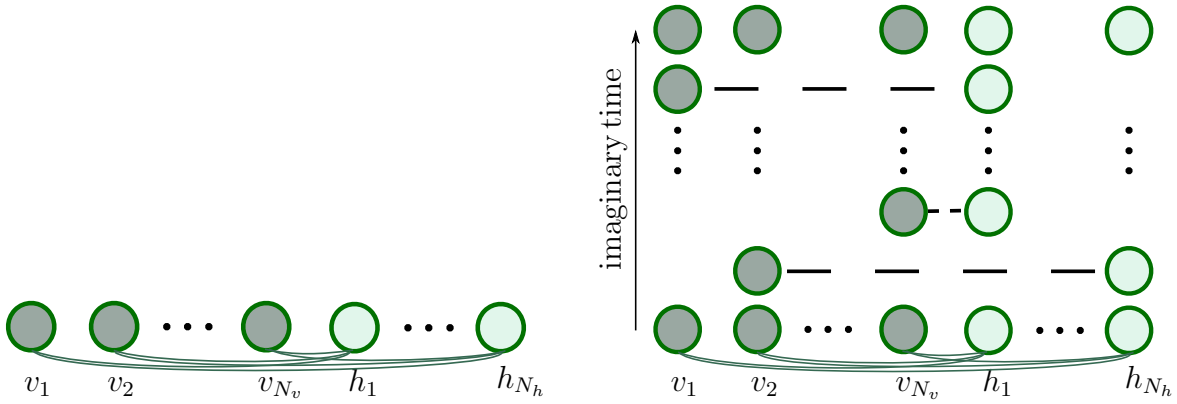
At a finite temperature $T = 1$, the statistical properties of QBM correspond to a Gibbs mixed state as described by the density matrix

$$\rho_\theta = e^{-H_\theta}. \tag{5.25}$$

In order to consider the density matrix as a generative model for an *empirical classical data*, the quantum nature of the underlying representation is collapsed with a projective measurement. The generative model is then determined by the Born rule which reads

$$p_\theta(\mathbf{v}) = \frac{\mathrm{Tr}\left[\rho_\theta M_{\mathbf{v}}\right]}{\mathrm{Tr}\left[\rho_\theta\right]}. \tag{5.26}$$

[3] We maintain the notation for a data vector $\mathbf{v}$. Its elements now are referred as $\mathbf{v}_v$.



(a) RBM (b) QBM

Figure 5.5: (a) A flattened out version of the classical restricted Boltzmann machine (RBM), conventionally displayed as in Fig. 5.2. This way of exposition facilitates a direct comparison with a restricted quantum Boltzmann Machine (QBM) on the right. (b) A cartoon depiction of a QBM which also has a restricted connectivity to hidden and visible units. However unlike RBM, QBM can be seen as having an extra dimension, for instance, corresponding to the operator list expansion in the stochastic series representation. This depiction makes manifest the potential advantage of a QBM over an RBM steaming from its extra dimension of flexibility.

Here, $M_{\mathbf{v}} = |\mathbf{v}\rangle\langle\mathbf{v}| \otimes I_{\mathbf{h}}$ is the projective measurement operator onto a visible vector $\mathbf{v}$ and $I_{\mathbf{h}}$ is the identity operator acting over the space of hidden units. The usefulness of studying such an extension can be motivated on different levels. In the first scenario, we consider the hidden variables removed and assume that the dataset originates from spin measurements of a quantum Hamiltonian with unknown parameters. In this case, the likelihood based training of the QBM is equivalent to the *inverse Ising problem* aiming to determine those unknown parameters. More specifically, if the unknown Hamiltonian is the TFIM, then the QBM can be seen as an algorithm to solve the inverse problem. More generally, the methodology developed further can be a useful starting point for the inverse problem of any Hamiltonian including the ones with the *sign* problem through a careful choice of the data measurements basis. Second, with the hidden units reinstated, the QBM training can be seen as a protocol for the quantum mixed state reconstruction with the benefit of a direct control over the model's expressiveness through the adjustment of the hidden unit number [22].

Leaving the realm of data generated via a quantum process, it is interesting to entertain a more speculative thought of quantum nature of classical data. That is, the idea that

purely quantum effects such as entanglement are beneficial to capture correlations present in classical data sets in a succinct way. The implication of such findings would mean that a quantum generative model would require a smaller number of parameters then its classical counterpart. While this idea can seem to be widely speculative, the dominance of the classical model over quantum ones is even less likely to be true. After all, the classical BM can be viewed as a particular instance of the quantum model and, therefore, there is no prior reason for its advantage. We will examine this hypothesis empirically in more detail in Sec. 5.3.4.

Lastly, the QBM might provide a computational edge over the classical BM as a generative model in the context of *quantum adiabatic computation* [122]. In order to explain this point, let us quickly review the core facts about this framework. Quantum adiabatic computation is a universal computing paradigm [123, 124, 125, 126] providing an alternative to the more established quantum circuit model. Unlike the *quantum gate computing* which is based on a wavefunction evolution under the action of discrete applications of local unitary operations called gates, the quantum adiabatic computation relies on a quantum adiabatic evolution [127] of a wavefunction under the action of *slowly* evolving Hamiltonian $H(t)$ as dictated by the Schrodinger's equation. Within thin framework, the Hamiltonian is traditionally decomposed into two non-commuting coevolving terms. The target piece, H_c represents the classical interacting term. It is designed in such a way that its ground state encodes the solution to a problem of interest. The second term is the driver Hamiltonian, H_d, with a simple ground state that can be easily prepared as the starting state. The prototypical example of $H(t)$ is the TFIM-based annealing,

$$-H(t) = [1 - c(t)] \overbrace{\sum_{ij} \sigma_i^z J_{ij} \sigma_j^z}^{H_c} + c(t) \overbrace{\sum_i \sigma_i^x}^{H_d}, \tag{5.27}$$

with the transverse field playing the role of the driver term. The parameter $c(t)$ can be any function interpolating between the boundary constraints, $c(t = 0) = 0$ and $c(t = T) = 1$. This function defines a path in the parameter-space of Hamiltonians undertaken during the physical time evolution $t \in [0, T]$. Initially, $c(t = 0) = 0$, only the driver term is present. The system is set to its ground state. According to the *quantum adiabatic theorem* [127] , if the evolution is slow enough (as defined below), the system will remain in its instantaneous ground state during the entire evolution path. Consequently, as the final Hamiltonian at $c(t = T) = 1$ is purely classical, the final state solves for the ground state of the interacting Hamiltonian. The limitation of this approach lies in the constraint on the evolution speed.

Namely [127],

$$T \gg \frac{\mathcal{E}_{max}}{\Delta_{min}}, \tag{5.28}$$

$$\text{where} \qquad \mathcal{E}_{max} \equiv \max_{0\leq t\leq T} \left\| \frac{\mathrm{d}H(t)}{\mathrm{d}t} \right\|,$$

$$\Delta_{min} \equiv \max_{0\leq t\leq T} \Delta(t).$$

with $\Delta(t)$ being the instantaneous energy gap between the ground state and the first excited state of $H(t)$. The nominator is not expected to grow faster then polynomially for a local Hamiltonian [122]. Therefore, the scaling efficiency of the quantum adiabatic computation is entirely determined by the minimal encountered gap, Δ_{min}, during the evolution.

Generalizations of the quantum adiabatic theorem to open systems also exist [128, 129]. In this context, the system of interest is described by a mixed state whose dynamics are coupled to the environment. Under similar restrictions on the evolution time T [128, 129], the quantum adiabatic evolution thus provides a tool for an efficient sampling of the quantum Gibbs state. In practice, it is common for the system to undergo phase transitions with the enclosing gap, $\Delta_{min} \to 0$, thus nullifying the theoretical guarantees of the adiabatic theorem. When this occurs, the relaxation time scale greatly exceeds the available time T, which leads for the system's dynamics to *freeze-out* [130] before reaching the target state. This fact brings us back to the consideration of QBM's advantage over BM. As both those models require sampling for the training, the quantum adiabatic annealing is naturally fit to help for this purpose and is applicable to both models. However, the classical model would require a longer adiabatic evolution since it lies further away in the parameter space from the $t = 0$ state. For this reason, it is more likely that for a given evolution time T, the quantum adiabatic evolution will sample the QBM better than the BM [104].

5.3.1 Exact log-likelihood gradient

With the new ansatz for the generative model, Eq. (5.26), we proceed with the derivation of the log-likelihood-based training procedure. The steps closely follow the classical RBM procedure of Sec. 5.2.1. The gradient-based optimization procedure requires calculation of the cost function gradient. The likelihood gradient for a single data vector $\mathbf{v}$ is

$$\begin{aligned} \frac{\partial \ln p_\theta(\mathbf{v})}{\partial \theta} &= \frac{\partial}{\partial \theta}\left(\ln \frac{\mathrm{Tr}\left[M_{\mathbf{v}}\rho_\theta\right]}{\mathrm{Tr}\left[\rho_\theta\right]} \right) \\ &= \frac{\mathrm{Tr}\left[M_{\mathbf{v}}\partial_\theta\rho_\theta\right]}{\mathrm{Tr}\left[M_{\mathbf{v}}\rho_\theta\right]} - \frac{\mathrm{Tr}\left[\partial_\theta\rho_\theta\right]}{\mathrm{Tr}\left[\rho_\theta\right]}. \end{aligned} \tag{5.29}$$

In the last expression, the partial derivative $\partial_\theta \rho_\theta$ requires a special treatment due to the non-commutative nature of terms contained in ρ_θ. In order to evaluate both of those terms, we employ the following identity [131]

$$\partial_\theta e^{-H_\theta} = -e^{-H_\theta} \int_0^1 \mathrm{d}\tau e^{\tau H_\theta} \partial_\theta H_\theta e^{-\tau H_\theta}. \tag{5.30}$$

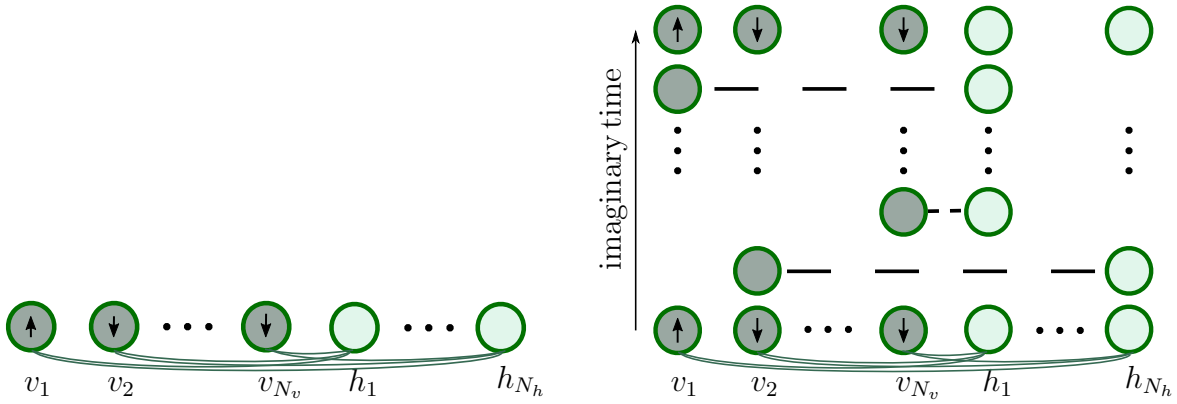
As a sanity check, we note that when $[\partial_\theta H_\theta, H_\theta] = 0$, this equation reduces to the classical expression $-e^{-H_\theta}\partial_\theta H_\theta$. Surprisingly, upon substitution of this identity in Eq. (5.29), a similar expression is obtained for the second term

$$\begin{aligned} \frac{\mathrm{Tr}\left[\partial_\theta \rho_\theta\right]}{\mathrm{Tr}\left[\rho_\theta\right]} &= -\frac{\mathrm{Tr}\left[e^{-H_\theta}\int_0^1 \mathrm{d}\tau e^{\tau H_\theta}\partial_\theta H_\theta e^{-\tau H_\theta}\right]}{\mathrm{Tr}\left[\rho_\theta\right]} \\ &= -\frac{\mathrm{Tr}\left[e^{-H_\theta}\partial_\theta H_\theta\right]}{\mathrm{Tr}\left[\rho_\theta\right]} \\ &= -\left\langle \partial_\theta H_\theta \right\rangle, \end{aligned} \tag{5.31}$$

where in the second line we have used the invariance of the trace under a cyclic permutation in order to move the term $e^{-\tau H_\theta}$ next to its conjugate so that they cancel out. Unfortunately, the non-commutativity of $M_\mathbf{v}$ with H_θ prevents us from obtaining an equivalently simple expression for the first term in Eq. (5.29),

$$\begin{aligned} \frac{\mathrm{Tr}\left[M_\mathbf{v}\partial_\theta \rho_\theta\right]}{\mathrm{Tr}\left[M_\mathbf{v}\rho_\theta\right]} &= -\frac{\mathrm{Tr}\left[M_\mathbf{v} e^{-H_\theta}\int_0^1 \mathrm{d}\tau e^{\tau H_\theta}\partial_\theta H_\theta e^{-\tau H_\theta}\right]}{\mathrm{Tr}\left[M_\mathbf{v}\rho_\theta\right]} \\ &\equiv \left\langle \int_0^1 \mathrm{d}\tau e^{\tau H_\theta}\partial_\theta H_\theta e^{-\tau H_\theta} \right\rangle_\mathbf{v}. \end{aligned} \tag{5.32}$$

Here, $\langle \ldots \rangle_\mathbf{v} = \langle \mathbf{v}| e^{-H_\theta} \ldots |\mathbf{v}\rangle$ defines an expectation with respect to a clamped quantum ensemble. It is visualized in Fig. 5.6 next to its classical counterpart. By contrasting this figure with the unclamped version Fig. 5.5, it becomes evident that the clamped quantum ensemble is less restrictive then the classical one. In the classical case, clamping leads to a full restriction on the state of the visible units, while in the quantum case, the restriction is imposed only on a single slice in the imaginary time. Therefore, while in the classical case, the clamped ensemble effectively represents an ensemble of disentangled hidden spins amenable for an analytic evaluation of expectations, in the quantum case, no such simplification occurs. The problem is aggravated by the ensuing prohibitive computational cost to the scaling as the calculation needs to be performed for every datapoint in the dataset. Unfortunately, the quantum clamped expectation value does not represent a physical process



(a) clamped RBM (b) clamped QBM

Figure 5.6: Comparison of (a) a restricted Boltzmann machine and (b) a restricted quantum Boltzmann machine in the corresponding clamped ensembles $\langle\cdot\rangle_{\mathbf{v}}$. The clamped RBM ensemble is very restrictive as it fully constrains the state of the visible units. Contrarily, the clamping of a QBM only constrains the state of a single slice in its extra dimension. The double occurrence of the clamped slice in (b) is due to the periodic boundary conditions.

that can be implemented on a quantum annealer, thus leaving no prospects for a quantum speedup. On mid-size systems, QMC-based techniques provide the only viable approach. We derive QMC estimators for both Eq. (5.31) and Eq. (5.32) in Sec. 2.3.5. In order to circumvent the computational dead-end on larger systems, we derive a semi-classical approximation.

5.3.2 Semi-classical log-likelihood gradient

The basis of our approximation is the well-known Golden-Thompson inequality,

$$\mathrm{Tr}\left[e^{A}e^{B}\right] \geq \left[e^{A+B}\right], \tag{5.33}$$

where A, B are Hermitian matrices. We apply this bound to the QBM probability $p_\theta(\mathbf{v})$, Eq. (5.26), which we write in a suitable form as

$$p_\theta(\mathbf{v}) = \frac{\mathrm{Tr}\left[M_{\mathbf{v}}e^{-H_\theta}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} = \lim_{\epsilon\to 0}\frac{\mathrm{Tr}\left[e^{\ln(M_{\mathbf{v}}+\epsilon\mathbb{1})}e^{-H_\theta}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]},$$

where ϵ is a small positive parameter coupled to an identity operator introduced to avoid the divergences in the values of $\ln M_{\mathbf{v}}$. The limit trivially follows from the fact that

$M_{\mathbf{v}} = e^{\ln M_{\mathbf{v}}}$. Now, applying the Golden-Thompson inequality to the right hand side with the substitutions $A = \ln(M_{\mathbf{v}} + \epsilon)$ and $B = -H_\theta$ we obtain

$$\frac{\mathrm{Tr}\left[e^{\ln(M_{\mathbf{v}}+\epsilon\mathbb{1})}e^{-H_\theta}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} \geq \frac{\mathrm{Tr}\left[e^{-[H_\theta-\ln(M_{\mathbf{v}}+\epsilon\mathbb{1})]}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]}. \tag{5.34}$$

In the resulting lower bound, the term $\ln(M_{\mathbf{v}} + \epsilon\mathbb{1})$ enters the expression in the same way as the Hamiltonian H_θ does. Therefore, they can be regarded as two separate contributions to the combined Hamiltonian

$$H_{\theta,\mathbf{v}}(\epsilon) \equiv H_\theta - \ln(M_{\mathbf{v}} + \epsilon\mathbb{1}) \tag{5.35}$$

For a descreasing ϵ, the second term introduces a diverging energy penalty to all states that have an overlap with basis states other then $|\mathbf{v}\rangle$. Those states therefore have a negligible contribution to the trace in Eq. (5.34) and in the limit $\epsilon \to 0$, the thermodynamics of $H_{\theta,\mathbf{v}}(\epsilon)$ become indistinguishable from its semi-classical approximation

$$\lim_{\epsilon\to 0} H_{\theta,\mathbf{v}}(\epsilon) = \langle\mathbf{v}| H_\theta |\mathbf{v}\rangle \otimes |\mathbf{v}\rangle\langle\mathbf{v}| \equiv H_{\theta,\mathbf{v}}. \tag{5.36}$$

This expression defines the clamped Hamiltonian $H_{\theta,\mathbf{v}}$. The latter effectively corresponds to the original Hamiltonian H_θ whose visible σ^z operators are replaced with classical spins clamped to $\mathbf{v}$ and the visible σ^x operators are turned off,

$$H_{\theta,\mathbf{v}} = H_\theta(\sigma^x_{\mathbf{v}} = 0, \sigma^z_{\mathbf{v}} = \mathbf{v}). \tag{5.37}$$

With the $\epsilon \to 0$ limits well understood on both sides of Eq. (5.34), we derive the desired lower bound $\widetilde{p}_\theta(\mathbf{v}) \leq p_\theta(\mathbf{v})$,

$$\begin{aligned} p_\theta(\mathbf{v}) \equiv \lim_{\epsilon\to 0} \frac{\mathrm{Tr}\left[e^{\ln(M_{\mathbf{v}}+\epsilon\mathbb{1})}e^{-H_\theta}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} &\geq \lim_{\epsilon\to 0} \frac{\mathrm{Tr}\left[e^{-[H_\theta-\ln(M_{\mathbf{v}}+\epsilon\mathbb{1})]}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} \\ &= \frac{\mathrm{Tr}\left[e^{-H_\theta(\mathbf{v})}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} \\ &\equiv \widetilde{p}_\theta(\mathbf{v}). \end{aligned} \tag{5.38}$$

Taking the gradient of the bound-based log-likelihood, we obtain

$$\begin{aligned} \frac{\partial \ln \widetilde{p}_\theta(\mathbf{v})}{\partial\theta} &= \frac{\mathrm{Tr}\left[\partial_\theta e^{-H_{\theta,\mathbf{v}}}\right]}{\mathrm{Tr}\left[e^{-H_{\theta,\mathbf{v}}}\right]} - \frac{\mathrm{Tr}\left[\partial_\theta e^{-H_\theta}\right]}{\mathrm{Tr}\left[e^{-H_\theta}\right]} \\ &= -\left\langle \partial_\theta H_{\theta,\mathbf{v}} \right\rangle_{\mathbf{v}} + \left\langle \partial_\theta H_\theta \right\rangle, \end{aligned} \tag{5.39}$$

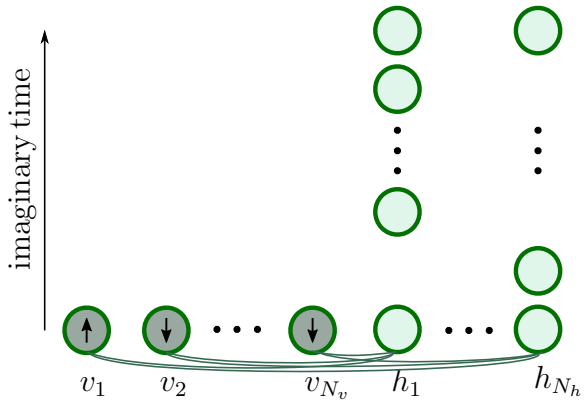


Figure 5.7: The clamped quantum Boltzmann machine in the semi-classical approximation Eq. (5.38). In this approximation, the visible units are effectively classical and are clamped to the vector $\mathbf{v}$ whose state is shown with the arrows. The hidden units are still described by quantum spins.

where $\langle \ldots \rangle_{\mathbf{v}}$ now represents the semi-classical ensemble based on the Boltzmann distribution of the clamped Hamiltonian, $e^{-H_{\theta,\mathbf{v}}}$. This distribution corresponds to a mix of classical and quantum spins. It is schematically represented on Fig. 5.7. Unlike the previous calculation for the exact log-likelihood gradient in Sec. 5.3.1, the expression for the gradient of the clamped term is easily evaluated since $[\partial_\theta H_{\theta,\mathbf{v}}, H_{\theta,\mathbf{v}}] = 0$. For the gradient of the unclamped term, we reused the previously derived result from Eq. (5.31).

5.3.3 Bound-based training

Bound-based approximations to the true training objective are routinely used in machine learning to trade off the accuracy of the model for its computational efficiency. A prominent example is the class of powerful generative models known as variational autoencoders (VAE) [132]. Having encountered a computational impasse in the evaluation of the clamped ensemble contribution to the gradient of the QBM-based log-likelihood, Eq. (5.32), we adopt a similar strategy. Namely, we employ the semi-classical approximation on the likelihood derived in Eq. (5.38). The bound-based training objective $\widetilde{\mathcal{L}}_\theta$ reads

$$-\mathcal{L}_\theta \leq -\widetilde{\mathcal{L}}_\theta = -\frac{1}{N_s}\sum_{\mathbf{v}\in V} \ln \widetilde{p}_\theta(\mathbf{v}) \tag{5.40}$$

The optimization of the parameters θ is still done with the help of the gradient ascent on $\widetilde{\mathcal{L}}_\theta$,

$$\theta \to \theta + \eta \frac{\partial \widetilde{\mathcal{L}}_\theta}{\partial \theta} \tag{5.41}$$

Unlike the gradient of the exact log-likelihood, the gradient of $\widetilde{\mathcal{L}}_\theta$ is amenable for the evaluation on a quantum hardware,

$$\begin{aligned}
\frac{\partial \widetilde{\mathcal{L}}_\theta}{\partial \theta} &= \frac{1}{N_s} \sum_{\mathbf{v} \in V} \frac{\partial \ln \widetilde{p}_\theta(\mathbf{v})}{\partial \theta} \\
&= -\frac{1}{N_s} \sum_{\mathbf{v} \in V} - \left\langle \partial_\theta H_{\theta,\mathbf{v}} \right\rangle_{\mathbf{v}} + \left\langle \partial_\theta H_\theta \right\rangle \\
&\equiv - \left\langle \partial_\theta H_{\theta,\mathbf{v}} \right\rangle_{\mathbf{v} \sim V} + \left\langle \partial_\theta H_\theta \right\rangle .
\end{aligned} \tag{5.42}$$

The last expression is strikingly similar to the log-likelihood gradient of the classical RBM in Eq. (5.3). Despite this apparent similarity, the evaluation of the expectations in the unclamped ensemble are drastically different between the classical and quantum BMs. For its samples, RBM relies on the computational efficiency of the blocked Gibbs update and global exploration techniques such as the contrastive divergence. On the other hand, QBM assumes an efficient quantum hardware implementation. For this reason, the architectural restrictions imposed by the classical sampling considerations can be lifted as far as the unclamped ensemble is concerned.

As for the evaluation of the clamped expectation values above, we require an analytic tractability in order to tolerate the scaling cost to large datasets. Therefore, the clamped ensemble must effectively represent a non-interacting system. Such requirement is satisfied for a semi-restricted QBM defined in Eq. (5.24). Indeed, its semi-classical version, $H_{\theta,\mathbf{v}}$, reduces to a collection of non-interacting qubits

$$\begin{aligned}
-H_{\theta,\mathbf{v}} &= \sum_{v,v'} \mathbf{v}_v W_{vv'} \mathbf{v}_{v'} + \sum_v b_v \mathbf{v}_v + \sum_{v,h} \mathbf{v}_v W_{vh} \sigma_h^z + \sum_h \Gamma_h \sigma_h^x + \sum_h b_h \sigma_h^z \\
&= C_{\mathbf{v}} + \sum_h b_h^{\mathbf{v}} \sigma_h^z + \sum_h \Gamma_h \sigma_h^x
\end{aligned}$$

where the zero energy offset, $C_{\mathbf{v}} = \sum_{v,v'} \mathbf{v}_v W_{vv'} \mathbf{v}_{v'} + \sum_v b_v \mathbf{v}_v$, and the clamped longitudinal bias, $b_h^{\mathbf{v}} = \mathbf{v}_v W_{vh} + b_h$, are dependent on the datavector $\mathbf{v}$. The corresponding distribution

is factorized and is fully characterized by the following expectations

$$\begin{aligned}\langle\sigma_h^z\rangle_{\mathbf{v}} &= \frac{b_h^{\mathbf{v}}}{D_h^{\mathbf{v}}}\tanh\left(D_h^{\mathbf{v}}\right) \\ \langle\sigma_h^x\rangle_{\mathbf{v}} &= \frac{\Gamma_h}{D_h^{\mathbf{v}}}\tanh\left(D_h^{\mathbf{v}}\right),\end{aligned} \tag{5.43}$$

where $D_h^{\mathbf{v}} \equiv \sqrt{\Gamma_h^2 + b_h^{\mathbf{v}2}}$. Writing out the gradient Eq. (5.42) explicitly for each type of parameter θ we obtain

$$\begin{aligned}\frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial W_{vv'}} &= \langle\mathbf{v}_v\mathbf{v}_{v'}\rangle_{\mathbf{v}\sim V} - \langle\sigma_v^z\sigma_{v'}^z\rangle \\ \frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial W_{vh}} &= \langle\mathbf{v}_v\sigma_h^z\rangle_{\mathbf{v}\sim V} - \langle\sigma_v^z\sigma_h^z\rangle \\ \frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial b_v} &= \langle\mathbf{v}_v\rangle_{\mathbf{v}\sim V} - \langle\sigma_v^z\rangle \\ \frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial b_h} &= \langle\sigma_h^z\rangle_{\mathbf{v}\sim V} - \langle\sigma_h^z\rangle \\ \frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial\Gamma_v} &= -\langle\sigma_v^x\rangle \\ \frac{\partial\widetilde{\mathcal{L}}_\theta}{\partial\Gamma_h} &= \langle\sigma_h^x\rangle_{\mathbf{v}\sim V} - \langle\sigma_h^x\rangle.\end{aligned} \tag{5.44}$$

The gradient with respect to Γ_v is self-cancelling. Indeed, a positive value Γ_v leads to a positive expectation $\langle\sigma_v^x\rangle$ which in turn leads to a decrease in Γ_v and vice versa. The fixed point of this process is $\Gamma_v = 0$. Such result seems to indicate that the QBM is naturally driven to its classical point. However, as we will see in the next section, the training based on the exact log-likelihood does not regress to the classical BM. Therefore such fixed point is an artifact of the semi-classical approximation.

So far we have relied on the assumption of an efficient quantum oracle that can produce samples from the unclamped ensemble. Here, we restrict ourselves to a communication protocol where the oracle only outputs measurements in the diagonal basis. Such scenario is natural to quantum annealers. This limitation makes it impossible to measure the off-diagonal expectations required to train Γ_h. For this reason and the reason mentioned in the previous paragraph, in the bound-based QBM training we treat Γ_s as hyperparameters.

5.3.4 Application to classical dataset

In this section, we illustrate the ideas described in the previous sections on a toy example. For the training data, we generate samples from a uniform mixture of M factorized distributions

$$P^{\text{data}}(\mathbf{v}) = \frac{1}{M}\sum_{k=1}^{M} P^k(\mathbf{v}) \tag{5.45}$$

Each mixture component P^k is defined by its mode $\mathbf{m}^k = [\mathbf{m}_1^k, \mathbf{m}_2^k, ..., \mathbf{m}_{N_v}^k]$ and the spin flip probability p^k. We choose to make the later equal across all the components, $p^k = (1-p)$. The mode is picked at random among all possible configurations of the visible spins. The mixture component distribution is then defined as independent spin fluctuations with probability p around its mode $\mathbf{m}^k$

$$P^k(\mathbf{v}) = \prod_{i=1}^{N_v}(1-p)^{\mathbf{v}_i \neq \mathbf{m}_i^k} p^{\mathbf{v}_i = \mathbf{m}_i^k}, \tag{5.46}$$

where we assume the convention 0(1) for the $False(True)$ boolean evaluation of the exponents. In all our examples, we choose $p = 0.9$. The modes are randomly and independently generated for each training set from a uniform distribution. Each dataset $\{\mathbf{v}_i\}^{N_s}$ contains a thousand training examples, $N_s = 1000$. Since we have a direct access to the data probability distribution, we show the performance as evaluated by the KL-divergence in our plots.

We start with a fully-connected architecture without hidden units in order to compare the expressiveness of a classical Boltzmann machine (BM) against quantum Boltzmann machine (QBM) and evaluate the quality of the bound (5.40) by training the semi-classical quantum Boltzmann machine to which we refer as bound-based QBM (bQBM). We consider a model with $N_v = 10$ qubits. In this case, the log-likelihood $\mathcal{L}$ is a convex function of its parameters and, therefore, has a unique global minimum. Classical BM has $N_v(N_v+1)/2$ trainable parameters. To make the comparison fair, we restrict the transverse field parameters of the QBM Hamiltonian to be the same ($\Gamma_s = \Gamma$). Due to the small size, we are able to evaluate all the expectations required for the BM training. The QBM is trained based on the exact log-likelihood using procedure developed in Sec. 5.3.1. This can be done with the help of the exact diagonalization for small systems. The bQBM is trained based on the bound $\widetilde{\mathcal{L}}$, Eq. (5.40), with a fixed Γ to some ad-hoc non-zero value $\Gamma = 2$. Comparing the training results of QBM with bQBM will give us an idea of the tightness of the bound $\widetilde{\mathcal{L}}$.

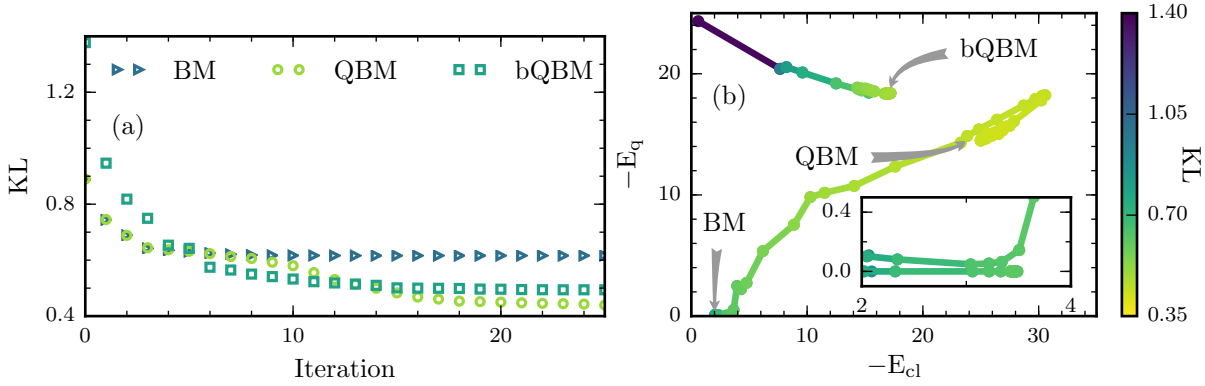


Figure 5.8: Training of a fully visible fully connected model with $N_v = 10$ qubits on artificial data from a mixture model (define in Eq. (5.45)) with the noise-parameter $p = 0.9$ and the number of modes $M = 8$. Training is done using second-order optimization routine BFGS. (a) KL-divergence of BM, QBM, bQBM models during training process. Both QBM and bQBM learn to KL values that are lower than that for BM. (b) Classical and quantum average energies during training process. The inset details the same data set on a finer scale close to the classical regime.

Since all expectations entering the gradients of the cost function are computed exactly, we use the second-order optimization routine BFGS [133]. The results of training BM, QBM and bQBM are given in Fig. 5.8(a). The x-axis in the figure corresponds to iterations of BFGS that does line search along the gradient (the learning rate η is selected automatically). QBM is able to learn the data noticeably better than BM, and bQBM approaches the value close to the one for QBM.

In order to visualize the training process, we keep track of the average values of classical E_{cl} and quantum E_q energy expectations of the Hamiltonian during the training. Fig 5.8(b) shows the learning trajectories in the space (E_{cl}, E_q). BM learns a model with average energy ≈ 3.5, and KL ≈ 0.62. One can see that QBM, which starts off with $\Gamma = 0.1$, initially lowers Γ and learns longitudinal parameters that are close to the best classical result (see the inset). Soon after, QBM increases its quantum as well as classical energies until it converges to a point with $\Gamma = 2.5$ and KL ≈ 0.42, which is better than classical BM value. Having a fixed transverse field, $\Gamma = 2$, bQBM starts with a large E_q and approaches the parameter learned by QBM (although does not reach the best value at $\Gamma = 2.5$ learned by QBM).

We note that QBM has an additional fit parameter as compared to BM and bQBM,

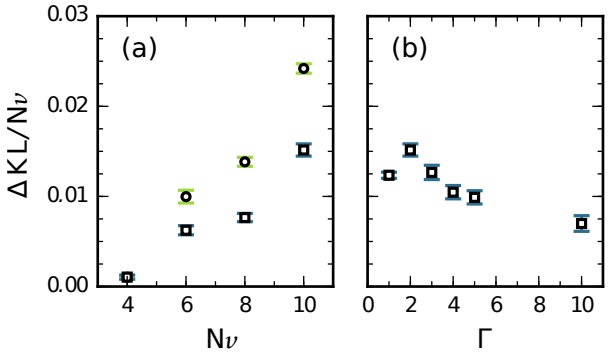


Figure 5.9: The advantage in the learning ability per spin of a quantum BM relative to the classical BM, $(\mathrm{KL} - \mathrm{KL}_{\mathrm{BM}})/N_v$ with circles/squares corresponding to a QBM/bQBM. This advantage is considered (a) as function of the system size with transverse-field $\Gamma = 2$ and (b) as function of the transverse-field strength on a system $N_v = 10$. For those plots, the number of modes in the mixture data model, Eq. (5.45), was kept equal to the number of qubits $N_v = M$ while the noise-parameter was always set to $p = 0.9$. In order to collect the statistics, the training was done for hundred dataset instances.

however its relative advantage should decrease as the total number of parameters increases. For this reason, we also study the dependence of the KL advantage as a function of the system size. Fig. 5.9(a) shows that both QBM and bQBM not only preserve their advantage over the classical BM but are able to increase the gap in their learning ability.

Since we cannot train the transverse field Γ for the bQBM, we have to treat it as a hyperparameter. We perform a scan of values of Γ and evaluate the KL advantage bQBM over BM based on 100 randomly generated datasets. Fig. 5.9(b) demonstrates a weak dependence of this advantage on the transverse field which justifies treating it as a hyperparameter.

We now consider a semi-restricted BM as per Eq. (5.24). Our toy model has 8 visible units and 2 hidden units. We allow full connectivity within the visible layer and all-to-all connectivity between the layers. The data is again generated using Eq. (5.45) for the visible variables, with $p = 0.9$ and $M = 8$. We present the results of training in Fig 5.10. Similarly to the fully visible model, QBM outperforms BM, and bQBM represents a good proxy for learning quantum distribution for a wide range of hyperparameter Γ.

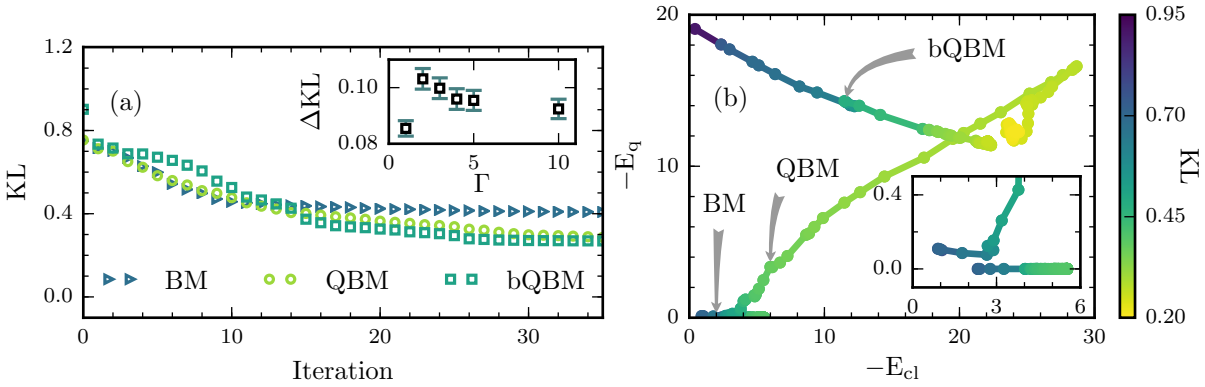


Figure 5.10: Training of a semi-restricted BM with 8 visible and 2 hidden units on artificial data from the mixture model (5.45) using second-order optimization routine. (a) KL-divergence of different models during training process. Again QBM and bQBM outperform BM. The inset studies the advantage in the learning ability of the bQBM relative to the classical BM, $\Delta\mathrm{KL} = \mathrm{KL}_{\mathrm{bQBM}} - \mathrm{KL}_{\mathrm{BM}}$, as function of the transverse-field strength. To collect the statistics, the training was done for a hundred different datasets generated with the noise parameter $p = 0.9$ and number of modes $M = 8$. (b) Classical and quantum average energies during training process. The inset details the same data set on a finer scale close to the classical regime.

Chapter 6

Conclusion and outlook

In this thesis, we pursued the program of entanglement entropy based classification of condensed matter ground states. At the core of this classification lies the geometry of the entangling region A which acts as a probe of universality. Different universal aspects of the underlying low-energy theory are revealed by varying different geometrical features of the region. We specialized in a cylindrical subregion A with a smooth entangling boundary confined to a toroidal system. Via a rigorous scaling analysis based on high-accuracy QMC datasets, we provided undeniable evidence for the universal signatures in the EE of the underlying ground states. More specifically, for the XY-model, we confirmed the full theoretical EE scaling prediction of Metlitski and Grover [23] for both of the universal EE contributions subleading to the area law [134]. Namely, the logarithmically divergent term was shown to be proportional to the number of Nambu-Goldstone modes arising due to a spontaneously broken continuous symmetry of the ground state. The universality of the geometric term subleading to the logarithm was also confirmed. This term is fully characterized by a universal function solely dependent on the aspect-ratio of the entangling region.

We recovered the trace of the same geometric function in an entirely different microscopic model. The same geometric EE contribution transpires for a massless free boson model regularized on a lattice. This model and this geometric function constitute the focus of the second part of our dive into the EE-driven exploration of strongly-interacting many-body systems [53]. The free boson model was used as a stepping stone to develop a rigorous fitting procedure to extract a universal number, κ_2, from a theoretically predicted behaviour of the geometric function in the limit when region A is shrunken to a thin slice. We then applied the benchmarked analysis to the critical transverse-field Ising model. The extracted value κ_2 was shown to be statistically different but similar between

the two microscopic models. This result is consistent with previously extracted values for another universal number that is based on corner-induced EE contributions. Theoretically, this similarity between the two sets of universal numbers characterizing those microscopic models is in line with the close relationship between their corresponding critical fixed points. Indeed, the Gaussian and Wilson-Fisher fixed points are smoothly connected via a perturbative ϵ-expansion. The absence of a physical interpretation for κ calls for further theoretical developments.

The experience gained from this work shows that a practical application of EE analysis as a tool for the identifying the underlying field theory of strongly-interacting systems hinges on many important technical ingredients. The importance of employing a proper fitting analysis is not to be underestimated as naive approaches easily lead to erroneous results. A successful extraction of the information contained in the universal part of the EE is contingent on a strong understanding of other contributions. In addition to the expected pollution from the area law and finite-size effects, the contribution from unexpected terms such as the conical singularity [135] and ultra-violet lattice effects in the EE signal need to be controlled in order to isolate the signal from the targeted universal terms. This isolation is often possible via a careful procedural design that takes full advantage of the flexibility provided by QMC techniques.

Each of the above projects required large amounts of computational resources estimated at over 300 CPU years per project. Fundamental to their success was the construction of the improved EE estimators. Without those estimators, the computational costs could have been a magnitude higher. Developed in stages, the theoretical procedure for the general construction of those estimators culminated in the *projected ensemble* framework based on a stochastic map onto a loop version of the underlying classical representation. The discovery of the connection of this method to the well-established Kandel-Domany framework [31] promises to be fruitful for the extension of the derived EE estimators to other models. The natural first extension would be the XXZ model.

The dramatic efficiency improvement witnessed in the performance of loop-based EE estimators warrants some discussion. As discussed in Sec. 3.2.4, those estimators effectively sample a large number of configurations conditioned on a sampled instance of a loop break up. In statistics, such a variance reduction technique is known as *conditioning* [24]. In turn, conditioning is related to *Rao-Blackwellization*, a more stringent technique requiring the conditioning variable to be a *sufficient statistic*. The latter is defined as the maximally informative statistic given a sample. This formulation hints at a deep connection between the physics of the loops and the EE. Exploration of this link could provide us with theoretical guiding principles for constructing numerical representations that inherently capture the physical EE properties of the underlying system and, ultimately,

lead to validation of EE as a tool for probing the universality of unknown systems.

As an alternative to the projected ensemble framework for estimating the Rényi entanglement entropy, a promising development is the out-of-equilibrium approach [136, 137]. This partition function ratio estimation method relies on measuring the work required to change one partition function into the other one. This approach has some similarity to thermodynamic integration but, in this case, the process does not need to be equilibrated at each intermediate step and is unbiased. This idea has seen an early adaptation in the machine learning community [138, 139] and more recently it has been applied to measuring EE [140]. More generally, framing the challenge of EE measurement as an estimation of the ratio of normalizing constants connects such endeavours to a vast body of literature in statistics and machine learning [141]. Not only does this theoretical bridge provide a new toolset of established methods such as the bridge-sampling [141] as well as an access to the state of the art [142], but it also provides a rigorous mathematical language for further exploration [143].

Concurrently, EE estimation can be phrased as an evaluation of the Swap operator [42]. Over the past few years, this viewpoint led to the first direct measurement of a Rényi entropy in an experiment [45, 144]. While being an exciting result, the setup has so far been limited to a few cold atoms. Extending the methodology to larger systems will unavoidably run into the same issues as those encountered in our work. Namely, as the size of the entangling region increases, the signal-to-noise ratio of the Swap operator statistics exponentially deteriorate due to the area law. Therefore, we anticipate that some of the variance-reduction tools used in QMC measurements will prove to be useful in the experimental setup. In particular, we underline the importance of the ratio trick [42]. While conceptually simple, this tool provides an exponential speed-up, making the detection of the EE in large systems a tractable problem. As the ratio trick relies on the ability to compute a quantity that does not take the form of an observable, it is not obvious how to apply the ratio trick to an experimental procedure. A competitive alternative might be based on the out-of-equilibrium framework mentioned above, as it relies on the measurement of the observed energy changes. As the core challenge posed by EE measurement is ultimately dictated by physics and is not dependent on a measuring setup, we envision a practical knowledge exchange between numerics and experiment for years to come.

Directly reconstructing the wavefunction from its projective measurements provides an alternative possibility for EE detection in an experiment. As a step towards this goal, we considered the framework of generative modeling built around the maximum-likelihood learning of classical probability distributions from data. As a concrete example, we adopted the restricted Boltzmann machine as a generative model. When trained on S_z measure-

ments of the transverse-field Ising model ground state in one dimension, we were able to reconstruct the EE of the state with good accuracy [50]. As the EE is a basis-independent quantity, this example serves as a proof-of-concept for the generalization ability of the RBM. Additionally, motivated by the development of physical annealers with a promise to sample from a quantum Boltzmann distribution, we considered a quantum generalization of the RBM termed the quantum Boltzmann machine [104]. Since it is not practical to train this model exactly, we developed a scalable bound-based training. On an artificially constructed classical data-set, our quantum Boltzmann machine demonstrated a better likelihood performance than its classical counterpart. Although this result needs to be verified on other datasets, the ability to efficiently learn a thermal quantum model from projective measurements makes this quantum Boltzmann machine a promising avenue for further exploration.

Unlike traditional tomography [145], the machine learning approach considered in this thesis is built on a principle of scalability to large datasets. Fundamental to this mission is the generalization ability of artificial neural networks. These composite highly non-linear structures promise to challenge the fundamentally linear construction of tensor networks. Combining learned patterns in novel non-linear ways is the central feature underlying the success of neural network based generative learning in the field of vision. To achieve ground-state reconstruction beyond the stoquastic Hamiltonians considered in this thesis, one would first need to extend this success to multi-basis datasets. In the era of Noisy Intermediate-Scale Quantum devices [146] designed to emulate physical models of interest, the question of the philosophical value of quantum data in this endeavour raises itself. Perhaps, as the area law indicated the way to the ground state for the tensor network community, data encodes the key to the ground state in the optimization landscape of artificial neural networks.

References

[1] D. Harlow, "Jerusalem lectures on black holes and quantum information", *Rev. Mod. Phys.* **88**, 015002 (2016), URL `https://link.aps.org/doi/10.1103/RevModPhys.88.015002`.

[2] X.-G. Wen, "Quantum orders and symmetric spin liquids", *Phys. Rev. B* **65**, 165113 (2002), URL `https://link.aps.org/doi/10.1103/PhysRevB.65.165113`.

[3] L. M. Balents, "Spin liquids in frustrated magnets", *Nature* **464**, 199–208 (2010).

[4] S. V. Isakov, M. B. Hastings and R. G. Melko, "Topological entanglement entropy of a Bose-Hubbard spin liquid", *Nature Physics* **7** (10), 772–775 (2011), `1102.1721`.

[5] C. Holzhey, F. Larsen and F. Wilczek, "Geometric and renormalized entropy in conformal field theory", *Nucl. Phys. B* **424** (3), 443–467 (1994).

[6] P. Calabrese and J. L. Cardy, "Entanglement entropy and quantum field theory", *J. Stat. Mech.* **0406**, P06002 (2004), `arXiv:hep-th/0405152`.

[7] A. B. Zamolodchikov, "Irreversibility of the Flux of the Renormalization Group in a 2D Field Theory", *JETP Lett.* **43**, 730–732 (1986), [Pisma Zh. Eksp. Teor. Fiz. 43, 565 (1986)].

[8] H. Casini and M. Huerta, "A c-theorem for entanglement entropy", *Journal of Physics A Mathematical General* **40** (25), 7031–7036 (2007), `cond-mat/0610375`.

[9] R. C. Myers and A. Sinha, "Holographic c-theorems in arbitrary dimensions", *Journal of High Energy Physics* **2011**, 125 (2011).

[10] H. Casini, M. Huerta and R. C. Myers, "Towards a derivation of holographic entanglement entropy", *JHEP* **05**, 036 (2011), `arXiv:1102.0440`.

[11] H. Casini and M. Huerta, "Renormalization group running of the entanglement entropy of a circle", **85** (12), 125016 (2012).

[12] L. Hayward Sierens, "Simulating quantum matter through lattice field theories", Ph.D. thesis, University of Waterloo (2017), URL http://hdl.handle.net/10012/11897.

[13] B. Swingle, "Mutual information and the structure of entanglement in quantum field theory", (2010), arXiv:1010.4038.

[14] G. Vidal, "Efficient classical simulation of slightly entangled quantum computations", *Phys. Rev. Lett.* **91**, 147902 (2003), URL https://link.aps.org/doi/10.1103/PhysRevLett.91.147902.

[15] S. R. White, "Density-matrix algorithms for quantum renormalization groups", *Phys. Rev. B* **48**, 10345–10356 (1993), URL https://link.aps.org/doi/10.1103/PhysRevB.48.10345.

[16] J. I. Latorre, E. Rico and G. Vidal, "Ground state entanglement in quantum spin chains", *arXiv e-prints* quant-ph/0304098 (2003), quant-ph/0304098.

[17] G. Vidal, "Entanglement Renormalization", .

[18] F. Verstraete and J. I. Cirac, "Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions", *arXiv e-prints* cond-mat/0407066 (2004).

[19] M. Troyer and U.-J. Wiese, "Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations", **94** (17), 170201 (2005), cond-mat/0408370.

[20] C. J. Umrigar, K. G. Wilson and J. W. Wilkins, "Optimized trial wave functions for quantum monte carlo calculations", *Phys. Rev. Lett.* **60**, 1719–1722 (1988), URL https://link.aps.org/doi/10.1103/PhysRevLett.60.1719.

[21] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks", *Science* **355**, 602–606 (2017), 1606.02318.

[22] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko and G. Carleo, "Neural-network quantum state tomography", *Nature Physics* **14**, 447–450 (2018).

[23] M. A. Metlitski and T. Grover, "Entanglement entropy of systems with spontaneously broken continuous symmetry", (2011), arXiv:1112.5166.

[24] A. B. Owen, *Monte Carlo theory, methods and examples* (2013).

[25] K. Werner, *Statistical Mechanics: Algorithms and Computations*, Oxford University Press (2006).

[26] S. C. Kapfer and W. Krauth, "Irreversible Local Markov Chains with Rapid Convergence towards Equilibrium", **119** (24), 240603 (2017), 1705.06689.

[27] D. Wilson, "Coupling from the past: a user's guide", (1998).

[28] C. Chanal and W. Krauth, "Renormalization Group Approach to Exact Sampling", **100** (6) (2008).

[29] J. G. Propp and D. B. Wilson, "Exact sampling with coupled markov chains and applications to statistical mechanics", *Random Structures & Algorithms* **9** (12), 223–252 (1996).

[30] V. Ambegaokar and M. Troyer **78** (2), 150–157 (2010), 0906.0943.

[31] H. G. Evertz, "The loop algorithm", *Advances in Physics* **52** (1), 1–66 (2003), cond-mat/9707221.

[32] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge Lecture Notes in Physics, Cambridge University Press (1996).

[33] S. Inglis and R. G. Melko, "Entanglement at a two-dimensional quantum critical point: a t = 0 projector quantum monte carlo study", *New Journal of Physics* **15** (7), 073048 (2013), arXiv:1305.1069.

[34] U. Wolff, "Collective monte carlo updating for spin systems", *Phys. Rev. Lett.* **62**, 361–364 (1989), URL https://link.aps.org/doi/10.1103/PhysRevLett.62.361.

[35] R. H. Swendsen and J.-S. Wang, "Nonuniversal critical dynamics in monte carlo simulations", *Phys. Rev. Lett.* **58**, 86–88 (1987), URL https://link.aps.org/doi/10.1103/PhysRevLett.58.86.

[36] C. Fortuin and P. Kasteleyn, "On the random-cluster model: I. introduction and relation to other models", *Physica* **57** (4), 536 – 564 (1972), ISSN 0031-8914, URL http://www.sciencedirect.com/science/article/pii/0031891472900456.

[37] L. Wang, "Exploring cluster monte carlo updates with boltzmann machines", *Phys. Rev. E* **96**, 051301 (2017), URL https://link.aps.org/doi/10.1103/PhysRevE.96.051301.

[38] G. Desjardins, *Improving Sampling, Optimization and Feature Extraction in Boltzmann Machines*, Universit de Montra (2014), https://papyrus.bib.umontreal.ca/xmlui/handle/1866/10550?locale-attribute=fr.

[39] G. Grimmett, "The Random-Cluster Model", *arXiv Mathematics e-prints* math/0205237 (2002), math/0205237.

[40] W. Janke and A. M. J. Schakel, "Fractal structure of spin clusters and domain walls in the two-dimensional ising model", *Phys. Rev. E* **71**, 036703 (2005), URL https://link.aps.org/doi/10.1103/PhysRevE.71.036703.

[41] O. F. Syljuåsen and A. W. Sandvik, "Quantum monte carlo with directed loops", *Phys. Rev. E* **66**, 046701 (2002), arXiv:cond-mat/0202316, URL https://link.aps.org/doi/10.1103/PhysRevE.66.046701.

[42] M. B. Hastings, I. González, A. B. Kallin and R. G. Melko, "Measuring renyi entanglement entropy in quantum monte carlo simulations", *Phys. Rev. Lett.* **104**, 157201 (2010), arXiv:1001.2335.

[43] S. Humeniuk and T. Roscilde, "Quantum monte carlo calculation of entanglement rényi entropies for generic quantum systems", *Phys. Rev. B* **86**, 235116 (2012).

[44] M. Caraglio and F. Gliozzi, "Entanglement Entropy and Twist Fields", *JHEP* **11**, 076 (2008), arXiv:0808.4094.

[45] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli and M. Greiner, "Measuring entanglement entropy through the interference of quantum many-body twins", *arXiv e-prints* arXiv:1509.01160 (2015), 1509.01160.

[46] S. Inglis and R. G. Melko, "Wang-Landau method for calculating Rényi entropies in finite-temperature quantum Monte Carlo simulations", **87** (1), 013306 (2013), 1207.5052.

[47] R. G. Melko, A. B. Kallin and M. B. Hastings, "Finite Size Scaling of Mutual Information: A Scalable Simulation", *arXiv e-prints* arXiv:1007.2182 (2010), 1007.2182.

[48] P. V. Buividovich and M. I. Polikarpov, "Numerical study of entanglement entropy in SU(2) lattice gauge theory", *Nuclear Physics B* **802** (3), 458–474 (2008).

[49] E. de Miguel, "Estimating errors in free energy calculations from thermodynamic integration using fitted data", *The Journal of Chemical Physics* **129** (21), 214112 (2008).

[50] M. J. S. Beach, I. De Vlugt, A. Golubeva, P. Huembeli, B. Kulchytskyy, X. Luo, R. G. Melko, E. Merali and G. Torlai, "QuCumber: wavefunction reconstruction with neural networks", *arXiv e-prints* arXiv:1812.09329 (2018).

[51] R. G. Melko, *Stochastic Series Expansion Quantum Monte Carlo*, (185) (2013).

[52] V. Alba, "Entanglement negativity and conformal field theory: a monte carlo study", *Journal of Statistical Mechanics: Theory and Experiment* **2013** (05), P05013 (2013).

[53] B. Kulchytskyy, L. E. Hayward Sierens and R. G. Melko, "Universal divergence of the Renyi entropy of a thinly sliced torus at the Ising fixed point", *arXiv e-prints* arXiv:1904.08955 (2019), 1904.08955.

[54] Inglis, Stephen, "Implementations and applications of renyi entanglement in monte carlo simulations of spin models", (2013), URL http://hdl.handle.net/10012/7940.

[55] A. W. Sandvik, "Stochastic series expansion method for quantum ising models with arbitrary interactions", *Phys. Rev. E* **68**, 056701 (2003).

[56] H. F. Song, N. Laflorencie, S. Rachel and K. Le Hur, "Entanglement entropy of the two-dimensional heisenberg antiferromagnet", *Phys. Rev. B* **83**, 224410 (2011), URL https://link.aps.org/doi/10.1103/PhysRevB.83.224410.

[57] A. B. Kallin, M. B. Hastings, R. G. Melko and R. R. P. Singh, "Anomalies in the entanglement properties of the square-lattice Heisenberg model", *Phys. Rev. B* **84** (16), 165134 (2011), ISSN 1098-0121.

[58] P. W. Anderson, "An approximate quantum theory of the antiferromagnetic ground state", *Phys. Rev.* **86**, 694–701 (1952), URL https://link.aps.org/doi/10.1103/PhysRev.86.694.

[59] C. Lhuillier, "Frustrated Quantum Magnets", *arXiv e-prints* cond-mat/0502464 (2005), cond-mat/0502464.

[60] G. Misguich and P. Sindzingre, "Detecting spontaneous symmetry breaking in finite-size spectra of frustrated quantum antiferromagnets", *Journal of Physics: Condensed Matter* **19** (14), 145202 (2007), URL https://doi.org/10.1088%2F0953-8984%2F19%2F14%2F145202.

[61] A. W. Sandvik, "Computational Studies of Quantum Spin Systems", (2010), volume 1297 of *American Institute of Physics Conference Series*, (135–338), 1101.3281.

[62] F. Kolley, S. Depenbrock, I. P. McCulloch, U. Schollwöck and V. Alba, "Entanglement spectroscopy of su(2)-broken phases in two dimensions", *Phys. Rev. B* **88**, 144426 (2013), URL https://link.aps.org/doi/10.1103/PhysRevB.88.144426.

[63] V. Alba, M. Haque and A. M. Läuchli, "Entanglement spectrum of the two-dimensional bose-hubbard model", *Phys. Rev. Lett.* **110**, 260403 (2013), URL https://link.aps.org/doi/10.1103/PhysRevLett.110.260403.

[64] B. I. HALPERIN and P. C. HOHENBERG, "Hydrodynamic theory of spin waves", *Phys. Rev.* **188**, 898–918 (1969), URL https://link.aps.org/doi/10.1103/PhysRev.188.898.

[65] F.-J. Jiang, "Method of calculating the spin-wave velocity of spin-$\frac{1}{2}$ antiferromagnets with $o(n)$ symmetry in a monte carlo simulation", *Phys. Rev. B* **83**, 024419 (2011).

[66] J. Iaconis, S. Inglis, A. B. Kallin and R. G. Melko, "Detecting classical phase transitions with renyi mutual information", *Phys. Rev. B* **87**, 195134 (2013), URL https://link.aps.org/doi/10.1103/PhysRevB.87.195134.

[67] I. Peschel, "Calculation of reduced density matrices from correlation functions", *J. Phys. A: Math. Gen.* **36**, L205 (2003), arXiv:cond-mat/0212631.

[68] A. B. Kallin, K. Hyatt, R. R. P. Singh and R. G. Melko, "Entanglement at a two-dimensional quantum critical point: A numerical linked-cluster expansion study", *Phys. Rev. Lett.* **110**, 135702 (2013), arXiv:1212.5269.

[69] E. M. Stoudenmire, P. Gustainis, R. Johal, S. Wessel and R. G. Melko, "Corner contribution to the entanglement entropy of strongly interacting o(2) quantum critical systems in 2+1 dimensions", *Phys. Rev. B* **90**, 235106 (2014), arXiv:1409.6327.

[70] A. B. Kallin, E. M. Stoudenmire, P. Fendley, R. R. P. Singh and R. G. Melko, "Corner contribution to the entanglement entropy of an O(3) quantum critical point in 2 + 1 dimensions", *J. Stat. Mech.* **1406**, P06009 (2014), 1401.3504.

[71] F. Kos, D. Poland and D. Simmons-Duffin, "Bootstrapping the O(N) vector models", *Journal of High Energy Physics* **2014** (6) (2014), arXiv:1406.4858.

[72] P. Bueno, R. C. Myers and W. Witczak-Krempa, "Universality of corner entanglement in conformal field theories", *Phys. Rev. Lett.* **115**, 021602 (2015), arXiv:1505.04804.

[73] P. Bueno, R. C. Myers and W. Witczak-Krempa, "Universal corner entanglement from twist operators", *JHEP* **09**, 091 (2015), 1507.06997.

[74] P. Bueno and W. Witczak-Krempa, "Bounds on corner entanglement in quantum critical states", *Phys. Rev. B* **93**, 045131 (2016), arXiv:1511.04077.

[75] T. Faulkner, R. G. Leigh and O. Parrikar, "Shape dependence of entanglement entropy in conformal field theories", *Journal of High Energy Physics* **2016** (4), 88 (2016), ISSN 1029-8479, arXiv:1511.05179, URL https://doi.org/10.1007/JHEP04(2016)088.

[76] J. Helmes, L. E. Hayward Sierens, A. Chandran, W. Witczak-Krempa and R. G. Melko, "Universal corner entanglement of dirac fermions and gapless bosons from the continuum to the lattice", *Phys. Rev. B* **94**, 125142 (2016), arXiv:1606.03096.

[77] H. Casini and M. Huerta, "Entanglement entropy in free quantum field theory", *Journal of Physics A: Mathematical and Theoretical* **42** (50), 504007 (2009), 0905.2562.

[78] I. A. Kovács and F. Iglói, "Universal logarithmic terms in the entanglement entropy of 2d, 3d and 4d random transverse-field ising models", *EPL (Europhysics Letters)* **97** (6), 67009 (2012), arXiv:1108.3942, URL http://stacks.iop.org/0295-5075/97/i=6/a=67009.

[79] I. R. Klebanov, T. Nishioka, S. S. Pufu and B. R. Safdi, "On Shape Dependence and RG Flow of Entanglement Entropy", *JHEP* **07**, 001 (2012), arXiv:1204.4160.

[80] R. C. Myers and A. Singh, "Entanglement Entropy for Singular Surfaces", *JHEP* **09**, 013 (2012), arXiv:1206.5225.

[81] T. Devakul and R. R. P. Singh, "Entanglement across a cubic interface in $3 + 1$ dimensions", *Phys. Rev. B* **90**, 054415 (2014), arXiv:1407.0084, URL https://link.aps.org/doi/10.1103/PhysRevB.90.054415.

[82] P. Bueno and R. C. Myers, "Universal entanglement for higher dimensional cones", *JHEP* **12**, 168 (2015), 1508.00587.

[83] L. E. Hayward Sierens, P. Bueno, R. R. P. Singh, R. C. Myers and R. G. Melko, "Cubic trihedral corner entanglement for a free scalar", *Phys. Rev. B* **96**, 035117 (2017), arXiv:1703.03413.

[84] G. Bednik, L. E. Hayward Sierens, M. Guo, R. C. Myers and R. G. Melko, "Probing trihedral corner entanglement for dirac fermions", (2018), arXiv:1810.02831.

[85] L. Chojnacki, C. Q. Cook, D. Dalidovich, L. E. Hayward Sierens, E. Lantagne-Hurtubise, R. G. Melko and T. J. Vlaar, "Shape dependence of two-cylinder rényi entropies for free bosons on a lattice", *Phys. Rev. B* **94**, 165136 (2016), arXiv:1607.05311, URL https://link.aps.org/doi/10.1103/PhysRevB.94.165136.

[86] W. Witczak-Krempa, L. E. Hayward Sierens and R. G. Melko, "Cornering gapless quantum states via their torus entanglement", *Phys. Rev. Lett.* **118**, 077202 (2017), arXiv:1603.02684.

[87] J.-M. Stéphan, H. Ju, P. Fendley and R. G. Melko, "Entanglement in gapless resonating valence bond states", *New. J. Phys.* **15**, 015004 (2013), arXiv:1207.3820.

[88] X. Chen, W. Witczak-Krempa, T. Faulkner and E. Fradkin, "Two-cylinder entanglement entropy under a twist", *Journal of Statistical Mechanics: Theory and Experiment* **4**, 043104 (2017), arXiv:1611.01847.

[89] X. Chen, G. Y. Cho, T. Faulkner and E. Fradkin, "Scaling of entanglement in 2 + 1-dimensional scale-invariant field theories", *Journal of Statistical Mechanics: Theory and Experiment* **2015** (2015), arXiv:1412.3546.

[90] J. Cardy and P. Calabrese, "Unusual corrections to scaling in entanglement entropy", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (04), P04023 (2010), arXiv:1002.4353.

[91] S. Sahoo, E. M. Stoudenmire, J.-M. Stéphan, T. Devakul, R. R. P. Singh and R. G. Melko, "Unusual corrections to scaling and convergence of universal renyi properties at quantum critical points", *Phys. Rev. B* **93**, 085120 (2016), arXiv:1509.00468.

[92] J. Helmes, "An entanglement perspective on phase transitions, conventional and topological order", Ph.D. thesis, University of Cologne (2017), URL http://www.thp.uni-koeln.de/trebst/thesis/PhD_JohannesHelmes.pdf.

[93] A. L. Talapov and H. W. J. Blöte, "The magnetization of the 3d ising model", *Journal of Physics A: Mathematical and General* **29** (17), 5727 (1996), arXiv:cond-mat/9603013, URL http://stacks.iop.org/0305-4470/29/i=17/a=042.

[94] L. Liu, A. W. Sandvik and W. Guo, "Typicality at quantum-critical points", *Chinese Physics B* **27** (8), 087501 (2018), arXiv:1805.04273, URL http://stacks.iop.org/1674-1056/27/i=8/a=087501.

[95] P. Bueno and R. C. Myers, "Corner contributions to holographic entanglement entropy", *JHEP* **08**, 068 (2015), 1505.07842.

[96] S. Whitsitt, W. Witczak-Krempa and S. Sachdev, "Entanglement entropy of the large N Wilson-Fisher conformal field theory", *Phys. Rev.* **B95** (4), 045148 (2017), 1610.06568.

[97] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science* **313** (5786), 504 – 507 (2006).

[98] A. Vahdat, E. Andriyash and W. G. Macready, "DVAE#: Discrete Variational Autoencoders with Relaxed Boltzmann Priors", *arXiv e-prints* arXiv:1805.07445 (2018), 1805.07445.

[99] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura and H. G. Katzgraber, "Physics-Inspired Optimization for Quadratic Unconstrained Problems Using a Digital Annealer", *arXiv e-prints* arXiv:1806.08815 (2018), 1806.08815.

[100] S. H. Adachi and M. P. Henderson, "Application of Quantum Annealing to Training of Deep Neural Networks", *arXiv e-prints* arXiv:1510.06356 (2015), 1510.06356.

[101] G. Torlai and R. G. Melko, "Learning thermodynamics with Boltzmann machines", *Physical Review B* **94**, 165134 (2016), 1606.02718.

[102] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez and J. I. Cirac, "Neural-network quantum states, string-bond states, and chiral topological states", *Phys. Rev. X* **8**, 011006 (2018), URL https://link.aps.org/doi/10.1103/PhysRevX.8.011006.

[103] J. Chen, S. Cheng, H. Xie, L. Wang and T. Xiang, "Equivalence of restricted Boltzmann machines and tensor network states", *Physical Review B* **97**, 085104 (2018), 1701.04831.

[104] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy and R. Melko, "Quantum Boltzmann Machine", *arXiv e-prints* arXiv:1601.02036 (2016), 1601.02036.

[105] Y. LeCun, S. Chopra, R. Hadsell, F. J. Huang and et al., "A tutorial on energy-based learning", in "PREDICTING STRUCTURED DATA", MIT Press (2006).

[106] R. Salakhutdinov, A. Mnih and G. Hinton, "Restricted boltzmann machines for collaborative filtering", in "Proceedings of the 24th International Conference on Machine Learning", ACM, New York, NY, USA (2007), ICML '07, (791–798), ISBN 978-1-59593-793-3, URL http://doi.acm.org/10.1145/1273496.1273596.

[107] M. Welling, M. Rosen-zvi and G. E. Hinton, "Exponential family harmoniums with an application to information retrieval", in L. K. Saul, Y. Weiss and L. Bottou (editors), "Advances in Neural Information Processing Systems 17", MIT Press, (1481–1488) (2005), URL http://papers.nips.cc/paper/2672-exponential-family-harmoniums-with-an-application-to-information-retrieval.pdf.

[108] R. S. Zemel, C. K. I. Williams and M. C. Mozer, "Directional-unit boltzmann machines", in S. J. Hanson, J. D. Cowan and C. L. Giles (editors), "Advances in Neural Information Processing Systems 5", Morgan-Kaufmann, (172–179) (1993), URL http://papers.nips.cc/paper/674-directional-unit-boltzmann-machines.pdf.

[109] R. Memisevic and G. E. Hinton, "Learning to represent spatial transformations with factored higher-order boltzmann machines", **22** (6), 1473–1492 (2010), exported from https://app.dimensions.ai on 2019/01/28, URL https://app.dimensions.ai/details/publication/pub.1046491641.

[110] N. Le Roux and Y. Bengio, "Representational power of restricted boltzmann machines and deep belief networks", *Neural Comput.* **20** (6), 1631–1649 (2008), ISSN 0899-7667, URL http://dx.doi.org/10.1162/neco.2008.04-07-510.

[111] P. Mehta and D. J. Schwab, "An exact mapping between the Variational Renormalization Group and Deep Learning", *arXiv e-prints* arXiv:1410.3831 (2014), 1410.3831.

[112] S. Iso, S. Shiba and S. Yokoo, "Scale-invariant feature extraction of neural network and renormalization group flow", *Phys. Rev. E* **97**, 053304 (2018), URL https://link.aps.org/doi/10.1103/PhysRevE.97.053304.

[113] M. Koch-Janusz and Z. Ringel, "Mutual information, neural networks and the renormalization group", *Nature Physics* **14**, 578–582 (2018), 1704.06279.

[114] Y. Zhang, Z. Ghahramani, A. J. Storkey and C. A. Sutton, "Continuous relaxations for discrete hamiltonian monte carlo", in F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (editors), "Advances in Neural Information Processing Systems 25", Curran Associates, Inc., (3194–3202) (2012), URL http://papers.nips.cc/paper/4652-continuous-relaxations-for-discrete-hamiltonian-monte-carlo.pdf.

[115] G. Desjardins, A. Courville, Y. Bengio, P. Vincent and O. Delalleau, "Tempered markov chain monte carlo for training of restricted boltzmann machines", in Y. W. Teh and M. Titterington (editors), "Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics", PMLR, Chia Laguna Resort, Sardinia, Italy (2010), volume 9 of *Proceedings of Machine Learning Research*, (145–152), URL http://proceedings.mlr.press/v9/desjardins10a.html.

[116] T. Tieleman and G. Hinton, "Using fast weights to improve persistent contrastive divergence", in "Proceedings of the 26th Annual International Conference on Machine Learning", ACM, New York, NY, USA (2009), ICML '09, (1033–1040), ISBN 978-1-60558-516-1, URL http://doi.acm.org/10.1145/1553374.1553506.

[117] G. E. Hinton, "Training products of experts by minimizing contrastive divergence", *Neural Comput.* **14** (8), 1771–1800 (2002), ISSN 0899-7667, URL http://dx.doi.org/10.1162/089976602760128018.

[118] T. Tieleman, "Training restricted boltzmann machines using approximations to the likelihood gradient", in "Proceedings of the 25th international conference on Machine learning", (2008), (1064–1071).

[119] L. Jian-wei, C. Guang-hui and L. Xiong-lin, "Contrastive divergence learning of restricted boltzmann machine", in "Proceedings of the 2012 Second International Conference on Electric Technology and Civil Engineering", IEEE Computer Society, Washington, DC, USA (2012), ICETCE '12, (712–715), ISBN 978-0-7695-4704-6, URL https://doi.org/10.1109/ICETCE.2012.173.

[120] G. E. Hinton, "A practical guide to training restricted boltzmann machines (version 1)", (2010).

[121] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko and G. Carleo, "Many-body quantum state tomography with neural networks", *arXiv e-prints* arXiv:1703.05334 (2017), 1703.05334.

[122] E. Farhi, J. Goldstone, S. Gutmann and M. Sipser, "Quantum Computation by Adiabatic Evolution", *arXiv e-prints* quant-ph/0001106 (2000), quant-ph/0001106.

[123] D. Aharonov, W. van Dam, J. Kempe, Z. Land au, S. Lloyd and O. Regev, "Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation", *arXiv e-prints* quant-ph/0405098 (2004), quant-ph/0405098.

[124] J. Kempe, A. Kitaev and O. Regev, "The Complexity of the Local Hamiltonian Problem", *arXiv e-prints* quant-ph/0406180 (2004), quant-ph/0406180.

[125] R. Oliveira and B. M. Terhal, "The complexity of quantum spin systems on a two-dimensional square lattice", *arXiv e-prints* quant-ph/0504050 (2005), quant-ph/0504050.

[126] D. Aharonov, D. Gottesman, S. Irani and J. Kempe, "The Power of Quantum Systems on a Line", *Communications in Mathematical Physics* **287**, 41–65 (2009), 0705.4077.

[127] T. Kato, "On the adiabatic theorem of quantum mechanics", *Journal of the Physical Society of Japan* **5** (6), 435–439 (1950), https://doi.org/10.1143/JPSJ.5.435, URL https://doi.org/10.1143/JPSJ.5.435.

[128] M. S. Sarandy and D. A. Lidar, "Adiabatic approximation in open quantum systems", *Phys. Rev. A* **71**, 012331 (2005), URL https://link.aps.org/doi/10.1103/PhysRevA.71.012331.

[129] L. C. Venuti, T. Albash, D. A. Lidar and P. Zanardi, "Adiabaticity in open quantum systems", *Phys. Rev. A* **93**, 032118 (2016), URL https://link.aps.org/doi/10.1103/PhysRevA.93.032118.

[130] M. H. Amin, "Searching for quantum speedup in quasistatic quantum annealers", *Phys. Rev. A* **92**, 052323 (2015), URL https://link.aps.org/doi/10.1103/PhysRevA.92.052323.

[131] R. M. Wilcox, "Exponential operators and parameter differentiation in quantum physics", *Journal of Mathematical Physics* **8** (4), 962–982 (1967), https://doi.org/10.1063/1.1705306, URL https://doi.org/10.1063/1.1705306.

[132] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", *arXiv e-prints* arXiv:1312.6114 (2013), 1312.6114.

[133] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, NY, USA, second edition (2006).

[134] B. Kulchytskyy, C. M. Herdman, S. Inglis and R. G. Melko, "Detecting goldstone modes with entanglement entropy", *Phys. Rev. B* **92**, 115146 (2015), arXiv:1502.01722, URL https://link.aps.org/doi/10.1103/PhysRevB.92.115146.

[135] J. Cardy and P. Calabrese, "Unusual corrections to scaling in entanglement entropy", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (4), 04023 (2010), 1002.4353.

[136] C. Jarzynski, "Rare events and the convergence of exponentially averaged work values", *Phys. Rev. E* **73**, 046105 (2006), URL https://link.aps.org/doi/10.1103/PhysRevE.73.046105.

[137] J. Cardy, "Measuring entanglement using quantum quenches", *Phys. Rev. Lett.* **106**, 150404 (2011), URL https://link.aps.org/doi/10.1103/PhysRevLett.106.150404.

[138] R. M. Neal, "Estimating Ratios of Normalizing Constants Using Linked Importance Sampling", *arXiv Mathematics e-prints* math/0511216 (2005).

[139] R. Salakhutdinov, "Learning and evaluating boltzmann machines", (2019).

[140] V. Alba, "Out-of-equilibrium protocol for rényi entropies via the jarzynski equality", *Phys. Rev. E* **95**, 062132 (2017), URL https://link.aps.org/doi/10.1103/PhysRevE.95.062132.

[141] A. Gelman and X.-L. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling", *Statist. Sci.* **13** (2), 163–185 (1998), URL https://doi.org/10.1214/ss/1028905934.

[142] Q. Liu, A. Ihler, J. Peng and J. Fisher, "Estimating the partition function by discriminance sampling", in "Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI 2015", AUAI Press (2015), (514–522).

[143] D. Carlson, P. Stinson, A. Pakman and L. Paninski, "Partition Functions from Rao-Blackwellized Tempered Sampling", *arXiv e-prints* arXiv:1603.01912 (2016).

[144] R. Islam, R. Ma, P. Preiss, M. Eric Tai, A. Lukin, M. Rispoli and M. Greiner, "Measuring entanglement entropy in a quantum many-body system", *Nature* **528**, 77–83 (2015).

[145] D. F. V. James, P. G. Kwiat, W. J. Munro and A. G. White, "Measurement of qubits", **64** (5), 052312 (2001).

[146] J. Preskill, "Quantum Computing in the NISQ era and beyond", *arXiv e-prints* arXiv:1801.00862 (2018), 1801.00862.

[147] F. Barahona, "On the computational complexity of ising spin glass models", (2001).

[148] R. M. Wilcox, "Exponential operators and parameter differentiation in quantum physics", *Journal of Mathematical Physics* **8** (4), 962–982 (1967), https://doi.org/10.1063/1.1705306, URL https://doi.org/10.1063/1.1705306.

[149] M. B. Haugh, "Simulation efficiency and an introduction to variance reduction methods", (2010).

[150] P. Young, "Everything you wanted to know about Data Analysis and Fitting but were afraid to ask", *arXiv e-prints* arXiv:1210.3781 (2012), 1210.3781.

[151] M. B. Hastings, I. González, A. B. Kallin and R. G. Melko, "Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations", **104** (15), 157201 (2010), 1001.2335.

[152] V. E. Korepin, "Universality of entropy scaling in one dimensional gapless models", *Phys. Rev. Lett.* **92** (9), 096402 (2004).

[153] G. Vidal, J. I. Latorre, E. Rico and A. Kitaev, "Entanglement in quantum critical phenomena", *Phys. Rev. Lett.* **90** (22), 227902 (2003).

[154] I. Peschel and V. Eisler, "Reduced density matrices and entanglement entropy in free lattice models", *J. Phys. A: Math. Gen.* **42**, 504003 (2009).

[155] L. Bombelli, R. K. Koul, J. Lee and R. D. Sorkin, "Quantum source of entropy for black holes", *Phys. Rev. D* **34**, 373 (1986).

[156] M. Srednicki, "Entropy and area", *Phys. Rev. Lett.* **71**, 666 (1993), arXiv:hep-th/9303048.

[157] I. Affleck, "Universal term in the free energy at a critical point and the conformal anomaly", *Phys. Rev. Lett.* **56**, 746–748 (1986), URL https://link.aps.org/doi/10.1103/PhysRevLett.56.746.

[158] I. Affleck and J. C. Bonner, "Logarithmic corrections in antiferromagnetic chains", *Phys. Rev. B* **42**, 954–957 (1990).

[159] H. W. J. Blöte, J. L. Cardy and M. P. Nightingale, "Conformal invariance, the central charge, and universal finite-size amplitudes at criticality", *Phys. Rev. Lett.* **56**, 742–745 (1986), URL https://link.aps.org/doi/10.1103/PhysRevLett.56.742.

[160] P. Calabrese and J. L. Cardy, "Entanglement entropy and conformal field theory", *J. Stat. Mech.* **42** (50), 504005 (2009), URL `http://stacks.iop.org/1751-8121/42/i=50/a=504005`.

[161] J. Cardy, "The ubiquitous ' c ': from the stefan boltzmann law to quantum information", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (10), P10004 (2010).

[162] H. Casini and M. Huerta, "Universal terms for the entanglement entropy in 2+1 dimensions", *Nuclear Physics B* **764** (3), 183 – 201 (2007).

[163] H. Casini and M. Huerta, "Renormalization group running of the entanglement entropy of a circle", *Phys. Rev. D* **85**, 125016 (2012).

[164] T. Devakul and R. R. P. Singh, "Quantum critical universality and singular corner entanglement entropy of bilayer heisenberg-ising model", *Phys. Rev. B* **90**, 064424 (2014).

[165] E. Fradkin and J. E. Moore, "Entanglement entropy of 2d conformal quantum critical points: Hearing the shape of a quantum drum", *Phys. Rev. Lett.* **97**, 050404 (2006).

[166] D. V. Fursaev, "Entanglement entropy in critical phenomena and analog models of quantum gravity", *Phys. Rev. D* **73**, 124025 (2006).

[167] T. Grover, "Entanglement monotonicity and the stability of gauge theories in three spacetime dimensions", *Phys. Rev. Lett.* **112**, 151601 (2014).

[168] J. Helmes and S. Wessel, "Entanglement entropy scaling in the bilayer heisenberg spin system", *Phys. Rev. B* **89**, 245120 (2014).

[169] J. Helmes and S. Wessel, "Correlations and entanglement in quantum critical bilayer and necklace xy models", *Phys. Rev. B* **92**, 125120 (2015).

[170] T. Hirata and T. Takayanagi, "Ads/cft and strong subadditivity of entanglement entropy", *Journal of High Energy Physics* **2007** (02), 042 (2007).

[171] M. Huerta, "Numerical determination of the entanglement entropy for free fields in the cylinder", *Physics Letters B* **710** (4–5), 691 – 696 (2012), ISSN 0370-2693.

[172] S. V. Isakov, R. G. Melko and M. B. Hastings, "Universal signatures of fractionalized quantum critical points", *Science* **335** (6065), 193–195 (2012).

[173] Calculations were performing using the ITensor Library: http://itensor.org/.

[174] J. I. Latorre, E. Rico and G. Vidal, "Ground state entanglement in quantum spin chains", *Quant. Inf. Comput.* **4**, 48–92 (2004), `quant-ph/0304098`.

[175] J. Lee, L. McGough and B. R. Safdi, "Rényi entropy and geometry", *Phys. Rev. D* **89**, 125016 (2014).

[176] M. Rigol, T. Bryant and R. R. P. Singh, "Numerical linked-cluster approach to quantum lattice models", *Phys. Rev. Lett.* **97**, 187202 (2006).

[177] M. Rigol, T. Bryant and R. R. P. Singh, "Numerical linked-cluster algorithms. i. spin systems on square, triangular, and kagomé lattices", *Phys. Rev. E* **75**, 061118 (2007).

[178] M. Rigol, T. Bryant and R. R. P. Singh, "Numerical linked-cluster algorithms. ii. $t-j$ models on the square lattice", *Phys. Rev. E* **75**, 061119 (2007).

[179] S. Sachdev, "The landscape of the hubbard model", (2010).

[180] A. W. Sandvik, "Stochastic series expansion method with operator-loop update", *Phys. Rev. B* **59** (22), R14157–R14160 (1999).

[181] A. W. Sandvik, "Continuous quantum phase transition between an antiferromagnet and a valence-bond solid in two dimensions: Evidence for logarithmic corrections to scaling", *Phys. Rev. Lett.* **104** (17), 177201 (2010).

[182] U. Schollwöck, "The density-matrix renormalization group", *Rev. Mod. Phys.* **77** (1), 259–315 (2005).

[183] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev and M. P. A. Fisher, "Deconfined quantum critical points", *Science* **303** (5663), 1490–1494 (2004).

[184] R. R. P. Singh, R. G. Melko and J. Oitmaa, "Thermodynamic singularities in the entanglement entropy at a two-dimensional quantum critical point", *Phys. Rev. B* **86**, 075106 (2012).

[185] S. N. Solodukhin, "Entanglement entropy, conformal invariance and extrinsic geometry", *Physics Letters B* **665** (4), 305 – 309 (2008), ISSN 0370-2693.

[186] B. Tang, E. Khatami and M. Rigol, "A short introduction to numerical linked-cluster expansions", *Computer Physics Communications* **184** (3), 557 – 564 (2013).

[187] The critical coupling $J_c^{\perp} = 5.460$ for the XY bilayer model H_2 was computed for the present paper by S. Wessel.

[188] Z. Zhang, K. Wierschem, I. Yap, Y. Kato, C. D. Batista and P. Sengupta, "Phase diagram and magnetic excitations of anisotropic spin-one magnets", *Phys. Rev. B* **87**, 174405 (2013).

[189] J. M. Maldacena, "The large n limit of superconformal field theories and supergravity", *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).

[190] B. Swingle and T. Senthil, "Structure of entanglement at deconfined quantum critical points", *Phys. Rev. B* **86**, 155131 (2012).

[191] E. Fradkin and J. E. Moore, "Entanglement entropy of 2d conformal quantum critical points: hearing the shape of a quantum drum", *Phys. Rev. Lett.* **97**, 050404 (2006).

[192] M. P. Zaletel, J. H. Bardarson and J. E. Moore, "Logarithmic terms in entanglement entropies of 2d quantum critical points and shannon entropies of spin chains", *Phys. Rev. Lett.* **107**, 020402 (2011).

[193] J.-M. Stéphan, G. Misguich and V. Pasquier, "Phase transition in the rényi-shannon entropy of luttinger liquids", *Phys. Rev. B* **84**, 195128 (2011).

[194] H. Ju, A. B. Kallin, P. Fendley, M. B. Hastings and R. G. Melko, "Entanglement scaling in two-dimensional gapless systems", *Phys. Rev. B* **85**, 165121 (2012).

[195] A. W. Sandvik and J. Kurkijärvi, "Quantum monte carlo simulation method for spin systems", *Phys. Rev. B* **43**, 5950–5961 (1991), URL `https://link.aps.org/doi/10.1103/PhysRevB.43.5950`.

[196] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantummechanical entropy", *Journal of Mathematical Physics* **14** (12), 1938–1941 (1973), `https://doi.org/10.1063/1.1666274`, URL `https://doi.org/10.1063/1.1666274`.

[197] E. M. Stoudenmire and S. R. White, "Studying two-dimensional systems with the density matrix renormalization group", *Annual Review of Condensed Matter Physics* **3** (1), 111–128 (2012).

[198] A. B. Kallin, "Computational methods for the measurement of entanglement in condensed matter systems", Ph.D. thesis, University of Waterloo (2014), URL `https://uwspace.uwaterloo.ca/handle/10012/8539`.

[199] M. S. Block, R. G. Melko and R. K. Kaul, "Fate of $\mathbb{C}\mathrm{p}^{N-1}$ fixed points with q monopoles", *Phys. Rev. Lett.* **111**, 137202 (2013).

[200] E. L. Pollock and D. M. Ceperley, "Path-integral computation of superfluid densities", *Phys. Rev. B* **36**, 8343–8352 (1987).

[201] M. A. Metlitski, C. A. Fuertes and S. Sachdev, "Entanglement entropy in the O(N) model", *Phys. Rev. B* **80**, 115122 (2009).

[202] A. Kitaev and J. Preskill, "Topological Entanglement Entropy", *Phys. Rev. Lett.* **96** (11), 110404 (2006).

[203] M. Levin and X.-G. Wen, "Detecting Topological Order in a Ground State Wave Function", *Phys. Rev. Lett.* **96** (11), 110405 (2006).

[204] M. M. Wolf, "Violation of the entropic area law for fermions", *Phys. Rev. Lett.* **96**, 010404 (2006).

[205] D. Gioev and I. Klich, "Entanglement entropy of fermions in any dimension and the widom conjecture", *Phys. Rev. Lett.* **96**, 100503 (2006).

[206] J. Cardy, "Measuring entanglement using quantum quenches", *Phys. Rev. Lett.* **106**, 150404 (2011).

[207] D. A. Abanin and E. Demler, "Measuring entanglement entropy of a generic many-body system with a quantum switch", *Phys. Rev. Lett.* **109**, 020504 (2012).

[208] H.-C. Jiang, R. R. P. Singh and L. Balents, "Accuracy of topological entanglement entropy on finite cylinders", *Phys. Rev. Lett.* **111**, 107205 (2013).

[209] R. G. Melko, A. B. Kallin and M. B. Hastings, "Finite-size scaling of mutual information in Monte Carlo simulations: Application to the spin-1/2 XXZ model", *Phys. Rev. B* **82** (10), 100409 (2010), ISSN 1098-0121, `arXiv:1007.2182`.

[210] R. R. P. Singh, M. B. Hastings, A. B. Kallin and R. G. Melko, "Finite-temperature critical behavior of mutual information", *Phys. Rev. Lett.* **106**, 135701 (2011).

[211] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge University Press (1996).

[212] M. Fagotti and P. Calabrese, "Entanglement entropy of two disjoint blocks in xy chains", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (04), P04016 (2010).

[213] M. Fagotti and P. Calabrese, "Universal parity effects in the entanglement entropy of xx chains with open boundary conditions", *Journal of Statistical Mechanics: Theory and Experiment* **2011** (01), P01017 (2011).

[214] P. Calabrese, J. Cardy and I. Peschel, "Corrections to scaling for block entanglement in massive spin chains", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (09), P09003 (2010).

[215] P. Calabrese and F. H. L. Essler, "Universal corrections to scaling for block entanglement in spin-1/2xx chains", *Journal of Statistical Mechanics: Theory and Experiment* **2010** (08), P08029 (2010).

[216] F. J. Wegner, "Corrections to scaling laws", *Phys. Rev. B* **5**, 4529–4536 (1972).

[217] A. Coser, L. Tagliacozzo and E. Tonni, "On renyi entropies of disjoint intervals in conformal field theory", *Journal of Statistical Mechanics: Theory and Experiment* **2014** (1), P01008 (2014).

[218] M.-C. Chung and I. Peschel, "Density-matrix spectra of solvable fermionic systems", *Phys. Rev. B* **64**, 064412 (2001).

[219] E. Lieb, T. Schultz and D. Mattis, "Two soluble models of an antiferromagnetic chain", *Ann. Phys.* **16**, 407 (1961).

[220] P. Di Francesco, P. Mathieu and D. Sénéchal, *Conformal Field Theory*, Springer Verlag (1997).

[221] M. E. Fisher and H. Au-Yang, "Inhomogeneous differential approximants for power series", *Journal of Physics A: Mathematical and General* **12** (10), 1677 (1979).

[222] D. L. Hunter and G. A. Baker, "Methods of series analysis. iii. integral approximant methods", *Phys. Rev. B* **19**, 3808–3821 (1979).

[223] J. Oitmaa, C. Hamer and W. Zheng, *Series Expansion Methods for strongly interacting lattice models*, Cambridge University Press (2006).

[224] D. J. Luitz, X. Plat, F. Alet and N. Laflorencie, "Universal logarithmic corrections to entanglement entropies in two dimensions with spontaneously broken continuous symmetries", *Phys. Rev. B* **91**, 155145 (2015).

[225] S. V. Isakov, M. B. Hastings and R. G. Melko, "Topological entanglement entropy of a bose-hubbard spin liquid", *Nat. Phys.* **7**, 772 (2011).

[226] A. J. Daley, H. Pichler, J. Schachenmayer and P. Zoller, "Measuring entanglement growth in quench dynamics of bosons in an optical lattice", *Phys. Rev. Lett.* **109**, 020505 (2012).

[227] P. Calabrese, M. Campostrini, F. Essler and B. Nienhuis, "Parity effects in the scaling of block entanglement in gapless spin chains", *Phys. Rev. Lett.* **104**, 095701 (2010).

[228] L. Wang, D. Poilblanc, Z.-C. Gu, X.-G. Wen and F. Verstraete, "Constructing a gapless spin-liquid state for the spin-1/2 J_1-J_2 heisenberg model on a square lattice", *Phys. Rev. Lett.* **111**, 037202 (2013).

[229] X. Chen, G. Y. Cho, T. Faulkner and E. Fradkin, "Scaling of entanglement in 2 + 1-dimensional scale-invariant field theories", *J. Stat. Mech.* **1502** (2), P02010 (2015), `arXiv:1412.3546`.

[230] T. Faulkner, R. G. Leigh and O. Parrikar, "Shape Dependence of Entanglement Entropy in Conformal Field Theories", *JHEP* **04**, 088 (2016), `arXiv:1511.05179`.

[231] T. Devakul and R. R. P. Singh, "Entanglement across a cubic interface in 3 + 1 dimensions", *Phys. Rev. B* **90**, 054415 (2014).

[232] M. Campostrini, A. Pelissetto and E. Vicari, "Finite-size scaling at quantum transitions", *Phys. Rev. B* **89**, 094516 (2014).

[233] H. E. Stanley, "Scaling, universality, and renormalization: Three pillars of modern critical phenomena", *Rev. Mod. Phys.* **71**, S358–S366 (1999).

[234] J. L. Cardy, "Is there a c-theorem in four dimensions?", *Physics Letters B* **215** (4), 749 – 752 (1988), ISSN 0370-2693.

[235] H. Elvang, D. Z. Freedman, L.-Y. Hung, M. Kiermaier, R. C. Myers and S. Theisen, "On renormalization group flows and the a-theorem in 6d", *Journal of High Energy Physics* **2012** (10), 1–43 (2012), ISSN 1029-8479.

[236] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from the anti˘de sitter space/conformal field theory correspondence", *Phys. Rev. Lett.* **96**, 181602 (2006).

[237] H. Casini, M. Huerta, R. C. Myers and A. Yale, "Mutual information and the F-theorem", *JHEP* **10**, 003 (2015), `arXiv:1506.06195`.

[238] H. Casini, C. D. Fosco and M. Huerta, "Entanglement and alpha entropies for a massive Dirac field in two dimensions", *J. Stat. Mech.* **0507**, P07007 (2005), arXiv:cond-mat/0505563.

[239] H. Casini and M. Huerta, "Remarks on the entanglement entropy for disconnected regions", *JHEP* **03**, 048 (2009), arXiv:0812.1773.

[240] R.-X. Miao, "A holographic proof of the universality of corner entanglement for CFTs", *JHEP* **10**, 038 (2015).

[241] H. Liu and M. Mezei, "A Refinement of entanglement entropy and the number of degrees of freedom", *JHEP* **04**, 162 (2013), arXiv:1202.2070.

[242] T. Zhou, X. Chen, T. Faulkner and E. Fradkin, "Entanglement Entropy and Mutual Information of Circular Entangling Surfaces in 2 + 1-dimensional Quantum Lifshitz Model", *J. Stat. Mech.* **1609** (9), 093101 (2016), arXiv:1607.01771.

[243] V. Alba, L. Tagliacozzo and P. Calabrese, "Entanglement entropy of two disjoint blocks in critical ising models", *Phys. Rev. B* **81**, 060411 (2010), arXiv:0910.0706.

[244] J.-M. Stéphan, S. Inglis, P. Fendley and R. G. Melko, "Geometric mutual information at classical critical points", *Phys. Rev. Lett.* **112**, 127204 (2014), arXiv:1312.3954.

[245] R. D. Sorkin, "On the Entropy of the vacuum outside a horizon", in B. Bertotti, F. de Felice and A. Pascolini (editors), "Tenth International Conference on General Relativity and Gravitation", Consiglio Nazionale Delle Ricerche, Roma (held in Padova, 4-9 July, 1983), (734–736), arXiv:1402.3589.