

Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of

Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Moritz Reh

born in: Mainz, Germany

Oral examination: 25.04.2024

FROM SPIN SYSTEMS TO BOSE-EINSTEIN CONDENSATES:
COMPUTATIONAL APPROACHES TO
STRONGLY CORRELATED QUANTUM MANY-BODY SYSTEMS

Referees: Prof. Dr. Martin Gärtner
Prof. Dr. Björn Malte Schäfer

Moritz Reh: *From Spin Systems to Bose-Einstein Condensates: Computational Approaches to Strongly Correlated Quantum Many-Body Systems*, © February 2024

To my Dad.

ABSTRACT

The numerical simulation of quantum many-body systems constitutes a long-standing and challenging problem, as the 'curse of dimensionality' restricts the applicability of exact methods to systems consisting of only a few particles. Thus approximative techniques that reduce the computational complexity are of high fundamental interest. Simultaneously, there exists a strong desire to benchmark the ever-growing capabilities of quantum simulators, thus strengthening the motivation to research tools that are capable of matching their increasing system sizes.

In this thesis, we, for one, develop and explore such new computational methods by exploiting the rapid developments in machine learning, allowing us to construct highly versatile ansatz functions to model quantum states based on deep artificial neural networks. Building on this, we establish a new numerical technique capable of modeling the dynamics of dissipative many-body quantum systems, relying on an accurate variational description of an informationally complete probability distribution that corresponds to the quantum system of interest. Additionally, we explore the differences in performance in ground state searches between a multitude of different network architectures and thereby shed light on the question of why some networks significantly outperform others.

Secondly, we adapt the developed techniques also for classical systems. This is possible as the only requirement is a probabilistic description with a (closed) evolution equation, thereby emphasizing the wide range of applicability.

Finally, we rely on existing approximative techniques to devise an experimental proposal aimed at observing an area to a volume law transition following a quench in a spin-1 Bose-Einstein condensate. Notably, we herein do not rely on quantum entropies but rather on differential entropies of the phase-space distribution describing the system. These quasi probability distributions are importantly readily accessible in experiments and we demonstrate that their entropies can be reliably estimated from a feasible number of samples without assuming a particular type of distribution, such as a Gaussian.

ZUSAMMENFASSUNG

Die numerische Simulation von Quantenvielteilchensystemen stellt ein seit langem bestehendes und herausforderndes Problem dar, da der 'Fluch der Dimensionalität' die Anwendbarkeit exakter Methoden auf Systeme mit wenigen Teilchen beschränkt. Deshalb sind approximative Techniken, welche die algorithmische Komplexität senken, von hohem wissenschaftlichen Interesse. Außerdem besteht ein großes Interesse daran, die stetig wachsenden Fähigkeiten von Quantensimulatoren zu verifizieren, was die Motivation, solche Methoden zu erforschen, weiter bestärkt.

In dieser Arbeit werden einerseits solche neuartigen computergestützten Methoden entwickelt und untersucht, wobei wir uns die rasanten Entwicklungen im Feld des maschinellen Lernens zunutze machen. Dies erlaubt, hochgradig vielseitige Ansatzfunktionen für Quantenzustände auf Basis von tiefen künstlichen neuronalen Netzen zu konstruieren. Darauf aufbauend etablieren wir eine neue numerische Methode die es ermöglicht die dissipative Dynamik von Quantenvielteilchensystemen zu modellieren, indem die informationstheoretisch vollständige Wahrscheinlichkeitsverteilung, welche zu dem interessierenden Quantensystem korrespondiert, variatonell dargestellt wird. Zusätzlich arbeiten wir für eine Vielzahl von neuronalen Netzwerkarchitekturen Unterschiede in den Fähigkeiten, Grundzustände zu finden, heraus und können so erklären, weshalb manche Architekturen in dieser Aufgabe deutlich besser abschneiden als andere.

Zweitens werden die entwickelten Techniken für klassische Systeme adaptiert. Dies ist möglich, da die einzige Voraussetzung hierfür eine probabilistische Beschreibung des Systems ist, welche einer (geschlossenen) Entwicklungsgleichung gehorcht, was die Vielzahl an Anwendungsmöglichkeiten hervorhebt.

Schließlich werden wir mit Hilfe von etablierten numerischen Methoden einen Vorschlag zur experimentellen Umsetzung ausarbeiten, welcher zum Ziel hat, den Übergang eines 'area laws' hin zu einem 'volume laws' in einem Spin-1 Bose-Einstein Kondensat nach einem Quench zu beobachten. Es ist erwähnenswert, dass wir hierfür keine Quantenentropien, sondern differentielle Entropien von Phasenraumverteilungen betrachten, welche das System beschreiben. Wichtig ist hierbei, dass diese Quasi-Wahrscheinlichkeitsverteilungen in Experimenten zugänglich sind und wir demonstrieren, dass ihre Entropien zuverlässig aus einer überschaubaren Anzahl an Samples geschätzt werden können, ohne dass wir dabei die Form der Verteilung auf Gauss'sche Verteilungen einschränken müssen.

PUBLICATIONS

This thesis is based on the following manuscripts and publications:

- [A] **Moritz Reh**, M. Schmitt, and M. Gärttner, “Time-Dependent Variational Principle for Open Quantum Systems with Artificial Neural Networks,” *Physical Review Letters* **127**, 230501 (2021).
- [B] **Moritz Reh** and M. Gärttner, “Variational Monte Carlo approach to partial differential equations with neural networks,” *Machine Learning: Science and Technology* **3**, 04LT02 (2022).
- [C] **Moritz Reh**, M. Schmitt, and M. Gärttner, “Optimizing design choices for neural quantum states,” *Physical Review B* **107**, 195115 (2023).
- [D] Y. Deller, M. Gärttner, T. Haas, M. K. Oberthaler, **Moritz Reh**, and H. Strobel, “Area laws for classical entropies in a spin-1 Bose-Einstein condensate,” In preparation (2024).
- [E] Y. Deller, M. Gärttner, T. Haas, M. K. Oberthaler, **Moritz Reh**, and H. Strobel, “Area laws and thermalization from classical entropies in spin-1 Bose-Einstein condensate,” In preparation (2024).

The author also contributed to the following publications:

- [F] T. Schmale, **Moritz Reh**, and M. Gärttner, “Efficient quantum state tomography with convolutional neural networks,” *npj Quantum Information* **8**, 115 (2022).
- [G] M. Schmitt and **Moritz Reh**, “jVMC: Versatile and performant variational Monte Carlo leveraging automated differentiation and GPU acceleration,” *SciPost Physics Codebases*, 2 (2022).
- [H] M. Rieger, **Moritz Reh**, and M. Gärttner, “Sample-efficient estimation of entanglement entropy through supervised learning,” *Phys. Rev. A* **109**, 012403 (2024).

CONTENTS

I Background

1	Introduction	3
1.1	Quantum Simulation	4
1.2	Neural Quantum States	9
2	Quantum Theory	13
2.1	Mathematical Formalism	14
2.2	Subsystems, Entanglement and Quantum Information	16
2.3	Dynamics of Open Quantum Systems	18
2.4	Positive Operator Valued Measures (POVMs)	20
2.5	Bosonic Modes	23
2.6	Phase-Space Representations of Bosonic Modes	25
2.7	Spin-1 Bose-Einstein Condensates	28
2.8	Thermalization of Closed Quantum Systems	31
3	Machine Learning	33
3.1	Generative Modeling	34
3.1.1	Sampling from high-dimensional Distributions	34
3.1.2	Expressivity	36
3.2	Neural Network Architectures	36
3.2.1	Restricted Boltzmann Machines (RBMs)	37
3.2.2	Convolutional Neural Networks (CNNs)	38
3.2.3	Recurrent Neural Networks (RNNs)	39
3.2.4	Graph Neural Networks (GNNs)	41
3.2.5	Invertible Neural Networks (INNs)	42
4	Classical Simulation Techniques	45
4.1	Neural Quantum States	46
4.1.1	Expectation Values of Operators	48
4.1.2	Generation of Samples	48
4.1.3	Ground State Optimization	49
4.1.4	Unitary Time Evolution	50
4.2	Truncated Wigner Approximation	51

II Results

5	Time-Dependent Variational Principle for Open Quantum Systems with Artificial Neural Networks	57
5.1	Contributions	57
5.2	Motivation	57
6	Variational Monte Carlo Approach to Partial Differential Equations with Neural Networks	71
6.1	Contributions	71
6.2	Motivation	71
7	Optimizing Design Choices for Neural Quantum States	85

7.1	Contributions	85
7.2	Motivation	85
8	Area laws for Classical Entropies in a Spin-1 Bose-Einstein Condensate	97
8.1	Contributions	97
8.2	Motivation	97
9	Area laws and Thermalization from Classical Entropies in a Spin-1 Bose-Einstein Condensate	127
9.1	Contributions	127
9.2	Motivation	127
10	Conclusion and Outlook	139
10.1	Neural Quantum States	139
10.2	Generalizations to Classical Dynamics	141
10.3	Area and Volume Laws in Bose-Einstein Condensates	141
	Bibliography	143

LIST OF FIGURES

Figure 1.1	Superfluid to Mott-Insulator Phase transition.	6
Figure 1.2	Thermalization in a one-dimensional Bose-Hubbard model.	7
Figure 1.3	Adiabatic ground state preparation in the transverse field Ising model.	8
Figure 1.4	Ground state energies for paradigmatic spin models obtained with neural quantum states.	10
Figure 1.5	Entanglement entropy estimation following a state reconstruction using neural quantum states.	11
Figure 2.1	Visualisation of a squeezed, Gaussian state.	13
Figure 2.2	Rabi rotations on the Bloch sphere.	15
Figure 2.3	Bipartition and entanglement growth in lattice systems.	17
Figure 2.4	Dissipative dynamics in the Monte Carlo wave function framework.	19
Figure 2.5	Choices of informationally complete POVMs for a single spin-1/2.	20
Figure 2.6	The probability simplex and the subset of physical states.	24
Figure 2.7	Wigner W - and Husimi Q -distribution of the vacuum and a Fock-state.	26
Figure 2.8	TWA simulation of BEC single well dynamics.	30
Figure 3.1	Popular activation functions.	37
Figure 3.2	Feed forward architecture and a restricted Boltzmann machine.	38
Figure 3.3	Depiction of a recurrent neural network.	40
Figure 3.4	Computations within a single coupling block of a normalizing flow.	43

NOTATION

We will work with natural units $\hbar = k_B = c = 1$. Furthermore, we will denote matrices as uppercase letters O , operators as boldface letters \mathbf{O} , vectors using arrows \vec{o} and scalars as lowercase letters o . In special instances, we will use a calligraphic font, such as \mathcal{H} to denote Hilbert spaces. All other notation is explained at the first point of occurrence.

Part I

BACKGROUND

INTRODUCTION

Nature seems predictable if judged from everyday experience. Given perfect knowledge of the state of a system and sufficient computational resources, one might argue, it should be possible to obtain unambiguous predictions with regards to measurement results of some system of interest. This is not so.

At the level of minuscule scales one finds that the world surrounding us is fundamentally non-deterministic and that even perfect knowledge of the state of a system does not suffice to give a definite answer on the outcome of an experiment¹ [2]. This observation was puzzling for those physicists who initially made contact with this ‘quantum world’ of small scales, famously leading Einstein to state that ‘*God does not play dice*’, a statement at odds with the observations made in quantum experiments. At present, one hundred years have passed since the Nobel prize for the discovery of the photoelectric effect was awarded to Einstein for demonstrating the quantization of light and marking the advent of a new era in physics [3].

Today, quantum mechanics is a well understood field and part of every physics curriculum. The non-deterministic nature of the universe is widely accepted and quantum properties are sought to be exploited to gain advantages over classical devices. Experimental progress regarding the design and control of quantum experiments over the last decades has been astonishing, allowing to probe various predictions of quantum theory in a wide variety of settings. Of particular theoretical and experimental interest, especially in the context of this thesis, is the emulation of toy-model Hamiltonians on suited platforms, allowing to demonstrate fundamental quantum many-body phenomena in a controlled setting. This idea is referred to as *quantum simulation* [4–7]. Quantum simulation is to be differentiated from the more general goal of *quantum computing* [2, 8, 9], in that quantum computing aims to implement *arbitrary* unitary transformations on quantum states, whereas quantum simulation aims to implement only such transformations that are *physically motivated*, meaning that they stem from an easy-to-implement Hamiltonian or similar. Depending on the system of interest, one may choose a quantum simulation platform from a wide variety of possible setups, including cold atoms [10–12], photonics [13, 14], trapped ions [15, 16] or nuclear or electronic spins as encountered for example in NMR experiments [17].

Although it was Einstein who had first observed glimpses of the quantum world, he remained a skeptic of quantum theory until his death, with a dislike for its indeterministic and thereby ‘incomplete’ nature (see EPR-paradox [1]).

¹ We here omit a discussion regarding the similarities and differences to classical chaotic systems.

This work is mainly concerned with the classical simulation of quantum many-body systems. A central motivation herein is that the ever-growing capabilities of quantum simulation experiments require classical benchmarks to assert their correct working. This motivates the exploration of more capable classical computation schemes that can keep up with the increasing system sizes of modern quantum simulators. Simultaneously, a parallel line of motivation is aimed at probing fundamental questions of interest to quantum many-body theory in otherwise inaccessible system size regimes. In the remainder of this chapter, we will therefore present and discuss milestones in the fields that are most relevant to this thesis — for one, quantum simulation and, secondly, a novel, classical computational technique referred to as neural quantum states. The following chapters will introduce these topics more formally, giving a more complete discussion of the theoretical background.

The results of this thesis can be separated into three categories: [Chapter 5](#) and [Chapter 7](#) are dedicated to the development and understanding of novel classical computation schemes of quantum systems, motivated by the aforementioned necessity to benchmark quantum simulation platforms as well as the desire to fundamentally push the boundaries of classical computation techniques to numerically probe the collective behavior in larger quantum systems.

The second part concerns the extension of the developed methodologies from [Chapter 5](#) to classical systems. These findings are summarized in [Chapter 6](#).

Finally, the third part consisting of [Chapter 8](#) and [Chapter 9](#) discusses a proposal on how one may test the presence of the so-called area-law in a particular quantum simulation platform, namely a Bose-Einstein condensate (BEC). Here, we will rely on classical computational methods to demonstrate the feasibility of the proposed experiment but do not attempt to further develop the utilized computational techniques.

1.1 QUANTUM SIMULATION

We herein give a (non-extensive) overview of some of the developments and theoretical predictions that were tested in quantum simulation experiments while explaining their physical significance. Simultaneously, this will give us the opportunity to introduce many of the physically relevant phenomena and models that we will come back to at later points in this thesis.

As quantum theory does not lend itself easily to classical computation, obtaining quantitative predictions of a system of interest too complicated for analytical solutions has always been challenging. This led Richard Feynman to the idea that one might be able to emulate this quantum system of interest using *another* quantum system, and

thus the idea of quantum simulation was born [4]. Since its conception, this idea has been a popular avenue to test various theoretical predictions in table-top experiments in the laboratory. While many of these experiments are concerned with testing predictions of fundamental interest in quantum theory [10, 12, 15, 18–66], there also exist strong efforts to answer questions of practical significance, such as material design and molecular research [67–71]. We will in the following limit our considerations to the former category, as it is of higher relevance to the work at hand.

One of the earliest examples, that showcased the potential of this idea was the implementation of the Bose-Hubbard model using a gas of ultracold rubidium atoms confined in a periodic lattice [18]. In their experiment, Greiner et. al. demonstrated the existence of a quantum phase transition in the Bose-Hubbard model, preparing ground states in both the superfluid and Mott-insulator phase in the low occupancy limit with $\langle \mathbf{n} \rangle = 1 - 3$ atoms per site. The Bose-Hubbard model features repulsive on-site interactions of strength U as well as tunneling between neighboring sites with strength J . The energy of the density-density interactions is minimized if atoms are distributed homogeneously over the lattice sites, leading to a ground state that is characterized by definite real-space occupation numbers and no phase coherence among lattice sites. In contrast, once the tunneling contribution becomes the dominant energy scale, the ‘good’ quantum numbers shift from real-space occupations to momentum-space occupations and phase coherence between sites is established, while atoms are delocalized over many lattice sites paying tribute to the Heisenberg uncertainty principle. While an analytical solution in the limit $U/J = 0$ or $J/U = 0$ is possible, obtaining the system’s ground state in the regime where the two contributions are approximately equally important is a challenging task. The critical value J_c for which the quantum phase transition occurs is dependent on the number of atoms in the system, or, equivalently, its chemical potential as well as its dimensionality and has been the subject of intense study [72–75]. Figure 1.1 shows the measured momentum space occupations in two dimensions, with the center of the image corresponding to momenta in x and y direction equal to zero, i.e. the kinetic eigenstates with the lowest energy. As the ratio U/J of on-site repulsion relative to tunneling strength is increased, phase coherence is reduced, causing a broadened distribution in momentum space.

While the phases of ground states of different models show a wide variety of captivating phenomena, another compelling research direction concerns the dynamic behavior of quantum systems. One typically prepares the system in an *easy* state² and then rapidly alters the parameters of the Hamiltonian, leaving no time for the system

² *Easy* typically means adjusting the control parameters, such as external fields for example, so that there is no entanglement between the system constituents.

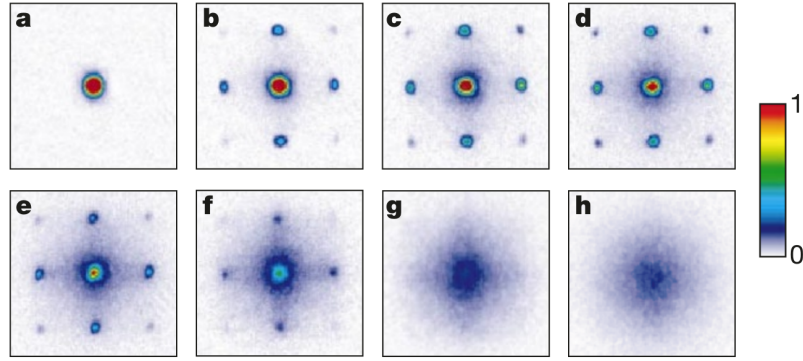


Figure 1.1: Absorption images following a free evolution of $t = 15\text{ms}$ after ground state preparation with different ratios J/U . The color-bar gives the occupations in k -space. The depth of the wells is increased from **a** to **h**, with all atoms in the $k = 0$ mode in **a** and no visible structure in k -space for high potential barriers in **h**. Adapted with permission from [18].

In the opposing limit, where one alters the parameters sufficiently slowly, the system continuously evolves to the new ground state. This is known as adiabatic state preparation [79].

If the system for example features sufficient disorder, the system stays 'close' to its original state and does not thermalize. This is referred to as localization [80].

to adjust to the new settings such that the sudden quench takes it far from equilibrium. Since the evolution of closed quantum systems obeys unitary transformations generated by the system's Hamiltonian, no information is lost during the evolution implying that it is reversible. During the evolution, the interaction among particles leads to a buildup of entanglement, such that initially localized information is coherently spread throughout the system. This means that a subsystem cannot be described by a single wave function but rather by a statistical mixture of wave functions. One finds that this statistical mixture becomes stationary in the long time limit under relatively mild constraints, raising the question as to what characterizes said statistical mixture and *how* the system approaches the stationary state. The first question has a surprisingly simple and beautiful answer. It turns out that the stationary state may oftentimes be described by a Gibbs ensemble, meaning that a single parameter, namely the temperature, is sufficient to describe the system in the long time limit. One therefore also refers to the process of approaching the equilibrated steady state as *thermalization*. The notion of temperature here arises due to our ignorance of the subsystem's complement, since we disregard its contained information, thereby increasing the (quantum) entropy. This is the same principle that leads to the notion of temperature in classical statistical mechanics, where one is ignorant about the individual positions and momenta of the macroscopically many particles, thereby making an ensemble description necessary. Quantum thermalization was first demonstrated using ultra-cold atoms in a BEC, again relying on an implementation of the Bose-Hubbard model [78], as sketched in Figure 1.2 A. Following a quench from deep inside Mott-insulator phase, Kaufman et. al. studied the single-site occupations and found them to be in agreement with the predictions from a thermal ensemble

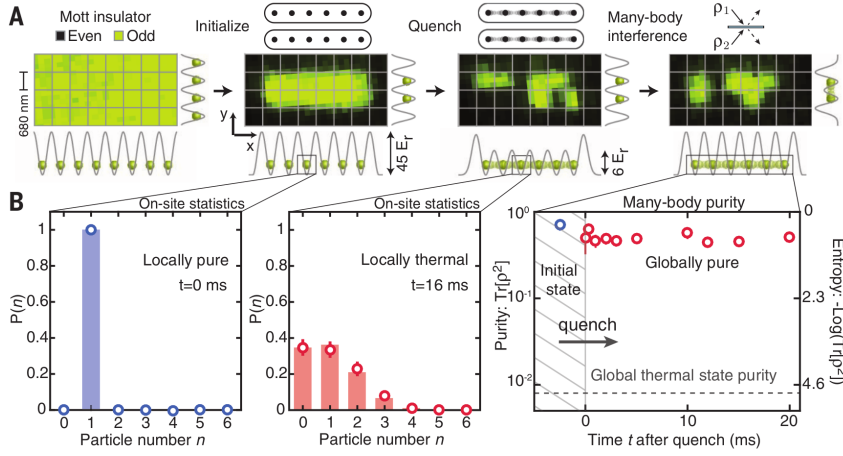


Figure 1.2: Thermalization in a one-dimensional Bose-Hubbard model following a quench from the Mott-insulator phase with a single atom per site. The quench is achieved by suddenly decreasing the potentials, such that atoms start tunneling between neighboring sites in x -direction (third panel in A). After a free evolution time of 16 ms, one finds the single site number occupations to be in agreement with a local thermal state (second panel in B), while the global state is still pure (third panel in C), as measured by two-copy interference in y -direction [20, 76, 77]. From [78]. Reprinted with permission from AAAS.

(second panel in Figure 1.2 B) while the purity of the full system is not in agreement with thermal predictions (third panel in Figure 1.2 B), thereby demonstrating the emergence of *local* thermal descriptions due to the generated entanglement between the single site and its complement.

Another phenomenon that is of paramount importance in the theory of many-body quantum systems is the emergence of magnetism in solid states. As quantum magnetism arises due to the internal spin degree of freedom of atoms that are located at fixed sites (such that all motional degrees of freedom are frozen out), any platform that allows to precisely control it may be deemed suited a priori. Various toy-model Hamiltonians exist to describe the wide range of quantum spin phases and phenomena [81–85]. The paradigmatic *classical* spin model is the Ising model, showing a classical phase transition driven by *thermal* fluctuations between an ordered and disordered phase. The paradigmatic *quantum* spin model is the *transverse-field* Ising model, featuring a quantum phase transition at zero temperature driven by *quantum* fluctuations between a ferromagnetic and paramagnetic phase. The transverse-field Ising model is a popular object of study, as it is integrable in one dimension by means of a Jordan-Wigner transformation, mapping the spin system to a system of non-interacting fermions [81]. It consists of two competing terms, namely interactions along the axis of quantization of nearest neighbors with strength J

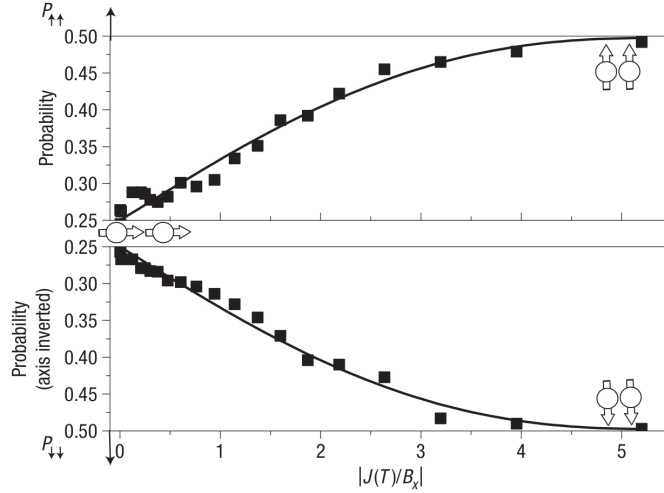


Figure 1.3: Probability of obtaining the outcome $|\uparrow\uparrow\rangle$ and $|\downarrow\downarrow\rangle$ following an adiabatic protocol from the initial paramagnetic state $|\psi\rangle = |\rightarrow\rightarrow\rangle$ to the final Hamiltonian parameters given on the x -axis. Adapted with permission from [79].

as well as an external magnetic field with strength B oriented orthogonally to the quantization axis. One of the earliest works that implemented the transverse-field Ising model experimentally is [79], demonstrating the preparation of (anti-) ferromagnetic ground states using an adiabatic protocol from the initial paramagnetic ground state for a system of two trapped ions, as shown in Figure 1.3. While the initial paramagnetic state ($B \gg J$) $|\psi\rangle = |\rightarrow\rightarrow\rangle = (|\uparrow\rangle + |\downarrow\rangle)(|\uparrow\rangle + |\downarrow\rangle)$ has uniform probability 0.25 for all 4 basis states, the ferromagnetic state that is obtained for $J \gg B$ only has non-vanishing probability for $|\psi\rangle = |\uparrow\uparrow\rangle$ and $|\psi\rangle = |\downarrow\downarrow\rangle$. Importantly, the authors demonstrate by means of parity measurements that the ferromagnetic configurations are in a coherent superposition instead of a statistical mixture, meaning that the adiabatic protocol indeed has generated entanglement between the two spins.

Also, the dynamics of spin systems is of high interest [86] and has recently been subject to experimental studies in a system consisting of 127 qubits [87]. Systems of such extent are no longer simulatable by means of exact computational schemes but require treatment with more advanced, approximative methods [88–92]. Of particular interest are variational schemes, which do not simplify the Hamiltonian of the system in question but rather try to find the best approximation to the state by adapting its variational parameters. With this motivation in mind, let us discuss machine learning inspired approaches to tackle such problems.

1.2 NEURAL QUANTUM STATES

Systems consisting of only a few dozen qubits are already sufficient to push the largest supercomputing clusters to their limits. This is due to the ‘*curse of dimensionality*’, meaning that each additional qubit that is added to the system, doubles the memory required to store its state. While this property of quantum mechanical systems may seem like an insurmountable barrier, there may be scenarios that are nevertheless within the reach of classical computation. While Hilbert space is vast, *typical* states, such as ground states or those obtained from quenches, are located in small regions thereof. Therefore, if one is able to *efficiently* parameterize these regions, there may still be hope to be able to simulate the system from first principles, meaning without a simplification of its Hamiltonian. The question that follows, of course, is what kind of ansatz function is a suitable choice to parameterize quantum states.

Efficiently here means without an exponential growth in computational cost with system size.

The first class of ansatz functions that were developed, were concerned with particle systems and were physically motivated in that they introduced factors to directly control physical properties, such as double occupancies of sites or similar. These approaches include Hartree-Fock wave functions, Gutzwiller wave functions, density-density Jastrow wave functions, and many more [93, 94]. Being derived from physical intuition, these classes of wave functions had considerable success due to their strong physical prior.

For spin systems, tensor network states were the unchallenged variational method [95, 96] until recently. Tensor networks assign a tensor of variable rank to each spin so that each spin has connections to its neighbors. Upon *contraction* of the tensor network, the tensors are turned into a wave function coefficient of a desired basis state. This will be explained in more detail in [Chapter 4](#). The parameters of tensor network states, i.e. the numbers that define the tensor entries, have no direct physical interpretation, unlike those of the Gutzwiller wave function for instance, which restrict double occupancies [93]. However, matrix product states are still physically motivated as they obey the principle of *locality*, meaning that only neighboring tensors are contracted and that correlations between spins that are more than one site apart must be mediated through the encompassed tensors, akin to how correlations spread in models with nearest-neighbor interactions. This makes tensor network states a powerful class of ansatz functions in one dimension. However, in two- or three-dimensional settings their performance is diminished [97–100], as will be discussed in greater detail in [Chapter 4](#). This motivates the exploration of more potent variational ansatz functions, that can also flexibly handle higher dimensional problems.

One such class of ansatz functions is given by artificial neural networks, whose exploration is motivated by *universal approximation*

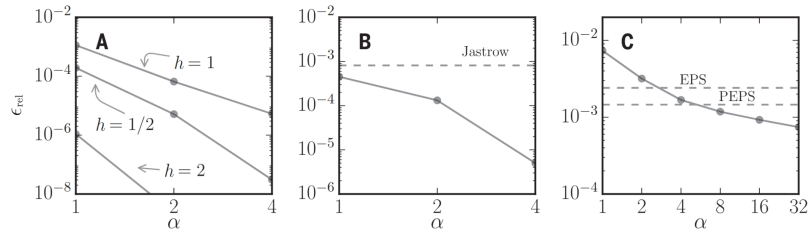


Figure 1.4: Ground state energy deviations obtained with the RBM as a function of its complexity α in the **A** 1D transverse field Ising model (80 sites) for different magnetic fields h , **B** 1D antiferromagnetic Heisenberg model (80 sites) and **C** 2D antiferromagnetic Heisenberg model (10×10 sites). From [121]. Reprinted with permission from AAAS.

Such UATs do however not simplify finding the optimal set of weights. One may still get stuck in local minima during optimization.

theorems (UAT) [101–119]. Universal approximation theorems guarantee the capability of approximating sufficiently smooth functions arbitrarily well in the appropriate limits of depth and width of specific network architectures. The ability to scale towards these limits was made possible by the availability of unprecedented GPU compute resources, leading to a revolution in deep learning, which attracted attention well beyond the computer science community ever since AlexNet won the ImageNet competition in 2012 [120]. Interest in these novel methods was quickly ignited within the scientific communities, with computational quantum physics being no exception. The possibility of utilizing neural networks as variational ansatz functions for quantum states was first explored by Carleo and Troyer in [121]. Neural quantum states assign complex wave function coefficients to basis configurations, with the latter being the input and the former being the output of the network. This general idea has been demonstrated to accurately describe many scenarios of physical interest, such as spin systems as well as bosonic and fermionic particles in both discrete and continuous space.

In their 2017 paper, Carleo and Troyer introduced the idea of neural quantum states [121], demonstrating the possibility of modeling ground states as well as time-evolved states in one- and two-dimensional spin systems using artificial neural networks to encode the wave function coefficients. They rely on the simple restricted Boltzmann machine (RBM) architecture and employ complex parameters for generality, facilitating, for example, phase accumulation during real-time evolution. By employing the second-order stochastic reconfiguration (SR) optimization scheme, the RBM outperforms the benchmarks given by the Jastrow wave function, as well as the projected entangled pair states in terms of ground state energy, as shown in Figure 1.4. Simultaneously, the same optimization routine also enables real-time evolution.

After their conception, neural quantum states quickly became popular and were adapted for further tasks, such as state reconstruction

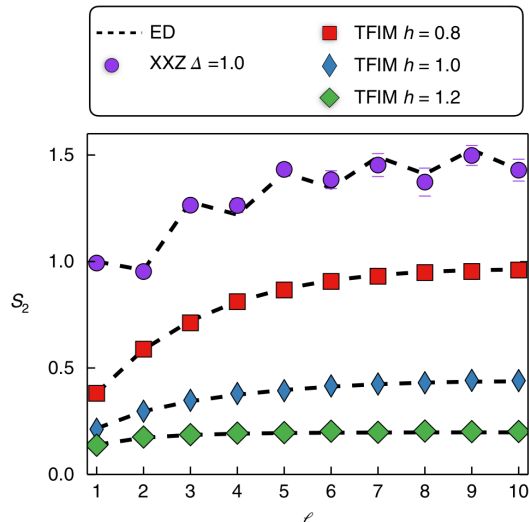


Figure 1.5: Estimation of the Rényi-2 entanglement entropy of the first l spins in a one-dimensional chain of length 20. The model is trained on synthetic samples from ground states of the XXZ and TFIM model at various field strengths. Adapted with permission from [122].

from experimental data, also known as *quantum state tomography*. A characterization of all present degrees of freedom in an experiment is a daunting task, since naïvely this number is exponentially large while the set of available measurement data usually is not. This apparent contradiction cannot be resolved easily, since the information contained in the set of measurement data is fundamentally limited and insufficient to uniquely determine the experimentally prepared state. However, if the state obeys a certain functional form, only the parameters defining this functional form need to be determined, significantly simplifying the task at hand as the number of parameters is typically much lower than the Hilbert space dimension. Additionally, not only the determination of these parameters becomes manageable, but also their storage, with the memory requirements given by the number of parameters. In the past, quantum state tomography has been carried out using various different variational ansatz functions, such as tensor network approaches [123–125], compressed sensing techniques [126–128] and permutationally invariant tomography schemes [129, 130]. The common feature that these approaches share is that the physical implications of their simplifying assumptions are well understood: tensor network states are biased towards lower entanglement, compressed sensing assumes a low-rank density matrix and permutationally invariant tomography requires a permutationally invariant state [F]. The same cannot be said about neural quantum states, as the bias introduced by the network is poorly understood in regard to its implications regarding physical properties. Nevertheless, quantum state tomography using neural networks is a promising application

Bias here refers to the tendency to prefer certain classes of states over others.

as the existence of universal approximation theorems guarantees that this bias can be systematically minimized by choosing a network of larger size. It has first been tested in Ref. [122], with applications to one- and two-dimensional lattice spin systems. Given a sample set, the authors optimize the parameters of the network in order to maximize the likelihood of observing it. Using only a fraction of the samples required for standard full-state tomography, accurate estimation of the entanglement entropy of a subsystem is demonstrated, as shown in Figure 1.5 by employing the replica trick [131].

Both data-free and data-driven neural quantum states have been in the focus of investigation over the past years, and even hybrid approaches have been explored [132, 133]. Today, neural quantum states give state-of-the-art results for ground states of various spin models [C, 97–99, 121, 134–153], their unitary dynamics [154–161] and have additionally been extended to dissipative systems [A, 162–168]. Fermionic [169–181] and bosonic [167, 168, 182–185] systems have also been explored, in both discrete and continuous settings. Of particular interest are works that study the entanglement properties of neural quantum states [186–188], promising an edge compared to tensor network states.

In this chapter, we will give the background in quantum theory that will be required in the remainder of this thesis. To give a comprehensive overview, we also discuss fundamental concepts, which the experienced reader can omit. Unless explicitly stated differently, we assume spin-1/2 systems for our discussion.

The perhaps most striking difference that separates classical and quantum mechanical systems is *definiteness* regarding the outcome of experiments, as the result of any experiment conducted on a classical system is uniquely determined by the initial configurations of its constituents. Even classical systems that only allow for statistical (ensemble) descriptions, such as an ideal gas, are made up of particles for which one can in principle give precise coordinates in phase space and solve the corresponding equations of motion; it is simply not practical to do so. This is in stark contrast to systems made up of sufficiently small constituents, for which the Heisenberg uncertainty principle states that the position x and momentum p of any particle cannot be determined precisely at the same time, leading to the Heisenberg uncertainty relation

$$\Delta x \Delta p \geq \frac{1}{2}, \quad (2.1)$$

where Δx and Δp are the standard deviations of position and momentum obtained through repeated measurements. This property (and its generalizations) is an inherent ingredient of quantum theory and is reflected therein as follows. In contrast to classical particles in one dimension, which are described by arbitrarily peaked distributions in the two-dimensional phase space, the wave function depends on

Uncertainty relations can be formulated for any pair of non-commuting observables.

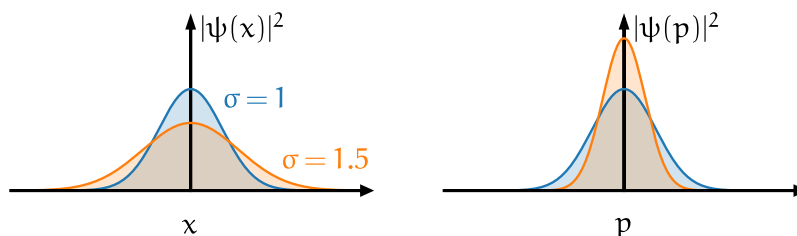


Figure 2.1: Demonstration of the Heisenberg uncertainty principle. A reduction in uncertainty in momentum space comes at the price of increasing uncertainty with regard to position, as indicated by the orange density profile.

The reduction and increase of uncertainty in a pair of non-commuting observables is also referred to as squeezing.

either position or momentum. Hence, given a wave function in real space, its properties in momentum space are implied (and vice versa) such that the Heisenberg uncertainty principle Eq. (2.1) is *automatically* fulfilled. If, for instance, the wave-function of a state with minimal uncertainty (meaning equality in Eq. (2.1)) produces a Gaussian probability density profile in real space that is stretched by a factor of σ , the associated density profile in momentum space is contracted by a factor $1/\sigma$, such that the product of the two is unchanged, as shown in Figure 2.1.

2.1 MATHEMATICAL FORMALISM

While continuous systems are ideally suited to visualize phenomena such as squeezing, we will be mainly concerned with discrete quantum systems in this thesis, such as spins or bosonic modes.

The arguably simplest discrete quantum system is a single particle with spin-1/2, which is characterized by a wave function $|\psi\rangle$ in a two-dimensional Hilbert space \mathcal{H} . The Hilbert space is spanned by the two orthogonal basis states $|\uparrow\rangle$ and $|\downarrow\rangle$ (usually understood to denote the z -axis $|\uparrow\rangle \equiv |\uparrow_z\rangle$ and $|\downarrow\rangle \equiv |\downarrow_z\rangle$), so that projections of the wave function onto these basis states uniquely identify $|\psi\rangle$. One obtains the x - and y -eigenstates as linear combinations of the z basis states

$$\begin{aligned} |\uparrow_x\rangle &= \frac{1}{\sqrt{2}} (|\uparrow\rangle + |\downarrow\rangle), & |\downarrow_x\rangle &= \frac{1}{\sqrt{2}} (|\uparrow\rangle - |\downarrow\rangle), \\ |\uparrow_y\rangle &= \frac{1}{\sqrt{2}} (|\uparrow\rangle + i|\downarrow\rangle), & |\downarrow_y\rangle &= \frac{1}{\sqrt{2}} (|\uparrow\rangle - i|\downarrow\rangle). \end{aligned} \quad (2.2)$$

The complex wave function coefficients $\langle\uparrow|\psi\rangle$ and $\langle\downarrow|\psi\rangle$ are not directly measurable experimentally, but rather determine the probability of the observations

$$P_\uparrow = |\langle\uparrow|\psi\rangle|^2, \quad P_\downarrow = |\langle\downarrow|\psi\rangle|^2 = 1 - P_\uparrow, \quad (2.3)$$

which are in practice inferred through repetition. To gather sufficient information to uniquely determine also the relative phases of the coefficients, measurements in additional bases are required, usually conveniently chosen as the mutually unbiased bases in x - and y -direction [2]. Equivalently, one can uniquely identify a spin-1/2 wave function by giving the expectation values of the three elementary spin-1/2 operators, the Pauli matrices,

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (2.4)$$

for which we identified

$$|\uparrow\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |\downarrow\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (2.5)$$

These expectation values are conveniently represented on the Bloch-sphere, as shown in Figure 2.2, allowing a very intuitive understanding of the simplest quantum system.

Importantly, the Pauli matrices do not commute, as

$$[\sigma_i, \sigma_j] = \sigma_i \sigma_j - \sigma_j \sigma_i = 2i \varepsilon_{ijk} \sigma_k, \quad (2.6)$$

with ε_{ijk} the Levi-Civita symbol, implying that angular momentum cannot be measured simultaneously with arbitrary precision in different directions.

As any hermitian operator, the Pauli matrices can be constructed by an *eigendecomposition*, yielding

$$\sigma_d = |\uparrow_d\rangle \langle \uparrow_d| - |\downarrow_d\rangle \langle \downarrow_d|, \quad (2.7)$$

with the outer product $|\cdot\rangle \langle \cdot|$ and $d \in \{x, y, z\}$. Expectation values thereof are obtained as

$$\langle \sigma_d \rangle = \langle \psi | \sigma_d | \psi \rangle, \quad (2.8)$$

with $\langle \sigma_d \rangle$ guaranteed to be real due to the hermiticity of σ_d .

Systems consisting of N spin-1/2 particles are described by wave functions on Hilbert spaces that are constructed as tensor products of the single particle Hilbert spaces,

$$\mathcal{H} = \mathcal{H}_{1/2} \otimes \mathcal{H}_{1/2} \otimes \dots \otimes \mathcal{H}_{1/2}. \quad (2.9)$$

Consequently, the wave function is uniquely determined by 2^N complex coefficients, demonstrating the curse of dimensionality of the quantum many-body problem.

The evolution of a closed quantum system is governed by its Hamiltonian \mathbf{H} through the Schrödinger equation

$$i\partial_t |\psi\rangle = \mathbf{H} |\psi\rangle. \quad (2.10)$$

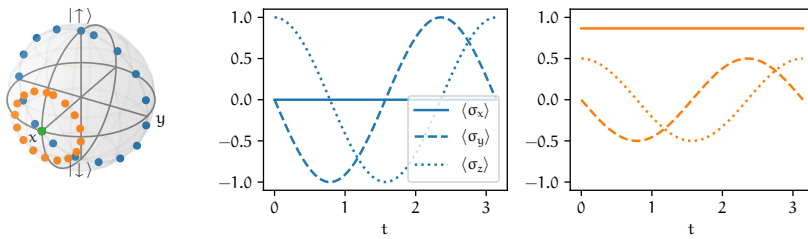


Figure 2.2: Unitary evolution under $\mathbf{H} = \sigma_x$ of a single spin-1/2 depicted on the Bloch sphere. The spin is initially prepared in the state $|\psi\rangle = \cos(\alpha) |\uparrow\rangle + \sin(\alpha) |\downarrow\rangle$, for $\alpha = 0$ ($\pi/6$, $\pi/4$) shown in blue (orange, green). The closer the initial state is to the eigenstate $|\uparrow\rangle_x$, the smaller the time-dependent signal becomes until it finally vanishes completely (green data point on the Bloch sphere).

For time-independent Hamiltonians, it is formally solved by the operator exponential

$$|\psi(t)\rangle = e^{-i\mathbf{H}t} |\psi(t=0)\rangle = \mathbf{U} |\psi(t=0)\rangle, \quad (2.11)$$

where \mathbf{U} is the unitary evolution operator generated by \mathbf{H} .

2.2 SUBSYSTEMS, ENTANGLEMENT AND QUANTUM INFORMATION

Wave functions describe *pure* quantum systems, which are perfectly shielded from the environment. Experimentally, this is an oftentimes unjustified assumption. Due to non-negligible interactions with its surroundings, the system of interest and the environment become entangled so that the total wave function cannot be written as a product state. Consequently, the (sub-)system of interest can no longer be described by a single wave function but rather by a statistical mixture of many wave functions. It should therefore be treated in the *density matrix* formalism.

The density operator ρ is given by

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|, \quad (2.12)$$

where p_i is the probability to find the system in state $|\psi_i\rangle$. Its time evolution is given by the von Neumann equation

$$\partial_t \rho = -i[\mathbf{H}, \rho]. \quad (2.13)$$

Density operators are normalized, hermitian, and positive [2]:

$$\text{Tr}[\rho] = 1, \quad \rho = \rho^\dagger, \quad \rho \geq 0, \quad (2.14)$$

so that we may in the following assume Eq. (2.12) to be the eigendecomposition of ρ in the orthonormal eigenbasis $\{|\psi_i\rangle\}$ without loss of generality.

The rank of ρ is upper bounded by the number of terms that appear in Eq. (2.12) and is one for pure states, for which only a single probability is one, while it is greater than one for thermal states

$$\rho = \frac{\exp(-\beta\mathbf{H})}{\text{Tr}[\exp(-\beta\mathbf{H})]}, \quad (2.15)$$

where $\beta \geq 0$ is the inverse temperature. This is reflected in the quantum purity $\text{Tr}[\rho^2]$, which vanishes for pure states and deviates from zero for mixed states [2].

Using the density matrix formalism, we can meaningfully describe a subsystem A of the system under consideration by *tracing out* its complement B ,

$$\rho_A = \text{Tr}_B[\rho] = \sum_{ij} \langle ij| \rho |ij\rangle |i\rangle \langle i|, \quad (2.16)$$

with $|i\rangle$ ($|j\rangle$) an element of the basis of subsystem A (B) and $|ij\rangle$ their tensor product. If ρ is pure, the von Neumann entropy

$$S_A = -\text{Tr}[\rho_A \ln(\rho_A)] \quad (2.17)$$

constitutes a measure of entanglement between the two subsystems, and so do their Rényi extensions

$$S_A^{(n)} = \frac{1}{1-n} \ln(\text{Tr}[\rho_A^n]), \quad (2.18)$$

for which the limit

$$\lim_{n \rightarrow 1} S^{(n)} = S \quad (2.19)$$

is understood [189]. Another information theoretical quantity of interest is the quantum mutual information,

$$I(A : B)^{(n)} = S_A^{(n)} + S_B^{(n)} - S_{AB}^{(n)} \quad (2.20)$$

which measures *all* correlations that are present between the subsystems A and B , in particular those that are not captured by second-order correlation functions.

A research question that has received a lot of attention is how the entanglement entropy of a subsystem changes with its size [190]. A central observation in quantum information theory and quantum many-body physics is that *typical* states, such as ground states of gapped Hamiltonians, show a so-called area law, given that only local interactions are present. This means that the entanglement entropy of that subregion, given a sufficient size, only grows with the size of the boundary and not its volume. In one-dimensional systems, this corresponds to a logarithmic growth of S_A , featuring an almost stationary value for large enough subsystem sizes after an initial

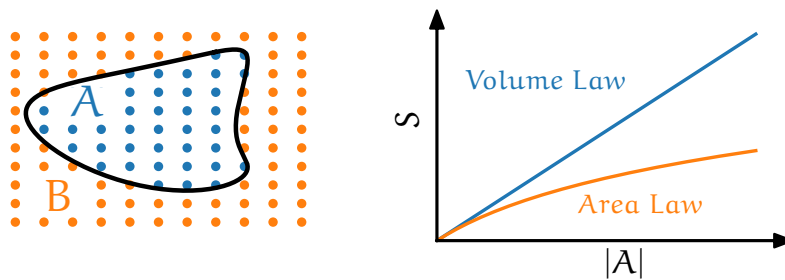


Figure 2.3: Left: Bipartition of a two dimensional system into subsystem A and B . Right: Depending on whether the entanglement entropy of a one-dimensional system grows linearly (logarithmically) with the size of the subsystem $|A|$, the system is said to exhibit a volume law (area law).

transient regime as shown on the right-hand side of [Figure 2.3](#). This is a striking observation, given that a random state that is *typical* according to the Haar-random measure obeys a volume-law of the entanglement entropy so that it is an extensive quantity for that subsystem [[191–193](#)]. Let us emphasize that the origin of area laws stems from the *locality* of the interactions in the Hamiltonian and is not present within all-to-all connected models, such as, for example, the Sachdev-Ye-Kitaev model [[194–197](#)]. The presence of area laws has important implications regarding the classical simulatability of quantum models, as we will discuss in more detail in [Chapter 4](#).

2.3 DYNAMICS OF OPEN QUANTUM SYSTEMS

While closed quantum systems evolve according to the Schrödinger equation ([2.10](#)) or von Neumann equation ([2.13](#)), the evolution of a system that is in contact with a bath is non-unitary, breaking the time-reversal symmetry as information flows out of the system and cannot be recovered [[198–200](#)]. In certain limits, the subsystem of interest obeys a Lindblad master equation

$$\partial_t \rho = -i[H, \rho] + \sum_i \gamma_i \left(L_i \rho L_i^\dagger - \frac{1}{2} \{L_i^\dagger L_i, \rho\} \right), \quad (2.21)$$

with the anticommutator $\{A, B\} = AB + BA$, the ‘jump’ operators L_i corresponding to different dissipation channels and γ_i denoting the dissipation rate.

The main assumptions that lead to the Lindblad master equation are a weak coupling between system and bath and that the bath is of *Markovian* type, meaning that correlations between system and bath decay on short time scales compared to the internal dynamics such that the bath has no memory of its previous states [[198, 201](#)].

The dynamics generated by Lindbladians typically lead to a unique steady state that the system asymptotically converges to in the long time limit irrespective of initial conditions [[202, 203](#)], such that

$$\lim_{t \rightarrow \infty} \rho(t) = \rho_{SS}. \quad (2.22)$$

The uniqueness property of the steady state is equivalent to the spectrum of the Lindbladian only having a single entry with a real part equal to zero. The real parts of all other entries are necessarily smaller than zero, as they are continuously projected out during the time evolution.

We may think about the Lindblad evolution equation also in a stochastic way, where dissipative jumps, such as the emission of a photon or the flipping of a spin, are discrete, random events in an otherwise continuous evolution [[204, 205](#)]. This bears the advantage, that the only objects that need to be modeled are wave functions,

meaning a quadratic reduction in complexity compared to the density matrix formalism, albeit one needs to average many such stochastic evolutions to obtain a reasonable signal-to-noise ratio. In this Monte Carlo wave function (MCWF) framework, which we use as a benchmark in [Chapter 5](#), the wave function obeys an evolution generated by the effective Hamiltonian

$$\mathbf{H}_{eff} = \mathbf{H} - \frac{i}{2} \sum_i \mathbf{L}_i^\dagger \mathbf{L}_i. \quad (2.23)$$

As \mathbf{H}_{eff} is not a hermitian operator, the wave function needs to be renormalized after every evolution step. This evolution is perturbed by projections that occur with probability

$$dp = dt \sum_i \langle \psi | \mathbf{L}_i^\dagger \mathbf{L}_i | \psi \rangle = dt \sum_i p_i, \quad (2.24)$$

while the evolution under \mathbf{H}_{eff} for a small time step dt takes place with probability $1 - dp$. The time step should be chosen such that the probability for a jump is low, $dp \ll 1$. The evolution equation for the wave function then reads

$$|\psi(t + dt)\rangle = e^{-i\mathbf{H}_{eff}dt} |\psi(t)\rangle / \sqrt{1 - dp} \quad (2.25)$$

if $dp < p$, with p a random number drawn uniformly from the interval $[0, 1]$. When $dp > p$, the system instead undergoes a jump event

$$|\psi(t + dt)\rangle = \mathbf{L}_i |\psi(t)\rangle / \sqrt{p_i}, \quad (2.26)$$

where the event i occurs with probability $dt p_i$ in a single step.

[Figure 2.4](#) shows the dynamics of a single spin-1/2 that is subject to an external magnetic field as well as dephasing in the MCWF

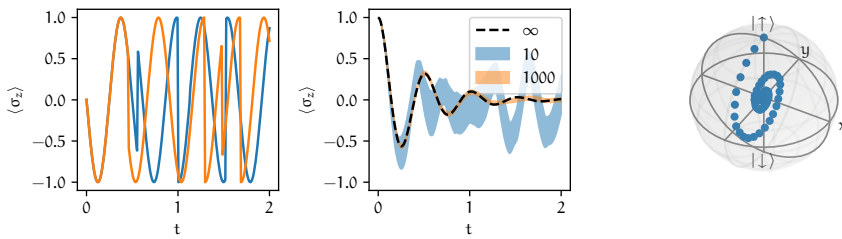


Figure 2.4: Monte Carlo wave function demonstration for a single spin-1/2 following a quench from $|\psi\rangle = |\uparrow\rangle$ under the Hamiltonian $\mathbf{H} = \sigma_x$ with dissipation $\mathbf{L} = \sigma_z$ with rate $\gamma = 1.5$. Left: Single wave function trajectories. Center: Averages over 10 (1000) trajectories with the shaded region corresponding to a 1σ confidence interval. Right: Evolution on the Bloch sphere. The dissipative evolution brings the state from the surface of the sphere to the origin, corresponding to the steady state $\rho_{SS} = \mathbb{1}/2$.

formalism. The left panel depicts two single trajectories for which the discrete jumps at random times are apparent. Upon averaging many trajectories, the correct ensemble description is recovered, as indicated in the center panel. As non-unitary evolutions change the purity of the quantum state, the dynamics on the Bloch-sphere are no longer restricted to its surface but finally end up in the steady state which is the origin for the case at hand.

2.4 POSITIVE OPERATOR VALUED MEASURES (POVMS)

Quantum mechanical systems are typically either described using wave functions, if the system is closed or density matrices if it is open. This section and Section 2.6 will introduce two alternative descriptions that may be advantageous in certain situations.

Suppose the following experimental setup: We are given access to a two-level system, i.e. a qubit, whose state $|\psi\rangle$ we wish to characterize. By carrying out projective measurements along the z-axis, we obtain statistics of ‘up’ and ‘down’ outcomes, which we can use to compute the expectation value $\langle\sigma_z\rangle$. However, unless $|\psi\rangle = |\uparrow\rangle$ or $|\psi\rangle = |\downarrow\rangle$, we have insufficient knowledge to associate a point on the Bloch sphere to the state. By carrying out projective measurements along the x- and y-axis and obtaining their respective expectation values, we can remove this ambiguity and uniquely identify the state. This process is known as *quantum state tomography* and is used to verify

Note that these are precisely the eigenstates of σ_z .

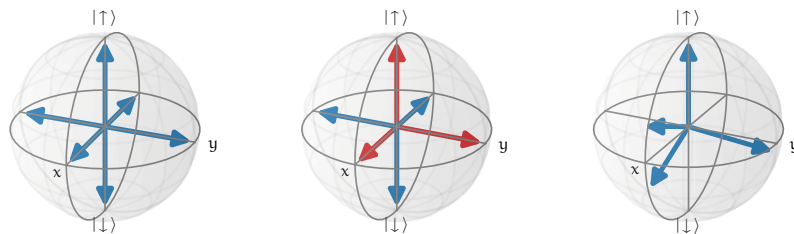


Figure 2.5: Different choices of informationally complete POVMs, with arrows indicating the directions of projective measurements. Left: The Pauli-6 POVM is an experimentally motivated example of an informationally complete POVM, where each projective measurement in the computational basis is carried out after the application of a random unitary that corresponds to a basis transformation to the x-, y- or z-basis. Middle: To transform the Pauli-6 POVM to a *minimal* informationally complete POVM, we may group the ‘up’ outcomes of the three different basis configurations (shown in red) to represent a new outcome. Note that this grouped outcome does not correspond to a rank-1 projector. Right: The *symmetric* informationally complete POVM (SIC-POVM) is characterized by pairwise equal Hilbert-Schmidt inner products among the projectors and is theoretically motivated.

the state preparation procedures in quantum devices. It is formalized using the concept of informationally complete positive operator valued measures (IC-POVMS), which assign a unique probability distribution over measurement outcomes to each state [2].

A general POVM is characterized by K measurement operators M_k that fulfill

$$M_k = M_k^\dagger, \quad \sum_k M_k = \mathbb{1}, \quad M_k \geq 0 \quad \forall k. \quad (2.27)$$

Informational completeness is guaranteed once the set of the M_k span the space of operators - if the number of operators is *just sufficient* to span that space, the IC-POVM is said to be *minimal*.

In the case of a single spin-1/2, three informationally complete POVMS are relevant, which are presented in Figure 2.5. Of particular interest to this thesis is the symmetric informationally complete POVM (SIC-POVM) [206] shown on the right, since it is used in our approach to model the dissipative dynamics using neural networks [A], which is discussed in Chapter 5. It is made up of the four ‘projectors’

$$M_k = \frac{1}{2} |\psi_k\rangle \langle \psi_k| \quad (2.28)$$

with the states $|\psi_k\rangle$ defined as

$$\begin{aligned} |\psi_0\rangle &= |\uparrow\rangle, & |\psi_1\rangle &= \frac{1}{\sqrt{3}} |\uparrow\rangle + \sqrt{\frac{2}{3}} |\downarrow\rangle, \\ |\psi_2\rangle &= \frac{1}{\sqrt{3}} |\uparrow\rangle + e^{i\frac{2\pi}{3}} \sqrt{\frac{2}{3}} |\downarrow\rangle, & |\psi_3\rangle &= \frac{1}{\sqrt{3}} |\uparrow\rangle + e^{i\frac{4\pi}{3}} \sqrt{\frac{2}{3}} |\downarrow\rangle. \end{aligned} \quad (2.29)$$

A POVM is referred to as symmetric if all its elements have the same pairwise Hilbert-Schmidt inner product,

$$\text{Tr}[M_k M_l] = \frac{1}{d^2} \frac{d\delta_{kl} + 1}{d + 1} \quad (2.30)$$

with d the dimension of the Hilbert space.

The other two POVMS shown in Figure 2.5 are experimentally motivated and are employed in [F] and [H]. One intuitive way to construct POVMS for a system of N spin-1/2 particles consists in building tensor products of the single spin measurement operators,

$$M_{\mathbf{k}} = M_{k_1 k_2 \dots k_N} = M_{k_1} \otimes M_{k_2} \otimes \dots \otimes M_{k_N}. \quad (2.31)$$

We note that the symmetry property of a POVM is lost upon using tensor products of SIC-POVMS to build POVMS on larger systems. The POVM distribution associated to a set of measurement operators $\{M_k\}$ is obtained by computing expectation values of the measurement operators

$$P^k = \text{Tr}[\rho M_k]. \quad (2.32)$$

Note that the term projector is commonly used but misplaced as $M_k^2 = M_k/2$, violating the property $P^2 = P$ a projector must obey. We will nevertheless use this nomenclature to be consistent with the literature.

If $\{M_k\}$ is informationally complete, the above probability distribution Eq. (2.32) contains all the information about the quantum state. In particular, it is possible to reconstruct the density operator from the probabilities using the inverse of Eq. (2.32) which is given by

$$\rho = \sum_{kl} M_k T_{kl}^{-1} P^l, \quad (2.33)$$

Note that this inverse does not exist for POVMs that are informationally complete but not minimal.

where T_{kl}^{-1} denotes the inverse of the overlap matrix

$$T_{kl} = \text{Tr} [M_k M_l]. \quad (2.34)$$

It is thus possible, to reformulate quantum mechanics using only the language of POVM probability distributions. In particular, we can write down expressions for expectation values and reformulate the master equation of the density operator ρ as a differential equation for the POVM distribution P .

In the POVM formalism, an expectation value of an operator O takes the form

$$\langle O \rangle = \text{Tr} [\rho O] = \sum_k P^k O^k, \quad (2.35)$$

where

$$O^k = \sum_l T_{kl}^{-1} \text{Tr} [M_l O]. \quad (2.36)$$

Given samples $\{k\}$ of the probability distribution P , it is thus sufficient to average the associated coefficients $\{O^k\}$ to arrive at an estimate for the expectation value, hinting at an efficient Monte Carlo estimation scheme of Eq. (2.35).

To arrive at an evolution equation of the probability distribution, we insert Eq. (2.21) into the time derivatives of the probabilities

$$\partial_t P^k = \partial_t \text{Tr} [M_k \rho] = \text{Tr} [M_k \partial_t \rho]. \quad (2.37)$$

For the unitary part contained in Eq. (2.21), we compute

$$\begin{aligned} \partial_t P^k &= \text{Tr} [-i [\mathbf{H}, \rho] M_k] \\ &= \text{Tr} \left[-i \left[\mathbf{H}, T_{U'}^{-1} M_{U'} \right] M_k \right] P^l \\ &= \text{Tr} \left[-i \mathbf{H} \left[T_{U'}^{-1} M_{U'}, M_k \right] \right] P^l \\ &= \mathcal{U}_{kl} P^l, \end{aligned} \quad (2.38)$$

where we inserted the definition of the density operator in terms of POVM probabilities (2.33) in the second line and used the cyclicity of the trace in the third line. For the dissipative part, we similarly find

$$\begin{aligned} \mathcal{D}_{kl} &= \text{Tr} \left[\sum_{U'} \sum_i \gamma_i \left(L_i T_{U'}^{-1} M_{U'} L_i^\dagger M_k \right. \right. \\ &\quad \left. \left. - \frac{1}{2} L_i^\dagger L_i \left\{ T_{U'}^{-1} M_{U'}, M_k \right\} \right) \right] \end{aligned} \quad (2.39)$$

and write the composite evolution equation

$$\partial_t P^k = \sum_l \mathcal{L}_{kl} P^l \quad (2.40)$$

with the POVM-Lindblad operator

$$\mathcal{L} = \mathcal{U} + \mathcal{D}. \quad (2.41)$$

For a meaningful evolution probability must be conserved, so that the matrix exponential of \mathcal{L} must be *stochastic*,

$$\sum_k \exp(\tau \mathcal{L})_{kl} = 1, \quad (2.42)$$

with τ denoting a small time step, implying

$$\sum_k \mathcal{L}_{kl} = 0. \quad (2.43)$$

While all (normalized) POVM probability distributions correspond to hermitian density operators with unit trace, the positivity constraint may be violated depending on the distribution, so that some POVM distributions correspond to non-physical density matrices [207]. For instance, SIC-POVM distributions that are delta-peaks are forbidden due to the non-orthogonality of the projectors in Eq. (2.29),

$$\langle \psi_i | \psi_j \rangle \neq \delta_{ij}, \quad (2.44)$$

which would exemplarily lead to

$$P = (1, 0, 0, 0)^T \iff \rho = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.45)$$

obviously violating the positivity constraint.

It is not easy to say whether a given POVM distribution is *positive*, meaning whether it corresponds to a density operator with $\rho \geq 0$, without explicitly constructing ρ using Eq. (2.33) and computing its eigenvalues. This point will also be discussed in some detail in [Chapter 5](#), where it would be desirable to define ansatz functions that only allow modeling *physical* distributions, which form a subset within the probability simplex as shown in [Figure 2.6](#).

2.5 BOSONIC MODES

One of the main results of this thesis is a proposal regarding the readout of an area to volume law transition in the classical differential entropies of phase-space distributions describing a Bose-Einstein condensate. We thus wish to describe bosonic modes in this section, their associated phase-space descriptions in [Section 2.6](#) and the details of the system studied in [Chapter 8](#) and [Chapter 9](#) in [Section 2.7](#).

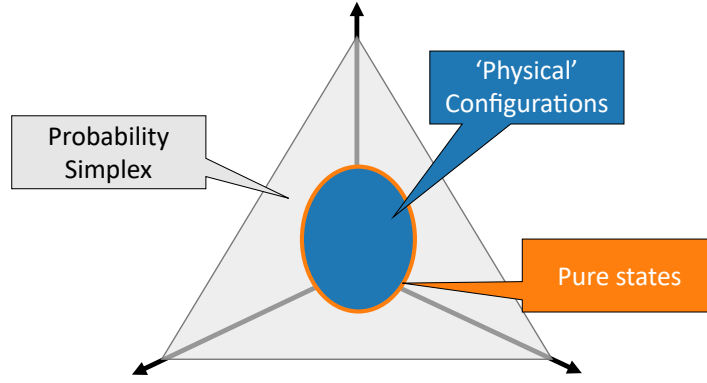


Figure 2.6: The probability simplex: The set of valid physical distributions is shown in blue, which is a subset of all (normalized) probability distributions shown in grey. The boundary of the subset is shown in orange, which is given by the set of pure states.

A single bosonic mode describes a harmonic oscillator with Hamiltonian

$$H = \frac{m\omega^2}{2}x^2 + \frac{1}{2m}p^2. \quad (2.46)$$

Its solutions are the Hermite functions

$$\psi_n(x) = \frac{1}{\sqrt{2^n n!}} \left(\frac{m\omega}{\pi}\right)^{1/4} \exp\left(-\frac{m\omega x^2}{2}\right) H_n(\sqrt{m\omega}x) \quad (2.47)$$

with n a non-negative integer and the Hermite polynomials

$$H_n(z) = (-1)^n e^{z^2} \frac{d^n}{dz^n} e^{-z^2}. \quad (2.48)$$

As eigenfunctions of a hermitian operator, the Hermite functions constitute an orthonormal basis

$$\langle \psi_n | \psi_m \rangle = \int dx \psi_n^*(x) \psi_m(x) = \delta_{nm}, \quad (2.49)$$

the harmonic oscillator lends itself to treatment in the discrete Hilbert space spanned by the Fock-basis

$$\mathcal{B} = \left\{ \frac{(a^\dagger)^n}{\sqrt{n!}} |0\rangle \mid \forall n \in \mathbb{N} \right\}. \quad (2.50)$$

Here $|0\rangle$ is the vacuum state and

$$a^\dagger = (x - ip) / \sqrt{2} \quad (2.51)$$

the creation operator, transitioning between the equidistant energy eigenlevels

$$E_n = \omega \left(n + \frac{1}{2} \right) \quad (2.52)$$

by creating an elementary excitation. Its adjoint \mathbf{a} instead annihilates an elementary excitation and we find

$$[\mathbf{a}, \mathbf{a}^\dagger] = 1 \quad (2.53)$$

and note that the creation and annihilation operators span the space of operators. We may write the Hamiltonian in its diagonal form

$$\mathbf{H} = \omega \left(\mathbf{a}^\dagger \mathbf{a} + \frac{1}{2} \right) = \omega \left(\mathbf{N} + \frac{1}{2} \right), \quad (2.54)$$

with \mathbf{N} the particle number operator.

Depending on the system under consideration, Fock states are difficult to prepare experimentally, particularly in the regime of high occupations. What is oftentimes more easily accessible are *coherent states* $|\alpha\rangle$, meaning eigenstates of the annihilation operator \mathbf{a} with eigenvalue α , which are given by the Fock-expansion

$$|\alpha\rangle = \exp\left(-\frac{\alpha^2}{2}\right) \sum_n \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (2.55)$$

In the following we associate the real part of α with position x and its imaginary part with momentum p , i.e.

$$\alpha = \frac{x + ip}{\sqrt{2}}. \quad (2.56)$$

Equivalently, we may define the displacement operator

$$\mathbf{D}(\alpha) = \exp\left(\alpha \mathbf{a}^\dagger - \alpha^* \mathbf{a}\right) \quad (2.57)$$

and generate a coherent state $|\alpha\rangle$ by acting on the vacuum with $\mathbf{D}(\alpha)$. Its average number occupation and associated standard deviation are given by

$$\langle \mathbf{N} \rangle = |\alpha|^2 \quad \sigma_{\langle \mathbf{N} \rangle} = \sqrt{\langle \mathbf{N} \rangle}, \quad (2.58)$$

so that the number occupation distribution becomes peaked in the (classical) limit of high occupations,

$$\frac{\sigma_{\langle \mathbf{N} \rangle}}{\langle \mathbf{N} \rangle} = 1/\sqrt{\langle \mathbf{N} \rangle}. \quad (2.59)$$

2.6 PHASE-SPACE REPRESENTATIONS OF BOSONIC MODES

Bosonic systems lend themselves to phase-space descriptions, allowing an arguably more intuitive treatment of physical phenomena than the description in Fock space [209].

The most relevant phase-space distributions for this thesis are the Wigner W - and Husimi Q -distribution. In contrast to phase-space

distributions of classical systems, these descriptions only constitute *quasi-probability* distributions, as we will see shortly.

The Wigner transform of an operator O defines its associated Weyl-symbol

$$O_W(x, p) = \int d\chi \left\langle x - \frac{\chi}{2} \left| O \right| x + \frac{\chi}{2} \right\rangle \exp(ip\chi). \quad (2.60)$$

The Weyl-symbol of the density operator gives the Wigner distribution $W(x, p) = W(\alpha)$ with α defined in (2.56). The Wigner distribution is normalized,

$$\int dx dp W(x, p) = 2 \int d^2\alpha W(\alpha) = 1 \quad (2.61)$$

but can take negative values, therefore not allowing a probabilistic interpretation for all states. Its marginals,

$$W(x) = \int dp W(x, p), \quad W(p) = \int dx W(x, p), \quad (2.62)$$

can however be shown to be strictly positive also for negative $W(x, p)$, as demonstrated in the right panel of Figure 2.7. They correspond to the marginal distributions that one would obtain from measuring either the position or momentum of the system.

Importantly, coherent states correspond to Gaussian Wigner distributions and are thus positive. Connecting to the discussion at the start

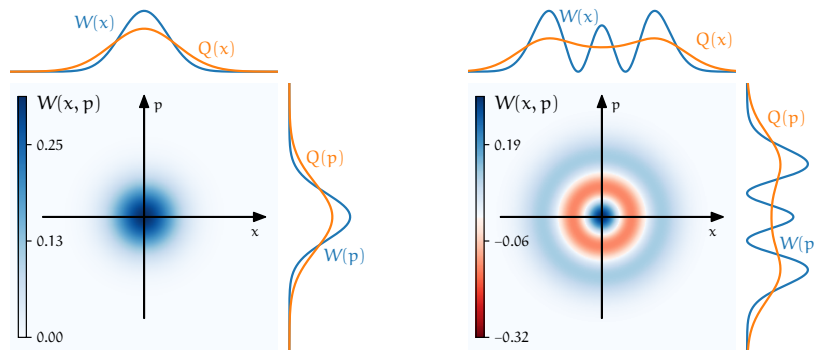


Figure 2.7: Left: Wigner W -distribution of the vacuum state $|0\rangle$, with marginalizations over p (x) shown on the top (right) in blue. If the two observables are read out simultaneously, the Wigner density profile is broadened, and the obtained marginals correspond to those of the Husimi Q -distribution (shown in orange). Right: Wigner W -distribution of the Fock state $|2\rangle$, which is no longer of Gaussian form but rather described by Laguerre polynomials [208], thus taking on negative values. Its marginals however are still positive, with the Husimi marginals now significantly more smeared out compared to the vacuum state.

of this chapter regarding the Heisenberg uncertainty principle, we are now faced with a quantity that, unlike the wave function, depends on *both* x and p , and hence the Wigner distribution must fulfill certain restrictions so that the Heisenberg uncertainty principle (2.1) is not violated. For Gaussian distributions, this translates into a lower bound on the determinant of the covariance matrix, which in turn translates into a lower bound on the differential entropy of W (also referred to as Wigner entropy)

$$S(W) = - \int dx dp W(x, p) \ln(W(x, p)) \geq 1 + \ln(\pi). \quad (2.63)$$

For general, i.e. non-Gaussian, distributions it has not yet been proven that (2.63) holds, with the hitherto best known *entropic uncertainty relation* reproducing the one obtained for a Gaussian [210, 211]. One may extend the definition of (2.63) to negative Wigner distributions, where $S(W)$ becomes complex-valued.

A strictly positive phase-space description is given by the Husimi Q -distribution. It is defined by the projection of the density matrix onto all coherent states

$$Q(x, p) = Q(\alpha) = \langle \alpha | \rho | \alpha \rangle. \quad (2.64)$$

The set of coherent states constitutes an overcomplete, non-orthogonal basis such that all information contained in the density operator is preserved in the Husimi Q -distribution. Its connection to the Wigner W -distribution can be made explicit, as it can also be obtained as a smoothed version of the Wigner W -distribution using a convolution with a Gaussian kernel [209]

$$Q(\alpha) = 4 \int d^2\beta W(\beta) \exp(-2|\alpha - \beta|^2). \quad (2.65)$$

With the appropriate integral measure, the Husimi Q -distribution is normalized,

$$\int \frac{dx dp}{2\pi} Q(x, p) = 2 \int \frac{d^2\alpha}{2\pi} Q(\alpha) = 1. \quad (2.66)$$

Its quasi-probabilistic nature is due to the interpretation of its marginals, which are broadened compared to those one would obtain if one were to measure either only position *or* momentum. This broadening arises since the projection onto coherent states has a clear physical interpretation as a measurement operation of *both* position and momentum following Eq. (2.56). Hence, in order to not violate the Heisenberg uncertainty principle, the distribution is broadened compared to the marginals that stem from non-simultaneous measurements, such as those of the Wigner distribution. The Husimi marginals can thus not be interpreted as those that are obtained from repeatedly measuring the position or momentum of the system. The differential entropy

For diagonal covariance matrices the constraint on the determinant is the same as (2.1).

that is associated with the Husimi Q -distribution is the Wehrl-entropy, satisfying

$$S(Q) = - \int \frac{dx dp}{2\pi} Q(x, p) \ln(Q(x, p)) \geq S(W) - \ln(\pi). \quad (2.67)$$

If we consider phase-space distributions of multiple modes, we may compute Wigner and Wehrl entropies of subsystems and hence also their mutual information

$$I(O_A : O_B) = S_A(O) + S_B(O) - S_{AB}(O) \quad (2.68)$$

with O either W or Q , much alike to their *quantum* counterpart in Eq. (2.20). We are motivated to explore these information theoretical phase-space quantities, as they also encode the area law similar to the discussion in Section 2.2, as we will show in Chapter 8 and Chapter 9.

2.7 SPIN-1 BOSE-EINSTEIN CONDENSATES

The bosonic system we wish to study in Chapter 8 and Chapter 9 using the introduced phase-space representations is a spin-1 Bose-Einstein condensate (BEC), that may be realized using Lithium-7 or Rubidium-87 atoms. In this section, we will give a brief overview of the system and link to more exhaustive sources for the interested reader.

Building on work from Bose on the introduction of quantum statistics [212], Einstein predicted the existence of a novel state of matter, the Bose-Einstein condensate [213]. It is characterized by a macroscopic occupation of the system's ground state at low densities, once the temperature falls below a critical value, which is typically close to absolute zero. Due to remarkable advances in the microscopic control over quantum systems, Bose-Einstein condensates have developed into an ubiquitous platform to implement and simulate various many-body phenomena ever since their first experimental realization in 1995 [214, 215].

In experimental setups, Bose-Einstein condensates are obtained using laser- and evaporative cooling techniques and confined to small spatial subregions by employing magneto-optical traps (MOTs) [216]. Using optical lattices the implementation of periodic potentials is feasible, enabling the realization of Hamiltonians such as Bose-Hubbard models in the lab [18], which will be of high interest to this thesis. Such setups may be used to study both ground state properties as well as dynamics in lattice systems. Information about the system is typically gained using projective imaging techniques that destroy the prepared state. Due to long state preparation cycle times (~ 40 seconds), efficiency with respect to the number of obtained samples is essential, a point which will be emphasized in Chapter 8 and Chapter 9.

Working in the single mode approximation¹ and the $F = 1$ manifold, each well j within the optical lattice is described by three bosonic modes \mathbf{a}_i^j , with i enumerating the three magnetic sublevels. These are coupled by the (internal) Hamiltonian

$$\begin{aligned} \mathbf{H}_{int}^j = & q \left(\mathbf{N}_1^j + \mathbf{N}_{-1}^j \right) + c_0 \mathbf{N}^j \left(\mathbf{N}^j - \mathbb{1} \right) \\ & + c_1 \left[\left(\mathbf{N}_0^j - (1/2)\mathbb{1} \right) \left(\mathbf{N}_1^j + \mathbf{N}_{-1}^j \right) \right. \\ & \left. + \mathbf{a}_0^{j\dagger} \mathbf{a}_0^j \mathbf{a}_1^j \mathbf{a}_{-1}^j + \mathbf{a}_1^{j\dagger} \mathbf{a}_{-1}^j \mathbf{a}_0^j \mathbf{a}_0^j \right]. \end{aligned} \quad (2.69)$$

Here, \mathbf{N}_i^j counts the number of atoms in mode i at site j and $\mathbf{N}^j = \mathbf{N}_{-1}^j + \mathbf{N}_0^j + \mathbf{N}_1^j$.

The terms proportional to c_1 and q are responsible for the internal spin dynamics, with c_1 giving the strength of the pair production process where two side-mode atoms are created upon the collision of two zero-mode atoms or vice versa. Depending on the value of q , this process may be suppressed as q serves as an energy penalty or detuning for the creation of side-mode populations. An illustration of the single well dynamics is given in [Figure 2.8](#).

The Hamiltonian of the full system is composed of the internal contributions of all wells and the couplings among them

$$\mathbf{H} = \sum_{j=1}^L \mathbf{H}_{int}^j + \sum_{j=1}^{L-1} \mathbf{H}_{tunnel}^j \quad (2.70)$$

where L denotes the number of wells in the lattice, and \mathbf{H}_{tunnel} is the nearest-neighbor tunnel Hamiltonian

$$\mathbf{H}_{tunnel}^j = -J \sum_{i=\pm 1} \mathbf{a}_i^{j\dagger} \mathbf{a}_i^{j+1} + \mathbf{a}_i^j \mathbf{a}_i^{(j+1)\dagger}, \quad (2.71)$$

where the restriction to side-mode hopping is motivated in [Chapter 8](#) and [Chapter 9](#).

In the limit of large values of q , all spin dynamics are suppressed and the remaining dynamics are those of the Bose-Hubbard model, with c_0 giving the on-site repulsion, thereby suppressing the creation of population imbalances between sites which are created by the tunnel Hamiltonian. Depending on their relative strengths, we thus face a competition of momentum-space ground states, characterized by phase coherence across all lattice sites favored by the tunnel Hamiltonian in contrast to definite real-space occupation numbers favored by the on-site repulsion. We note that while both J and q are experimentally tunable parameters, the ratio c_0/c_1 is determined by the atomic species and thus not tunable.

¹ The single mode approximation assumes that all atoms are described by the same real-space wave function throughout the experiment [216].

The spin physics we aim to study within the described system is accessible by measuring the two non-commuting spin observables of our interest

$$\begin{aligned}\phi^j &= \frac{S_x^j}{\sqrt{2\langle N^j \rangle}} \\ &= \frac{1}{\sqrt{2}} \left[a_0^{j\dagger} (a_1^j + a_{-1}^j) + a_0^j (a_1^{j\dagger} + a_{-1}^{j\dagger}) \right] / \sqrt{2\langle N^j \rangle},\end{aligned}\quad (2.72)$$

and

$$\begin{aligned}\pi^j &= -\frac{Q_{yz}^j}{\sqrt{2\langle N^j \rangle}} \\ &= \frac{-i}{\sqrt{2}} \left[a_0^{j\dagger} (a_1^j + a_{-1}^j) - a_0^j (a_1^{j\dagger} + a_{-1}^{j\dagger}) \right] / \sqrt{2\langle N^j \rangle}.\end{aligned}\quad (2.73)$$

This can be done successively, such that a single shot gives a sample of the marginal Wigner distribution associated with either ϕ or π . Another option is their simultaneous readout, leading back to the prior discussion regarding the Husimi Q -distribution (see Figure 2.7 and Eq. (2.64)).

For a discussion on how this can be done experimentally, see [216] and [217].

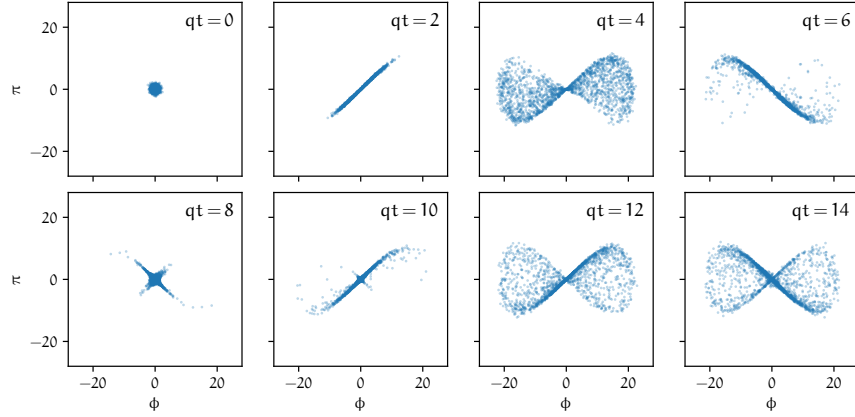


Figure 2.8: Truncated Wigner simulation of the single well dynamics generated by the internal Hamiltonian Eq. (2.69) after reading out ϕ and π given in Eqs. (2.72) and (2.73). The Gaussian distribution of (ϕ, π) that is associated with the initial polar state $|\psi\rangle = |0, \alpha, 0\rangle$, constructed from side mode vacua and a coherent state in the zero mode, is first squeezed (roughly corresponding to $qt = 2$) before depletion effects cause the distribution to develop more complex features. Note that the evolution equations of (ϕ, π) are not closed and hence no flow field can be given. Here we chose $q = \langle N \rangle c_1$.

2.8 THERMALIZATION OF CLOSED QUANTUM SYSTEMS

As we will also consider the long time limit of the dynamics generated by the Hamiltonian in Eq. (2.70) in Chapter 9, we want to briefly discuss the mechanisms that cause quantum systems to thermalize.

It is a priori unclear what leads to the equilibration and thermalization of generic, interacting closed quantum systems, as their unitary dynamics do not feature any attractors. *Locally*, the situation at hand is however qualitatively different as a spatial subregion A with reduced density matrix $\rho_A = \text{Tr}_B[\rho]$ does not obey a unitary evolution, since the initially localized information flows out of subsystem A into subsystem B as the two become entangled. Thus, in the long time limit, local subregions may approach a steady state which is oftentimes well described by a thermal ensemble [80, 218–220].

One such ensemble is the quantum microcanonical ensemble

$$\rho_A^{MC} = \text{Tr}_B \left[\frac{1}{D} \sum_{E_i \in I} |i\rangle \langle i| \right] \quad (2.74)$$

which attributes equal probability to all eigenstates $|i\rangle$ that fall into the energy interval $I = [E, E + \Delta E]$, with ΔE small compared to the macroscopic energy scale but large enough so that many energy eigenstates are contained in the interval.

Surprisingly, the local predictions for subsystem A typically do not strongly depend on the details of the chosen energy interval I , meaning its precise location E and the size of the energy window ΔE . One idea that postulates an explanation for this, is the *eigenstate thermalization hypothesis (ETH)* [221], which assumes that eigenstates that are close in energy yield similar local descriptions so that the different terms in the sum of Eq. (2.74) give similar results despite their global orthogonality.

Another ensemble that may be used to describe the emergent thermal features of subsystems is the canonical ensemble, which we also define in analogy to classical statistical mechanics

$$\rho_A^C = \frac{\text{Tr}_B [\exp(-\beta \mathbf{H})]}{\text{Tr} [\exp(-\beta \mathbf{H})]}, \quad (2.75)$$

where the inverse temperature β must fulfill the self-consistency equation

$$\frac{\text{Tr} [\mathbf{H} \exp(-\beta \mathbf{H})]}{\text{Tr} [\exp(-\beta \mathbf{H})]} = \langle \mathbf{H} \rangle, \quad (2.76)$$

with the energy expectation value on the right-hand side given by the pure initial state.

For completeness, let us also mention that there exist quantum systems which do not thermalize and instead remain ‘close’ to their initial conditions: such systems are said to exhibit *localization* [222–227].

It is challenging to give a definition of machine learning (ML) that does justice to the vastness of the field. Instead, we will describe the underlying aims, which is an arguably easier task.

The high complexity of the world surrounding us is matched by our equally high ability to filter out only those bits of information that are relevant to us at a specific point in time. While this filtering may seem like a trivial task, defining strict rules for what is relevant in a specific situation is almost never straightforward. In order to have machines that can make sense of heaps of complex data, such as images or natural language for example, explicit, rule-based approaches therefore seem like a dead end. Instead, a new paradigm that breaks with this 'if-else' logic is required. This paradigm is machine learning.

The shortcomings of explicit programming become apparent for comparably easy tasks. A frequently cited example is image classification, where one aims to assign images to their corresponding categories. While defining the precise characteristics, or rules, because of which one reached a particular decision is difficult, such tasks are usually trivial to any human above a certain age. What makes this task trivial is our ability to relate new impressions to prior experience. The question thus becomes how we may build machines or algorithms that are capable of learning from prior experience.

In standard machine learning instances, this is achieved by defining loss functions that measure the performance of the algorithm on some training and test set. Then, elaborate gradient descent schemes are used to optimize these performance measures. For our applications, we need to alter these procedures, since the neural quantum state applications we have in mind are not data-driven, meaning that no training set exists. These novel approaches are therefore neither well placed in the category of supervised nor unsupervised learning. Instead, they form a *standalone* approach that is not restricted to scenarios that have already been studied by means of other methods. One may therefore hope to achieve results that go beyond the state-of-the-art that is set by competing computational techniques.

This chapter starts by introducing the challenges that are most relevant in the context of neural quantum states, namely those associated with generative modeling, before discussing the different network architectures that are employed in the remainder of this work.

3.1 GENERATIVE MODELING

The most relevant branch of machine learning for the following chapters is generative modeling. Generative modeling aims to model high-dimensional objects, such as probability distributions, which are analytically intractable and which are not associated with a particular distribution family. This description also fits the requirements when trying to model quantum states, so that applying the tools of generative modeling to problems from quantum many-body physics seems plausible.

3.1.1 Sampling from high-dimensional Distributions

As we will see in the following chapters, generating samples from the encoded distributions is essential in many applications. The fundamental problem associated with this task is the intractability of the partition function

$$Z = \sum_{\{\vec{x}\}} p(\vec{x}) \quad (3.1)$$

in discrete or

$$Z = \int d\vec{x} p(\vec{x}) \quad (3.2)$$

in continuous settings. Focusing on the case of an N -particle spin-1/2 system in the following, we see that the exponential complexity associated with the naïve normalization of the probabilities would render any computational approach inefficient. We thus cannot generate samples by drawing uniformly distributed random numbers between zero and one and check to which configuration they correspond. Instead, we are forced to resort to Markov chain Monte Carlo techniques.

Here we have
 $p(\vec{x}) = |\psi(\vec{x})|^2$.

3.1.1.1 Markov chain Monte Carlo

Markov chain Monte Carlo algorithms generate a set of samples in an iterative fashion. First, an initial sample \vec{x} is defined as the start of the chain before a new sample \vec{x}' is proposed. The new sample is accepted and added to the chain with probability

$$r = \min \left(1, \frac{p(\vec{x}')}{p(\vec{x})} \right). \quad (3.3)$$

If the new sample is not accepted, \vec{x} is added to the chain instead. Then the procedure starts anew and compares the next proposed sample to the one that was added most recently. Notice, that the acceptance probability only relies on the *ratio* of the probabilities such that working with unnormalized probabilities is not an issue, as the normalization cancels out.

Eq. (3.3) constitutes a valid Markov chain Monte Carlo scheme if the proposal probabilities $p(\vec{x} \rightarrow \vec{x}')$ and $p(\vec{x}' \rightarrow \vec{x})$ are identical. If this condition of *detailed balance* is not met, Eq. (3.3) must be adapted accordingly. For most applications, $p(\vec{x} \rightarrow \vec{x}')$ is zero unless \vec{x} and \vec{x}' differ in exactly one or two positions. This proposal scheme is used to maintain high acceptance rates within the Markov chain but also implies that neighboring samples within the chain may be highly correlated. To obtain sample sets that are less correlated, one has to disregard a significant portion of the sample set by considering only every N -th sample or so, increasing computational complexity.

Furthermore, the Markov chain scheme does not guarantee an ergodic exploration of the configuration space. In the case of the classical Ising model, for example, the two fully magnetized configurations are highly unlikely to occur within the same Markov chain if we use local updates, as they are connected only by passing through regions of highly energetic states. This means that despite the energetic equivalence of the two magnetized states, the chain would spontaneously break the spin-flip symmetry and stay either within the sector of positive or negative magnetization. To resolve incorrect expectation values that are an artifact of this phenomenon, one may run multiple chains in parallel, which will approximately distribute equally into the two regions.

These shortcomings reduce the appeal of Markov chain Monte Carlo techniques so that the exploration of network architectures that support *exact* sampling from high dimensional distributions was met with great interest.

3.1.1.2 Exact sampling

Some specific model architectures, such as recurrent neural networks or normalizing flows support exact sampling in discrete and continuous settings by design. This means that samples are perfectly uncorrelated and that sampling from multimodal distributions is no issue.

In discrete settings, exact sampling can be achieved using the autoregressive property of multivariate probability distributions. This property guarantees the existence of conditionals, that uniquely define the distribution, such that

$$p(\vec{x}) = p(x_1) \cdot p(x_2|x_1) \cdot \dots \cdot p(x_N|x_{N-1}\dots x_1). \quad (3.4)$$

To obtain a sample \vec{x} , we thus first sample x_1 from $p(x_1)$. Normalizing $p(x_1)$ is tractable as x_1 can only take the two values that correspond to the up and down spin configurations so that we do not need Markov chains for this task. We can then proceed similarly through the lattice, sampling one spin at a time conditioned on all previously sampled spins. Autoregressive sampling can be achieved with various network

architectures and we will revisit this task in the context of recurrent neural networks in [Section 3.2.3](#).

Normalizing flows achieve exact sampling of continuous distributions by transforming a simple distribution, such as a Gaussian, from which exact samples may be generated, into a more complex one by means of a learnable coordinate transformation. This architecture will be discussed in [Section 3.2.5](#).

3.1.2 Expressivity

Most distributions that are of interest in the context of this thesis do not stem from a particular distribution family. Thus using generative modeling with artificial neural networks as ansatz functions is a highly promising avenue in this case. Here, the motivation is again given by universal approximation theorems, which allow to increase the representable range of functions in a controlled way. In the following section, we discuss various network architectures that can be used for the described tasks.

3.2 NEURAL NETWORK ARCHITECTURES

Inspired by the research into the neural circuits of the human brain, the idea of synthetically recreating similar structures in software was developed. While a lot of theoretical considerations regarding these *artificial neural networks* (or neural networks for short) were developed in the 20th century, they only achieved major breakthroughs at the beginning of the last decade, owing to the high requirements of (previously unavailable) compute power. Since then, their employment led to state-of-the-art results in computer vision [[120](#), [228–230](#)], natural language processing [[231–233](#)] and various other fields that range from playing Go [[234](#)] to quantum physics [[235](#)]. Their vast applicability is explained by *universal approximation theorems* [[101–119](#)], which guarantee that the network may approximate (reasonably well-behaved) functions arbitrarily well given a sufficient number of parameters.

In the simplest case of a fully connected feed-forward neural network, as shown in the left panel of [Figure 3.2](#), the network is a map ϕ consisting of a composition of elementary functions ϕ_i

$$\phi(\vec{x}) = \phi_n \circ \phi_{n-1} \circ \dots \circ \phi_1(\vec{x}) \quad (3.5)$$

with ϕ_i given by an element-wise non-linear *activation function* σ_i , such as those shown in [Figure 3.1](#), which acts on an *affine transformation*

$$\phi_i : \mathbb{R}^n \rightarrow \mathbb{R}^m : \vec{x} \mapsto \phi_i(\vec{x}) = \sigma_i(W_i \cdot \vec{x} + \vec{b}_i), \quad (3.6)$$

so that each ϕ_i constitutes a single *feed-forward layer*. The affine transformation consists of a linear map given by the (not necessarily square)

weight matrix $W \in \mathbb{R}^{m \times n}$ and an offset vector $\vec{b} \in \mathbb{R}^m$, referred to as bias. The set of all weight matrices and bias vectors constitute the learnable, or variational, parameters of the neural network, also simply referred to as *weights*, which we summarize in the weight vector $\vec{\theta}$. We omit this explicit dependence of the map $\phi(\vec{x})$ on the variational parameters $\vec{\theta}$ for readability here and in the following. The first and last layers of the network are also referred to as the input and output layers, while the layers in between are called hidden layers. The computation of the network is typically only interpretable at the input and output layers and we refer to the representation of the input at the intermediate stages as a *latent representation*.

Note that this definition of a feed-forward neural network is completely agnostic to the task it will be used for; there is no underlying motivation for its architecture other than that upon increasing the network size the sought-after function becomes representable, which is guaranteed by the aforementioned theorems regarding universal approximation.

Obviously, there exist scenarios where we know that the map we aim to find can be restricted with prior knowledge about the problem at hand. For image classification tasks, for example, it does not matter *where* on the image an object is located, it is only important *that* it is located somewhere. We may thus use this knowledge regarding the present translation symmetry to only define such maps that respect it. The construction of general neural networks whose outputs are confined to submanifolds by various symmetries is an active field of research referred to as geometric deep learning [236].

3.2.1 Restricted Boltzmann Machines (RBMs)

A restricted Boltzmann machine (RBM) [237–239], in the context of this thesis, is a shallow two-layer network as shown in the right

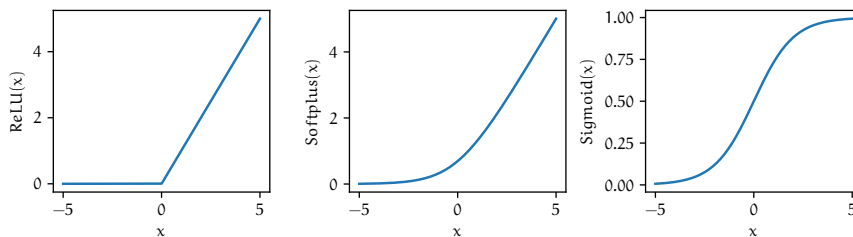


Figure 3.1: Popular choices of activation functions. Left: The rectified linear unit (ReLU) is zero for negative input values and the identity otherwise, $f(x) = \max(0, x)$. Center: The softplus activation function, $f(x) = \ln(1 + \exp(x))$, which has the same asymptotic behavior of the ReLU but is differentiable everywhere. Right: The sigmoid activation function, bounded to the range $(-1, 1)$ and given by $f(x) = 1/(1 + \exp(-x))$.

panel of Figure 3.2. Its output is a single number, that is obtained by summing over the neurons in the second layer after the application of the activation function

$$f(\vec{x}) = \sum \sigma(W \cdot \vec{x} + \vec{b}). \quad (3.7)$$

Typically, σ is chosen to be the hyperbolic cosine.

Historically, restricted Boltzmann machines were developed to model binary probability distributions over the visible neurons \vec{v} in the first layer and the hidden neurons \vec{h} in the second layer, giving rise to the joint probability distribution

$$p(\vec{v}, \vec{h}) = \frac{1}{Z} \exp\left(-\left(\vec{v}^T \cdot W \cdot \vec{h} + \vec{b}_1^T \cdot \vec{v} + \vec{b}_2^T \cdot \vec{h}\right)\right), \quad (3.8)$$

with the partition function $Z = \sum_{\{\vec{v}, \vec{h}\}} p(\vec{v}, \vec{h})$. As the computation of the partition function is generally intractable, we must resort to Markov chain Monte Carlo methods to generate samples from $p(\vec{v}) = \sum_{\{\vec{h}\}} p(\vec{v}, \vec{h})$. Restricted Boltzmann machines are to be differentiated from general Boltzmann machines, which also allow connections within each layer, making an analytical marginalization over the hidden units impossible. These probabilistic considerations break down, when the weights and biases become complex-valued, as is required for the modeling of complex wave function coefficients. The term restricted Boltzmann machine in the sense of Eq. (3.7) is therefore potentially misleading in those applications; we will nevertheless use it to connect to the existing literature.

3.2.2 Convolutional Neural Networks (CNNs)

Convolutional networks were originally developed for computer vision tasks, such as image recognition [120, 229, 240], as their (approximate [241]) translation invariance is aligned with the aim of recognizing whether an object is present in an image or not, without caring for its precise location.

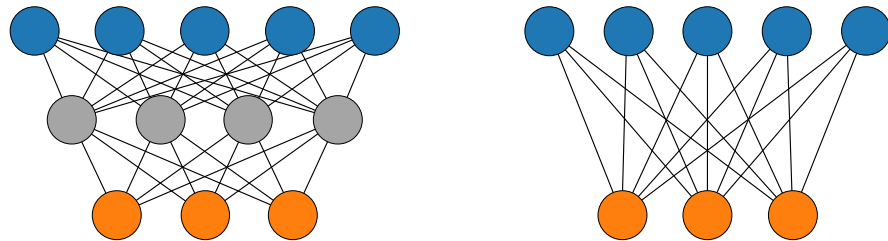


Figure 3.2: Left: Feed-forward architecture with an input (hidden, output) layer consisting of 3 (4, 5) neurons. Right: Restricted Boltzmann machine with 3 input and 5 output neurons.

Convolutional networks are inspired by the mathematical convolution operation between two functions f and g

$$(f * g)(x) = \int dx' f(x')g(x' - x). \quad (3.9)$$

Of particular interest to our considerations are convolutional kernels f which have highly localized support, so that $(f * g)(x)$ only depends on a small neighborhood interval centered around x . This setup is akin to what is typically encountered in deep convolutional networks for computer vision tasks, albeit in a discretized and two-dimensional fashion.

These discrete convolutional kernels are also referred to as *filters* and are scanned over the image, creating *local* latent representations of pixel neighborhoods. As the same filters are used independent of the position within the image, convolutional networks have considerably fewer weights than comparable feed-forward architectures, making them computationally more efficient. Each filter may pick up on different characteristics. For example, edges are detectable as local gradients within images. Thus, a discretized gradient is a common filter in the first convolutional network layers [242], and might take the following forms

Using the same weights for different parts of the input is known as weight sharing.

$$F_h = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \quad F_v = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad F_d = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}, \quad (3.10)$$

for horizontal, vertical, and diagonal edges respectively. In this example, the discrete support of the convolutional kernels only consists of three-by-three pixel patches, so that longer range correlations spanning many pixels may only be picked up in deep architectures. Of course, such filters are never hardcoded but learned by the network by minimization of a suited cost function.

After parallel application of N filters (resulting in N images), all values are fed into an element-wise non-linear activation function, before the next layer of filters is applied to the N images. One may freely choose the network depth, filter size as well as their number per layer (also referred to as channels).

3.2.3 Recurrent Neural Networks (RNNs)

In order to work with *sequential* data, such as time-series [243–245] or natural language [246–248], feed-forward based architectures present a suboptimal choice, as they are not capable of working with inputs of variable length. Hence, an architecture that can work with inputs of different shapes is required. This led to the development of recurrent

RNNs are, however, not restricted to modeling probability distributions.

networks, in which the input is treated sequentially in the order of occurrence within the sequence while treating each occurring item with the same network [249–251].

Recurrent networks are particularly suited to model discrete probability distributions, due to their *autoregressive* property, which allows us to assign *normalized* probabilities to an event \vec{x} by partitioning the probability into the product of its conditionals

$$p(\vec{x}) = p(x_1) \cdot p(x_2|x_1) \cdot p(x_3|x_2, x_1) \cdot \dots \cdot p(x_N|x_{N-1}\dots x_1), \quad (3.11)$$

as already stated in Section 3.1.1.2. This is achieved by scanning a recurrent cell iteratively over the input as shown in Figure 3.3 and keeping a memory vector of latent context information of the encountered items in a so-called hidden vector \vec{h} that is updated after each item

$$\vec{h}_i = f(\vec{h}_{i-1}, x_{i-1}). \quad (3.12)$$

This hidden vector is used to compute the conditional probabilities as

$$p(x_i|x_{i-1}\dots x_1) = g(\vec{h}_i). \quad (3.13)$$

In the simplest case, both f and g are shallow feed-forward networks [249]. However, they can also be more complex, featuring so-called ‘gating mechanisms’, as in the long short-term memory (LSTM) [250] or gated recurrent unit (GRU) [251] architectures.

Notice the advantages when generating samples from the encoded probability distribution using such architectures: Instead of having to

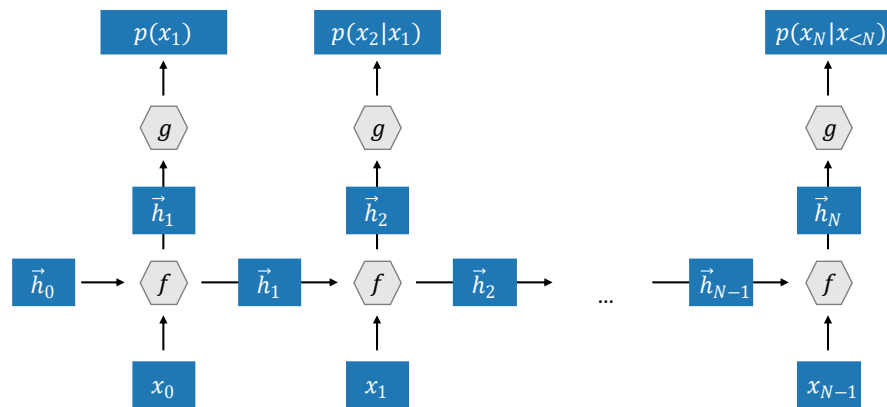


Figure 3.3: Treatment of an input sequence using a recurrent neural network. In each step, the RNN-cell f computes a new hidden state based on the previous hidden state and the data input x_i . As the RNN-cell f uses the same parameters for each operation, it is agnostic to the length of the sequence so that RNNs can, for example, be trained on one sequence length and deployed on another. Using the hidden state, another network g computes the conditional probabilities $p(x_i|x_{<i})$. This is possible since information flows strictly from left to right so that a causal interpretation is possible.

await thermalization of a Markov chain in order to get an uncorrelated sample, we are *guaranteed* to obtain uncorrelated samples as we sample from the probability distribution *directly*. This is possible since we generate each sample iteratively, only requiring computations of *local* partition functions, given by two terms corresponding to the two possible spin configurations in the case of the Ising model and the size of the dictionary in natural language tasks. The price we pay for the sequential treatment is our inability to parallelize the evaluation, thus increasing computation time. Another challenge is the possible loss of information when working with long sequences, such that the dependence of $p(x_N|x_{N-1}\dots x_1)$ on the first elements of the sequence may not be captured correctly. This was the main motivation for the development of the LSTM and GRU architectures. Dependencies in long sequences are not an issue for transformers [231].

While MCMC algorithms have been shown to work well in Ising-like models, imagine the overhead this would introduce in tasks such as natural language processing, where the local dimension is the size of the considered dictionary instead of only the two spin configurations. This underlines the necessity of being able to generate exact samples in certain scenarios.

Autoregressive sampling is also referred to as direct or exact sampling.

Random sentences are rarely meaningful, in contrast to random spin configurations.

3.2.4 Graph Neural Networks (GNNs)

If we know the input data to feature a graph-like structure, we may restrict the search space to permutation invariant functions by choosing a suited architecture. Scenarios in which the input data features such a structure are ubiquitous: The positions of atoms inside a molecule can be viewed as a graph, just like social networks; even sets can be understood as disconnected and trivial or all-to-all connected graphs.

Graphs consist of nodes that are connected through edges to their neighbors; the set of neighbors of a node i is referred to as the neighborhood \mathcal{N}_i . Each node is attributed a latent representation $\vec{h}_i \in \mathbb{R}^F$, that is updated in every layer of the graph network. This can be achieved in different ways, as discussed in [236]. Since we employed the graph attention variant [252] in [H], we discuss it here.

A graph attention layer updates the node features \vec{h}_i of node i as

$$\vec{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W \cdot \vec{h}_j \right), \quad (3.14)$$

with an activation function σ , a linear learnable transformation $W \in \mathbb{R}^{F \times F}$ and the self-attention coefficient $\alpha_{ij} \in \mathbb{R}$ between nodes i and j

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}, \quad (3.15)$$

where

$$e_{ij} = \vec{a}^T \cdot (W \cdot \vec{h}_i \parallel W \cdot \vec{h}_j). \quad (3.16)$$

Here, $\vec{a} \in \mathbb{R}^{2F}$ is a vector of learnable parameters, and \parallel denotes the concatenation of the vectors on its left and right-hand side.

So far, the graph network is permutation *equivariant*, meaning that upon changing the order of the inputs the order of the output after some layers changes accordingly. To build a permutation *invariant* map, we merely need a transformation that is agnostic to the ordering, such as a sum, so that we can map to a suitable output using a network ϕ of our choice

$$f(\vec{h}_1, \dots, \vec{h}_N) = \phi \left(\sum_i \vec{h}_i \right). \quad (3.17)$$

3.2.5 Invertible Neural Networks (INNs)

Modeling probability densities in continuous space constitutes another goal with numerous applications, such as modeling the density of fluids [253], the Brownian motion of particles [254] or derivative-pricing in finance [255] as well as sampling in lattice field theories [256–258] and modeling quasi-probability distributions in phase space [167, 209, 259] and many more. Thus, various neural network based approaches have been developed to encode complex density distributions, going beyond simpler previous machine learning approaches such as Gaussian mixture models.

The challenge when constructing such densities lies in parameterizing a preferably general probability density, from which we also wish to be able to generate exact samples. This can be achieved by using an *invertible neural network (INN)*, which constitutes a (unique) coordinate transform [260–266]. Invertible networks are also referred to as *normalizing flows (NFs)*.

An invertible neural network uses the change of variables formula

$$p(\vec{x}) = p(\vec{z}) \left| \frac{d\vec{z}}{d\vec{x}} \right|, \quad (3.18)$$

to construct a probability density $p(\vec{x})$ in the d -dimensional real space \mathcal{X} from a given latent space distribution $p(\vec{z})$ in the d -dimensional latent space \mathcal{Z} . The latent space is to be understood as an auxiliary space, in which we define $p(\vec{z})$ to be a simple probability density of our choosing, such as a d -dimensional Gaussian, while the real space is the interpretable space where the data is located. The Jacobian $|d\vec{z}/d\vec{x}|$ quantifies how the value of the probability density changes due to the stretching that is given by the coordinate transformation. The network f implements this learnable coordinate transformation

$$\vec{x} = f(\vec{z}), \quad (3.19)$$

where a particular challenge is to only construct such maps that are one-to-one and that allow for an efficient evaluation of the Jacobian, meaning sub-cubic computational cost in the problem dimension d .

One option, which we utilize in [Chapter 6](#), is a coordinate transformation that is composed of many small, elementary coordinate transformations referred to as ‘coupling blocks’

$$f(\vec{z}) = \varphi_1 \circ \varphi_2 \circ \dots \circ \varphi_N(\vec{z}). \quad (3.20)$$

It was originally introduced in [265], where each coupling block is defined as depicted in [Figure 3.4](#).

In the forward/left-to-right evaluation shown in the top of [Figure 3.4](#), the input vector \vec{u} is split in a random but fixed way into the vectors \vec{u}_1 and \vec{u}_2 so that their dimension is approximately half of the dimension of \vec{u} . Using the standard non-invertible feed-forward networks s_1 , s_2 , t_1 and t_2 , \vec{u}_1 and \vec{u}_2 are transformed into the output \vec{v}_1 and \vec{v}_2 by assigning

$$\begin{aligned} \vec{v}_1 &= \vec{u}_1 \odot \exp(s_2(\vec{u}_2)) + t_2(\vec{u}_2), \\ \vec{v}_2 &= \vec{u}_2 \odot \exp(s_1(\vec{v}_1)) + t_1(\vec{v}_1), \end{aligned} \quad (3.21)$$

where \odot means element-wise multiplication. Notably, although the networks s_i and t_i are not invertible, the full transformation is, so that \vec{u}_1 and \vec{u}_2 may be recovered from \vec{v}_1 and \vec{v}_2 as

$$\begin{aligned} \vec{u}_2 &= (\vec{v}_2 - t_1(\vec{v}_1)) \odot \exp(-s_1(\vec{v}_1)), \\ \vec{u}_1 &= (\vec{v}_1 - t_2(\vec{u}_2)) \odot \exp(-s_2(\vec{u}_2)), \end{aligned} \quad (3.22)$$

shown in the bottom of [Figure 3.4](#).

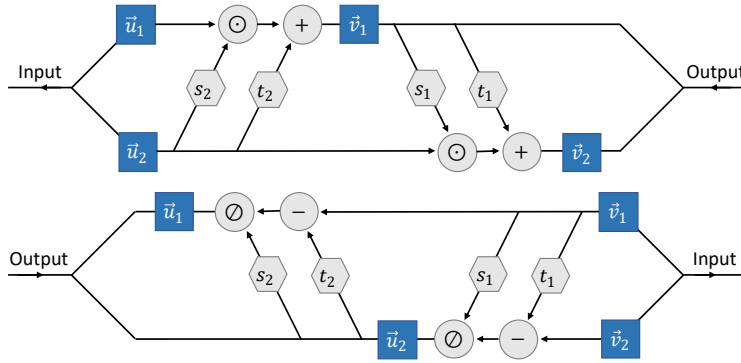


Figure 3.4: Computations carried out in a single coupling block for the forward-evaluation (top) and backward-evaluation (bottom). The data passing through the coupling block is randomly split into the vectors \vec{u}_1 and \vec{u}_2 of approximately equal size, and changed by the actions of the neural networks s_i and t_i as described in Eq. (3.21). As a typical invertible neural network consists of many coupling blocks, it is ensured that the full architecture allows very general interactions between the data, thereby constituting a powerful coordinate transform and an expressive architecture for encoding continuous probability densities.

The Jacobian of the first step of the transformation (3.21), mapping $\vec{u}_1 \rightarrow \vec{v}_1$ and $\vec{u}_2 \rightarrow \vec{u}_2$, becomes trivial to compute as

$$J = \frac{\partial(\vec{v}_1, \vec{u}_2)}{\partial(\vec{u}_1, \vec{u}_2)^T} = \begin{pmatrix} \text{diag}[\exp(s_2(\vec{u}_2))] & 0 \\ \partial\vec{v}_1/\partial\vec{u}_2^T & \mathbf{1} \end{pmatrix} \quad (3.23)$$

is observed to be of triangular form, such that its determinant is given by

$$|J| = \exp\left(\sum_i s_2(\vec{u}_2)_i\right), \quad (3.24)$$

thus incurring only linear cost in the problem dimension, instead of the generic cubic complexity [265]. Importantly, the Jacobian of the full transformation f as defined in (3.20), is obtained by multiplication of all 'elementary' Jacobians associated with both \vec{u}_1 and \vec{u}_2 . The same considerations hold true for the inverse step, as the Jacobian determinant in this case is simply the inverse one of the forward step.

Let us briefly mention, that it is also possible to construct normalizing flows with *continuous* depth, where depth is understood as the final integration time of the ODE defined by a neural network. These so-called *Neural ODEs (NODEs)* [267] use a neural network f to define the infinitesimal change of a hidden state h in time

$$\dot{\vec{h}}(t) = f(\vec{h}(t), t). \quad (3.25)$$

It is thus possible to obtain the state $\vec{h}(t^*)$ at some time of interest t^* using black-box integrators. As shown in [267], the computation of gradients of loss functions can be made efficient by using the *adjoint sensitivity method*, without needing to differentiate *through* the ODE-solver.

Neural ODEs are also referred to as continuous normalizing flows.

The simulation of quantum systems on classical computers is difficult, due to the exponential growth of the Hilbert space dimension with the number of system constituents N . This NP-hardness renders exact approaches to problems within the many-body regime futile. However, one may oftentimes employ suited approximations to make the problem of interest tangible.

For the purpose of this thesis, we differentiate between approximations that simplify the Hamiltonian of interest and such that work with the full quantum Hamiltonian and rather approximate the wave function. In the former case, one can imagine instances where being in the high occupation limit justifies working with the classical rather than the quantum Hamiltonian. This approximation will be discussed in detail in the context of the truncated Wigner Approximation (TWA) in [Section 4.2](#). The latter case which is concerned with approximating the wave function, is highly interesting, as it holds the potential for accurate solutions to the full quantum system given that we can find an efficient and sufficiently expressive parameterization of the wave function. Such encodings are based on the optimization of variational parameters that characterize this so-called *ansatz function*. Such ansatz functions can, but do not need to be based on physical insights.

For spin-1/2 systems in one dimension, one such physical insight is that ground states of local gapped Hamiltonians are known to satisfy an area law with respect to the entanglement entropy [190]. Hence, the entanglement entropy of a large enough region only grows with its boundary, implying that spins are only entangled with those spins in their neighborhood, but not with those that are far away. This is a remarkable observation: If one were to pick a random state in Hilbert space, it would *almost always* show a volume law scaling of the entanglement entropy, meaning a growth that is proportional to the size of the considered region [190]. To accurately model the ground state, we can thus target this so-called *physical corner of Hilbert space* and may safely ignore its complement, posing a striking decrease in complexity. The remaining question is *how* we accurately target this corner numerically. One can show that matrix product states are sufficient to efficiently encode area-law states in one dimension [95, 96]. One-dimensional matrix product states are constructed by assigning a rank-three tensor K to each spin, where two of the indices, ranging from one to χ , serve to mediate correlations between the spin and its neighbors while the last index chooses the spin orientation. To com-

χ is referred to as the bond dimension and may be raised to increase the expressivity of the ansatz.

pute the wave function coefficient $\psi(\vec{s})$ of a given spin configuration \vec{s} we must contract the tensor network

$$\psi(\vec{s}) = \langle s_1 s_2 \dots s_N | \psi \rangle = \sum_{l_1, \dots, l_N} K_1^{s_1 l_1 l_2} K_2^{s_2 l_2 l_3} \dots K_N^{s_N l_N l_1}, \quad (4.1)$$

where we assumed periodic boundary conditions so that both K_1 and K_N are also tensors of rank three. The total number of variational parameters is $2\chi^2 N$, an immense reduction compared to the naïve exponential scaling of 2^N . Thus, both storage and evaluation of the tensor network are efficient as well as its optimization, as discussed in [95, 96]. The guiding principle in the construction of matrix product states is the locality of the Hamiltonian, which served as intuition for how the tensor network shall be contracted: Those tensors corresponding to neighboring spins are multiplied directly so that their correlations become easy to represent. In contrast, correlations between spins that are far away must be mediated through the enclosed tensors, so that they are naturally harder to represent, which makes a restriction to area-law states plausible.

While the tremendous success of tensor network methods for one-dimensional systems is a significant achievement, generalizations to two dimensions are not straightforward. While many approaches have been put forward [96, 249, 268–270], their success never matched that of their one-dimensional counterparts. A more recent variational technique based on neural networks thus caused significant interest, as it does not rely on the dimension of the underlying system in the same fundamental way tensor networks do. These approaches are coined *neural quantum states (NQS)*, are central to this thesis, and will be discussed in the following section.

4.1 NEURAL QUANTUM STATES

Neural quantum states are a class of variational methods, that can be used to approximate wave functions of many-body quantum systems [121]. While fermionic and bosonic systems in both discrete and continuous space are applicable to the treatment within the NQS framework [167–185], we will here focus on the arguably simpler case of spin-1/2 systems [A, C, 97–99, 121, 134–153, 162–166], as it is most relevant to this thesis.

Neural quantum states approximate wave function coefficients by assigning

$$\ln(\psi(\vec{s})) = f(\vec{s}), \quad (4.2)$$

where f is a neural network with trainable parameters and \vec{s} a spin configuration. For numerical stability, one usually interprets the network output as the logarithm of the wave-function coefficient, as written in Eq. (4.2).

While the initial proposal of neural quantum states [121] chose f to be an RBM (Section 3.2.1), subsequent works explored other architectures, such as CNNs (Section 3.2.2) [F, 99, 148, 154], RNNs (Section 3.2.3) [A, 134, 158, 175, 271] and GNNs (Section 3.2.4) [149, 170–173, 182, 183], each with their own set of advantages and disadvantages [C]. We note that a priori any differentiable function that maps spin configurations \vec{s} onto complex numbers can be treated within the variational Monte Carlo framework used to optimize neural quantum states.

Therefore the choice of network does not necessarily require a physical motivation. In the case of restricted Boltzmann machines, for instance, such a motivation is lacking since there exists no sense of locality as it is a simple, fully connected feed-forward architecture. This non-locality is the cause for the perhaps somewhat surprising observation that RBMs are capable of efficiently encoding volume law states [186–188], in stark contrast to tensor network states. This observation generated a lot of interest in neural quantum states, as it renders them potential solutions in instances where the system is expected to exhibit volume-law entanglement, such as the Sachdev-Ye-Kitaev model [194–197]. Simultaneously, this raises the question, of what fundamentally limits the capabilities of neural quantum states, given that it is not the encoded entanglement.

To approximate generic wave functions, the output of the network must be complex-valued. This can either be achieved by using complex parameters or by employing two networks with real-valued parameters so that $f(\vec{s})$ is set together from an amplitude and a phase network. We can, however, treat both approaches on equal footing, by interpreting any occurring complex network parameter as the two real parameters that define its real and imaginary part, and will thus assume that all parameters are real-valued, allowing for a unified discussion in the following.

In contrast to tensor network states, the optimization of neural quantum states is immensely challenging due to the highly non-linear nature of neural networks [157, 272]. An additional challenge lies in the high precision that is required to obtain faithful estimates of ground states, a prerequisite that is atypical for standard machine learning applications. This oftentimes forces one to go beyond simple stochastic gradient descent schemes, as we will show in the following.

As variational ansatz functions, neural quantum states fall into the larger framework of variational Monte Carlo techniques. In variational Monte Carlo, we generally deal with an unnormalized variational wave function $|\psi\rangle$, whose expansion in the computational basis $\{\vec{s}\}$ is given by

$$|\psi\rangle = \sum_{\{\vec{s}\}} \psi(\vec{s}) |\vec{s}\rangle. \quad (4.3)$$

The following discussion is based on [G].

The challenge in the following consists of ensuring that all occurring expressions, such as expectation values or gradients, are obtainable as Monte Carlo expectation values so that at no point an evaluation of the exponentially large sum over all basis configurations is required.

4.1.1 Expectation Values of Operators

For instance, given an operator \mathbf{O} whose expectation value we are interested in, we compute

$$\begin{aligned}
\langle \mathbf{O} \rangle &= \frac{\langle \psi | \mathbf{O} | \psi \rangle}{\langle \psi | \psi \rangle} \\
&= \sum_{\{\vec{s}\}, \{\vec{s}'\}} \frac{\psi(\vec{s})^* \psi(\vec{s}')}{\langle \psi | \psi \rangle} \langle \vec{s} | \mathbf{O} | \vec{s}' \rangle \\
&= \sum_{\{\vec{s}\}} \frac{|\psi(\vec{s})|^2}{\langle \psi | \psi \rangle} \sum_{\{\vec{s}'\}} \frac{\psi(\vec{s}')}{\psi(\vec{s})} \langle \vec{s} | \mathbf{O} | \vec{s}' \rangle \\
&\approx \mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\sum_{\{\vec{s}'\}} \frac{\psi(\vec{s}')}{\psi(\vec{s})} \langle \vec{s} | \mathbf{O} | \vec{s}' \rangle \right] \\
&= \mathbb{E}_{\vec{s} \sim p(\vec{s})} [O_{\text{loc}}(\vec{s})]
\end{aligned} \tag{4.4}$$

where we multiplied and divided by $\psi(\vec{s})$ in the third line, defined the probability distribution of the basis configuration

$$p(\vec{s}) = \frac{|\psi(\vec{s})|^2}{\langle \psi | \psi \rangle} \tag{4.5}$$

and replaced the exponentially costly sum over all basis configurations $\{\vec{s}\}$ by a Monte-Carlo expectation value that only requires a sub-exponential number of samples obtained from $p(\vec{s})$ to converge. We still have a sum over the complete basis in the definition of the so-called 'local energies'

$$O_{\text{loc}}(\vec{s}) = \sum_{\{\vec{s}'\}} \frac{\psi(\vec{s}')}{\psi(\vec{s})} \langle \vec{s} | \mathbf{O} | \vec{s}' \rangle. \tag{4.6}$$

However, as we are usually interested in the expectation values of *local* operators \mathbf{O} , or their sums, the non-vanishing entries of $\langle \vec{s} | \mathbf{O} | \vec{s}' \rangle$ are only linear in the system size so that the evaluation of $O_{\text{loc}}(\vec{s})$ becomes efficient.

4.1.2 Generation of Samples

As described in [Section 3.1.1.1](#), samples with regards to $p(\vec{s})$ are obtained either in an autoregressive fashion if the architecture supports direct sampling [134, 135] or through Monte Carlo Markov chains. In the latter case, the most common choice is Metropolis-Hastings sampling, where the ratio of the probability of the proposed configuration

\vec{s}' divided by the probability of the current configuration \vec{s} gives the acceptance probability

$$r = \min \left(1, \frac{p(\vec{s}')}{p(\vec{s})} \right). \quad (4.7)$$

The generation of the proposal configurations \vec{s}' typically is achieved locally, meaning that \vec{s}' and \vec{s} differ in only a single position, however, various proposal strategies are possible. We refer the interested reader to [G] for technical details.

4.1.3 Ground State Optimization

By replacing \mathcal{O} with the Hamiltonian \mathbf{H} of the system in Eq. (4.4) and taking gradients with respect to the variational parameters $\vec{\theta}$, we can optimize the variational wave function towards lower energies. Analytically, one obtains

$$\begin{aligned} \nabla_{\vec{\theta}} \langle \mathbf{H} \rangle &= 2\text{Re} \left(\mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\vec{\mathcal{O}}(\vec{s}) H_{\text{loc}}(\vec{s}) \right] - \right. \\ &\quad \left. \mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\vec{\mathcal{O}}(\vec{s}) \right] \mathbb{E}_{\vec{s} \sim p(\vec{s})} [H_{\text{loc}}(\vec{s})] \right) \\ &= 2\text{Re} \left(\vec{F} \right) \end{aligned} \quad (4.8)$$

where we defined the so-called force-vector \vec{F} and introduced the logarithmic derivatives

$$\vec{\mathcal{O}}(\vec{s}) = \nabla_{\vec{\theta}} \ln(\psi(\vec{s})) \quad (4.9)$$

as well as the local energies $H_{\text{loc}}(\vec{s})$ belonging to the Hamiltonian \mathbf{H}

$$H_{\text{loc}}(\vec{s}) = \sum_{\{\vec{s}'\}} \frac{\psi(\vec{s}')}{\psi(\vec{s})} \langle \vec{s} | \mathbf{H} | \vec{s}' \rangle. \quad (4.10)$$

With this, we can define the iterative gradient descent scheme

$$\vec{\theta}^{n+1} = \vec{\theta}^n - \eta \nabla_{\vec{\theta}} \langle \mathbf{H} \rangle, \quad (4.11)$$

where η denotes a small learning rate.

The requirement of highly accurate representations of the state of interest is in stark contrast to traditional machine learning applications. One thus finds that simple, first-order optimization, is in many instances insufficient to obtain state-of-the-art results. Fortunately, by insertion of the so-called *quantum geometric tensor*, we can turn the first-order optimization into a second-order optimization, coined *natural gradient descent*. Effectively, we herein take the curvature of the variational manifold into account and optimize the network parameters in dependence on one another. It can also be understood as an

imaginary time evolution that iteratively projects out higher energy contributions [93] and is given by

$$\vec{\theta}^{n+1} = \vec{\theta}^n - \eta \text{Re}(S)^{-1} \nabla_{\vec{\theta}} \langle \mathbf{H} \rangle, \quad (4.12)$$

with S the connected correlator of the outer product of the logarithmic derivatives

$$S = \mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\vec{\mathcal{O}}(\vec{s}) \cdot \vec{\mathcal{O}}(\vec{s})^T \right] - \mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\vec{\mathcal{O}}(\vec{s}) \right] \cdot \mathbb{E}_{\vec{s} \sim p(\vec{s})} \left[\vec{\mathcal{O}}(\vec{s})^T \right]. \quad (4.13)$$

As S is often rank-deficient, care must be taken when computing its inverse. Therefore, different regularization schemes have been put forward, including the addition of a small identity

$$S \rightarrow S + \epsilon \mathbb{1} \quad (4.14)$$

and only inverting those eigenvalues that are larger than a given threshold [G]. Recently, it has also been shown that regularization with respect to the signal-to-noise ratio of the noisy estimates of S and $\nabla_{\vec{\theta}} \langle \mathbf{H} \rangle$ can significantly enhance the performance in applications concerned with real-time evolution [154].

As the size of S is dictated by the number of network parameters N_p , and the computational cost of matrix inversions scales approximately cubically in the matrix size, one is restricted to networks with few parameters, far away from universal approximation regimes. However, recently, progress has been made to reduce the computational cost in those cases where the number of samples N_s used to estimate S and $\nabla_{\vec{\theta}} \langle \mathbf{H} \rangle$ is significantly lower than the number of network parameters N_p . By a suited rearrangement of the terms in Eq. (4.12), the to be inverted matrix is only of size $N_s \times N_s$ and has full rank [98, 99], so that even some of the regularization techniques may no longer be required.

4.1.4 Unitary Time Evolution

To turn the imaginary time evolution into a real-time evolution, it is sufficient to replace the real-valued learning rate η by $i\tau$, with τ a real-valued time step. The time-dependent variational principle that is obtained in this way admits two different formulations [G]. The principle of least action leads to

$$\text{Im}(S) \cdot \partial_t \vec{\theta} = \text{Im}(-i\vec{F}), \quad (4.15)$$

while the minimization of a suited distance measure results in

$$\text{Re}(S) \cdot \partial_t \vec{\theta} = \text{Re}(-i\vec{F}). \quad (4.16)$$

Only in the former case can the energy be shown to be conserved during the real-time evolution. Similar to the imaginary time evolution, S is typically ill-conditioned and thus needs to be regularized in order to obtain a stable solution [G].

While the projective character of the imaginary time evolution does not penalize errors during the evolution, such safeguards do not exist for the case of real-time evolution, where errors will accumulate as time progresses. To minimize these effects, regularization should only be as high as necessary, and more sophisticated integrators in comparison to Euler steps (4.12) should be chosen.

A pseudo-code overview of the natural gradient scheme for both imaginary and real-time evolution is given in [G].

In the case of imaginary time evolution, the dynamics lead to a fixed point. This is a crucial difference compared to the case of real-time evolution.

4.2 TRUNCATED WIGNER APPROXIMATION

We now introduce the truncated Wigner approximation. The following introduction is a slightly modified excerpt from [D], which we here present for completeness.

The high occupation limit of a bosonic system of interest is usually too complex to admit computational solutions to the full quantum dynamics. Instead, as the populations are increased, the role of quantum fluctuations compared to the mean-field dynamics grows smaller. This emergent simplicity implies that the predictions made by approximate techniques become more and more accurate, justifying a semiclassical approach to the problem. One such semiclassical technique is the truncated Wigner approximation (TWA). At its core, all mode operators are replaced by complex numbers [209]

$$\mathbf{a}^{(\dagger)} \rightarrow \alpha^{(*)}, \quad (4.17)$$

which amounts to a lowest-order expansion of the Wigner-Weyl correspondence rules. Consequently, the Hamiltonian operator $\mathbf{H}(\mathbf{a}, \mathbf{a}^\dagger)$ reduces to a classical Hamiltonian function $H(\alpha, \alpha^*)$ of the complex-valued phase-space coordinates (α, α^*) .

In general, the von Neumann equation

$$i\partial_t \mathbf{O}(t) = [\mathbf{O}(t), \mathbf{H}], \quad (4.18)$$

for some time-dependent operator \mathbf{O} , can be translated into an equation of motion for the operator's Weyl symbol O in phase space via the Moyal bracket

$$\{A, B\}_{\text{MB}} = \frac{A}{2} \sin \left[2 \left(\overleftarrow{\partial}_\alpha \overrightarrow{\partial}_{\alpha^*} - \overleftarrow{\partial}_{\alpha^*} \overrightarrow{\partial}_\alpha \right) \right] B, \quad (4.19)$$

which results in [209]

$$i\partial_t O(t) = \{O(t), H\}_{\text{MB}}. \quad (4.20)$$

In TWA, one is interested in the evolution of the elementary c -numbers α representing the mode operators to leading order, for which Eq. (4.20) simplifies to the classical Poisson bracket

$$i\partial_t\alpha(t) = \{\alpha(t), H\}_{\text{PB}} = \partial_{\alpha^*(t)}H, \quad (4.21)$$

since all higher-order derivatives vanish.

In the case when multiple modes are coupled Eq. (4.21) defines a system of coupled differential equations that may be solved numerically using a suited integrator. The initial conditions of the system $\vec{\alpha}(0)$ are obtained as Monte Carlo samples from the initial Wigner W -distribution, for which two cases are relevant. When considering the vacuum $|0\rangle$, the Wigner W -distribution takes the form of a two-dimensional Gaussian, which is centered at the origin and contains no correlations, i.e.

$$|\psi\rangle = |0\rangle \iff \alpha \sim \mathcal{N}\left[0, \frac{1}{2}\right] + i\mathcal{N}\left[0, \frac{1}{2}\right]. \quad (4.22)$$

The samples of coherent states $|\gamma\rangle$ have the same variance but are displaced by the square root of their mean particle number $\langle \mathbf{n} \rangle = |\gamma|^2$ from the origin

$$|\psi\rangle = |\gamma\rangle \iff \alpha \sim \mathcal{N}\left[\text{Re}(\gamma), \frac{1}{2}\right] + i\mathcal{N}\left[\text{Im}(\gamma), \frac{1}{2}\right], \quad (4.23)$$

since $|\gamma\rangle$ itself is obtained by acting with the displacement operator $D(\gamma)$ on the vacuum, see (2.57).

In the presence of thermal fluctuations, the pure vacuum state $|0\rangle$ is replaced by the thermal ensemble $\rho \propto \exp(-\beta \mathbf{a}^\dagger \mathbf{a})$ with the inverse temperature $\beta = 1/T$. Accordingly, the standard deviations of the Wigner W -distribution are rescaled as

$$\frac{1}{2} \rightarrow \sqrt{1 + 2n_{\text{BE}}(\beta)} \frac{1}{2}, \quad (4.24)$$

which is equivalent to adding $n_{\text{BE}}(\beta)$ to both variances, where

$$n_{\text{BE}}(\beta) = \frac{1}{e^\beta - 1} \quad (4.25)$$

denotes the Bose-Einstein distribution.

The expectation value of some observable \mathbf{O} is obtained as the stochastic average over all generated samples

$$\langle \mathbf{O} \rangle_{\text{TWA}} = \frac{1}{|\mathcal{S}|} \sum_{\alpha \in \mathcal{S}} O(\alpha, \alpha^*), \quad (4.26)$$

where \mathcal{S} denotes the set of all samples. Note here that we again rely on the correspondence between \mathbf{O} and O given in Eq. (4.17).

Having introduced all necessary tools in TWA, we want to briefly comment on its regimes of applicability. In the limit of high occupations $\langle \mathbf{n} \rangle \gg 1$ and initial states with minimal fluctuations, i.e. coherent

states, the operator to c -number correspondence Eq. (4.17) is a justified simplification, since the relative fluctuations scale as $1/\sqrt{\langle n \rangle}$. The fluctuations of the initial state are herein represented accurately, as they correspond to exact samples of the Wigner W -distribution of the initial state $|\psi\rangle$, while only those fluctuations and inference effects are due to the unitary evolution under H are not captured. In the limit $\hbar \rightarrow 0$, the dynamics generated by TWA become exact. For a more thorough picture on this matter, see Ref. [273], in which the impossibility of quantum phase-space trajectories for generic (i.e. anharmonic) quantum systems is discussed. For further reading regarding TWA and other semiclassical techniques, we refer the reader to [209] and [274].

Part II

RESULTS

TIME-DEPENDENT VARIATIONAL PRINCIPLE FOR OPEN QUANTUM SYSTEMS WITH ARTIFICIAL NEURAL NETWORKS

5.1 CONTRIBUTIONS

This chapter presents the publication in [A]. Martin Gärttner proposed the project and I developed the code and ran all simulations in collaboration with Markus Schmitt. All authors contributed to the writing of the text. A GPU- and cluster-ready implementation of the utilized algorithms can be found as part of the published jVMC library [G].

5.2 MOTIVATION

As discussed in [Chapter 2](#) and [Chapter 4](#), the numerical simulation of many-body quantum systems is a highly challenging endeavor, even more so if we consider the case of open quantum systems (see [Section 2.3](#)). Neural quantum states were quickly adapted for those instances, relying on purification schemes of the density matrix [162–165]. We herein formulate another way to use neural quantum states to encode the dissipative evolution of the density matrix building on the POVM formalism that was discussed in [Section 2.4](#). Our work is similar in this regard to [168], but while we rely on an explicit update rule, Luo et. al. employ an iterative optimization scheme.

© 2024 American Physical Society

Time-Dependent Variational Principle for Open Quantum Systems with Artificial Neural Networks

Moritz Reh^{1,*}, Markus Schmitt², and Martin Gärtner^{1,3,4}

¹*Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*

²*Institut für Theoretische Physik, Universität zu Köln, 50937 Köln, Germany*

³*Physikalisches Institut, Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany*

⁴*Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

 (Received 26 April 2021; revised 16 July 2021; accepted 5 November 2021; published 1 December 2021)

We develop a variational approach to simulating the dynamics of open quantum many-body systems using deep autoregressive neural networks. The parameters of a compressed representation of a mixed quantum state are adapted dynamically according to the Lindblad master equation by employing a time-dependent variational principle. We illustrate our approach by solving the dissipative quantum Heisenberg model in one dimension for up to 40 spins and in two dimensions for a 4×4 system and by applying it to the simulation of confinement dynamics in the presence of dissipation.

DOI: 10.1103/PhysRevLett.127.230501

Introduction.—Solving the quantum many-body problem where it is analytically intractable constitutes a formidable challenge due to the inherent curse of dimensionality with growing system size. Today, two main routes are pursued to address this issue. On the one hand, the boundaries of classical computation are pushed by the development of tailored numerical techniques that build on the inherent structure of the quantum state of interest to find compressed representations using a subexponential number of variational parameters [1–4]. On the other hand, recent years have brought tremendous progress in the realization of quantum simulators as originally envisioned by Feynman [5,6], which emulate paradigmatic quantum many-body models using precisely controlled synthetic quantum systems of ultracold atoms in optical lattices [7–13], trapped ions [14–16], Rydberg atoms [17–19], and many more [20–31]. These “noisy intermediate-scale quantum” (NISQ) simulators [32] already present a valuable expansion of our scientific toolbox, enabling the discovery of new physical phenomena [12,18,19,31,33–35]. In particular, they challenge the numerical state of the art and open up largely uncharted terrain, e.g., nonequilibrium quantum matter in two spatial dimensions. As the term NISQ implies, the openness of these quantum systems will play a central role for near-term applications, and accounting for it appropriately is one of the key challenges.

In this work, we present a novel way to simulate the dynamics of open quantum systems (OQS) using a neural network encoding of the quantum state, which is relevant for two reasons: In view of the recent experimental developments, computational tools that can keep up with the system sizes of quantum simulators also in intermediate spatial dimensions are highly desired as they allow us to certify experimental observations and provide a link to

theoretical models. Simultaneously, the exploration of phenomena associated with driven dissipative systems is a major aim in itself, for which our approach opens new possibilities.

The state of an OQS is described by the density operator $\hat{\rho}$, whose dynamics, for Markovian systems, is governed by a Lindblad master equation. For a system of N spin-1/2 particles considered here the curse of dimensionality manifests in the 4^N coefficients necessary to fully represent $\hat{\rho}$, which limits exact numerical treatments to small N . Various numerical methods have been developed to reduce this complexity [36], each coming with different strengths and limitations. Stochastic Monte Carlo wave function (MCWF) methods [37–41] achieve a quadratic improvement of the N scaling at the cost of requiring statistical averaging. Semiclassical [42–45] and mean-field-like methods [46,47] provide a polynomial scaling in N but often suffer from uncontrolled approximations and numerical instabilities. Tensor network based approaches [1,2,48–56] are limited to weakly entangled states and require further approximations if applied in dimensions $d > 1$ [57–59]. A recently introduced class of methods, that can potentially resolve many of these issues are neural network quantum states (NQS) [60–75]. NQS have been applied successfully to OQS [76–79]. A natural approach is to employ a latent state purification [80]; however, this procedure has so far been restricted to shallow neural network architectures. A more recent work uses a probabilistic representation of the quantum state [64] which allows the use of deeper, more expressive networks but has the drawback of being forced to globally optimize the network parameters in each time step.

Here we introduce a numerical approach, summarized graphically in Fig. 1, that is not restricted in terms of

network architectures and operates based on explicit second-order local updates, thus overcoming structural and technical limitations of previously proposed methods. The derivation of a first-order differential equation for the time dependence of variational parameters in the context of a probabilistic formulation of quantum mechanics is a central result of our work. Thereby, our method expands the capabilities of previous approaches [64,76] in terms of system sizes and timescales reached reliably. This is demonstrated by the application to benchmark problems of spin systems in 1D and 2D geometries and by showing a first physics-motivated application.

Probabilistic representation.—Any quantum state $\hat{\rho}$ can be represented equivalently as a probability distribution P over measurement outcomes using positive operator valued measures (POVMs) [61,62,64,81]:

$$P^{\mathbf{a}} = \text{tr}(\hat{\rho}\hat{M}^{\mathbf{a}}), \quad (1)$$

where $\hat{M}^{\mathbf{a}} = \hat{M}^{a_1} \otimes \dots \otimes \hat{M}^{a_N}$ are measurement operators associated with the outcome $\mathbf{a} = a_1 \dots a_N$ of a tomographically complete measurement on N spins. We choose \hat{M}^i to be the symmetric informationally complete-POVM (SIC-POVM), or tetrahedral POVM [61]. Its elements are obtained from the definition $M^{\mathbf{a}} = (\mathbb{1} + \vec{s}^{\mathbf{a}} \cdot \vec{\sigma})/4$, in which the $\vec{s}^{\mathbf{a}}$ form a tetrahedron on the surface of the Bloch sphere and $\vec{\sigma}$ denotes the vector of Pauli matrices. Inverting Eq. (1) gives

$$\hat{\rho} = P^{\mathbf{a}} T^{-1\mathbf{a}\mathbf{a}'} \hat{M}^{\mathbf{a}'} \quad (2)$$

with the overlap matrix $T^{\mathbf{a}\mathbf{a}'} = \text{tr}(\hat{M}^{\mathbf{a}}\hat{M}^{\mathbf{a}'})$, where implicit summation over repeated indices is assumed from here on. Since the POVM elements $\hat{M}^{\mathbf{a}}$ form an operator basis, observables can be decomposed as $\hat{O} = \Omega^{\mathbf{a}}\hat{M}^{\mathbf{a}}$ and their expectation values become $\langle \hat{O} \rangle = P^{\mathbf{a}}\Omega^{\mathbf{a}}$. Compared to the complex-valued density matrix or its purification, the probabilistic representation has the advantage that it allows us to directly leverage the highly sophisticated toolbox for generative models developed in recent years by the machine learning (ML) community [82–84].

The dynamics of Markovian OQS is described by the Lindblad master equation [36]

$$\dot{\hat{\rho}} = -i[\hat{H}, \hat{\rho}] + \gamma \sum_i (\hat{L}^i \hat{\rho} \hat{L}^{i\dagger} - \frac{1}{2} \{\hat{L}^{i\dagger} \hat{L}^i, \hat{\rho}\}), \quad (3)$$

with $[\dots]$ ($\{\dots\}$) denoting the (anti-)commutator. The operators \hat{L}^i are commonly referred to as jump operators and are representative of the dissipative processes that the system is subject to. Differentiating Eq. (1) and inserting Eqs. (3) and (2) allows us to state the master equation in the probabilistic formulation:

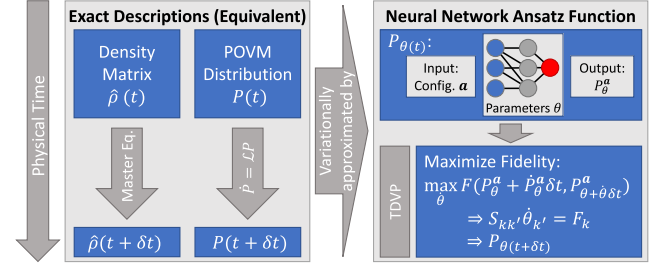


FIG. 1. Illustration of the variational approach to OQS dynamics. On the left, the standard density matrix formalism is shown along with the equivalent probabilistic formulation using POVMs. The right-hand side shows the variational approach, in which an artificial neural network is used as an ansatz function for the probability distribution over POVM outcomes. The main contribution of this work is the TDVP illustrated in the bottom box which leads to a general and accurate scheme for updating the network parameters according to the dynamics dictated by the master equation.

$$\dot{P}^{\mathbf{a}} = \mathcal{L}^{\mathbf{a}\mathbf{b}} P^{\mathbf{b}}. \quad (4)$$

The full expression for the Lindbladian \mathcal{L} is given in the Supplemental Material [85]. Crucially, \mathcal{L} is sparse since the restriction to one- and two-body interactions in Eq. (3) is preserved in the probabilistic reformulation, allowing us to evaluate its action efficiently.

Nonetheless, Eq. (4) is numerically intractable for many-body systems, because of the exponentially large number of coefficients $P^{\mathbf{a}}$. In the following we employ a variational approximation by introducing a trial distribution $P_{\theta}^{\mathbf{a}}$ with variational parameters θ , see Fig. 1. The compressed representation of the state in a polynomial number of variational parameters renders the approach numerically feasible.

Time dependent variational principle for POVMs.—The main theoretical contribution of our work is a time dependent variational principle (TDVP) for POVM-probability distributions which dictates the time dependence of the network parameters $\theta(t)$ by determining the closest approximation of the Lindbladian dynamics within the variational manifold. The starting point is a distance measure $\mathcal{D}(P, Q)$ for probability distributions P, Q . Assuming a small time step τ at time t with associated network parameters $\theta(t)$, the aim is to minimize the distance between the updated POVM-probability distribution $P_{\theta(t)+\dot{\theta}\tau}^{\mathbf{a}}$ and the time-propagated distribution $P_{\theta(t)}^{\mathbf{a}} + \tau \mathcal{L}^{\mathbf{a}\mathbf{b}} P_{\theta(t)}^{\mathbf{b}}$ that follows from Eq. (4). We found that two natural choices for the distance measure \mathcal{D} are equivalent for this purpose, because they describe locally identical geometries. The first is the Hellinger distance $\mathcal{D}_H(P, Q) = 1 - F(P, Q)$, which is defined via the Bhattacharyya coefficient (or classical fidelity) $F(P, Q) = \sum_{\mathbf{a}} \sqrt{P^{\mathbf{a}} Q^{\mathbf{a}}}$. The second is the Kullback-Leibler divergence $\mathcal{D}_{\text{KL}}(P, Q) = \sum_{\mathbf{a}} P^{\mathbf{a}} \log(P^{\mathbf{a}}/Q^{\mathbf{a}})$. In both cases, a second order

consistent small- τ expansion of $\mathcal{D}_{H/KL}(P_{\theta(t)+\dot{\theta}\tau}^a, P_{\theta(t)}^a + \tau \mathcal{L}^{ab} P_{\theta(t)}^b)$ and subsequently demanding stationarity to find the optimal parameter update $\dot{\theta}$ yields the TDVP equation

$$S_{kk'} \dot{\theta}_{k'} = F_k. \quad (5)$$

Here, S denotes the Fisher-metric $S_{kk'} = \langle O_k^a O_{k'}^a \rangle_{\mathbf{a} \sim P}^c$, $F_k = \langle O_k^a \mathcal{L}^{ab} (P^b / P^a) \rangle_{\mathbf{a} \sim P}^c$ and the repeated indices \mathbf{a} inside the brackets are not summed over. The brackets denote connected correlation functions $\langle AB \rangle^c = \langle AB \rangle - \langle A \rangle \langle B \rangle$ of expectation values with respect to the POVM-distribution P and $O_k^a = \partial_{\theta_k} \log P^a$. We provide a detailed derivation of Eq. (5) in the Supplemental Material [85]. It is worth noting that for models that are normalized by default [such as recurrent neural networks (RNNs)] $\langle O_k^a \rangle = 0$. Usually, Eq. (5) is ill conditioned and needs to be regularized. Here, advanced regularization schemes such as described in Ref. [63] are applicable but they did not turn out to be crucial for the test cases we consider.

The TDVP equation (5) exhibits a number of features beneficial for the numerical time evolution of $\theta(t)$. As a result of the employed short-time expansion, the variational optimization problem becomes convex and information about the local geometry of the variational manifold is taken into account in the form of the Fisher-metric S . Upon inverting S , the differential equation can be solved straightforwardly with explicit integration schemes. Importantly, Monte Carlo sampling is only required once per time step. These features are in contrast to the implicit integration scheme presented in Ref. [64], where Monte Carlo sampling is required at each optimization step performed for the iterative global minimization of a nonconvex cost function.

Network architecture.—Neural networks are highly non-linear universal function approximators in the limit of large networks [86–88]. For the purpose of generative modeling autoregressive networks are advantageous, because they enable direct generation of uncorrelated samples; therefore, various autoregressive architectures have recently been explored for NQS [64,65,89–91]. In the following, we employ RNNs, which belong to this family of network architectures (see Supplemental Material [85] for details).

Numerical results.—To illustrate the accuracy and scalability of our method we apply it to the anisotropic Heisenberg model,

$$\hat{H} = \sum_{\langle ij \rangle} (J_x \hat{X}_i \hat{X}_j + J_y \hat{Y}_i \hat{Y}_j + J_z \hat{Z}_i \hat{Z}_j) + \sum_i h_z \hat{Z}_i, \quad (6)$$

with nearest neighbor interactions and periodic boundary conditions, which was also used in Ref. [64] as a benchmark system. The considered decoherence channel is spontaneous decay given by the jump operator $\hat{L} = \hat{\sigma}^- = (\hat{X} - i\hat{Y})/2$ acting on each spin. We obtain benchmark data using exact simulations for $N = 10$ spins and test our approach in the case of $N = 10$ [85] and $N = 40$ [Figs. 2(a) and 2(b)] spins, where we compare magnetizations and next-nearest neighbor correlators. Since finite-size effects are negligible to good approximation for systems with more than $N = 10$ spins, we can use the exact data for $N = 10$ spins as comparison for the case of $N = 40$ spins studied in the main text. The noise in the correlation signal is due to the finite number of samples that are used to evaluate the observables. One observes slight deviations in the correlation functions, which may be attributed to both an imperfect choice of hyperparameters and the stochastic

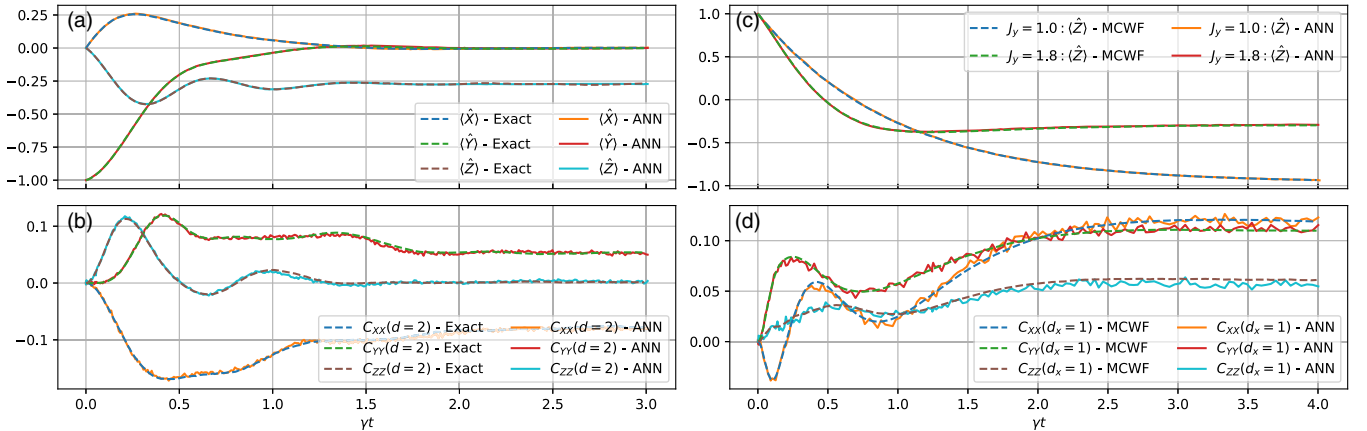


FIG. 2. (a) and (b) Mean magnetizations and next-nearest neighbor connected correlation functions [e.g., $C_{XX}(d=2) = \sum_i \langle \hat{X}_i \hat{X}_{i+2} \rangle^c / N$] as a function of time in the anisotropic 1D Heisenberg model for $N = 40$ spins starting in the product state $\langle \hat{Y} \rangle = -1$. Nearest neighbor couplings are given by $\vec{J}/\gamma = (2, 0, 1)$, $h_z/\gamma = 1$ and the dissipation channel is $\hat{L} = \hat{\sigma}^- = \frac{1}{2}(\hat{X} - i\hat{Y})$. The exact data are obtained for $N = 10$ spins. (c) and (d) Mean z magnetizations and nearest neighbor connected correlation functions (for $J_y/\gamma = 1.8$) in a 4×4 anisotropic 2D Heisenberg lattice with nearest neighbor couplings $\vec{J}/\gamma = [0.9, 1.0(1.8), 1.0]$ and the same decay as in (a) and (b), starting in the product state $\langle \hat{Z} \rangle = 1$.

nature of the proposed method. We found that the sample size needed to reach a given precision does not need to be increased when transitioning to larger systems, as the overall noise decreases thanks to a self-averaging effect in the translationally invariant system.

Figures 2(c) and 2(d) show results for a 4×4 lattice initialized in a product state with $\langle \hat{Z} \rangle = 1$. In panel (c) we compare the magnetization for two different parameter choices to results from MCWF stochastic integration with 500 trajectories, showing perfect agreement. Exact integration of Eq. (3) would be exceedingly costly in this case. We provide a comparison of a 3×3 lattice to exact dynamics in the Supplemental Material [85], which shall serve as a numerically exact benchmark. Nearest neighbor correlations shown in panel (d) for the case of $J_y/\gamma = 1.8$ show small deviations at late times, which we attribute to the finite number of samples used for estimating the updates $\hat{\theta}$ [85].

Having benchmarked our approach on generic spin models, we now apply it to a physical scenario to gauge the effect of decoherence for large problem instances. It was recently shown that confinement dynamics, as found for quarks in quantum chromodynamics, can be realized in the Ising spin model with transverse and longitudinal fields [92]

$$\hat{H} = \sum_{\langle ij \rangle} J_z \hat{Z}_i \hat{Z}_j + \sum_i (h_z \hat{Z}_i + h_x \hat{X}_i). \quad (7)$$

In this system pairs of domain walls form after a quench. For $h_z = 0$ these domain walls can propagate freely while for finite h_z the separation between them comes with an energy cost leading to confinement. This phenomenon manifests in a buildup of dominant spin-spin correlations that is limited to short distances and a much weaker light-cone spreading due to the propagation of bound domain-wall pairs. Signatures of this effect have been observed recently for moderate system sizes on IBMQ [93].

Here we study in what way dissipation influences the signature spreading of spatio-temporal correlations. We consider single particle dephasing $\hat{L} = \hat{Z}$ as the dissipation channel, significantly altering the nature of the spreading on timescales $\gamma t \gtrsim 1$.

Figure 3 shows results for a typical scenario with a dephasing rate of $\gamma = 0.25J$ and $N = 32$. The magnetization [panel (a)] initially shows coherent oscillations (inset) which are quickly damped out followed by a slow relaxation of all magnetizations towards zero. The dashed lines show MCWF simulations for $N = 16$.

The top half of panel (b) shows the correlation dynamics in the considered dissipative scenario using the hitherto described numerical approach. For comparison, we show the corresponding unitary dynamics simulated using matrix product states (MPS) [1,2] on the bottom half of panel (b). In the unitary case correlations initially show a light-cone

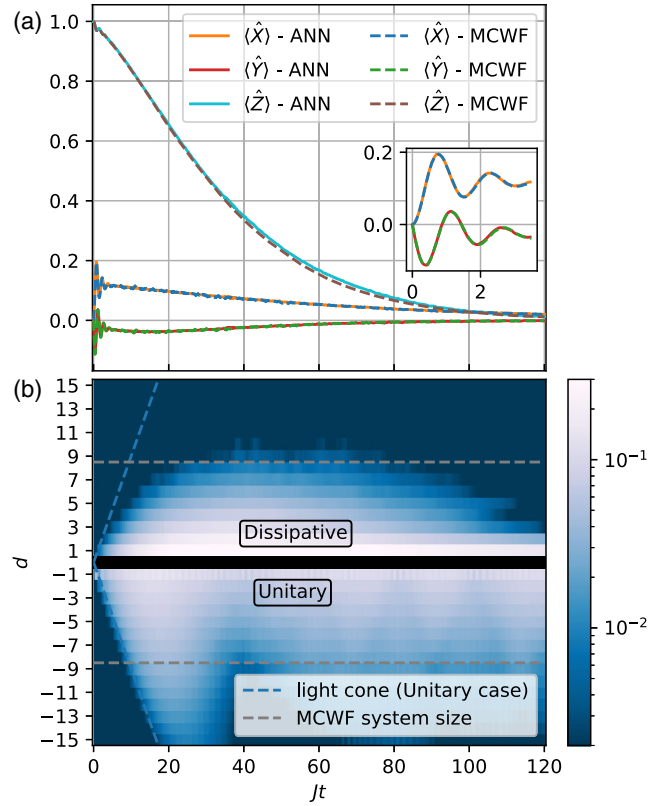


FIG. 3. (a) Mean magnetizations in a spin chain of length $N = 32$ with the quench parameters $h_x/J_z = 0.25$, $h_z/J_z = 0.05$ and the dissipation channel $\hat{L} = \hat{Z}$ with relative strength $\gamma/J_z = 0.25$ compared to MCWF-data for $N = 16$ spins starting in the product state $\langle \hat{Z} \rangle = 1$. (b) Spreading of correlations in the spin chain. Top: Dissipative system with $\gamma = 0.25$, Bottom: MPS simulation of the unitary system where $\gamma = 0.0$. After an initial linear light-cone spreading, the nature of the dissipative propagation grows more diffusive, before all correlations eventually vanish. Notice that the slight deviations in panel (a) coincide with the time at which the dissipative correlations cross the MCWF system size boundary.

spreading. In contrast, the dissipative dynamics deviates from this light cone even for short times as the dissipation results in correlation growth that we find to be consistent with diffusive spreading on intermediate timescales [85]. At long times all correlations decay and the system approaches the featureless steady state $\hat{\rho}(t \rightarrow \infty) \propto 1$.

The ability to simulate these dynamics is a direct consequence of the polynomial scaling of the described ansatz. The system size of the MCWF-approach ($N = 16$) is plotted in panel (b) as a dashed gray line. As is obvious from the chosen color-scale cutoff, the MCWF-approach suffers from finite-size effects at around $Jt = 40$ when correlations beyond $d = 8$ build up. This is also the time at which the z magnetization in panel (a) deviates, suggesting that this deviation is due to finite-size effects present in the MCWF simulation.

Discussion and outlook.—We have introduced a novel method that allows the variational simulation of open quantum dynamics based on the efficient encoding of the quantum state in an artificial neural network and demonstrated its potential. Our method scales approximately cubically with the number of spins N , since the computationally intense part is obtaining F in Eq. (5). For that purpose, the nonvanishing entries of \mathcal{L}^{ab} , the cost of evaluating a plain RNN, and the incorporation of translational symmetry each contribute a factor of N to the computational cost. Importantly, however, the algorithm admits massive parallelization, e.g., on GPU clusters [63,94], which allows for extensive control of the total compute wall time. Here, different levels of parallelization are exploited: The algorithm permits treating the samples independently from each other, allowing us to employ multiple accelerators which communicate via MPI. At the same time, the batched network evaluation of the configurations allows for convenient vectorization of the computations on the single accelerator level.

For future research it will be crucial to better understand the limitations of finite network architectures to represent physically relevant quantum states. An obvious shortcoming of the probabilistic state representation is that the positivity of the density operator is not guaranteed; understanding the consequences will be key for further progress.

The presented combination of the probabilistic formulation of mixed quantum states with a TDVP opens new possibilities for the investigation of driven-dissipative many-body systems in regimes that are challenging for other approaches, for example, to study two-dimensional systems or the propagation of information across large distances [95,96]. Future work employing the TDVP for OQS could address the emergence of glassy dynamics [97–100] or self-organization in OQS [101,102]. Furthermore, the developed technique is not restricted to solving the Lindblad equation; it could be generalized for other use cases of master equations with large discrete configuration space, such as disease dynamics models [103,104] or the chemical master equation [105].

The MCWF data was obtained using QuTip [106]. The TDVP algorithm was implemented using the jVMC codebase [107] and the JAX library [108]. We acknowledge contributions by Thomas Gasenzer and Felix Behrens at early stages of this work. We thank Johannes Schachenmayer for support with MPS simulations. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster) and within the Collaborative Research Center SFB1225 (ISOQUANT). This work was partially financed by the Baden-Württemberg Stiftung gGmbH. The authors acknowledge support by the state of Baden-Württemberg

through bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/575-1 FUGG (JUSTUS 2 cluster). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [109] at Jülich Supercomputing Centre (JSC).

*moritz.reh@kip.uni-heidelberg.de

- [1] U. Schollwöck, *Ann. Phys. (Amsterdam)* **326**, 96 (2011).
- [2] R. Orús, *Ann. Phys. (Amsterdam)* **349**, 117 (2014).
- [3] I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete, [arXiv:2011.12127](https://arxiv.org/abs/2011.12127).
- [4] J. Carrasquilla, *Adv. Phys. X* **5**, 1797528 (2020).
- [5] R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [6] S. Lloyd, *Science* **273**, 1073 (1996).
- [7] N. R. Cooper, J. Dalibard, and I. B. Spielman, *Rev. Mod. Phys.* **91**, 015005 (2019).
- [8] D. Mitra, P. T. Brown, E. Guardado-Sanchez, S. S. Kondov, T. Devakul, D. A. Huse, P. Schauß, and W. S. Bakr, *Nat. Phys.* **14**, 173 (2018).
- [9] P. Bordia, H. Lüschen, S. Scherg, S. Gopalakrishnan, M. Knap, U. Schneider, and I. Bloch, *Phys. Rev. X* **7**, 041047 (2017).
- [10] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, *Science* **349**, 842 (2015).
- [11] I. Bloch, J. Dalibard, and W. Zwerger, *Rev. Mod. Phys.* **80**, 885 (2008).
- [12] J. y. Choi, S. Hild, J. Zeiher, P. Schauss, A. Rubio-Abadal, T. Yefsah, V. Khemani, D. A. Huse, I. Bloch, and C. Gross, *Science* **352**, 1547 (2016).
- [13] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, *Nature (London)* **462**, 74 (2009).
- [14] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, *Nature (London)* **534**, 516 (2016).
- [15] J. I. Cirac and P. Zoller, *Phys. Rev. Lett.* **74**, 4091 (1995).
- [16] M. Gärtner, J. G. Bohnet, A. Safavi-Naini, M. L. Wall, J. J. Bollinger, and A. M. Rey, *Nat. Phys.* **13**, 781 (2017).
- [17] A. Browaeys and T. Lahaye, *Nat. Phys.* **16**, 132 (2020).
- [18] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletic, and M. D. Lukin, *Nature (London)* **595**, 227 (2021).
- [19] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, *Nature (London)* **595**, 233 (2021).
- [20] D. I. Tsomokos, S. Ashhab, and F. Nori, *Phys. Rev. A* **82**, 052311 (2010).
- [21] M. J. Hartmann, F. G. S. L. Brandão, and M. B. Plenio, *Nat. Phys.* **2**, 849 (2006).
- [22] M. Ortner, A. Micheli, G. Pupillo, and P. Zoller, *New J. Phys.* **11**, 055045 (2009).

- [23] G. Roumpos, C. P. Master, and Y. Yamamoto, *Phys. Rev. B* **75**, 094415 (2007).
- [24] T. Byrnes, N. Y. Kim, K. Kusudo, and Y. Yamamoto, *Phys. Rev. B* **78**, 075320 (2008).
- [25] A. D. Greentree, C. Tahan, J. H. Cole, and L. C. L. Hollenberg, *Nat. Phys.* **2**, 856 (2006).
- [26] X. Peng, J. Du, and D. Suter, *Phys. Rev. A* **71**, 012307 (2005).
- [27] S. Mostame and R. Schützhold, *Phys. Rev. Lett.* **101**, 220501 (2008).
- [28] D. G. Angelakis, M. F. Santos, and S. Bose, *Phys. Rev. A* **76**, 031805(R) (2007).
- [29] J. Zhang, F. M. Cucchietti, C. M. Chandrashekar, M. Laforest, C. A. Ryan, M. Ditty, A. Hubbard, J. K. Gamble, and R. Laflamme, *Phys. Rev. A* **79**, 012305 (2009).
- [30] J. Cai, A. Retzker, F. Jelezko, and M. B. Plenio, *Nat. Phys.* **9**, 168 (2013).
- [31] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, C. von Keyserlingk, N. Y. Yao, E. Demler, and M. D. Lukin, *Nature (London)* **543**, 221 (2017).
- [32] J. Preskill, *Quantum* **2**, 79 (2018).
- [33] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, *Nature (London)* **551**, 579 (2017).
- [34] M. Prüfer, P. Kunkel, H. Strobel, S. Lannig, D. Linnemann, C.-M. Schmied, J. Berges, T. Gasenzer, and M. K. Oberthaler, *Nature (London)* **563**, 217 (2018).
- [35] S. Erne, R. Bücker, T. Gasenzer, J. Berges, and J. Schmiedmayer, *Nature (London)* **563**, 225 (2018).
- [36] H. Weimer, A. Kshetrimayum, and R. Orús, *Rev. Mod. Phys.* **93**, 015008 (2021).
- [37] K. Mølmer, Y. Castin, and J. Dalibard, *J. Opt. Soc. Am. B* **10**, 524 (1993).
- [38] M. B. Plenio and P. L. Knight, *Rev. Mod. Phys.* **70**, 101 (1998).
- [39] J. Dalibard, Y. Castin, and K. Mølmer, *Phys. Rev. Lett.* **68**, 580 (1992).
- [40] R. Dum, P. Zoller, and H. Ritsch, *Phys. Rev. A* **45**, 4879 (1992).
- [41] M. Kornyik and A. Vukics, *Comput. Phys. Commun.* **238**, 88 (2019).
- [42] I. Carusotto and C. Ciuti, *Rev. Mod. Phys.* **85**, 299 (2013).
- [43] F. Vicentini, F. Minganti, A. Biella, G. Orso, and C. Ciuti, *Phys. Rev. A* **99**, 032115 (2019).
- [44] F. Vicentini, F. Minganti, R. Rota, G. Orso, and C. Ciuti, *Phys. Rev. A* **97**, 013853 (2018).
- [45] G. Dagvadorj, J. M. Fellows, S. Matyjaśkiewicz, F. M. Marchetti, I. Carusotto, and M. H. Szymańska, *Phys. Rev. X* **5**, 041028 (2015).
- [46] R. Liboff, *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions* (Springer-Verlag, Berlin, 2003).
- [47] P. Navez and R. Schützhold, *Phys. Rev. A* **82**, 063603 (2010).
- [48] C. J. Wood, J. D. Biamonte, and D. G. Cory, *Quantum Inf. Comput.* **15**, 759 (2015).
- [49] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov, *Phys. Rev. Lett.* **122**, 160401 (2019).
- [50] C. M. Keever and M. H. Szymańska, *Phys. Rev. X* **11**, 021035 (2021).
- [51] D. Kilda, A. Biella, M. Schiro, R. Fazio, and J. Keeling, *SciPost Phys. Core* **4**, 5 (2021).
- [52] A. H. Werner, D. Jaschke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero, *Phys. Rev. Lett.* **116**, 237201 (2016).
- [53] D. Jaschke, S. Montangero, and L. D. Carr, *Quantum Sci. Technol.* **4**, 013001 (2018).
- [54] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, *New J. Phys.* **12**, 025012 (2010).
- [55] J. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, *Ann. Phys. (Amsterdam)* **378**, 100 (2017).
- [56] E. Mascarenhas, H. Flayac, and V. Savona, *Phys. Rev. A* **92**, 022116 (2015).
- [57] J. Haferkamp, D. Hangleiter, J. Eisert, and M. Gluza, *Phys. Rev. Research* **2**, 013010 (2020).
- [58] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Phys. Rev. Lett.* **98**, 140506 (2007).
- [59] A. Kshetrimayum, H. Weimer, and R. Orús, *Nat. Commun.* **8**, 1291 (2017).
- [60] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [61] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nat. Mach. Intell.* **1**, 155 (2019).
- [62] J. Carrasquilla, D. Luo, F. Pérez, A. Milsted, B. K. Clark, M. Volkovs, and L. Aolita, *Phys. Rev. A* **104**, 032610 (2021).
- [63] M. Schmitt and M. Heyl, *Phys. Rev. Lett.* **125**, 100503 (2020).
- [64] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, *arXiv:2009.05580*.
- [65] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, *Phys. Rev. Research* **2**, 023358 (2020).
- [66] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. X* **7**, 021021 (2017).
- [67] S. Czischek, M. Gärtner, and T. Gasenzer, *Phys. Rev. B* **98**, 024311 (2018).
- [68] M. Neugebauer, L. Fischer, A. Jäger, S. Czischek, S. Jochim, M. Weidemüller, and M. Gärtner, *Phys. Rev. A* **102**, 042604 (2020).
- [69] X. Gao and L.-M. Duan, *Nat. Commun.* **8**, 662 (2017).
- [70] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. B* **96**, 195145 (2017).
- [71] Z.-A. Jia, Y.-H. Zhang, Y.-C. Wu, L. Kong, G.-C. Guo, and G.-P. Guo, *Phys. Rev. A* **99**, 012307 (2019).
- [72] Y.-H. Zhang, Z.-A. Jia, Y.-C. Wu, and G.-C. Guo, *arXiv:1809.08631*.
- [73] Y. Huang and J. E. Moore, *Phys. Rev. Lett.* **127**, 170601 (2021).
- [74] S. Lu, X. Gao, and L.-M. Duan, *Phys. Rev. B* **99**, 155136 (2019).
- [75] W.-C. Gan and F.-W. Shu, *Int. J. Mod. Phys. D* **26**, 1743020 (2017).
- [76] M. J. Hartmann and G. Carleo, *Phys. Rev. Lett.* **122**, 250502 (2019).
- [77] N. Yoshioka and R. Hamazaki, *Phys. Rev. B* **99**, 214306 (2019).
- [78] A. Nagy and V. Savona, *Phys. Rev. Lett.* **122**, 250501 (2019).
- [79] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, *Phys. Rev. Lett.* **122**, 250503 (2019).
- [80] G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **120**, 240503 (2018).

- [81] A. Peres, *Quantum Theory: Concepts and Methods* (Springer, Dordrecht, 2002).
- [82] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations* (MIT Press, Cambridge, MA, USA, 1986), p. 318–362.
- [83] S. Hochreiter and J. Schmidhuber, *Neural Comput.* **9**, 1735 (1997).
- [84] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, [arXiv:1706.03762](https://arxiv.org/abs/1706.03762).
- [85] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.127.230501> for details on the POVM formalism and the derivation of the TDVP equation.
- [86] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [87] K. Hornik, *Neural Netw.* **4**, 251 (1991).
- [88] A. Pinkus, *Acta Numer.* **8**, 143 (1999).
- [89] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, *Phys. Rev. Lett.* **124**, 020503 (2020).
- [90] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, [arXiv:2101.07243](https://arxiv.org/abs/2101.07243).
- [91] S.-H. Lin and F. Pollmann, [arXiv:2104.10696](https://arxiv.org/abs/2104.10696).
- [92] M. Kormos, M. Collura, G. Takács, and P. Calabrese, *Nat. Phys.* **13**, 246 (2017).
- [93] J. Vovrosh and J. Knolle, *Sci. Rep.* **11**, 11577 (2021).
- [94] L. Yang, W. Hu, and L. Li, [arXiv:2011.12453](https://arxiv.org/abs/2011.12453).
- [95] M. J. Kastoryano and J. Eisert, *J. Math. Phys. (N.Y.)* **54**, 102201 (2013).
- [96] R. Sweke, J. Eisert, and M. Kastner, *J. Phys. A* **52**, 424003 (2019).
- [97] D. Poletti, P. Barmettler, A. Georges, and C. Kollath, *Phys. Rev. Lett.* **111**, 195301 (2013).
- [98] G. Kucsko, S. Choi, J. Choi, P. C. Maurer, H. Zhou, R. Landig, H. Sumiya, S. Onoda, J. Isoya, F. Jelezko, E. Demler, N. Y. Yao, and M. D. Lukin, *Phys. Rev. Lett.* **121**, 023601 (2018).
- [99] J. Choi, S. Choi, G. Kucsko, P. C. Maurer, B. J. Shields, H. Sumiya, S. Onoda, J. Isoya, E. Demler, F. Jelezko, N. Y. Yao, and M. D. Lukin, *Phys. Rev. Lett.* **118**, 093601 (2017).
- [100] B. Everest, I. Lesanovsky, J. P. Garrahan, and E. Levi, *Phys. Rev. B* **95**, 024310 (2017).
- [101] B. Zhu, J. Schachenmayer, M. Xu, F. Herrera, J. G. Restrepo, M. J. Holland, and A. M. Rey, *New J. Phys.* **17**, 083063 (2015).
- [102] S. Helmrich, A. Arias, G. Lochead, T. M. Wintermantel, M. Buchhold, S. Diehl, and S. Whitlock, *Nature (London)* **577**, 481 (2020).
- [103] G. Jenkinson and J. Goutsias, *PLoS One* **7**, e36160 (2012).
- [104] M. J. Keeling and J. V. Ross, *J. R. Soc. Interface* **5**, 171 (2008).
- [105] V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger, *BMC Syst. Biol.* **4**, 42 (2010).
- [106] J. Johansson, P. Nation, and F. Nori, *Comput. Phys. Commun.* **183**, 1760 (2012).
- [107] M. Schmitt and M. Reh, [arXiv:2108.03409](https://arxiv.org/abs/2108.03409).
- [108] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python + NumPy programs (2018), [http://github.com/google/jax](https://github.com/google/jax).
- [109] Jülich Supercomputing Centre, *J. Large-Scale Res. Facil.* **5** (2019).

Supplementary Materials

Derivation of the TDVP equation

The basic idea of a TDVP is to minimize the distance

$$\mathcal{D} \left(P_{\theta(t)} + \dot{P}_{\theta(t)}\tau, P_{\theta(t)} + \sum_k \frac{\partial P_{\theta(t)}}{\partial \theta_k} \dot{\theta}_k \tau \right), \quad (8)$$

between the evolved state at time $t + \tau$ and the network with a set of yet unknown update parameters $\dot{\theta}$ at each time t . Here, we exemplarily derive Eq. (5) from the Hellinger distance $\mathcal{D}_H(P, Q)$, i.e. by maximizing the classical fidelity $F(P, Q) = 1 - \mathcal{D}_H(P, Q)$. As noted in the main text, an equivalent derivation is possible using the Kullback-Leibler divergence \mathcal{D}_{KL} . For better readability, we drop the time index and continue with the optimality condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial \dot{\theta}_k} F \left(P + \dot{P}\tau, P + \sum_{k'} \frac{\partial P}{\partial \theta_{k'}} \dot{\theta}_{k'} \tau \right) \\ &= \frac{\partial}{\partial \dot{\theta}_k} \sum_{\mathbf{a}} P^{\mathbf{a}} \sqrt{1 + a\tau + b\tau^2}, \end{aligned} \quad (9)$$

where a and b are given by

$$\begin{aligned} a &= \frac{\partial \log P^{\mathbf{a}}}{\partial t} + \sum_{k'} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'}, \\ b &= \frac{\partial \log P^{\mathbf{a}}}{\partial t} \sum_{k'} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'}. \end{aligned} \quad (10)$$

Next we perform a second order expansion of the square root in the time step τ :

$$\sqrt{1 + a\tau + b\tau^2} = 1 + \frac{a\tau}{2} + \frac{\tau^2}{8}(4b - a^2) + \mathcal{O}(\tau^3). \quad (11)$$

Using that the normalization of P is conserved under the time evolution one finds that the term linear in τ vanishes:

$$\begin{aligned} P^{\mathbf{a}} a &= \sum_{\mathbf{a}} \left(\dot{P}^{\mathbf{a}} + \sum_{k'} \frac{\partial P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \right) \\ &= \sum_{k'} \dot{\theta}_{k'} \frac{\partial}{\partial \theta_{k'}} \sum_{\mathbf{a}} P^{\mathbf{a}} \\ &= \sum_{k'} \dot{\theta}_{k'} \frac{\partial}{\partial \theta_{k'}} 1 \\ &= 0. \end{aligned} \quad (12)$$

Thus, the optimality condition becomes

$$\begin{aligned} 0 &= \frac{\partial}{\partial \dot{\theta}_k} \sum_{\mathbf{a}} \frac{P^{\mathbf{a}}}{P^{\mathbf{a}^2}} \left(4\dot{P}^{\mathbf{a}} \sum_{k'} \frac{\partial P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} - \left(\dot{P}^{\mathbf{a}} + \sum_{k'} \frac{\partial P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \right)^2 \right) \\ &= -\frac{\partial}{\partial \dot{\theta}_k} \sum_{\mathbf{a}} \frac{P^{\mathbf{a}}}{P^{\mathbf{a}^2}} \left(\dot{P}^{\mathbf{a}} - \sum_{k'} \frac{\partial P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \right)^2 \\ &= -\frac{\partial}{\partial \dot{\theta}_k} \sum_{\mathbf{a}} P^{\mathbf{a}} \left(\frac{\partial \log P^{\mathbf{a}}}{\partial t} - \sum_{k'} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \right)^2 \\ &= 2 \sum_{\mathbf{a}} P^{\mathbf{a}} \frac{\log P^{\mathbf{a}}}{\partial \theta_k} \left(\frac{\partial \log P^{\mathbf{a}}}{\partial t} - \sum_{k'} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \right). \end{aligned} \quad (13)$$

Dropping the factor of 2 we obtain an equation for the optimal parameter update $\dot{\theta}$:

$$\begin{aligned} 0 &= \underbrace{\sum_{\mathbf{a}} P^{\mathbf{a}} \frac{\partial \log P^{\mathbf{a}}}{\partial t} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k}}_{=F_k} \\ &\quad - \underbrace{\sum_{k'} \sum_{\mathbf{a}} P^{\mathbf{a}} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'}}_{=S_{kk'}}. \end{aligned} \quad (14)$$

Importantly we can now tackle the sum over the exponentially many indices \mathbf{a} by sampling according to the encoded probabilities $P^{\mathbf{a}}$ since both F and S are proportional to $P^{\mathbf{a}}$. This is a unique property of \mathcal{D}_H and \mathcal{D}_{KL} while other distance measures, as for example the L^2 norm, do not lead to expressions of a form that can be efficiently evaluated from Monte Carlo samples. Further, inserting the probabilistic form of the Lindblad master equation leads to

$$\begin{aligned} F_k &= \sum_{\mathbf{a}} P^{\mathbf{a}} \frac{\partial \log P^{\mathbf{a}}}{\partial t} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \\ &= \left\langle \mathcal{L}^{\mathbf{a}b} \frac{P^{\mathbf{b}}}{P^{\mathbf{a}}} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \right\rangle_{\mathbf{a} \sim P} \end{aligned} \quad (15)$$

and

$$\begin{aligned} S_{kk'} &= \sum_{\mathbf{a}} P^{\mathbf{a}} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \\ &= \left\langle \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_{k'}} \right\rangle_{\mathbf{a} \sim P}. \end{aligned} \quad (16)$$

The same derivation can be carried out without assuming normalization. In this case the form of S and F is altered

to

$$\begin{aligned}
P^{\mathbf{a}} &\rightarrow \frac{P^{\mathbf{a}}}{\sum_{\mathbf{b}} P^{\mathbf{b}}} \\
\log P^{\mathbf{a}} &\rightarrow \log P^{\mathbf{a}} - \log \sum_{\mathbf{b}} P^{\mathbf{b}} \\
\frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} &\rightarrow \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} - \left\langle \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \right\rangle_{\mathbf{a} \sim P} \\
\frac{\partial \log P^{\mathbf{a}}}{\partial t} &\rightarrow \frac{\partial \log P^{\mathbf{a}}}{\partial t} - \left\langle \frac{\partial \log P^{\mathbf{a}}}{\partial t} \right\rangle_{\mathbf{a} \sim P},
\end{aligned} \tag{17}$$

where the last two lines are obtained using

$$\begin{aligned}
&\frac{\partial}{\partial \theta_k} \left(\log P^{\mathbf{a}} - \log \sum_{\mathbf{b}} P^{\mathbf{b}} \right) \\
&= \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} - \frac{\sum_{\mathbf{b}} \frac{\partial P^{\mathbf{b}}}{\partial \theta_k}}{\sum_{\mathbf{c}} P^{\mathbf{c}}} \\
&= \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} - \sum_{\mathbf{b}} \frac{P^{\mathbf{b}}}{\sum_{\mathbf{c}} P^{\mathbf{c}}} \frac{\partial \log P^{\mathbf{b}}}{\partial \theta_k} \\
&= \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} - \left\langle \frac{\partial \log P^{\mathbf{a}}}{\partial \theta_k} \right\rangle_{\mathbf{a} \sim P}.
\end{aligned} \tag{18}$$

Here, the log derivative trick was used in the third line and we renamed the dummy indices \mathbf{b} and \mathbf{c} in the last step. One may proceed similarly for the time derivative. Overall, this leaves us with the connected correlator structure described in the main text

$$\begin{aligned}
S_{kk'} &= \langle O_k^{\mathbf{a}} O_{k'}^{\mathbf{a}} \rangle_{\mathbf{a} \sim P} - \langle O_k^{\mathbf{a}} \rangle_{\mathbf{a} \sim P} \langle O_{k'}^{\mathbf{a}} \rangle_{\mathbf{a} \sim P} \\
F_k &= \left\langle \mathcal{L}^{\mathbf{ab}} \frac{P^{\mathbf{b}}}{P^{\mathbf{a}}} O_k^{\mathbf{a}} \right\rangle_{\mathbf{a} \sim P} - \langle O_k^{\mathbf{a}} \rangle_{\mathbf{a} \sim P} \left\langle \mathcal{L}^{\mathbf{ab}} \frac{P^{\mathbf{b}}}{P^{\mathbf{a}}} \right\rangle_{\mathbf{a} \sim P}.
\end{aligned} \tag{19}$$

We finally arrive at

$$\dot{\theta}_k = \tilde{S}_{kk'}^{-1} F_{k'} \tag{20}$$

where the tilde is due to the fact that we cannot invert S directly but rather need to regularize it because it is usually ill-conditioned. One can easily show that the updates that were found are indeed maxima of the fidelity:

$$\begin{aligned}
&\frac{\partial^2}{\partial \theta_k^2} F(P^{\mathbf{a}} + \dot{P}^{\mathbf{a}} \tau, P^{\mathbf{a}} + \sum_{k'} \frac{\partial P^{\mathbf{a}}}{\partial \theta_{k'}} \dot{\theta}_{k'} \tau) \\
&= \frac{\partial}{\partial \theta_k} (F_k - S_{kk'} \dot{\theta}_{k'}) \\
&= -S_{kk'} \delta_{k'k} \\
&= -S_{kk} \\
&= -\left\langle (O_k^{\mathbf{a}} - \langle O_k^{\mathbf{a}} \rangle)^2 \right\rangle_{\mathbf{a} \sim P} \\
&\leq 0.
\end{aligned} \tag{21}$$

Observables and Operators in the POVM-formalism

As described in the main text, the POVM-distribution P is obtained as expectation values of the respective POVM-operators \hat{M} ,

$$P^{\mathbf{a}} = \text{tr} \left(\hat{\rho} \hat{M}^{\mathbf{a}} \right), \tag{22}$$

where $\hat{M}^{\mathbf{a}} = \hat{M}^{a_1} \otimes \dots \otimes \hat{M}^{a_N}$ are product operators. For IC-POVMs with the minimal number of $(d^2)^N$ elements, where d is the local Hilbert space dimension ($d = 2$ for spins), this relation can be inverted:

$$\hat{\rho} = P^{\mathbf{a}} T^{-1 \mathbf{aa}'} \hat{M}^{\mathbf{a}'}, \tag{23}$$

with the overlap matrix $T^{\mathbf{aa}'} = \text{tr} \left(\hat{M}^{\mathbf{a}} \hat{M}^{\mathbf{a}'} \right)$. We note that not every normalized probability distribution inserted into Eq. (23) results in a physical density matrix, as the positivity of $\hat{\rho}$ is not ensured. The equation describing the dynamics of P can be obtained from Eq. (23) and the Lindblad master equation according to

$$\dot{P}^{\mathbf{a}} = \text{tr} \left(\dot{\hat{\rho}} \hat{M}^{\mathbf{a}} \right) = \mathcal{L}^{\mathbf{ab}} P^{\mathbf{b}}. \tag{24}$$

The the part of the linear map \mathcal{L} resulting from the von-Neumann term, i.e. the part accounting for the unitary evolution, is

$$\begin{aligned}
\dot{P}^{\mathbf{a}} &= \text{tr} \left(-i \left[\hat{H}, \hat{\rho} \right] \hat{M}^{\mathbf{a}} \right) \\
&= \text{tr} \left(-i \left[\hat{H}, T^{-1 \mathbf{bb}'} \hat{M}^{\mathbf{b}'} \right] \hat{M}^{\mathbf{a}} \right) P^{\mathbf{b}} \\
&= \text{tr} \left(-i \hat{H} \left[T^{-1 \mathbf{bb}'} \hat{M}^{\mathbf{b}'}, \hat{M}^{\mathbf{a}} \right] \right) P^{\mathbf{b}} \\
&= U^{\mathbf{ab}} P^{\mathbf{b}},
\end{aligned} \tag{25}$$

where the cyclicity of the trace was used in the last line. A similar expression can be found for the dissipative part

$$\begin{aligned}
D^{\mathbf{ab}} &= \gamma \text{tr} \left(\sum_i \hat{L}^i T^{-1 \mathbf{bb}'} \hat{M}^{\mathbf{b}'} \hat{L}^{i\dagger} \hat{M}^{\mathbf{a}} \right. \\
&\quad \left. - \frac{1}{2} \hat{L}^i \hat{L}^{i\dagger} \left\{ T^{-1 \mathbf{bb}'} \hat{M}^{\mathbf{b}'}, \hat{M}^{\mathbf{a}} \right\} \right),
\end{aligned} \tag{26}$$

from which we set \mathcal{L} together according to

$$\mathcal{L}^{\mathbf{ab}} = U^{\mathbf{ab}} + D^{\mathbf{ab}}. \tag{27}$$

The expectation value of any observable in physical index space may be correspondingly expressed in the POVM-formalism replacing $\langle \hat{O} \rangle = \text{tr} \left(\hat{\rho} \hat{O} \right)$ by $\langle \hat{O} \rangle = P^{\mathbf{a}} \Omega^{\mathbf{a}}$. The numerical values of the coefficients $\Omega^{\mathbf{a}}$ are obtained in similar fashion as the Lindbladian operator \mathcal{L} , namely by substituting $\hat{\rho}$ according to Eq. (23)

$$\langle \hat{O} \rangle = \text{tr} \left(\hat{\rho} \hat{O} \right) = P^{\mathbf{a}} T^{-1 \mathbf{aa}'} \text{tr} \left(\hat{M}^{\mathbf{a}'} \hat{O} \right) = P^{\mathbf{a}} \Omega^{\mathbf{a}}. \tag{28}$$

Details of the RNN-architecture

As described in the main text, the RNN encodes the probability distribution $P^{\mathbf{a}}$ as a product of conditionals, $P^{\mathbf{a}} = \prod_i P(a_i|a_{<i})$. The formula implies that the network’s knowledge of previous POVM outcomes $a_{<i}$ may alter the estimation of POVM outcome probabilities at site i . In the network architecture, this is ensured by passing a hidden state to the next lattice site where it enters the computation of the probability output. This hidden state may be regarded as a latent embedding of physical contextual information, and is required to accurately encode correlations in the physical system. Our results are obtained using standard RNN cells which are known to have exponentially decaying correlation length [110]. In scenarios, where this is expected to be insufficient, more advanced cells, such as the Long Short Term Memory (LSTM) [83], whose correlation length decays algebraically [110], or the transformer [84] may be used instead.

Since the RNN architecture was originally developed to tackle tasks associated with serial data, some changes are required in order to make it suitable for quantum applications. For one, to allow the treatment of 2D systems the RNN evaluation and sampling schemes need to be generalized. Here, we adapt a scheme, introduced in [65], that treats correlations along both spatial direction on equal footing. Additionally, we enforce all symmetries present in the Lindbladian \mathcal{L} by averaging all symmetry-invariant outcome configurations [65]. These include translational symmetries as well as point symmetries. We emphasize, that explicitly restoring these symmetries in our ansatz improved the accuracies of observables substantially.

Furthermore, we here lay out how the network is initialized. Product states, which form typical initial states in non-equilibrium time evolution, may be encoded to numerical precision in the network, by setting the biases of the output layer to the logarithm of the to be encoded 1-particle probability distribution while simultaneously setting all weights connecting to the output layer to zero. We may therefore attribute all accumulated error to imprecise updates during time-evolution and note that no preceding computations are required.

For our simulations we use RNNs implemented in the open source machine learning library JAX [108]. An RNN is a generative model that works on sequential data, in which the bits of the sequential data are processed in an iterative fashion. The RNN fulfills two tasks: It assigns probabilities to a given POVM outcome configuration and, as a generative model, is capable of exact sampling, meaning that it can be programmed to output sample POVM configurations in agreement with the assigned probabilities. This is a major advantage of autoregressive networks compared to other network architectures, in which the sampling step is carried out using

Markov Chain Monte Carlo schemes, which potentially may be plagued by long autocorrelation times.

An RNN-cell is the basic building block of an RNN; RNN-cells may be stacked to form the complete RNN, increasing the representational power of the network. Let us first limit our considerations to RNNs with one layer, i.e. single RNN-cells. The input to every RNN-cell consists of two parts: For one, the physical POVM outcomes $\mathbf{a} = a_1..a_N$ are fed into the model piece by piece. Here, each outcome is transformed to a one-hot encoded vector of length 4. Simultaneously, a hidden state of length l which is initialized to zero, i.e. $\mathbf{h}_0 = 0$ is fed into the model. The first step consists of finding the first probability appearing in $P^{\mathbf{a}} = \prod_i P(a_i|a_{<i})$, i.e. $P(a_1)$. First, a new hidden state is computed

$$\mathbf{h}_1 = \phi(W_h \cdot \mathbf{h}_0 + W_a \cdot \mathbf{a}_0 + \mathbf{b}_h). \quad (29)$$

\mathbf{a}_0 is an input of length 4 carrying zeros, similar to the empty input \mathbf{h}_0 of length l . The parameters $W^{l(a)}$ consequently are matrices with shape $l \times l$ ($l \times 4$), while the bias vector \mathbf{b}_h has length l . We choose the element-wise activation function ϕ to be the Exponential Linear Unit (ELU)

$$\phi(x) = \begin{cases} x & x > 0, \\ \alpha(e^x - 1) & x \leq 0. \end{cases} \quad (30)$$

Two more sets of parameters W_s (\mathbf{b}_s) with shape $4 \times l$ (4) enter the computation of the output of the RNN-cell,

$$P(a_1) = \sigma(W_s \cdot \mathbf{h}_1 + \mathbf{b}_s). \quad (31)$$

Here σ denotes the softmax-activation,

$$\sigma(\mathbf{x})_i = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad (32)$$

and the summation includes the four possible POVM-outcomes, allowing to interpret P^{a_1} as a proper discrete probability distribution. Depending on the task at hand, one may either store the probability of a POVM outcome of interest or sample the first POVM outcome from P^{a_1} .

Obtaining an expression for $P(a_2|a_1)$ is identical to the hitherto described procedure, by substituting \mathbf{h}_0 for \mathbf{h}_1 and \mathbf{a}_0 for \mathbf{a}_1 and, more generally, \mathbf{h}_i for \mathbf{h}_{i+1} and \mathbf{a}_i for \mathbf{a}_{i+1} in the following steps. Here, the ‘recurrent’ nature becomes apparent, since the same parameters, i.e. the same network, is used in every computation step.

If one desires to use deeper networks with K layers, Eq. (29) changes to

$$\mathbf{h}_i^k = \phi(W_h^k \cdot \mathbf{h}_{i-1}^k + W_a^k \cdot \mathbf{h}_i^{k-1} + \mathbf{b}_h^k). \quad (33)$$

\mathbf{h}_i^k is then called the hidden state at layer k at lattice site i . The computation of $P(a_i|a_{<i})$ is still analogous to Eq. (31), i.e.

$$P(a_i|a_{<i}) = \sigma(W_s \cdot \mathbf{h}_i^K + \mathbf{b}_s). \quad (34)$$

As the product of these probabilities becomes exponentially small in the system size N , one stores the logarithm of the conditional probability instead of the probability itself.

In two-dimensional systems, the situation is slightly more involved. One might be tempted to map the 2D system in a snake-like fashion to a one-dimensional system. However, using this method one observes that correlators of vertical neighbours are not encoded accurately as information may potentially get lost upon long traversing times in horizontal direction [65]. Instead, we opt to pass hidden states in a two-dimensional fashion, incorporating the dimensionality of the system as shown in Ref. [65]. Herein, we once again change Eq. (33) to read

$$\mathbf{h}_{ij}^k = \phi \left(W_h^k \cdot \mathbf{h}_{i-1j}^k + W_h^k \cdot \mathbf{h}_{ij-1}^k + W_a^k \cdot \mathbf{h}_{ij}^{k-1} + \mathbf{b}_h^k \right). \quad (35)$$

This method can in principle be extended to three dimensional systems.

Comparison to exact numerical simulations for small system sizes

To obtain uncontroversial benchmarks, we test our method in system size regimes where exact dynamics is feasible. As benchmark systems we choose the 1D and 2D systems described in the main text in Fig. 2 and reduce the system size to $N = 10$ spins in the 1D case and

a 3×3 lattice in the 2D case.

Dissipative confinement correlations

One question that arises in relation with Fig. 3 in the main text is how the spreading of correlations is to be described in the dissipative setting. As decoherence generically leads to classical transport dynamics one may expect diffusive growth of correlation that is proportional to \sqrt{t} in contrast to the unitary linear light-cone proportional to t . This intuition, however, can only hold in an intermediate time regime since at long times the system will relax to its steady state prohibiting an indefinite growth of correlations. In the present case of single particle dephasing noise the steady state is given by $\hat{\rho}(t \rightarrow \infty) \propto \mathbb{1}$, as is easily verified by the observation that the unity operator commutes with \hat{H} and similar for the dissipative part of the evolution Eq. (3) of the main text. This means that all correlations will eventually decay to zero again in the long-time limit. One observes that correlations indeed start to disappear again at around $Jt \sim 70$ in the considered setting. Nevertheless, the spreading of correlations at intermediate times is consistent with a square-root, as shown in Fig. 5 where the dashed line is a fit to the first passage data points of a given threshold.

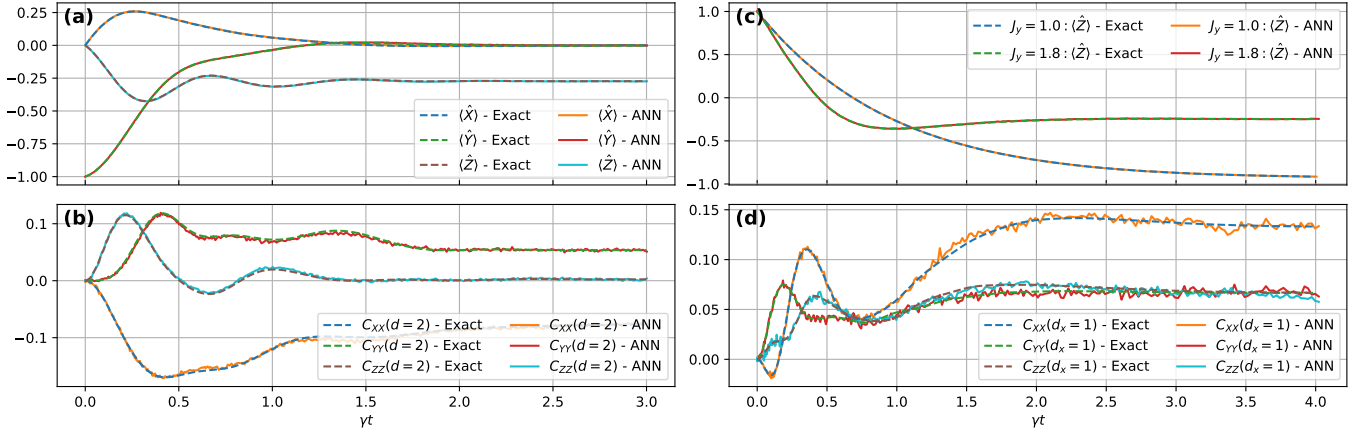


FIG. 4. (a) and (b): Mean magnetizations and next-nearest neighbour connected correlation functions (e.g. $C_{XX}(d=2) = \sum_i \langle \hat{X}_i \hat{X}_{i+2} \rangle^c / N$) as a function of time in the anisotropic 1D Heisenberg model for $N = 10$ spins starting in the product state $\langle \hat{Y} \rangle = -1$. Nearest neighbor couplings are given by $\vec{J}/\gamma = (2, 0, 1)$, $h_z/\gamma = 1$ and the dissipation channel is $\hat{L} = \hat{\sigma}^- = \frac{1}{2}(\hat{X} - i\hat{Y})$. The exact data is obtained for $N = 10$ spins. (c) and (d): Mean z -magnetizations and nearest neighbour connected correlation functions (for $J_y/\gamma = 1.8$) in a 3×3 anisotropic 2D Heisenberg lattice with nearest neighbor couplings $\vec{J}/\gamma = (0.9, 1.0(1.8), 1.0)$ and the same decay as in (a) and (b), starting in the product state $\langle \hat{Z} \rangle = 1$.

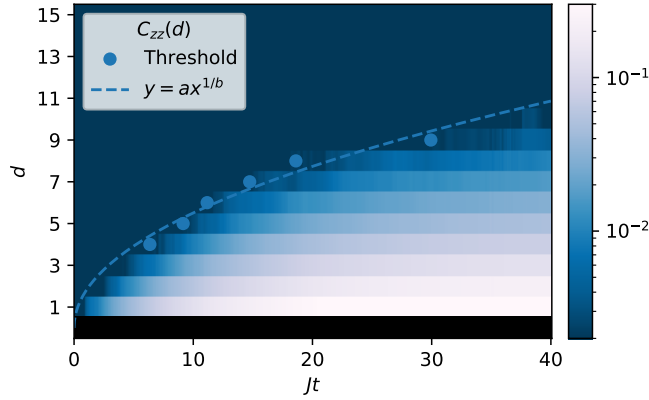


FIG. 5. Spreading of correlations in the dissipative confinement model discussed in the main text (Fig. 3). The data points are obtained as the first passages where $C_{ZZ}(d) \geq 0.002$ and the fitted curve $y = ax^{1/b}$ yields $b = 2.04$.

Figure	Number of layers	Layer size	Number of parameters	Number of Samples	Integration tol. ϵ
Fig. 2 (1D)	3	20	2224	80.000	1e-05
Fig. 2 (2D, $J_y/\gamma = 1.0$)	5	12	2224	8.000	1e-02
Fig. 2 (2D, $J_y/\gamma = 1.8$)	3	20	3504	80.000	5e-03
Fig. 3	5	12	1456	160.000	1e-03

TABLE I. Hyperparameters that were used for the different figures in the main text. The integration tolerance is with respect to the S -matrix scheme proposed in [63].

VARIATIONAL MONTE CARLO APPROACH TO PARTIAL DIFFERENTIAL EQUATIONS WITH NEURAL NETWORKS

6.1 CONTRIBUTIONS

This chapter presents the publication in [B]. I proposed the project, developed the code, and ran all simulations, with Martin Gärtner acting as a supervisor. All authors contributed to the writing of the text. An implementation based on the jVMC library [G] can be found on [GitHub](#).

6.2 MOTIVATION

The evolution equation of the variational parameters that encode the probability distribution in [A] is not tailored to POVM distributions but is generally applicable to probabilistic dynamics. Thus we can extend its application also to continuous scenarios that become challenging to model in high-dimensional settings when relying on discretization schemes, due to the curse of dimensionality that is inherent to high-dimensional grids. Instead, we may employ a suited network architecture, such as a normalizing flow discussed in [Section 3.2.5](#), to build a variational model of the probability density, thereby circumventing the need for high-dimensional grids.

LETTER • OPEN ACCESS

Variational Monte Carlo approach to partial differential equations with neural networks

To cite this article: Moritz Reh and Martin Gärtner 2022 *Mach. Learn.: Sci. Technol.* **3** 04LT02

View the [article online](#) for updates and enhancements.

You may also like

- [Variational quantum one-class classifier](#)
Gunhee Park, Joonsuk Huh and Daniel K Park
- [Robust and scalable uncertainty estimation with conformal prediction for machine-learned interatomic potentials](#)
Yuge Hu, Joseph Musielewicz, Zachary W Ulissi et al.
- [Incompleteness of graph neural networks for points clouds in three dimensions](#)
Sergey N Pozdnyakov and Michele Ceriotti



LETTER

Variational Monte Carlo approach to partial differential equations with neural networks

OPEN ACCESS

RECEIVED
22 July 2022

REVISED
28 October 2022

ACCEPTED FOR PUBLICATION
15 November 2022

PUBLISHED
1 December 2022

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Moritz Reh^{1,*} and Martin Gärtner^{1,2,3,*}

¹ Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany

² Physikalisches Institut, Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany

³ Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, Heidelberg, 69120, Germany

* Authors to whom any correspondence should be addressed.

E-mail: moritz.reh@kip.uni-heidelberg.de and martin.gaertner@kip.uni-heidelberg.de

Keywords: normalizing flows, partial differential equations, Fokker–Planck equation, dynamical systems, statistical physics

Supplementary material for this article is available [online](#)

Abstract

The accurate numerical solution of partial differential equations (PDEs) is a central task in numerical analysis allowing to model a wide range of natural phenomena by employing specialized solvers depending on the scenario of application. Here, we develop a variational approach for solving PDEs governing the evolution of high dimensional probability distributions. Our approach naturally works on the unbounded continuous domain and encodes the full probability density function through its variational parameters, which are adapted dynamically during the evolution to optimally reflect the dynamics of the density. In contrast to previous works, this dynamical adaptation of the parameters is carried out using an explicit prescription avoiding iterative gradient descent. For the considered benchmark cases we observe excellent agreement with numerical solutions as well as analytical solutions for tasks that are challenging for traditional computational approaches.

1. Introduction

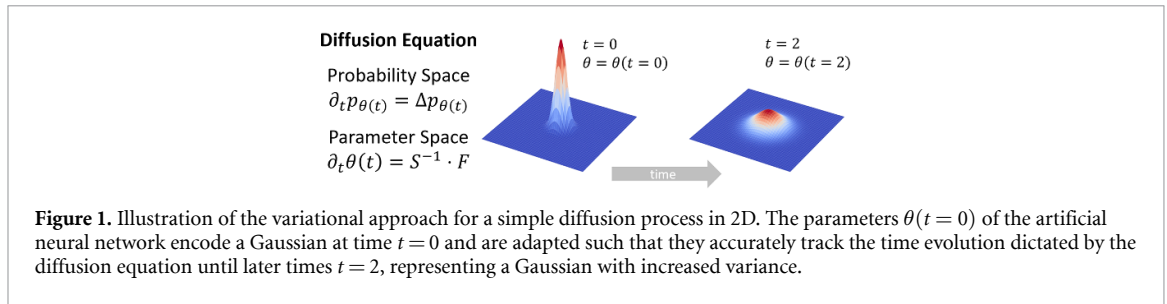
The description of nearly all processes in nature is formalized and modelled by means of differential equations, which dictate the evolution of a system given its initial state. Examples include the Navier–Stokes equation in fluid mechanics [1–4], the Schrödinger equation in quantum mechanics [5–7], and the Fokker–Planck equation governing diffusive processes [8–15]. Analytical solutions of these equations are only available in special cases and, generally, one is forced to resort to numerical techniques. A significant effort during the last century was made to improve the numerical solutions of differential equations [16–18]. There are numerous properties a numerical solver should ideally fulfil, rendering the field quite diverse, with many specialized solvers being developed [19].

Here, we focus on modelling the dynamics of d -dimensional probability density functions (PDFs) by means of an ansatz function, which in our case is given by an artificial neural network (ANN), as illustrated in figure 1. We consider evolution equations of Fokker–Planck form

$$\partial_t p = - \sum_i^d \partial_{x_i} \mu_i p + \sum_{ij}^d \partial_{x_i} \partial_{x_j} D_{ij} p, \quad (1)$$

where $\mu \in \mathbb{R}^d$ is the drift and $D \in \mathbb{R}^{d \times d}$ is the positive semi-definite diffusion matrix and it is understood that p , μ and D are evaluated at position \mathbf{x} and time t .

PDFs arise naturally across many disciplines, describing, for example, the phase space evolution of (quantum) matter [20, 21], the positions of particles subject to Brownian motion [11], the density of fluids [1] or stock prices in finance [15]. For many of these scenarios the PDF evolution is described by a diffusion process, meaning that the path of a single sampled point evolves according to a stochastic differential



equation (SDE) [22]. In the limit of averaging infinitely many stochastic trajectories one recovers the evolution of the PDF.

Consequently, the temporal evolution of probability densities can be obtained by either directly solving equation (1) via spatial discretization (grid based solvers), or by solving the corresponding stochastic dynamics for a large number of sample points (particle based solvers). The former approach, while allowing to control the discretization error via the grid spacing, suffers from the curse of dimensionality [23, 24] as the computational cost scales exponentially in the spatial dimension, restricting its applicability to low dimensional cases. The latter approach solves the SDE associated to the Fokker–Planck equation through the Feynman–Kac formula for an ensemble of points sampled from the initial distribution [25, 26]. While suited to compute observables, such as moments of the distribution, in high dimensions, there is no direct way to obtain estimates for functionals of the distribution as an expression for p is lacking [27–29].

In this work, we present a new tool that overcomes the aforementioned limitations of traditional methods by combining variational Monte-Carlo (VMC) with normalizing flows (NFs). While VMC is a long established technique in quantum many-body physics [30–33], NFs are a relatively novel class of ANNs also known as invertible neural networks (INNs) [34]. They have been applied with remarkable success to long standing problems in statistical physics [35], inference and data generation [34, 36–40], as well as quantum field theories [41, 42]. Here, we understand the NF as an ansatz function for the time-dependent density. The choice of the ansatz-function is a degree of freedom in our approach and can be adapted to the problem at hand exploiting prior knowledge about the function class the time-dependent density belongs to. Among the possible choices, ANNs are a promising class of ansatz functions, as they may become *universal* function approximators in the infinite parameter limit, which applies to lesser extent to NFs [43, 44]. Adjusting the parameters of the ansatz function to the dynamics dictated by equation (1) is achieved by a time-dependent variational principle (TDVP), which maps the dynamics of the PDF onto the variational manifold generated by the ansatz function [32, 33, 45]. Crucially, the approach is self-contained and at no point relies on data generated from other solvers, in contrast to prior works using neural networks to solve partial differential equations (PDEs) [46–49], allowing us to obtain numerical solutions for tasks that are challenging for grid-based or particle-based solvers. Our approach differs from the popular physics informed neural networks (PINNs) [50, 51] in that we do not carry out a costly global gradient-descent based optimization in each time step to update the models’ parameters, but rather follow an explicit, analytically derived time derivative of the network parameters which is given by the TDVP. We are particularly interested in high-dimensional scenarios which are infeasible to solve with grid-based methods and in quantities which are not easily obtainable by modelling many stochastic processes, such as functionals of the PDF. Indeed we show that, using the developed approach, we can reliably estimate differential entropies in a Monte Carlo fashion requiring only a few thousand samples. We benchmark our approach for the case of an eight-dimensional heat equation and a six-dimensional dissipative phase space evolution.

2. Normalizing flows

While we employ neural networks as ansatz functions, we emphasize that the derived TDVP is applicable to any parameterized density, such as Gaussian mixture models or energy-based estimators. We use NFs [34, 36] to model densities as they have many desirable properties, among which are (a) a guarantee of normalization for any set of parameters θ , (b) a tractable likelihood and (c) the ability to generate independent samples without the need to resort to Markov Chains. NFs parameterize densities by assuming a latent distribution π which is transformed into the distribution of interest by a trainable and invertible map \mathbf{f}_θ ,

$$\mathbf{x} = \mathbf{f}_\theta(\mathbf{z}) \text{ with } \mathbf{z} \sim \pi. \quad (2)$$

Usually, π is chosen to be a ‘simple’ distribution, e.g. a Gaussian, such that its samples \mathbf{z} can be generated easily. The probability associated with the point \mathbf{x} is proportional to $\pi(\mathbf{f}_\theta^{-1}(\mathbf{x}))$ times the determinant of the Jacobian of the transformation,

$$p_\theta(\mathbf{x}) = \pi(\mathbf{f}_\theta^{-1}(\mathbf{x})) \left| \det \left(\frac{\partial \mathbf{f}_\theta^{-1}(\mathbf{x})}{\partial \mathbf{x}} \right) \right|. \quad (3)$$

The function \mathbf{f}_θ is composed from a series of invertible transformations $\mathbf{f}_\theta = \varphi_\theta^1 \circ \dots \circ \varphi_\theta^N$ which are explained in detail in the supplemental material (SM). Importantly, the Jacobian of the function is tractable meaning that its determinant is efficiently inferred when computing a forward pass, an operation carried out whenever the real space probability is evaluated at some point of interest. By stacking many of these ‘coupling blocks’ φ^i , the function \mathbf{f}_θ becomes an expressive coordinate transform, that is, however, incapable of changing the tail behaviour of the latent space distribution [52]. We overcome this problem by dynamically adapting the latent space distribution π to reflect dynamical changes in the tails of the distribution. This is explained in more detail below and in the SM.

3. Time-dependent variational principle

The idea of the TDVP originated in the context of VMC [30] where it has been applied extensively to solve problems in quantum-many-body physics, with a growing interest in the use of neural networks as variational ansatz functions [32, 33, 45, 53]. Its aim is to locally search for the closest approximation to the dynamics of the density within the variational manifold. Concretely, one aims to solve

$$\operatorname{argmin}_{\dot{\theta}} \mathcal{D}(p_{\theta(t)+\tau\dot{\theta}}, p_{\theta(t)} + \tau\dot{p}_{\theta(t)}) \quad (4)$$

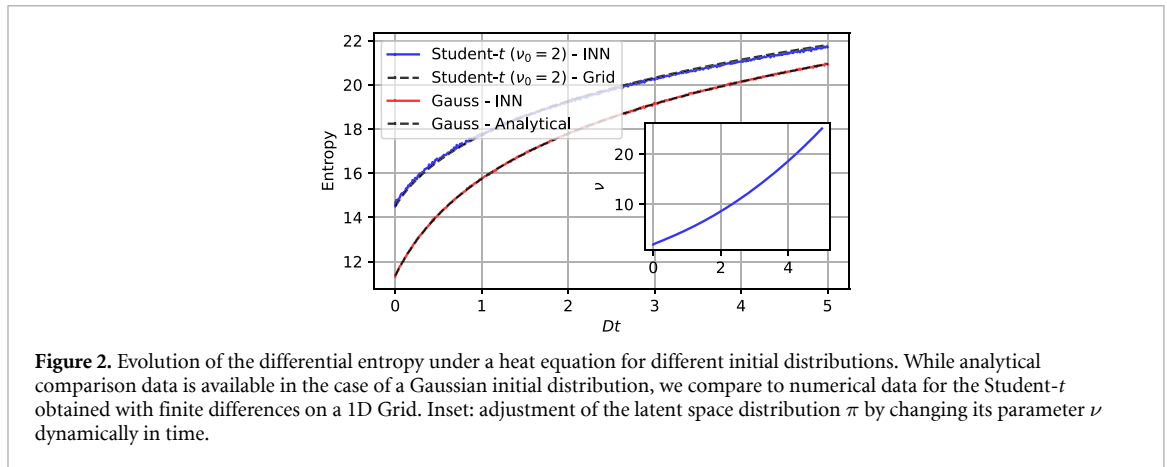
where \mathcal{D} is a suitable distance measure between probability distributions, τ denotes a small time step, $\dot{p}_{\theta(t)}$ is the derivative given by equation (1), and $\dot{\theta}$ is the unknown corresponding parameter time derivative. The solution to equation (4) can be found by requiring the derivative with respect to $\dot{\theta}$ to be zero. By expanding equation (4) to second order in τ one finds

$$S_{kk'} \dot{\theta}_{k'} = F_k. \quad (5)$$

We defer the details of this derivation to the SM. Here $S_{kk'} = \langle O_k(\mathbf{x}) O_{k'}(\mathbf{x}) \rangle_{\mathbf{x} \sim p_{\theta(t)}}$ denotes the Fisher information metric and $F_k = \langle O_k(\mathbf{x}) \partial_t \log(p_{\theta(t)}(\mathbf{x})) \rangle_{\mathbf{x} \sim p_{\theta(t)}}$ is a force term, where O_k denotes the (logarithmic) variational derivative $O_k(\mathbf{x}) = \partial_{\theta_k} \log(p_{\theta(t)}(\mathbf{x}))$ and $\partial_t \log(p_{\theta(t)}(\mathbf{x}))$ is given by the RHS of the to be solved PDE. Here $\langle \cdot \rangle_{\mathbf{x} \sim p_{\theta(t)}}$ denotes an expectation value evaluated through Monte Carlo sampling from the model distribution $p_{\theta(t)}$. Notice, that we heavily rely on the differentiability of the ansatz function $p_{\theta(t)}$ with respect to both variational parameters and spatial coordinates. The latter frequently appear on the RHS of equation (1) and are thus required for computing $\dot{p}_{\theta(t)}$. This is in striking contrast to grid-based techniques which require making grid cells finer for higher accuracy. Here, instead, we have access to the exact derivatives through automatic differentiation. The choice of distance measure to compare the two probability distributions is not arbitrary as the form of S and F directly depends on it. In order to obtain expressions of S and F that can be efficiently estimated through a finite number of samples, we found that both the Hellinger distance $\mathcal{D}_H(p, q) = 1 - F(p, q) = 1 - \int \sqrt{pq} d\mathbf{x}$ and the Kullback–Leibler (KL) divergence $\mathcal{D}_{KL}(p, q) = \int p \log(p/q) d\mathbf{x}$ yield the same result of the desired form. Care has to be taken when solving equation (5) for $\dot{\theta}$, as the inverse of S may not exist. This is the case if directions in parameter space are present along which the probabilities are stationary, which can be dealt with by regularization procedures [33, 53].

4. Problem setup

We are interested in solving initial value problems, for which the initial density distribution $p(0, \mathbf{x}) = u(\mathbf{x})$ is given along with the RHS of equation (1) which governs its evolution. To exactly encode the initial distribution $u(\mathbf{x})$ in the model $p_{\theta(t=0)}$, the latent distribution is set to $u(\mathbf{x})$ and the parameters of the map $\mathbf{f}_{\theta(t=0)}$ are chosen such that it represents the identity map $\mathbf{f}_{\theta(t=0)}(\mathbf{x}) = \mathbf{x}$. If the initial distribution cannot be given in closed form and therefore cannot be set analytically as the latent space distribution π , the network may be trained on its samples to approximately encode it at time $t = 0$. Then a solver is used which integrates the parameters according to equation (5).



5. Application 1: diffusion in high dimensions

As a first benchmark scenario we consider the heat equation in $d = 8$ dimensions. The heat equation appears across many disciplines ranging from engineering [54, 55] and molecular motion [11] to the pricing of financial derivatives given by the famous Black–Scholes equation [56, 57] and reads

$$\partial_t p(t, \mathbf{x}) = D\Delta_{\mathbf{x}} p(t, \mathbf{x}). \tag{6}$$

Importantly, an analytical solution exists against which we can benchmark, making the described scenario a good showcase of the proposed approach. The solution is given by a convolution of the initial distribution $p(0, \mathbf{x})$ with the ‘heat kernel’ $\Phi(t, \mathbf{x}) = (4\pi t)^{-(d/2)} \exp(-\mathbf{x}^2/4Dt)$ [58], which is the Green’s function to equation (6), such that

$$p(t, \mathbf{x}) = \int p(0, \mathbf{y}) \Phi(t, \mathbf{x} - \mathbf{y}) d\mathbf{y}. \tag{7}$$

We aim to observe the growth of the differential entropy

$$S(t) = - \int p(t, \mathbf{x}) \log(p(t, \mathbf{x})) d\mathbf{x} = - \langle \log(p(t, \mathbf{x})) \rangle_{\mathbf{x} \sim p(t, \mathbf{x})} \tag{8}$$

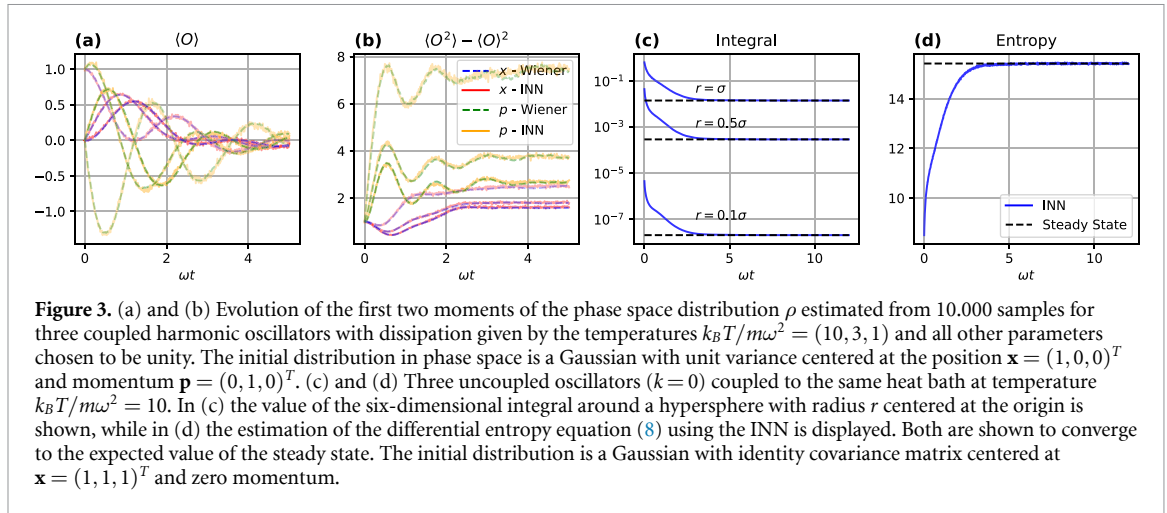
with time, a task, which is challenging or even intractable using other numerical techniques in high dimensions for the reasons mentioned above [23, 24, 27–29]. In the case of a Gaussian distribution for $p(0, \mathbf{x})$ with zero mean and unit covariance matrix, we obtain a Gaussian of larger variance at later points in time, in which case we observe perfect agreement between the analytical solution and the one obtained using the INN as shown in figure 2. If we choose a Student- t distribution as initial distribution, i.e.

$$p(0, \mathbf{x}) \propto \left(1 + \frac{\mathbf{x}^2}{\nu} \right)^{-(\nu+d)/2} \tag{9}$$

with $\nu = 2$ we can no longer compare to the analytical solution as the involved integrals become infeasible to solve. However, by exploiting the spherical symmetry of the problem, we can map the evolution to an effective 1D problem of the radial dependency of p which we can approximately solve on a grid using finite differences. The grid based solution and that obtained using the INN are generally in good agreement. We observe a slight difference which we attribute to technical challenges of the grid-based approach, which we discuss more elaborately in the SM.

6. Application 2: diffusion in classical phase space

As a second demonstration of the proposed approach we consider classical Hamiltonian dynamics in phase space with additional diffusion. Concretely, we choose the Hamiltonian H to represent coupled harmonic oscillators (coupling strength k) which are in contact with heat baths of different temperatures T_i , such that the solution does not factorize in the eigenbasis of H . We provide the Hamiltonian and its generated phase space flow in the SM. The heat baths lead to diffusion in phase space, which implies that sampled points of the distribution evolve according to an SDE. We show that the INN faithfully estimates moments of the



distribution, probabilities (i.e. integrals over finite domains) as well as functionals of the PDF that correspond to integrals over the entire domain.

The described system obeys the following Fokker–Planck equation [59]

$$\partial_t \rho(t, \mathbf{x}, \mathbf{p}) = \left[-\partial_{\mathbf{p}} H \cdot \partial_{\mathbf{x}} + \partial_{\mathbf{x}} H \cdot \partial_{\mathbf{p}} + \gamma \left(\mathbf{p} \cdot \partial_{\mathbf{p}} + mk_B \sum_i T_i \partial_{p_i}^2 \right) \right] \rho(t, \mathbf{x}, \mathbf{p}), \tag{10}$$

whose corresponding SDE is given by [59]

$$\begin{aligned} dx_i &= \partial_{p_i} H dt, \\ dp_i &= -[\gamma p_i + \partial_{x_i} H] dt + \sqrt{\lambda_i} dw_i. \end{aligned} \tag{11}$$

Here, $\lambda_i = \sqrt{2m\gamma k_B T_i}$, $dw_i = \Pi_i \sqrt{dt}$ is the Wiener process with zero average $\langle dw_i \rangle = 0$ and standard scaling $\langle dw_i^2 \rangle = dt$ implying that Π is drawn from a standard Gaussian $\Pi \sim \mathcal{N}(0, 1)$. For simplicity we choose all quantities except T_i equal to unity.

In the case of heat baths of equal temperatures $T_i = T$ and vanishing coupling ($k = 0$) the system assumes a thermal steady state of Gaussian form in the long time limit given by the Gibbs-ensemble

$$\begin{aligned} \rho_{SS} &= \exp(-H/k_B T) / Z \\ &= \exp\left(-\frac{1}{2}(m\omega^2 \mathbf{x}^2 + \mathbf{p}^2/m)/k_B T\right) / Z, \end{aligned} \tag{12}$$

with $Z = \int \exp(-H/k_B T) d\mathbf{x} d\mathbf{p}$ the partition function, where the Gaussian form allows to compare against analytical results.

We consider four quantities of interest which we evaluate by drawing 10,000 samples from the INN, see figure 3. The first two quantities are the means and variances of the distribution evolved for the case of different T_i and $k = 1$. Here, comparison against estimates from solving the SDE for the same number of sampled points is straight forward and one observes excellent agreement between both methods. To obtain an easy benchmark case for integral and entropy estimation, we choose $k = 0$ and $T_i = T$ such that the steady state is Gaussian, see equation (12). We choose the integration volumes to be hyperspheres of radius r centered at the origin allowing for analytical evaluation of the Gaussian integral. The values of these integrals correspond to the probability of finding the system inside the hypersphere. Using the INN, we can estimate such integrals in a Monte-Carlo fashion by uniformly sampling points \mathbf{x}_i from inside the integration domain and average the associated probabilities $p_\theta(\mathbf{x}_i)$, which are shown to converge to the analytically obtained steady-state value in figure 2(c).

Finally, we again focus on the differential entropy (equation (8)), where figure 3(d) shows that our method succeeds to predict the differential entropy with low noise while converging to the expected steady state value.

7. Conclusion and outlook

We have introduced a variational approach to the dynamics of continuous probability distributions using NFs and demonstrated its power by applying it to paradigmatic benchmark problems. Our method is widely applicable, even beyond the Fokker–Planck form (1), e.g. to cases with non-local terms [60]. Its unique strength lies in estimating functionals of probability densities in high dimensions enabled by the availability of exact samples with tractable likelihood. We emphasize that other approaches such as PINN [50] require solving a large-scale non-convex optimization problem in each time step, which the TDVP replaces by the explicit update rule (5) (see SM for further discussion). The form of the ansatz function can be chosen flexibly and is not required to be a neural network. The only restrictions are that (a) samples from its distribution may be obtained and (b) derivatives with respect to inputs and parameters are computable. While building NFs using stacked coupling blocks is a popular approach, other flow architectures exist and it would be interesting to investigate their potential in solving PDEs in the future. Since the TDVP can also work with non-normalized probabilities, also energy based models would be viable ansatz functions although this would mean that samples would have to be obtained by resorting to Markov-chains.

For the utilized architecture we found that challenges exist when trying to solve chaotic dynamics. We believe this to be caused by the high amount of information of the phase space distribution which needs to be encoded using comparably few parameters. Additionally, we found it challenging to model distributions whose tail behaviour deviated from that of the latent space distribution. In the example shown in figure 2 this could be dealt with by elevating ν to be a variational parameter, which would tend to infinity for late times, representing the exact tail behaviour of the real space distribution. However, if the real space tail behaviour cannot be accurately modelled in latent space, e.g. because its form is not known beforehand, one cannot expect to accurately model the distribution on the entire domain.

Data availability statement

The code used for this project is based on the jVMC library [61], making use of flax [62] and jax [63] and is available under [GitHub: RehMoritz/vmc_pde](https://github.com/RehMoritz/vmc_pde). The repository also contains the data from figures 2 and 3.

Acknowledgments

The authors thank Markus Schmitt, Julian Urban, Ullrich Köthe and Peter Sorrenson for helpful discussions. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC2181/1-39 090 0948 (the Heidelberg STRUCTURES Excellence Cluster) and within the Collaborative Research Center SFB1225 (ISOQUANT). This work was partially financed by the Baden-Württemberg Stiftung gGmbH. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/575-1 FUGG (JUSTUS 2 cluster). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [64] at Jülich Supercomputing Centre (JSC).

ORCID iDs

Moritz Reh  <https://orcid.org/0000-0002-8408-7558>

Martin Gärtner  <https://orcid.org/0000-0003-1914-7099>

References

- [1] Ferziger J H and Perić M 2002 *Computational Methods for Fluid Dynamics* (Berlin: Springer)
- [2] Tropea C, Yarin A and Foss J 2007 *Springer Handbook of Experimental Fluid Mechanics* (Berlin: Springer)
- [3] Wesseling P 2001 *Principles of Computational Fluid Dynamics* (Berlin: Springer)
- [4] Spurk J H and Aksel N 2020 *Fluid Mechanics* (Cham: Springer)
- [5] Sakurai J J and Napolitano J 2017 *Modern Quantum Mechanics* (Cambridge: Cambridge University Press)
- [6] Griffiths D J and Schroeter D F 2018 *Introduction to Quantum Mechanics* (Cambridge: Cambridge University Press)
- [7] Schwabl F 2008 *Advanced Quantum Mechanics* (Berlin: Springer)
- [8] Kampen N V 2007 *Stochastic Processes in Physics and Chemistry* (Amsterdam: Elsevier)
- [9] Coffey W T and Kalmykov Y P 2011 *The Langevin Equation* (Singapore: World Scientific)
- [10] Sornette D 2001 *Physica A* **290** 211
- [11] Freedman D 1983 *Brownian Motion and Diffusion* (New York: Springer)
- [12] Shen X, Chen H, Dai J and Dai W 2002 *Queueing Syst.* **42** 33

- [13] Rouse P E 1953 *J. Chem. Phys.* **21** 1272
- [14] Prakash J R and Öttinger H C 1999 *Macromolecules* **32** 2028
- [15] Reisinger C 2004 *Numerische Methoden für Hochdimensionale Parabolische Gleichungen am Beispiel von Optionspreisaufgaben* (Heidelberg: Heidelberg University Library)
- [16] Thomas J W 1995 *Numerical Partial Differential Equations: Finite Difference Methods* (New York: Springer)
- [17] Hairer E 2006 *Geometric Numerical Integration* (Berlin: Springer)
- [18] Kress R 1998 *Graduate Texts in Mathematics* (New York: Springer) pp 189–224
- [19] Quarteroni A and Valli A 1994 *Numerical Approximation of Partial Differential Equations* (Berlin: Springer)
- [20] Schwabl F 2006 *Statistical Mechanics* (Berlin: Springer)
- [21] Zachos C K, Fairlie D B and Curtright T L 2005 *Quantum Mechanics in Phase Space* (Singapore: World Scientific)
- [22] Tomé T and de Oliveira M J 2015 *Stochastic Dynamics and Irreversibility* (Cham: Springer International Publishing)
- [23] Kang W and Wilcox L C 2017 *Comput. Optim. Appl.* **68** 289
- [24] Griebel M 2006 *Sparse Grids and Related Approximation Schemes for Higher Dimensional Problems* (Cambridge: Cambridge University Press) pp 106–61
- [25] Moral P D 2004 *Feynman-Kac Formulae* (New York: Springer)
- [26] Oksendal B 1998 *Stochastic Differential Equations* (Berlin: Springer)
- [27] Kraskov A, Stögbauer H and Grassberger P 2004 *Phys. Rev. E* **69** 066138
- [28] Singh S and Póczos B 2016 Finite-sample analysis of fixed-k nearest neighbor density functional estimators (arXiv:1606.01554 [math.ST])
- [29] Ao Z and Li J 2022 Entropy estimation via normalizing flow (available at: https://aaai-2022.virtualchair.net/poster_aaai8184)
- [30] McMillan W L 1965 *Phys. Rev.* **138** A442
- [31] Ceperley D, Chester G V and Kalos M H 1977 *Phys. Rev. B* **16** 3081
- [32] Carleo G, Cevolani L, Sanchez-Palencia L and Holzmann M 2017 *Phys. Rev. X* **7** 031026
- [33] Carleo G and Troyer M 2017 *Science* **355** 602
- [34] Dinh L, Sohl-Dickstein J, and Bengio S 2016 Density estimation using Real NVP
- [35] Noé F, Olsson S, Köhler J and Wu H 2019 *Science* **365** eaaw1147
- [36] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S and Lakshminarayanan B 2019 Normalizing flows for probabilistic modeling and inference (arXiv:1912.02762 [stat.ML])
- [37] Grathwohl W, Chen R T Q, Bettencourt J, Sutskever I, and Duvenaud D 2018 Fjord: free-form continuous dynamics for scalable reversible generative models (arXiv:1810.01367 [cs.LG])
- [38] Dinh L, Krueger D and Bengio Y 2014 Nice: non-linear independent components estimation (arXiv:1410.8516 [cs.LG])
- [39] Kingma D P and Dhariwal P 2018 Glow: generative flow with invertible 1x1 convolutions (arXiv:1807.03039 [stat.ML])
- [40] Ardizzone L, Kruse J, Wirkert S, Rahner D, Pellegrini E W, Klessen R S, Maier-Hein L, Rother C and Köthe U 2018 Analyzing inverse problems with invertible neural networks (arXiv:1808.04730 [cs.LG])
- [41] Albergo M S, Kanwar G, Racanière S, Rezende D J, Urban J M, Boyda D, Cranmer K, Hackett D C and Shanahan P E 2021 *Phys. Rev. D* **104** 114507
- [42] Pawłowski J M and Urban J M 2022 Flow-based density of states for complex actions (<https://doi.org/10.48550/arXiv.2203.01243>)
- [43] Kong Z and Chaudhuri K 2020 The expressive power of a class of normalizing flow models (arXiv:2006.00392 [cs.LG])
- [44] Teshima T, Ishikawa I, Tojo K, Oono K, Ikeda M and Sugiyama M 2020 Coupling-based invertible neural networks are universal diffeomorphism approximators (arXiv:2006.11469 [cs.LG])
- [45] Reh M, Schmitt M and Gärtner M 2021 *Phys. Rev. Lett.* **127** 230501
- [46] Brandstetter J, Worrall D and Welling M 2022 Message passing neural PDE solvers (arXiv:2202.03376 [cs.LG])
- [47] Lu L, Jin P, Pang G, Zhang Z and Karniadakis G E 2021 *Nat. Mach. Intell.* **3** 218–29
- [48] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A and Anandkumar A 2020 Fourier neural operator for parametric partial differential equations (arXiv:2010.08895 [cs.LG])
- [49] Beck C, Becker S, Grohs P, Jaafari N and Jentzen A 2021 *J. Sci. Comput.* **88** 73
- [50] Raissi M, Perdikaris P and Karniadakis G 2019 *J. Comput. Phys.* **378** 686
- [51] Feng X, Zeng L and Zhou T 2022 *SSRN Electron. J.* **32** 401–23
- [52] Jaini P, Selby K A, and Yu Y 2019 Sum-of-squares polynomial flow (arXiv:1905.02325 [cs.LG])
- [53] Schmitt M and Heyl M 2020 *Phys. Rev. Lett.* **125** 100503
- [54] Poirier D R and Geiger G H 2016 *Transport Phenomena in Materials Processing* (Cham: Springer)
- [55] Annaratone D 2010 *Engineering Heat Transfer* (Berlin: Springer)
- [56] Black F and Scholes M 1973 *J. Political Econ.* **81** 637
- [57] Guillaume T 2018 *Ann. Oper. Res.* **281** 229
- [58] Davies E B 1989 *Heat Kernels and Spectral Theory* (Cambridge: Cambridge University Press)
- [59] Presilla C, Onofrio R and Patriarca M 1997 *J. Phys. A: Math. Gen.* **30** 7385
- [60] Lin L, Duan J, Wang X and Zhang Y 2021 *Chaos* **31** 051105
- [61] Schmitt M and Reh M 2022 *SciPost Physics Codebases* p 2
- [62] Heek J, Levskaya A, Oliver A, Ritter M, Rondepierre B, Steiner A and vanZee M 2020 Flax: a neural network library and ecosystem for JAX (available at: <http://github.com/google/flax>)
- [63] Bradbury J, Frostig R, Hawkins P, Johnson M J, Leary C, Maclaurin D, Necula G, Paszke A, VanderPlas J, Wanderman-Milne S and Zhang Q 2018 JAX: composable transformations of Python+NumPy programs (available at: <http://github.com/google/jax>)
- [64] Krause D 2019 JUWELS: modular tier-0/1 supercomputer at the Jülich supercomputing centre *J. Large-Scale Res. Facilities* **5**

Supplementary Materials

Derivation of the TDVP Equation

The basic idea of a TDVP is to minimize the distance

$$\mathcal{D} \left(p_{\theta(t)} + \dot{p}_{\theta(t)}\tau, p_{\theta(t)} + \sum_k \frac{\partial p_{\theta(t)}}{\partial \theta_k} \dot{\theta}_k \tau \right), \quad (13)$$

between the evolved state at time $t + \tau$ and the updated network state with respect to the parameter updates $\dot{\theta}$ at each time t . Here, we exemplarily derive Eq. (5) from the Hellinger distance $\mathcal{D}_H(p, q) = \int \sqrt{p(\mathbf{x})q(\mathbf{x})} d\mathbf{x}$, i.e. by maximizing the classical fidelity $F(p, q) = 1 - \mathcal{D}_H(p, q)$. As noted in the main text, an equivalent derivation is possible using the Kullback-Leibler divergence \mathcal{D}_{KL} which leads to the same result. For better readability, we drop the time index and continue with the optimality condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial \dot{\theta}_k} F \left(p + \dot{p}\tau, p + \sum_{k'} \frac{\partial p}{\partial \theta_{k'}} \dot{\theta}_{k'} \tau \right) \\ &= \frac{\partial}{\partial \dot{\theta}_k} \int p(\mathbf{x}) \sqrt{1 + a\tau + b\tau^2} d\mathbf{x}, \end{aligned} \quad (14)$$

where a and b are given by

$$\begin{aligned} a &= \frac{\partial \log p(\mathbf{x})}{\partial t} + \sum_{k'} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'}, \\ b &= \frac{\partial \log p(\mathbf{x})}{\partial t} \sum_{k'} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'}. \end{aligned} \quad (15)$$

Next we perform a second order expansion of the square root in the (small) time step τ :

$$\sqrt{1 + a\tau + b\tau^2} = 1 + \frac{a\tau}{2} + \frac{\tau^2}{8}(4b - a^2) + \mathcal{O}(\tau^3). \quad (16)$$

Using that the normalization of p is conserved under the time evolution one finds that the term linear in τ vanishes:

$$\begin{aligned} p(\mathbf{x})a &= \int \left(\dot{p}(\mathbf{x}) + \sum_{k'} \frac{\partial p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'} \right) d\mathbf{x} \\ &= \sum_{k'} \dot{\theta}_{k'} \frac{\partial}{\partial \theta_{k'}} \int p(\mathbf{x}) d\mathbf{x} \\ &= \sum_{k'} \dot{\theta}_{k'} \frac{\partial}{\partial \theta_{k'}} 1 \\ &= 0. \end{aligned} \quad (17)$$

Thus, the optimality condition becomes

$$\begin{aligned} 0 &= \frac{\partial}{\partial \dot{\theta}_k} \int \frac{p(\mathbf{x})}{p(\mathbf{x})^2} \left(4\dot{p}(\mathbf{x}) \sum_{k'} \frac{\partial p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'} \right. \\ &\quad \left. - (\dot{p}(\mathbf{x}) + \sum_{k'} \frac{\partial p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'})^2 \right) d\mathbf{x} \\ &= -\frac{\partial}{\partial \dot{\theta}_k} \int \frac{p(\mathbf{x})}{p(\mathbf{x})^2} \left(\dot{p}(\mathbf{x}) - \sum_{k'} \frac{\partial p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'} \right)^2 d\mathbf{x} \\ &= -\frac{\partial}{\partial \dot{\theta}_k} \int p(\mathbf{x}) \left(\frac{\partial \log p(\mathbf{x})}{\partial t} - \sum_{k'} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'} \right)^2 d\mathbf{x} \\ &= 2 \int p(\mathbf{x}) \frac{\log p(\mathbf{x})}{\partial \theta_k} \left(\frac{\partial \log p(\mathbf{x})}{\partial t} - \sum_{k'} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} \dot{\theta}_{k'} \right) d\mathbf{x}. \end{aligned} \quad (18)$$

Dropping the factor of 2 we obtain an equation for the optimal parameter update $\dot{\theta}$:

$$\begin{aligned} 0 &= \underbrace{\int p(\mathbf{x}) \frac{\partial \log p(\mathbf{x})}{\partial t} \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} d\mathbf{x}}_{=F_k} \\ &\quad - \sum_{k'} \underbrace{\int p(\mathbf{x}) \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} d\mathbf{x}}_{=S_{kk'}} \dot{\theta}_{k'} \\ &= \left\langle \frac{\partial \log p(\mathbf{x})}{\partial t} \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} \right\rangle_{\mathbf{x} \sim p} \\ &\quad - \sum_{k'} \left\langle \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} \frac{\partial \log p(\mathbf{x})}{\partial \theta_{k'}} \right\rangle_{\mathbf{x} \sim p} \dot{\theta}_{k'}. \end{aligned} \quad (19)$$

Importantly, we can now evaluate the integral by sampling according to the encoded probabilities $p(\mathbf{x})$ since the integrand is proportional to $p(\mathbf{x})$ for both F and S . This is a unique property of the distance measures \mathcal{D}_H and \mathcal{D}_{KL} while other distance measures, as for example the L^2 norm, do not lead to expressions of a form that can be efficiently evaluated from Monte Carlo samples. The same derivation can be carried out without assuming normalization. In this case the form of S and F is altered to

$$\begin{aligned} p(\mathbf{x}) &\rightarrow \frac{p(\mathbf{x})}{\int p(\mathbf{x}) d\mathbf{x}} \\ \log p(\mathbf{x}) &\rightarrow \log p(\mathbf{x}) - \log \int p(\mathbf{x}) d\mathbf{x} \\ \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} &\rightarrow \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} - \left\langle \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} \right\rangle_{\mathbf{x} \sim p} \\ \frac{\partial \log p(\mathbf{x})}{\partial t} &\rightarrow \frac{\partial \log p(\mathbf{x})}{\partial t} - \left\langle \frac{\partial \log p(\mathbf{x})}{\partial t} \right\rangle_{\mathbf{x} \sim p}, \end{aligned} \quad (20)$$

where the last two lines are obtained using

$$\begin{aligned}
& \frac{\partial}{\partial \theta_k} \left(\log p(\mathbf{x}) - \log \int p(\mathbf{x}) d\mathbf{x} \right) \\
&= \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} - \frac{\int \frac{\partial p(\mathbf{x})}{\partial \theta_k} d\mathbf{x}}{\int p(\mathbf{x}') d\mathbf{x}'} \\
&= \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} - \int \frac{p(\mathbf{x})}{\int p(\mathbf{x}') d\mathbf{x}'} \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} d\mathbf{x} \\
&= \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} - \left\langle \frac{\partial \log p(\mathbf{x})}{\partial \theta_k} \right\rangle_{\mathbf{x} \sim p}.
\end{aligned} \tag{21}$$

Here, the log derivative trick was used in the third line. One may proceed similarly for the time derivative. Overall, this leaves us with a connected correlator structure instead of a simple correlator

$$\begin{aligned}
S_{kk'} &= \langle O_k(\mathbf{x}) O_{k'}(\mathbf{x}) \rangle_{\mathbf{x} \sim p_\theta(t)} \\
&\quad - \langle O_k(\mathbf{x}) \rangle_{\mathbf{x} \sim p_\theta(t)} \langle O_{k'}(\mathbf{x}) \rangle_{\mathbf{x} \sim p_\theta(t)}, \\
F_k &= \langle O_k(\mathbf{x}) \partial_t \log(p_\theta(t)(\mathbf{x})) \rangle_{\mathbf{x} \sim p_\theta(t)} \\
&\quad - \langle O_k(\mathbf{x}) \rangle_{\mathbf{x} \sim p_\theta(t)} \langle \partial_t \log(p_\theta(t)(\mathbf{x})) \rangle_{\mathbf{x} \sim p_\theta(t)}
\end{aligned} \tag{22}$$

with O_k the (logarithmic) variational derivative

$$O_k(\mathbf{x}) = \partial_{\theta_k} \log(p_\theta(t)(\mathbf{x})). \tag{23}$$

We finally arrive at

$$\dot{\theta}_k = \tilde{S}_{kk'}^{-1} F_{k'} \tag{24}$$

where the tilde is due to the fact that we cannot invert S itself but rather need to regularize it because it is usually rank-deficient. One can show that the updates that were found are indeed maxima of the fidelity:

$$\begin{aligned}
& \frac{\partial^2}{\partial \dot{\theta}_k^2} F(p + \tau \dot{p}, p + \tau \sum_{k'} \frac{\partial p}{\partial \theta_{k'}} \dot{\theta}_{k'}) \\
&= \frac{\partial}{\partial \dot{\theta}_k} \left(F_k - \sum_{k'} S_{kk'} \dot{\theta}_{k'} \right) \\
&= - \sum_{k'} S_{kk'} \delta_{k'k} \\
&= - S_{kk} \\
&= - \left\langle \left(O_k(\mathbf{x}) - \langle O_k(\mathbf{x}) \rangle \right)^2 \right\rangle_{\mathbf{x} \sim p} \\
&\leq 0.
\end{aligned} \tag{25}$$

Approximation Error

The adjustment of the parameters to reflect changes in the probability carries an associated error, as the parameters can usually not be changed to perfectly reflect

the time derivatives of all the sampled points used to estimate F . The TDVP allows to quantify this error by estimating the residual

$$r(t) = \frac{1}{N_s} \sum_i \left| \dot{p}(t, \mathbf{x}_i) - \sum_k \frac{\partial p(t, \mathbf{x}_i)}{\partial \theta_k} \dot{\theta}_k \right|^2. \tag{26}$$

Computational Complexity

Here we compare the computational complexity of the explicit variational method that we are proposing to an iterative gradient descent based technique.

The operations carried out in the gradient descent based procedure are given in Alg. 1. To summarize, the algorithm computes time derivatives at the sampled points using the differential operator \mathcal{F} , which depends on the PDE under scrutiny (e.g. $\mathcal{F} = D\Delta_{\mathbf{x}}$ in the case of the heat equation, see Eq. (6)). The time derivatives, weighted with some small time step τ , are added to the current probability values $p_\theta(\mathbf{x}_i)$ and define the new regression targets. Using a loss function that is minimal when the encoded distribution agrees with the new regression targets, one searches for a new solution in the parameter space. Once a convergence criterion is met, the search stops and continues with the next time step. This search can become costly, since the optimization problem is in general non-convex without convergence guarantees. It is therefore beneficial to avoid the iterative search using a closed form, as lined out in Alg. 2.

Algorithm 1 Iterative Gradient Descent

```

1: procedure TIMESTEP( $p_\theta$ )
2:    $K \leftarrow \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$   $\triangleright$  obtain sample set  $K$ 
3:    $\partial_t p_\theta(\mathbf{x}_i) \leftarrow \mathcal{F}(p_\theta)(\mathbf{x}_i)$   $\triangleright$  get time derivatives at each  $\mathbf{x}$ 
4:
5:   while convergence criteria is not met do
6:      $L \leftarrow \sum_i \mathcal{D}(p_\theta(\mathbf{x}_i), p_\theta(\mathbf{x}_i) + \tau \partial_t p_\theta(\mathbf{x}_i))$   $\triangleright$  Define
       Loss function, Derivative acts on first argument in  $\mathcal{D}$ 
7:      $\theta \leftarrow \theta + \eta \nabla_\theta L$   $\triangleright$  Gradient descent step
8:
9:   return  $\theta$ 

```

Algorithm 2 Explicit second order scheme

```

1: procedure TIMESTEP( $p_\theta$ )
2:    $K \leftarrow \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$   $\triangleright$  obtain sample set  $K$ 
3:    $\partial_t p_\theta(\mathbf{x}_i) \leftarrow \mathcal{F}(p_\theta)(\mathbf{x}_i)$   $\triangleright$  get time derivatives at each  $\mathbf{x}$ 
4:
5:    $S_{kk'} \leftarrow 1/N \sum_{i \in S} O_k(\mathbf{x}_i) O_{k'}(\mathbf{x}_i)$ 
6:    $F_k \leftarrow 1/N \sum_{i \in S} O_k(\mathbf{x}_i) \partial_t \log(p_\theta(\mathbf{x}_i))$ 
7:    $\dot{\theta}_k \leftarrow S_{kk'}^{-1} F_{k'}$ 
8:    $\theta \leftarrow \theta + \tau \dot{\theta}$ 
9:
10:  return  $\theta$ 

```

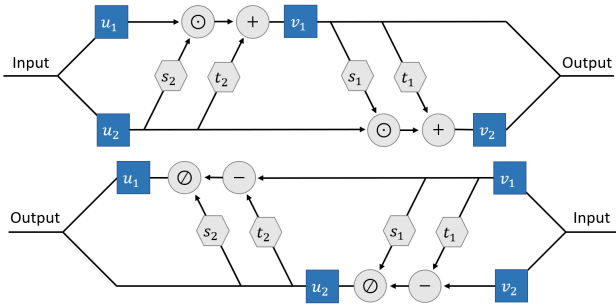


FIG. 4. Illustration of a coupling block. Top: Forward pass, corresponding to Eq. (27). Bottom: Inverse pass, corresponding to Eq. (28).

The proposed algorithm also relies on obtaining samples from the distribution for which then time derivatives are computed. Beyond that, we require the logarithmic variational derivatives of the probabilities $O_k(\mathbf{x}) = \partial_{\theta_k} \log(p_\theta(\mathbf{x}))$. Obtaining these derivatives is of similar computational cost compared to a single gradient descent step in Alg. 1, as such a step also requires the differentiation of the probabilities at all sample positions. However, using the explicit scheme, this operation needs to be carried out only once. Additionally, we need to add these derivatives together, as shown in lines 5 and 6 of Alg. 2, an operation with negligible computational cost. To arrive at the time derivatives of the parameters θ , we need to invert the S matrix, which has cubic computational cost in the number of network parameters. In practice, there are many more computationally efficient ways than actually computing the inverse, which allow to reduce the computational cost of this step. While this step limits the number of network parameters that can practically be used, and thus the expressivity of the ansatz, we have not found this to be a limiting factor in the application considered in this work.

Figure	Input Dim.	# Coupling Blocks	# Layers	Net t	# Parameters	# Samples	π
Fig. 2	8	4	2	No	392	10.000	Student- t / Gauss
Fig. 3	6	4	2	Yes	234	10.000	Gauss

TABLE I. Hyperparameters that were used for the different figures in the main text.

Isotropic Heat Equation as a 1D Problem

Here we describe the procedure with which the reference data for Fig. 2 in the main text was obtained in the case of a Student- t initial distribution and give an explanation for the slight discrepancies observed in Fig. 2 in the main text.

The heat equation

$$\partial_t p(t, \mathbf{x}) = D \Delta p(t, \mathbf{x}) \quad (30)$$

can be recast as a 1D problem if the initial condition $p(t, \mathbf{x}) = u(\mathbf{x})$ features a spherical symmetry. This is the

The runtimes of the examples we presented all lie below half an hour on a single NVIDIA A100 GPU.

Normalizing Flows

For the definition of the normalizing flow we use the Real-NVP type coupling blocks introduced in [34] and depicted in Fig. 4. Each coupling block φ_i splits the input into two parts u_1 and u_2 which is done in a random but fixed way. The transformations in a single coupling block are defined by four networks s_1, t_1, s_2, t_2 that change the input as follows:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 \odot \exp(\mathbf{s}_2(\mathbf{u}_2)) + \mathbf{t}_2(\mathbf{u}_2), \\ \mathbf{v}_2 &= \mathbf{u}_2 \odot \exp(\mathbf{s}_1(\mathbf{v}_1)) + \mathbf{t}_1(\mathbf{v}_1). \end{aligned} \quad (27)$$

The inverse of this transformation is given by:

$$\begin{aligned} \mathbf{u}_2 &= (\mathbf{v}_2 - \mathbf{t}_1(\mathbf{v}_1)) \odot \exp(-\mathbf{s}_1(\mathbf{v}_1)), \\ \mathbf{u}_1 &= (\mathbf{v}_1 - \mathbf{t}_2(\mathbf{u}_2)) \odot \exp(-\mathbf{s}_2(\mathbf{u}_2)). \end{aligned} \quad (28)$$

Here \odot means element-wise multiplication. The networks \mathbf{s} and \mathbf{t} are built equivalently as two layer feed-forward networks with half as many nodes in each layer as there are dimensions. In some cases we found it useful to not include the additive \mathbf{t} networks. Additionally, we allowed the network to adjust the mean $\boldsymbol{\mu}$ and the covariance matrix Σ of the distribution in latent space directly, potentially along with parameters ϑ of the distribution, e.g. ν in the case of the Student- t , such that

$$\pi = \pi(\boldsymbol{\mu}_\theta, \Sigma_\theta, \vartheta_\theta). \quad (29)$$

We parameterize Σ using either the Cholesky decomposition or by setting $\Sigma = \mathbf{1} + AA^T$, where we found the latter to be more stable numerically for simulating the heat equation. Network details are listed in in Table I.

case if it is fully described by a mean $\boldsymbol{\mu}$ and a covariance matrix Σ , as this allows to rescale coordinates such that the new distribution obeys $\boldsymbol{\mu} = 0$ and $\Sigma = \mathbf{1}$ enabling us to write $p(t, \mathbf{x})$ as $p(t, r)$, where $r = |\mathbf{x}|$.

Then, the spherical form of the Laplacian may be exploited

$$\Delta = \partial_r^2 + \frac{d-1}{r} \partial_r, \quad (31)$$

where d is the dimension of the distribution. The evolution of the distribution can then be solved using finite differences on a 1D grid. Note however, that there are

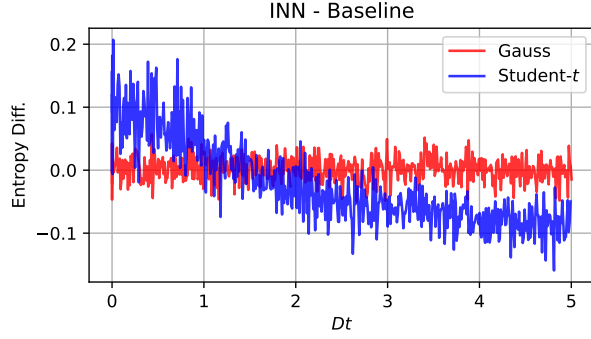


FIG. 5. Differences between entropy estimates of the INN compared to the baseline methods from Fig. 2 in the main text. Systematic deviations are visible for the case of the Student- t , which we attribute to problems with the baseline method as layed out below.

caveats associated with this procedure as the second term of Eq. (31) has a divergence for $r \rightarrow 0$. For the diffusion cases we considered, the distribution $p(t, r)$ has a maximum at $r = 0$, irrespective of the time t , implying that also the numerator $\partial_t p(t, r)$ vanishes. This necessitates the use of L'Hôpital's rule to write

$$\lim_{r \rightarrow 0} \frac{\partial_r p}{r} = \frac{\partial_r^2 p}{\partial_r r} = \partial_r^2 p. \quad (32)$$

We work with equidistant grid cells of size $\delta = 4 \cdot 10^{-3}$ and set a cutoff at $r = 100$. We employ L'Hôpital approximation for the first 10 grid cells, i.e. for $r \in [0, 10\delta]$,

which we found to be necessary for numerical stability. This implies however, that also the reference data is not free of approximations, which may be particularly interesting as we did not observe the INN curve to come closer to the reference data when increasing the network size. This could be viewed as an indication that it is not necessarily the INN whose curve is deviating from the true entropy, but rather the data obtained from the 1D grid-method described in this section. The difference between INN and grid-based result is shown in Fig. 5.

Phase Space Evolution

The phase space evolution of the example discussed in Fig. 3 of the main text is governed by the Hamiltonian H , which we choose to be

$$H = \sum_i \frac{1}{2} (m\omega^2 x_i^2 + p_i^2/m) + k \sum_i (x_i - x_{(i+1)\%N})^2, \quad (33)$$

such that k gives the strength of the coupling between oscillators. The resulting phase space flow is given by

$$\begin{aligned} \dot{x}_i &= \partial_{p_i} H, \\ \dot{p}_i &= -\partial_{x_i} H. \end{aligned} \quad (34)$$

If one considers damping in phase space, the following Fokker-Planck equation is obtained [60]

$$\begin{aligned} \partial_t \rho(t, \mathbf{x}, \mathbf{p}) &= [-\partial_{\mathbf{p}} H \cdot \partial_{\mathbf{x}} + \partial_{\mathbf{x}} H \cdot \partial_{\mathbf{p}} + \\ &\quad \gamma (\mathbf{p} \cdot \partial_{\mathbf{p}} + mk_B \sum_i T_i \partial_{p_i}^2)] \rho(t, \mathbf{x}, \mathbf{p}). \end{aligned} \quad (35)$$

OPTIMIZING DESIGN CHOICES FOR NEURAL QUANTUM STATES

7.1 CONTRIBUTIONS



This chapter presents the publication in [C]. The proposal of the project stemmed from joint discussions between Martin Gärttner, Markus Schmitt, and me. I developed the code and ran all simulations, with Martin Gärttner and Markus Schmitt acting as supervisors. All authors contributed to the writing of the text. A GPU- and cluster-ready implementation of the utilized algorithms can be found as part of the published jVMC library [G].

7.2 MOTIVATION

While various different architectures have been proposed for neural quantum state applications (see [Section 3.2.1](#), [Section 3.2.2](#), [Section 3.2.3](#)), there is little insight into which architectures work best in what circumstances. We here set out to test the capabilities of the most popular neural quantum state architectures when trying to find ground states of stoquastic Hamiltonians as well as frustrated spin systems that feature a non-trivial sign structure. We choose ground state searches as a benchmark task as it is a problem of high relevance, while simultaneously featuring a clear performance metric, namely the lowest achieved energy.

© 2024 American Physical Society

Optimizing design choices for neural quantum states

Moritz Reh ^{1,*}, Markus Schmitt,² and Martin Gärtner ^{1,3,4}

¹*Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*

²*Forschungszentrum Jülich GmbH, Peter Grünberg Institute, Quantum Control (PGI-8), 52425 Jülich, Germany*

³*Physikalisches Institut, Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany*

⁴*Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*



(Received 18 January 2023; revised 30 March 2023; accepted 1 May 2023; published 9 May 2023)

Neural quantum states are a new family of variational *Ansätze* for quantum-many body wave functions with advantageous properties in the notoriously challenging case of two spatial dimensions. Since their introduction, a wide variety of different network architectures have been employed to study paradigmatic models in quantum many-body physics with a particular focus on quantum spin models. Nonetheless, many questions remain about the effect that the choice of architecture has on the performance on a given task. In this work, we present a unified comparison of a selection of popular network architectures and symmetrization schemes employed for ground-state searches of prototypical spin Hamiltonians, namely, the two-dimensional transverse-field Ising model and the J_1 - J_2 model. In the presence of a nontrivial sign structure of the ground states, we find that the details of symmetrization crucially influence the performance. We describe this effect in detail and discuss its consequences, especially for autoregressive models, as their direct sampling procedure is not compatible with the symmetrization procedure that we found to be optimal.

DOI: [10.1103/PhysRevB.107.195115](https://doi.org/10.1103/PhysRevB.107.195115)

I. INTRODUCTION

Many intriguing phenomena in condensed matter physics emerge in strongly interacting many-body systems, which rarely allow for analytical solutions. Theoretical attempts to understand a problem of interest therefore usually rely on numerical techniques. However, the exponential scaling of required computational resources with the number of system constituents implies that true many-body settings involving hundreds of qubits or more are not amenable to naive classical computational approaches [1].

Therefore, variational approaches have become popular [2–4] aiming to efficiently parametrize the physically relevant “corner” of Hilbert space [5–7]. Such approaches reduce the number of required parameters to be subexponential, circumventing the curse of dimensionality, at the expense of generality. Various methods have been developed that fall into this category. Tensor network states (TNS) [2,3] and neural quantum states (NQS) [4] are particularly popular and important in the case of many-body spin systems because both are versatile and numerically exact in the sense that the accuracy can be systematically controlled with self-consistent convergence checks. Tensor network approaches comprise many different *Ansätze* [8–15], which are generally composed of a number of tensors with a specific bond dimension, that regulates their expressivity. Matrix product states (MPS) [2] form the best studied example of TNS, and allow to encode weakly entangled states in one dimension with great success, especially in conjunction with the celebrated density ma-

trix renormalization group (DMRG) algorithm [16,17]. MPS constitute an ideal tool to efficiently represent ground states of one-dimensional gapped Hamiltonians, which feature an area law of entanglement [6]. Their limitation is set by the exponential growth of the required bond dimension with entanglement entropy and more involved TNS were devised to mitigate this issue [18]. While instances of TNS exist which deal with two-dimensional (2D) settings [3,13,18], this regime remains particularly challenging.

Neural quantum states therefore received a lot of attention upon their introduction as they presented a potential remedy to some of the aforementioned shortcomings of TNS, due to several factors. For one, opposed to the DMRG algorithm, the design of NQS and its associated optimization algorithms do not rely on any spatial structure, rendering them ideal candidates for 2D settings [4,19]. Second, NQS allow for unprecedented flexibility; in principle, any network that maps a spin configuration, i.e., a computational basis state, to an associated complex wave-function coefficient presents a valid *Ansatz* [19–25]. Finally, it has been demonstrated that NQS can indeed encode volume-law entangled states without introducing exponential cost [26–29] in stark contrast to TNS approaches. This observation serves as a strong motivation for the further exploration of NQS and its yet undetermined limitations.

In this work we aim to shed light on the latter question, focusing mainly on the differences introduced by different network architectures and symmetrization schemes. To this end, we optimize various networks that have been proposed for the use as *Ansatz* wave functions to represent the ground state of prototypical 2D spin Hamiltonians. We focus on intermediate network sizes, which allows us to comparatively

*moritz.reh@kip.uni-heidelberg.de

study intrinsic network biases. In order to connect to the existing literature, we choose the 12×12 transverse-field Ising model (TFIM) [20,21] and J_1 - J_2 model [30–32] at lattice size 6×6 and 10×10 .

The paper is organized as follows: At first we give an introduction to variational Monte Carlo (VMC) and NQS, by introducing the employed optimization strategy (Sec. II), the different architectures we benchmarked (Sec. III), as well as the different symmetrization options (Sec. IV). We then present the results on the 2D TFIM as well as the 2D J_1 - J_2 model, where we put an emphasis on the interplay of the symmetrization strategy and the learned sign structure (Sec. V) and discuss implications for autoregressive networks (Sec. VI).

II. VARIATIONAL MONTE CARLO

In variational Monte Carlo one parametrizes a wave function as

$$|\psi_\theta\rangle = \sum_{\mathbf{s}} \psi_\theta(\mathbf{s}) |\mathbf{s}\rangle, \quad (1)$$

where $\mathbf{s} = (s_1, \dots, s_N)$ labels the computational basis states of the system composed of N degrees of freedom and θ denotes the vector of variational parameters. Using this *Ansatz* function, one can estimate the expectation values of operators \hat{O} according to

$$\begin{aligned} \langle \hat{O} \rangle &= \langle \psi_\theta | \hat{O} | \psi_\theta \rangle = \sum_{\mathbf{s}\mathbf{s}'} \psi_\theta(\mathbf{s})^* O_{\mathbf{s}\mathbf{s}'} \psi_\theta(\mathbf{s}') \\ &= \sum_{\mathbf{s}\mathbf{s}'} |\psi_\theta(\mathbf{s})|^2 O_{\mathbf{s}\mathbf{s}'} \frac{\psi_\theta(\mathbf{s}')}{\psi_\theta(\mathbf{s})} = \left\langle \sum_{\mathbf{s}\mathbf{s}'} O_{\mathbf{s}\mathbf{s}'} \frac{\psi_\theta(\mathbf{s}')}{\psi_\theta(\mathbf{s})} \right\rangle, \end{aligned} \quad (2)$$

where $\langle \cdot \rangle$ denotes the Monte Carlo (MC) average with respect to the probability distribution $|\psi_\theta(\mathbf{s})|^2$ from here on. The local estimator $O_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} O_{\mathbf{s}\mathbf{s}'} \psi_\theta(\mathbf{s}') / \psi_\theta(\mathbf{s})$ can be evaluated efficiently for local operators \hat{O} , meaning that the sample size required to reach a certain precision does not scale with the total system size. This renders the estimation of $\langle \hat{O} \rangle$ efficient [33]. For variationally approximating ground states, we can take \hat{O} to be the system's Hamiltonian \hat{H} , compute its expectation value, and, since we assumed differentiability with respect to θ , optimize by gradient descent. A more elaborate approach uses information about the local curvature of the variational manifold, as measured by the Fubini-Study metric, in form of the quantum geometric tensor $S = \langle \langle \Gamma_k^*(\mathbf{s}) \Gamma_{k'}(\mathbf{s}) \rangle \rangle$, where $\Gamma_k(\mathbf{s}) = \partial_{\theta_k} \ln \psi_\theta(\mathbf{s})$ and $\langle \langle AB \rangle \rangle = \langle AB \rangle - \langle A \rangle \langle B \rangle$. This results in the commonly used stochastic reconfiguration (SR) algorithm [34], in which the parameter update rule is given by

$$\theta_k^{(n+1)} = \theta_k^{(n)} - \tau \sum_{k'} \text{Re}(S^{-1})_{kk'} F_{k'} |_{\theta=\theta^{(n)}}. \quad (3)$$

Here, $F_k = \partial_{\theta_k} \langle E_{\text{loc}}(\mathbf{s}) \rangle = 2 \text{Re}(\langle \langle \Gamma_k^*(\mathbf{s}) E_{\text{loc}}(\mathbf{s}) \rangle \rangle)$ and τ is the update step size. As S can be rank deficient, regularization techniques need to be applied upon computing the inverse [4,19]. In this work, we achieve regularization by scaling all diagonal entries of S by a factor of $1 + \delta_1$ and adding an identity matrix scaled with δ_2 to S . During optimization a second-order Runge-Kutta integrator with adaptive (imaginary-) time step τ is employed to obtain the integrated

evolution of the network parameters. Further details are given in Appendix A.

III. NEURAL QUANTUM STATES

Neural quantum states form a particular class of functions that can be used as an *Ansatz*, ψ_θ , in the VMC framework described in Sec. II. In this case, the *Ansatz* function is given by an artificial neural network (ANN), which defines a nonlinear differentiable map from spin configurations \mathbf{s} to the associated (generally complex) wave-function coefficient $\psi_\theta(\mathbf{s})$. In fact, it is common practice to have the ANN produce the logarithmic wave-function coefficient $\chi_\theta(\mathbf{s})$ such that $\psi_\theta(\mathbf{s}) \equiv \exp[\chi_\theta(\mathbf{s})]$ can accurately capture coefficients over multiple orders of magnitude. Typically, neural networks are built up in layers which iteratively transform the input to a desired output, constituting a very general class of function approximators that are particularly attractive due to their flexibility in construction and the existence of universal approximation theorems, stating that neural networks can approximate any function given sufficiently many parameters [35–38].

Various architectures have been proposed for the use in NQS applications, but a broad comparative benchmark within a unified study is so far missing. We attribute this to both the technical difficulties of implementation and the many details concerning the structure and optimization of the networks, which can have a strong effect on performance. Especially the choice of hyperparameters as well as architecture details, e.g., whether amplitude and phase are treated by separate networks or in a unified scheme, can be challenging.

We design the networks to encode both phase and amplitude simultaneously, except for the case of the TFIM where no phase is modeled since the ground state is known to be positive [39]. When the network models both phase and amplitude, we design the feed-forward-based architectures as holomorphic maps, using complex parameters and outputting a single complex number $\chi_\theta(\mathbf{s})$. The recurrent architectures, in contrast, do not define holomorphic maps and therefore utilize real parameters. For detailed explanations on the utilized networks see Appendix B.

Within this work we consider the following network architectures:

Restricted Boltzmann machines. The earliest NQS architectures relied on a dense single-layer feed-forward network, usually referred to as restricted Boltzmann machine (RBM) in the NQS context [4,27,30,40–43]. Due to the dense connectivity, the spatial structure of the input is not natively represented by the network architecture. To add a notion of locality to the RBM *Ansatz* one can include the product of physically coupled spins as input features to obtain a correlator RBM (CorrRBM), as proposed in [22].

Convolutional neural networks. A further generalization of RBMs are convolutional neural networks (CNN) [19,32,44,45]. The layers in deep CNNs are defined by a number of filters of a certain width, thereby allowing great flexibility in the design of the network. By choosing the depth, i.e., the number of layers to one, and the filter size such that it spans the entire system one obtains a translationally invariant RBM, meaning that symmetrized RBMs are a strict subclass of CNNs. By using smaller filters and multiple layers, an

intrinsic representation of locality in the network is restored based on a hierarchical representation of features, loosely reminiscent of tree tensor networks (TTN) [11] or the multiscale entanglement renormalization Ansatz (MERA) [12]. In analogy to the CorrRBM, adding correlations of the spin configurations to the input extends the CNN to a correlator CNN (“CorrCNN”).

Recurrent networks. A very different form of Ansatz is obtained by substituting the feed-forward-based network with a recurrent architecture, which assigns the coefficient $\psi_\theta(\mathbf{s})$ as a product $\psi_\theta(\mathbf{s}) = \psi_\theta(s_1) \cdot \psi_\theta(s_2|s_1) \cdot \dots \cdot \psi_\theta(s_N|s_{N-1} \dots s_1)$ [20,21,46–49]. A central motivation for this construction is the fact that such architectures allow for autoregressive sampling. This means that a new independent sample can be generated with a single network evaluation, such that no Markov chain Monte Carlo (MCMC) with potentially long autocorrelation or thermalization times is required. A recurrent architecture is defined by a cell containing the variational parameters. It is then iteratively scanned over the input, i.e., the spin configuration $\mathbf{s} = (s_1, \dots, s_N)$, while storing information about previous sites in a hidden state h_t that fulfills the recurrence relation $h_{t+1} = f(h_t, s_t)$. Since it is always the same cell that is scanned over the input sequence, the number of variational parameters does not grow with the system size; this can enhance efficiency, but, on the other hand, the inherently sequential design does not allow for parallelization. The architecture of the recurrent cell is highly customizable; we refer to the vanilla recurrent neural network with a single layer as “RNN,” and use the abbreviations “LSTM” and “GRU” for the long short-term memory [50] and gated recurrent unit [51], respectively.

IV. SYMMETRIES

When the Hamiltonian describing the spin system exhibits certain symmetries, such as invariance under translations, rotations, reflections, or parity symmetry, its ground state exhibits the same symmetries. This *a priori* knowledge of properties that the ground-state wave function must fulfill can be used to restrict the variational optimization to the correct symmetry sector and thereby to enhance the performance of the algorithm. Since not all of the considered network architectures can be composed of equivariant layers to incorporate symmetry [52], we instead follow the common approach to symmetrize the wave function by averaging procedures of all symmetry-equivalent configurations as is commonly done [19,20,30]. The symmetrized coefficient $\psi_\theta^S(\mathbf{s})$ will then read as $\psi_\theta^S(\mathbf{s}) = \text{Avg}\{\psi_\theta(\sigma(\mathbf{s})) | \sigma \in \mathcal{S}\}$ with \mathcal{S} the set of all symmetry operations and Avg denoting one of the three symmetrization methods listed below. In fact, we will show that the details of this symmetrization procedure are crucial for the performance of the algorithm and that not all architectures are amenable to the symmetrization procedure we found to be optimal. We differentiate between three different procedures for defining symmetrized wave-function coefficients $\psi_\theta^S(\mathbf{s})$ from the network outputs $\chi_\theta(\mathbf{s})$.

(i) Bare-symmetry:

$$\psi_\theta^S(\mathbf{s}) = \exp\left(\frac{1}{|\mathcal{S}|} \sum_{s' \in \mathcal{S}(\mathbf{s})} \chi_\theta(s')\right). \quad (4)$$

(ii) Exp-symmetry:

$$\psi_\theta^S(\mathbf{s}) = \frac{1}{|\mathcal{S}|} \sum_{s' \in \mathcal{S}(\mathbf{s})} \exp(\chi_\theta(s')). \quad (5)$$

(iii) Sep-symmetry:

$$\psi_\theta^S(\mathbf{s}) = \sqrt{\frac{1}{|\mathcal{S}|} \sum_{s' \in \mathcal{S}(\mathbf{s})} \exp(2 \text{Re}[\chi_\theta(s')])} \times \exp\left[i \arg\left(\sum_{s' \in \mathcal{S}(\mathbf{s})} \exp(i \text{Im}[\chi_\theta(s')])\right)\right]. \quad (6)$$

Equation (5) is a natural choice when the ANN output is considered to be the logarithmic wave-function coefficient. Another option to proceed is to exponentiate the logarithmic coefficients prior to averaging, as done in Eq. (5), resulting in a symmetrization procedure with the potential that phases may positively or negatively interfere, as opposed to Eqs. (4) and (6). Finally, the symmetrization procedure in Eq. (6) is designed to be compatible with an autoregressive property, which is lost when choosing one of the other options. The reason for this is that the relation

$$\sum_{s' \in \mathcal{S}(\mathbf{s})} |\psi_\theta^S(s')|^2 = \sum_{s' \in \mathcal{S}(\mathbf{s})} |\psi_\theta(s')|^2 \quad (7)$$

has to be fulfilled for direct sampling. The only way to generate a new configuration from the distribution encoded in an autoregressive network is to sequentially sample the local configurations s_i from the conditional probabilities $|\psi_\theta(s_i|s_{i-1} \dots s_1)|^2$. Following this procedure, symmetry-equivalent configurations will in general not be generated with identical probability. However, if Eq. (7) holds, the frequency of samples from one equivalence class matches its probability given by the symmetrized Ansatz $\psi_\theta^S(\mathbf{s})$, which is sufficient for a representative set of samples.

All presented symmetrization options have been applied in previous works: The bare option was, for example, used in [19], the exponential option in [30], and the separate option in [20]. Our contribution is a direct comparison between these options, applied to the same physical problem with otherwise identical network architectures.

V. RESULTS

In the following we will examine the performance of different networks on the TFIM and J_1 - J_2 model on a two-dimensional square lattice. Since we want to learn about the distinct representational power of the networks, they are designed to have approximately the same number of parameters (see Table II in Appendix C). Note that this can imply different computational costs for the different networks, as for example the evaluation of autoregressive architectures is more demanding than for a feed-forward architecture. We trained the networks until convergence was reached, irrespective of computational cost. Details of the network architectures and obtained energies can be found in Appendix B.

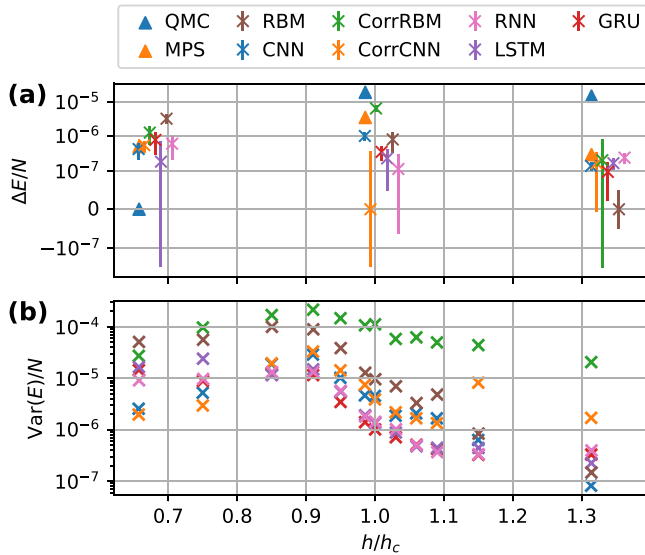


FIG. 1. Performance of the tested network architectures for finding the ground state of the 2D 12×12 TFIM at different magnetic field strengths. Here all networks use the bare symmetrization as no phase is modeled. In (a) the deviation from the lowest observed energy density is shown for the magnetic fields $h \in [2, 3, 4]$. The data points are artificially shifted horizontally such that differences are more easily visible. Error bars quantify statistical fluctuation of the estimator. In (b) the energy variance per spin is plotted, which serves as an additional performance indicator. The QMC and MPS data were taken from [20,21].

A. Transverse-field Ising model

The transverse-field Ising model

$$\hat{H} = -J \sum_{\langle ij \rangle} \hat{\sigma}_z^i \hat{\sigma}_z^j - h \sum_i \hat{\sigma}_x^i \quad (8)$$

on a 2D square lattice features an Ising interaction between nearest neighbors and an external field in the x direction, giving rise to a quantum phase transition from a ferromagnetic to a paramagnetic phase at $h = h_c \approx 3.044J$ [53]. The stoquastic nature of the Hamiltonian [39] make it an ideal benchmark case for various approximative methods, such as quantum Monte Carlo (QMC) and MPS. We compare to both techniques with data taken from [20,21].

The stoquasticity implies a positive ground-state wave function, meaning that the NQS networks do not need to encode a complex phase. Here, we limit ourselves to the bare symmetry option. We employ open boundary conditions, to allow for comparisons with [20,21].

We present our results in Fig. 1. Figure 1(a) shows the deviation of the predicted ground-state energy density from the lowest value attained by the various networks and QMC and MPS at each field strength $h \in [2, 3, 4]$. In order to ease the legibility we shifted the field strengths slightly. In addition, since no exact numerical benchmark value for the energy itself can be given, we consider the variance of the energy estimate as a performance indicator, that vanishes when the network represents an eigenstate of the system, such as the ground state. Figure 1(b) shows the energy variance as a function of

field strength for the various network architectures. The first observation is that for all architectures, performance depends significantly on the strength of the magnetic field. In the region of the critical value performance diminishes as one might expect, as correlations grow longer ranged. The worst performance is, however, found at $h = 0.91h_c$ slightly away from the critical point, which might be an indicator of finite-size effects being present. For the differences between the networks it is more difficult to make out clear trends. While the RBM and correlator RBM architecture seem to give the worst performance, CNN and correlator CNN perform better, which indicates that increasing the depth of the network at the expense of smaller filter sizes is beneficial, especially in the ferromagnetic phase. Autoregressive nets perform reasonably well for all values of h , especially in the paramagnetic regime where they outperform the other architectures. Interestingly, the details of the recurrent cell do not seem to have an influence in this scenario as RNN, LSTM, and GRU show almost identical performance. Here, we also want to note that the employed stochastic reconfiguration method in conjunction with RNNs allows to obtain lower energies compared to their original introduction in [20], that utilized larger networks while giving up on second-order accurate gradients.

We conclude that the ground state of the stoquastic TFIM can be approximated well, without major differences among network architectures. In the following, we therefore shift our attention to the 2D J_1 - J_2 model, i.e., a nonstoquastic Hamiltonian with a nontrivial sign structure. At the point $J_2 = 0.5J_1$ the ground state is maximally frustrated, giving rise to exotic phases such as a spin liquid and valence bond solid state [30], which are characterized by high entanglement and the absence of an energy gap.

B. J_1 - J_2 model

The J_1 - J_2 model features competing antiferromagnetic Heisenberg couplings between nearest- and next-nearest neighbors, with respective strengths J_1 and J_2 , such that

$$\hat{H} = J_1 \sum_{\langle ij \rangle} \hat{\sigma}_i \cdot \hat{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\sigma}_i \cdot \hat{\sigma}_j. \quad (9)$$

It has been the focus of numerous works, especially in the context of neural quantum states [30,32,54–56], due to its challenging nature that is characteristic for 2D frustrated ground states. Depending on the relative strength of interactions, one finds the system in the Néel ($J_2 \lesssim 0.49J_1$) or striped ($J_2 \gtrsim 0.61J_1$) phase, while a spin liquid and valence bond solid phase exists between the two [54]. All of our experiments are carried out at the point $J_2/J_1 = 0.5$, where frustration is maximal and we employ periodic boundaries. The abundance of symmetries reduces the size of the relevant Hilbert space sector, allowing to exactly compute the ground-state energy for a lattice of size 6×6 , which serves as a reference value.

Due to the nonstoquastic nature of the Hamiltonian, the ground state is not positive and therefore a major challenge consists in finding the ground-state sign structure [31,57,58]. At $J_1 = 0$ or $J_2 = 0$, the phase follows the Marshall-Peierls sign rule, which states that the sign of each coefficient $\psi(\mathbf{s})$ is $(-1)^{N_{\uparrow \in A(\mathbf{s})}}$, where A denotes one of the two sublattices

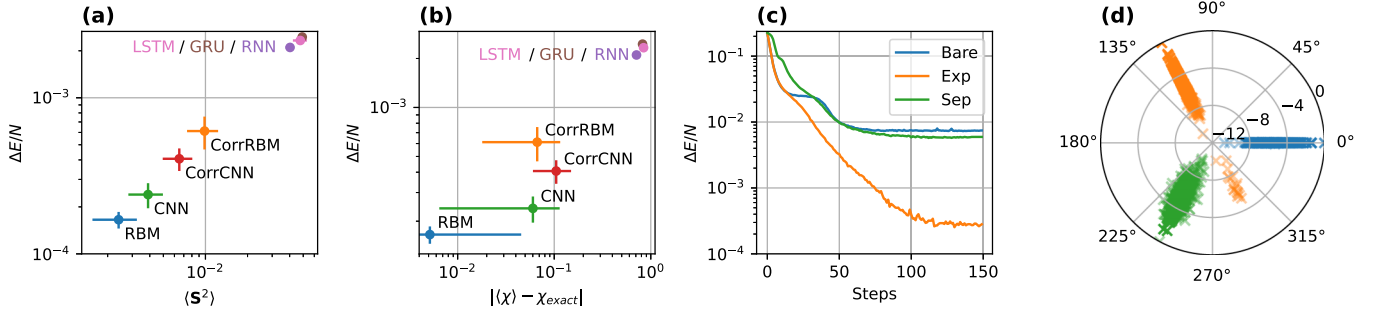


FIG. 2. Results on the 2D 6×6 J_1 - J_2 model. In (a) and (b) the performance of the different architectures is depicted as a function of observables of interest. The error bars which are smaller than the filled circles are omitted. In (c), the effect of the different symmetrization procedures is shown for the case of the RBM architecture, leaving all other specifications of the network unchanged. In (d), sampled wave-function coefficients of the network are depicted in the complex plane for the symmetrizations shown in (c), with a logarithmic scale for the absolute value. Only in the exponential case were we able to observe a nontrivial sign structure with deviations from the Marshall sign rule, in which there were samples on the opposite site of the bulk.

and $N_{\uparrow \in A}(\mathbf{s})$ is the number of up spins in configuration \mathbf{s} on sublattice A . When going to finite ratios J_2/J_1 , the Marshall sign rule does no longer hold exactly, but presents a good approximation. Hence, we choose to hard code the sign rule in form of a transformed Hamiltonian, which is also explained in Appendix D, and let the network learn deviations from it. We therefore change the design of the feed-forward-based architectures to holomorphic maps with complex network parameters, while the autoregressive architectures stay non-holomorphic maps using real parameters but are allowed to model a phase.

Additionally, the ground state is known to lie in the zero magnetization sector, implying that the wave-function coefficients are only nonzero for configurations \mathbf{s} with the same number of up and down spins. We exploit this by only constructing sample configurations that fulfill this condition.

As was demonstrated in [30], supplying the network with physical information in form of symmetries improves performance dramatically. We therefore symmetrize our *Ansatz* function with all four present symmetries, i.e., translations, the C_4 point group, and the \mathbb{Z}_2 spin-flip symmetry. As described previously, different options regarding the details of symmetrization exist, and the data presented below reveal that these details have significant effect on the performance of the algorithm.

Let us, however, first examine the performance differences between the different network architectures, shown in Figs. 2(a) and 2(b). We show the deviation of the energy density from the exact ground-state value as a function of the total magnetization

$$\hat{S}^2 = \sum_{ij} \hat{\sigma}_i \cdot \hat{\sigma}_j \quad (10)$$

and as a function of the structure factor

$$\hat{\chi} = \sum_{ij} \hat{\sigma}_i \cdot \hat{\sigma}_j e^{i\mathbf{q} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \quad (11)$$

for $\mathbf{q} = (\pi, \pi)$, where the sum runs over all pairs of lattice sites. The choice of these observables is physically motivated: The Hamiltonian's $SU(2)$ symmetry implies that the ground state fulfills $\langle \hat{S}^2 \rangle = 0$, and thus the deviation from zero serves

as another figure of merit for the approximation quality. The structure factor is a crucial observable providing insight into the magnetization structure, i.e., whether the striped or Néel phase is observed. The observables are estimated using 10^4 samples. The error bars shown in the figure are obtained by estimating the observables based on 10 independently drawn sample sets and determining the fluctuations between them. The feed-forward architectures, employing the exponential symmetrization strategy, perform significantly better compared to the autoregressive networks using the separate (sep-)symmetrization strategy. In Fig. 2(c) we compare the effect of the three different strategies for the RBM. The exponential option, utilized in [30], performs best by far. Other strategies, that were, for example, used in [19,20], perform considerably worse for the given task.

In a next step, we provide an analysis revealing the origin of the observed performance differences. Concretely, we examine the capability of the different symmetrization methods to capture deviations from the Marshall sign rule. In Fig. 2(d) we show argument and amplitude of the complex wave-function coefficients that are obtained by sampling the NQS; notice that all coefficients from one wave function can be rotated by an arbitrary angle in the polar plot, corresponding to an irrelevant global phase. Importantly, deviations from the Marshall sign rule are only observed in the case of exp-symmetrization, while the other methods only optimize the wave-function amplitudes, thus not being able to reveal the nontrivial physics in this problem. This explains the inferior performance of the alternative symmetrization schemes observed in Fig. 2(c). This, of course, does not mean that we cannot model such behavior with the other strategies: it can, however, hint at the fact that finding such sets of variational parameters using the stochastic reconfiguration algorithm is challenging.

We now discuss the implications of this finding. In autoregressive architectures, samples are drawn from the unsymmetrized distribution since we cannot average over all symmetry-invariant configurations while sampling. One therefore needs to ensure that the symmetrized probability measure of a configuration is equal to the nonsymmetrized one, ruling out the exponential symmetrization strategy in

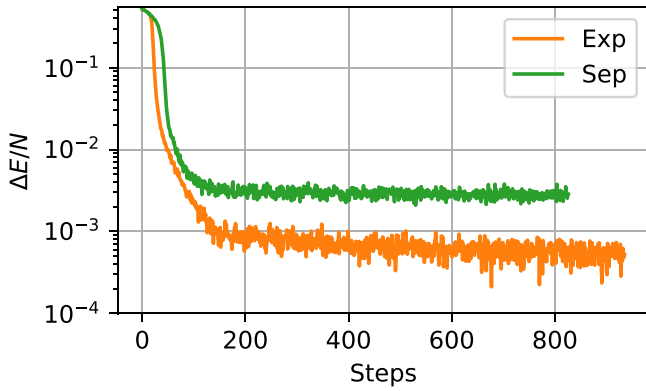


FIG. 3. Learning curve of an RNN with the sep-symmetry and exact samples (green) compared to the same network using the exponential symmetrization strategy with MCMC samples (orange). As is clearly visible, the network profits greatly from the exponential symmetrization strategy which is, however, incompatible with exact sampling, thereby taking away the major benefit of autoregressive architectures.

conjunction with autoregressive samples, since it violates Eq. (7) and similarly for the bare symmetrization. To verify the influence of the chosen symmetrization on the performance also for recurrent architectures, we compare the performance of a sep-symmetrized RNN optimized using autoregressive sampling to its exp-symmetrized counterpart with MCMC sampling in Fig. 3. The performance difference again reveals that the exponential symmetrization procedure significantly improves the quality of the results, although it prohibits autoregressive samples. Therefore, the limited flexibility in choosing the symmetrization scheme constitutes a significant drawback for autoregressive architectures when trying to learn the nontrivial sign structures of the frustrated ground state.

Having found clear indication that the details of symmetrization influence the results, we now give an intuitive understanding as to why this may be the case. One immediate observation is the fact that only in the exponential case the phase can influence the final amplitude since the coefficients may positively or negatively interfere, yielding a potential gain in expressivity. Similarly, if the prediction of the phase were random (or noisy) for a given spin configuration, the sum of these noisy coefficients would be small in absolute value. Thereby, it seems plausible that in the case of bad generalization of the ANN, the resulting coefficients are automatically suppressed. The interplay between complex phases and amplitudes can also affect the optimization: The logarithmic derivatives $\Gamma_k(\mathbf{s}) = \partial_{\theta_k} \ln |\psi_{\theta}^S| + \partial_{\theta_k} \arg[\psi_{\theta}^S]$ of the symmetrized coefficients consist of the sum of the gradients of the (logarithm) amplitude and the complex phase. If only the amplitudes of the nonsymmetrized coefficients $\psi_{\theta}(s)$ contribute to the symmetrized amplitude, as is the case for the bare- and sep-symmetrizations, its gradients with respect to the complex phases of $\psi_{\theta}(s)$ vanish. This is in contrast to the exp-symmetry case, where the gradients of the symmetric amplitude with respect to the nonsymmetric phases generally take finite values since they influence the symmetrized amplitude.

TABLE I. Ground-state energies achieved with neural network quantum states on the 10×10 J_1 - J_2 model with periodic boundary conditions at the point $J_2/J_1 = 0.5$. References [56] and [54] incorporate projection procedures, thereby providing physical information beyond symmetries which are not present in the other approaches.

Architecture	Energy	Reference
CNN	-0.473591	[55]
CNN	-0.49516(1)	[32]
RBM + GP	-0.49575(2)	[56]
CNN	-0.49586(4)	This work
RBM + PP	-0.497629(1)	[54]

C. Large system sizes

Having gained insight into network specific advantages and disadvantages, we finally want to answer the question which network architectures can be deemed suitable to scale up to larger system sizes. To consider an architecture scalable we demand that both the number of its parameters and the network evaluation cost should at most grow mildly with system size N . The RBMs and its variants violate the first requirement as the number of parameters grows quadratically given a constant ratio α between the number of hidden and visible neurons. For recurrent networks, one finds a quadratic scaling of the evaluation cost, as the cost of evaluating a single (unsymmetrized) configuration grows linear with system size since the cell needs to be scanned over longer inputs, while symmetrization gives another factor N .

This leaves CNNs as the only viable option, as they allow for subquadratic growth of parameters using small filter sizes in deep architectures, still allowing for system-wide correlations, and evaluations that approximately scale linearly in system size since the filters need to be scanned over larger inputs. Since we saw in Fig. 2 that adding correlation input features does not add to performance, we test the CNN on a 10×10 lattice and compare to previous works in Table I. All listed energies stem from NQS approaches, where those that used physically informed input in addition to the NQS are marked with a “+.” The additional input is using variational parameters that define the weight of the projection of the basis state onto a physically motivated state, such as a pair state in the pair-projection (PP) algorithm or a Gutzwiller-projected (GP) state. The performance of the CNN presented here is comparable to [56], without using any physically informed input.

VI. CONCLUSION AND OUTLOOK

In this work, we systematically compared the performance of different network architectures and symmetrization strategies when trying to find ground states of many-body spin Hamiltonians. While there are no clear trends visible in the case of the 2D TFIM model, the disparities in the J_1 - J_2 model are pronounced, where the feed-forward architectures clearly outperform the autoregressive architectures. We identified the details of the symmetrization strategy as a key to higher performance, which hinders learning the correct sign structure if not done correctly. Specifically, it is imperative to add the

wave-function coefficients of all symmetry-invariant configurations such that they may interfere positively or negatively as laid out in Eq. (5). It is thus not sufficient to restrict the search space to the right symmetry sector; the details of the symmetrization procedure play a crucial role.

This finding has direct implications for autoregressive architectures since their sampling strategy is not amenable to the exponential symmetrization strategy described in Eq. (5), as it violates Eq. (7). We make this explicit in Fig. 3, by showing that an autoregressive model using MCMC samples outperforms the same model with exact samples if the symmetrization strategy is such that interference among the coefficients is possible. Therefore, learning quantum states which feature a nontrivial sign structure appears at least challenging with autoregressive models, if symmetrization is required at the same time. However, it is to be noted that more elaborate optimization strategies, such as [59,60], as well as slightly altered design choices [61] may potentially yield improved results. We conclude that the RNN's performance strongly depends on the physical model in question and add that there may of course be physical models in which an autoregressive network outperforms its competitors.

We conjecture that further progress will rely on symmetrical wave-function *Ansätze* that support good scaling characteristics to larger system sizes. The symmetry-invariant map should add up the generated coefficients using the exp-symmetry operation allowing for readily learning a nontrivial sign structure as opposed to the other cases. We have found the CNN architecture to be suitable for this purpose and demonstrated comparable performance to previous works, which amended NQS wave functions with additional physical bias.

Note added. Recently, Ref. [62] appeared, which also studies the J_1 - J_2 model on the square lattice using group convolutional neural networks (GCNNs). The findings of that work support our claim that the exponential operation prior to summation is crucial for optimizing performance.

The code developed for this project relies on the jVMC library [63], that can be found on [64]. Data are available upon request.

ACKNOWLEDGMENTS

The authors would like to thank O. Sharir, M. Hibat-Allah, and F. Ferrari for providing data from their respective works. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy Grant No. EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster) and within the Collaborative Research Center SFB1225 (ISOQUANT). This work was partially financed by the Baden-Württemberg Stiftung gGmbH. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through Grant No. INST 40/575-1 FUGG (JUSTUS 2 cluster). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. [65] for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [66] at Jülich Supercomputing Centre (JSC).

APPENDIX A: REGULARIZATION, INVERSION, AND IMAGINARY-TIME EVOLUTION

As laid out in the main text, the stochastic reconfiguration algorithm culminates in the parameter evolution equation

$$\sum_{k'} \text{Re}(S^{-1})_{kk'} \dot{\theta}_{k'} = F_k. \quad (\text{A1})$$

Since S can be singular, rendering its inverse ill defined, one needs to regularize S in order to solve for $\dot{\theta}$. We achieve this in two steps, first multiplying all diagonal elements with a factor $1 + \delta_1$ and then adding δ_2 to all diagonal entries. We choose δ_1 on the order of 10 and let it decay exponentially during optimization, while δ_2 is fixed around 10^{-4} . We then find θ using a linear solver (we used the JAX implementation of the SCIPY linear solver `jax.scipy.linalg.solve`) before integrating θ using a second-order Runge-Kutta scheme with adaptive step size [19].

APPENDIX B: NEURAL NETWORK ARCHITECTURES

Restricted Boltzmann machines (RBMs). A dense, single-layer feed-forward network with N input and M hidden neurons is, somewhat imprecisely [67], referred to as RBM in the NQS literature. Its output is the sum of the hidden layer with an activation function applied. Usually, this function is chosen to be the logarithm of the hyperbolic cosine. In the complex plane, this function has poles at $z = i\pi(1/2 + n)$ for integer n , which is why we choose to approximate the activation function with its first two nonvanishing Taylor series terms, mitigating this problem. The number of parameters for a network with weights and biases are given by

$$n_p = 2NM + 2M, \quad (\text{B1})$$

where the factor of 2 arises since all parameters are complex. In varying the number of hidden neurons M , one can control the complexity and expressivity of the network. The dense layer bears no physical motivation, as it is irrespective of locality, begging the question for more physically motivated architectures.

Convolutional neural networks (CNNs). In contrast to RBMs, CNNs allow to keep a notion of locality, by using filter sizes that are smaller than the system size. These are scanned over the input, automatically respecting translational equivariance. CNNs allow to be made more expressive by tuning the depth (D), filter size (S_F), and the number of filters and channels (N_F). Without biases, the resulting number of parameters is

$$n_p = 2 \sum_{d=1}^D S_F^d N_F^i N_F^{i-1}, \quad (\text{B2})$$

with $N_F^0 = 1$ for the input spin configuration. In the end, the output is summed over the channel dimension before exponentiating and summation over the remaining dimensions, in order to implement the exp-symmetrization option.

Correlation networks. In order to guide the network towards the important features of the input configuration, [22] proposes to add correlations of coupled spins explicitly to

TABLE II. Characteristics of the networks that were utilized throughout the paper.

Figure or Table	Name	No. of Parameters	Symmetry	Specifications	Energies
1	RBM	2592	Bare	$M = 18$	$h = 2: -2.4095998(9)$ $h = 3: -3.1739016(5)$ $h = 4: -4.12179829(5)$
1	CNN	2594	Bare	$D = 2, N_F = [24, 2], S_F = [6 \times 6, 6 \times 6]$	$h = 2: -2.4096025(2)$ $h = 3: -3.1739014(3)$ $h = 4: -4.12179815(3)$
1	CorrRBM	2645	Bare	$M = 5$	$h = 2: -2.4096017(7)$ $h = 3: -3.173896(1)$ $h = 4: -4.1217980(6)$
1	CorrCNN	2483	Bare	$D = 2, N_F = [20, 3], S_F = [7 \times 7, 5 \times 5]$	$h = 2: -2.4096024(2)$ $h = 3: -3.1739024(4)$ $h = 4: -4.1217981(2)$
1	RNN	2411	Sep	$n_h = 33$	$h = 2: -2.4096024(4)$ $h = 3: -3.1739023(2)$ $h = 4: -4.12179805(8)$
1	LSTM	2463	Sep	$n_h = 19$	$h = 2: -2.4096028(5)$ $h = 3: -3.1739022(2)$ $h = 4: -4.12179812(6)$
1	GRU	2490	Sep	$n_h = 21$	$h = 2: -2.4096024(4)$ $h = 3: -3.1739023(2)$ $h = 4: -4.12179805(8)$
2(a) + 2(b)	RBM	3996	Exp	$M = 54$	$-0.50364(2)$
2(a) + 2(b)	CNN	3962	Exp	$D = 2, N_F = [60, 1], S_F = [4 \times 4, 4 \times 4]$	$-0.50357(4)$
2(a) + 2(b)	CorrRBM	4060	Exp	$M = 14$	$-0.5032(1)$
2(a) + 2(b)	CorrCNN	3872	Exp	$D = 2, N_F = [15, 1], S_F = [8 \times 8, 8 \times 8]$	$-0.50340(7)$
2(a) + 2(b)	RNN	4138	Sep	$n_h = 44$	$-0.5017(1)$
2(a) + 2(b)	LSTM	4010	Sep	$n_h = 24$	$-0.5014(1)$
2(a) + 2(b)	GRU	4160	Sep	$n_h = 27$	$-0.5015(1)$
2(c) + 2(d)	RBM	3996	Bare	$M = 54$	$-0.4965(1)$
2(c) + 2(d)	RBM	3996	Exp	$M = 54$	$-0.50355(3)$
2(c) + 2(d)	RBM	3996	Sep	$M = 54$	$-0.4981(1)$
3	RNN	3442	Exp	$n_h = 40$	$-0.5036(2)$
3	RNN	3442	Sep	$n_h = 40$	$-0.5017(3)$
I	CNN	10952	Exp	$D = 2, N_F = [75, 1], S_F = [6 \times 6, 6 \times 6]$	$-0.49586(4)$

the input of the network. In order to test the proposal, we add a new input for every link between interacting spins and define its value to be the product of the interacting spins in the $\{+1, -1\}$ representation, such that features from aligned (antialigned) spins enter as $+1$ (-1). As this procedure simply alters the input configuration, it is in principle applicable to any (nonautoregressive) architecture. This is because the autoregressive sampling procedure cannot use correlations between spins that were not sampled yet. The number of parameters the correlation networks use is given by the same formulas as previously, but with N substituted by $N + N_c$, where N_c are the number of couplings between spins in the Hamiltonian.

Recurrent neural networks (RNNs). Recurrent networks scan a cell f over the input configuration \mathbf{s} , while maintaining a memory in a hidden state vector h with length n_h . This allows to compute and sample the wave function based on conditionals, $\psi_\theta(\mathbf{s}) = \psi_\theta(s_1) \cdot \psi_\theta(s_2|s_1) \cdot \dots \cdot \psi_\theta(s_N|s_{N-1} \dots s_1)$ [20,21,46–49]. The conditionals are obtained using $h_{t+1} = f(h_t, s_t)$, before computing

$\psi_\theta(s_{t+1}|s_t \dots s_1) = g(h_{t+1})$, where g is a dense layer with two outputs that we interpret as $\ln(|\psi|)$ and $\arg(\psi)$ and from which we compute both $\psi_\theta(\uparrow|s_t \dots s_1)$ and $\psi_\theta(\downarrow|s_t \dots s_1)$. Since the conditional wave functions are normalized to allow for autoregressive sampling, they do not correspond to holomorphic functions, which is why we employ real parameters for all recurrent models to predict amplitude and phase as two distinct outputs of the network. In two dimensions, the hidden states h_t and s_t are formed by concatenating the hidden states and spin configurations of the top and right (left) neighbors of the currently considered site, depending on whether the site is located on an even (odd) row [20].

Different architectures have been proposed for the cell f . In the easiest case, a “vanilla” RNN cell is used (referred to as RNN in the main text), which is defined as a single affine map between inputs and the output with an additional activation function. More complicated cells, such as the LSTM (long short-term memory [50]) or GRU (gated recurrent unit [51]) cell use additional “gates” that allow to keep or discard information from previous sites. As the number of parameters

in these architectures is not easily given, we omit a formula for its calculation at this point and refer the reader to [50,51] for further details.

APPENDIX C: NETWORK SPECIFICATIONS

We summarize the details of the networks, in particular network sizes and symmetries, for all networks that were used in the paper in Table II.

APPENDIX D: MARSHALL SIGN RULE

The Marshall sign rule gives the exact sign for each coefficient $\psi(\mathbf{s})$ in the J_1 - J_2 model for $J_1 = 0$ or $J_2 = 0$. As defined in the main text, it assigns the sign $(-1)^{N_{\uparrow \in A}(\mathbf{s})}$, where A

denotes one of the two sublattices and $N_{\uparrow \in A}(\mathbf{s})$ is the number of up spins in configuration \mathbf{s} on sublattice A . It can equivalently be understood as a gauge transformation of the Hamiltonian, which then reads as

$$H' = J_1 \sum_{\langle ij \rangle} (-\hat{\sigma}_x^i \hat{\sigma}_x^j - \hat{\sigma}_y^i \hat{\sigma}_y^j + \hat{\sigma}_z^i \hat{\sigma}_z^j) + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\sigma}_i \cdot \hat{\sigma}_j. \quad (\text{D1})$$

Away from the points $J_1 = 0$ and $J_2 = 0$ the sign rule does no longer hold exactly but presents a good approximation, which we use as a starting point from which we aim to learn the correct deviations. Recovering the physics of the original Hamiltonian from H' is achieved by flipping all x and y correlators with support on both sublattices.

-
- [1] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [2] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys. (Amsterdam)* **326**, 96 (2011).
- [3] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Ann. Phys.* **349**, 117 (2014).
- [4] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [5] M. B. Hastings and T. Koma, Spectral gap and exponential decay of correlations, *Commun. Math. Phys.* **265**, 781 (2006).
- [6] M. B. Hastings, An area law for one-dimensional quantum systems, *J. Stat. Mech. Theory Exp.* (2007) P08024.
- [7] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [8] M. Zwolak and G. Vidal, Mixed-State Dynamics in One-Dimensional Quantum Lattice Systems: A Time-Dependent Superoperator Renormalization Algorithm, *Phys. Rev. Lett.* **93**, 207205 (2004).
- [9] G. Evenbly and G. Vidal, Class of Highly Entangled Many-Body States that can be Efficiently Simulated, *Phys. Rev. Lett.* **112**, 240502 (2014).
- [10] G. Vidal, Entanglement Renormalization, *Phys. Rev. Lett.* **99**, 220405 (2007).
- [11] Y.-Y. Shi, L.-M. Duan, and G. Vidal, Classical simulation of quantum many-body systems with a tree tensor network, *Phys. Rev. A* **74**, 022320 (2006).
- [12] L. Cincio, J. Dziarmaga, and M. M. Rams, Multiscale Entanglement Renormalization Ansatz in Two Dimensions: Quantum Ising Model, *Phys. Rev. Lett.* **100**, 240603 (2008).
- [13] F. Verstraete and J. I. Cirac, Renormalization algorithms for quantum-many body systems in two and higher dimensions, [arXiv:cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066).
- [14] A. Klümper, A. Schadschneider, and J. Zittartz, Matrix product ground states for one-dimensional spin-1 quantum antiferromagnets, *Europhys. Lett.* **24**, 293 (1993).
- [15] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, *Commun. Math. Phys.* **144**, 443 (1992).
- [16] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [17] U. Schollwöck, The density-matrix renormalization group, *Rev. Mod. Phys.* **77**, 259 (2005).
- [18] R. Orús, Tensor networks for complex quantum systems, *Nat. Rev. Phys.* **1**, 538 (2019).
- [19] M. Schmitt and M. Heyl, Quantum Many-Body Dynamics in Two Dimensions with Artificial Neural Networks, *Phys. Rev. Lett.* **125**, 100503 (2020).
- [20] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, *Phys. Rev. Res.* **2**, 023358 (2020).
- [21] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems, *Phys. Rev. Lett.* **124**, 020503 (2020).
- [22] A. Valenti, E. Greplöva, N. H. Lindner, and S. D. Huber, Correlation-enhanced neural networks as interpretable variational quantum states, *Phys. Rev. Res.* **4**, L012010 (2022).
- [23] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, Neural-network quantum states for periodic systems in continuous space, *Phys. Rev. Res.* **4**, 023138 (2022).
- [24] J. Carrasquilla, D. Luo, F. Pérez, A. Milsted, B. K. Clark, M. Volkovs, and L. Aolita, Probabilistic simulation of quantum circuits using a deep-learning architecture, *Phys. Rev. A* **104**, 032610 (2021).
- [25] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Reconstructing quantum states with generative models, *Nat. Mach. Intell.* **1**, 155 (2019).
- [26] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, *Phys. Rev. B* **106**, 205136 (2022).
- [27] D.-L. Deng, X. Li, and S. Das Sarma, Quantum Entanglement in Neural Network States, *Phys. Rev. X* **7**, 021021 (2017).
- [28] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, *Nat. Commun.* **8**, 662 (2017).
- [29] G. Passetti, D. Hofmann, P. Neitemeier, L. Grunwald, M. A. Sentef, and D. M. Kennes, Can neural quantum states learn volume-law ground states? [arXiv:2212.02204](https://arxiv.org/abs/2212.02204).

- [30] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, *J. Phys.: Condens. Matter* **33**, 174003 (2021).
- [31] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, *SciPost Phys.* **10**, 147 (2021).
- [32] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J_1 - J_2 model studied with neural network quantum states, *Phys. Rev. B* **100**, 125124 (2019).
- [33] M. Van den Nest, Simulating quantum computers with probabilistic methods, *Quantum Inf. Comput.* **11**, 784 (2011).
- [34] S. Sorella, M. Casula, and D. Rocca, Weak binding between two aromatic rings: Feeling the van der waals attraction by quantum monte carlo methods, *J. Chem. Phys.* **127**, 014105 (2007).
- [35] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control. Signals, Syst.* **2**, 303 (1989).
- [36] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4**, 251 (1991).
- [37] T. Kim and T. Adalı, Approximation by fully complex multilayer perceptrons, *Neural Comput.* **15**, 1641 (2003).
- [38] N. L. Roux and Y. Bengio, Representational power of restricted boltzmann machines and deep belief networks, *Neural Comput.* **20**, 1631 (2008).
- [39] S. Bravyi, D. P. Divincenzo, R. Oliveira, and B. M. Terhal, The complexity of stoquastic local hamiltonian problems, *Quantum Inf; Comput* **8**, 361 (2008).
- [40] S. Czischek, M. Gärttner, and T. Gasenzer, Quenches near ising quantum criticality as a challenge for artificial neural networks, *Phys. Rev. B* **98**, 024311 (2018).
- [41] X.-Q. Sun, T. Nebabu, X. Han, M. O. Flynn, and X.-L. Qi, Entanglement features of random neural network quantum states, *Phys. Rev. B* **106**, 115138 (2022).
- [42] L. L. Viteritti, F. Ferrari, and F. Becca, Accuracy of restricted Boltzmann machines for the one-dimensional J_1 - J_2 Heisenberg model, *SciPost Phys.* **12**, 166 (2022).
- [43] D. Hofmann, G. Fabiani, J. H. Mentink, G. Carleo, and M. A. Sentef, Role of stochastic noise and generalization error in the time propagation of neural-network quantum states, *SciPost Phys.* **12**, 165 (2022).
- [44] T. Schmale, M. Reh, and M. Gärttner, Efficient quantum state tomography with convolutional neural networks, *npj Quantum Inf.* **8**, 115 (2022).
- [45] I. L. Gutiérrez and C. B. Mendl, Real time evolution with neural-network quantum states, *Quantum* **6**, 627 (2022).
- [46] M. Reh, M. Schmitt, and M. Gärttner, Time-Dependent Variational Principle for Open Quantum Systems with Artificial Neural Networks, *Phys. Rev. Lett.* **127**, 230501 (2021).
- [47] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, Autoregressive Neural Network for Simulating Open Quantum Systems via a Probabilistic Formulation, *Phys. Rev. Lett.* **128**, 090501 (2022).
- [48] K. Donatella, Z. Denis, A. L. Boité, and C. Ciuti, Dynamics with autoregressive neural quantum states: application to critical quench dynamics, [arXiv:2209.03241](https://arxiv.org/abs/2209.03241).
- [49] F. Vicentini, R. Rossi, and G. Carleo, Positive-definite parametrization of mixed quantum states with deep neural networks, [arXiv:2206.13488](https://arxiv.org/abs/2206.13488).
- [50] S. Hochreiter and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.* **9**, 1735 (1997).
- [51] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, [arXiv:1412.3555](https://arxiv.org/abs/1412.3555).
- [52] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, [arXiv:2104.13478](https://arxiv.org/abs/2104.13478).
- [53] H. W. J. Blöte and Y. Deng, Cluster monte carlo simulation of the transverse ising model, *Phys. Rev. E* **66**, 066110 (2002).
- [54] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy, *Phys. Rev. X* **11**, 031034 (2021).
- [55] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Solving frustrated quantum many-particle models with convolutional neural networks, *Phys. Rev. B* **98**, 104426 (2018).
- [56] F. Ferrari, F. Becca, and J. Carrasquilla, Neural gutzwiller-projected variational wave functions, *Phys. Rev. B* **100**, 125131 (2019).
- [57] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, *Nat. Commun.* **11**, 1593 (2020).
- [58] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, *Phys. Rev. Res.* **2**, 033075 (2020).
- [59] C. Roth, Iterative retraining of quantum spin models using recurrent neural networks, [arXiv:2003.06228](https://arxiv.org/abs/2003.06228).
- [60] M. Hibat-Allah, R. G. Melko, and J. Carrasquilla, Supplementing recurrent neural network wave functions with symmetry and annealing to improve accuracy, [arXiv:2207.14314](https://arxiv.org/abs/2207.14314).
- [61] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, Gauge invariant autoregressive neural networks for quantum lattice models, *Phys. Rev. Res.* **5**, 013216 (2023).
- [62] C. Roth, A. Szabó, and A. MacDonald, High-accuracy variational monte carlo for frustrated magnets with deep neural networks, [arXiv:2211.07749](https://arxiv.org/abs/2211.07749).
- [63] M. Schmitt and M. Reh, jVMC: Versatile and performant variational Monte Carlo leveraging automated differentiation and GPU acceleration, *SciPost Phys. Codebases* **2**, 2 (2022).
- [64] See [GitHub:markusschmitt/vmc_jax](https://github.com/markusschmitt/vmc_jax)
- [65] See www.gauss-centre.eu
- [66] Jülich Supercomputing Centre, JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre, *J. Large-Scale Res. Facilities* **5**, A171 (2019).
- [67] G. E. Hinton, A practical guide to training restricted Boltzmann machines, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer, Berlin, Heidelberg, 2012), pp. 599–619.

AREA LAWS FOR CLASSICAL ENTROPIES IN A SPIN-1 BOSE-EINSTEIN CONDENSATE

8.1 CONTRIBUTIONS

This chapter presents a finalized draft of the work in [D]. Tobias Haas originally proposed the project, and the proposal was refined in joint discussions between Tobias Haas, Martin Gärttner, Yannick Deller, Helmut Strobel, Markus Oberthaler, and me. Tobias Haas and I ran all simulations in close collaboration, with the analytical model implemented by Tobias Haas and the truncated Wigner simulation, discussed in [Section 4.2](#), implemented by me. Tobias Haas and I wrote the initial draft of the manuscript and all other authors helped in refining it.

8.2 MOTIVATION

The area-law is a central aspect of many-body quantum theory with important implications regarding classical simulatability, see [Section 2.2](#) and the discussion in [Chapter 4](#). It is typically observed in quantum entropies, such as the von Neumann or Rényi entropy of the density operator [190]. We will herein show that this quantum feature is preserved when transitioning to a phase-space description of the system and provide a detailed proposal of how this can be tested experimentally in a Spin-1 Bose-Einstein condensate as discussed in [Section 2.7](#).

Area laws for classical entropies in a spin-1 Bose-Einstein condensate

Yannick Deller,^{1,*} Martin Gärttner,^{2,†} Tobias Haas,^{3,‡} Markus
K. Oberthaler,^{1,§} Moritz Reh,^{1,¶} and Helmut Strobel^{1,**}

¹*Kirchhoff-Institut für Physik, Universität Heidelberg,
Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*

²*Institut für Festkörperteorie und Optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany*

³*Centre for Quantum Information and Communication, École polytechnique de Bruxelles,
CP 165, Université libre de Bruxelles, 1050 Brussels, Belgium*

We investigate the information extractable from measurement distributions of two non-commuting spin observables in a multi-well spin-1 Bose-Einstein condensate. We provide a variety of analytic and numerical evidence that suitably chosen classical entropies and classical mutual informations thereof contain the typical feature of quantum entropies known in quantum field theories, that is, the area law, even in the non-Gaussian regime and for a non-zero temperature. Towards a feasible experimental implementation, we estimate entropic quantities from a finite number of samples without any additional assumptions on the underlying quantum state using k -nearest neighbor estimators.

I. INTRODUCTION

The scaling of entropic measures associated with the quantum state of a spatial subregion is one of the central characteristics describing how quantum information and entanglement are encoded in spacetime. Originally discovered by Bekenstein in the context of black hole physics [1, 2] (see also [3–6]), the entropy of a subregion is, to leading order, proportional to the area of its enclosing surface – rather than the subregion’s volume – which is conveniently referred to as the *area law*. Over the last two decades, much theoretical evidence has been gathered for its appearance in various contexts, including, for instance, quantum field theory [7–12], quantum many-body systems [13–15], tensor networks [16] and thermalization [17]. Also, it has been shown to occur for measures of correlations between the subregion of interest and its complement, e.g. the quantum mutual information [18].

Considerably less evidence is available on the experimental side, which can be traced back to the notoriously difficult task of extracting the local quantum state, i.e. performing full quantum state tomography. So far, the area law has been observed only for a handful of degrees of freedom for which entropic measures can be reduced to direct observables. This includes, for example, a study of a six-site Bose-Hubbard system where the Rényi-2 quantum entropy is read out via two-copy interference [19, 20] and a trapped ion quantum simulator of ten qubits [21], in which the same quantity was extracted using random measurements [22].

The task of accessing the quantum state is even more challenging for continuous quantum many-body systems, which are in principle described by quantum fields over continuous positions as well as infinite-dimensional local Hilbert spaces. A recent ultracold atom experiment reported an area law for the quantum mutual information under the assumption that the underlying quantum state is of Gaussian form, in which case measurements of two-point correlation functions suffice to calculate entropic quantities [23]. Nevertheless, studies beyond the Gaussian case have remained elusive so far.

What all aforementioned approaches, theoretical as well as experimental, have in common, is their reliance on a *quantum* entropy as an indicator for the area law-like behavior of quantum correlations. However, it has been shown recently that the appearance of the area law is by no means restricted to such quantum entropies: when considering phase-space representations and measurement distributions of the quantum state instead, their corresponding *classical* entropies reveal the area law in the *next-to-leading* order terms, i.e. when classical contributions are subtracted properly [24].

Prominent examples of such distributions are the Wigner W -distribution [25], its marginals and the Husimi Q -distribution [26, 27] (see [28–30] for reviews on phase-space methods). Especially the latter is of particular interest as it is a non-negative and normalized function in phase space, which allows for well-defined entropic descriptions [31–36]. Further, its usefulness for witnessing entanglement has already been demonstrated theoretically in terms of entropic measures [37–39] as well as experimentally [40, 41].

Although all phase-space distributions contain the very same information as the density operator, estimating their associated classical differential entropies is a significantly simpler task than reconstructing the full many-body quantum state and computing its quantum entropy. The outcome of a typical experimental setting can be thought of as a sample drawn from a given distribution (usually the marginals of the Wigner W - or the full Husimi Q -

* yannick.deller@kip.uni-heidelberg.de

† martin.gaerttner@uni-jena.de

‡ tobias.haas@ulb.be

§ markus.oberthaler@kip.uni-heidelberg.de

¶ moritz.reh@kip.uni-heidelberg.de

** helmut.strobel@kip.uni-heidelberg.de

distribution are considered), see e.g. [40–48]. In principle, one could then attempt to infer the corresponding distribution from these samples, which is equivalent to state tomography when considering a full phase-space distribution, and calculate its entropy thereafter. However, the need for reconstructing the underlying distribution can be bypassed by employing sophisticated methods to estimate the entropy from the sampled data *directly*. In this context, a widely-used approach is the k -nearest neighbor (k NN) method, which produces an asymptotically unbiased estimate for the entropy without any assumptions on the underlying distribution [49–63].

In this work, we consider an experimentally friendly scenario in which the built-up of an area law over time is expected: the post-quench dynamics of a multi-well spin-1 Bose-Einstein condensate (BEC). In what follows, we simulate the dynamics using the truncated Wigner approximation (TWA) and employ the k NN method to estimate subtracted classical entropies as well as classical mutual informations of phase-space distributions from sampled data. We provide detailed numerical evidence for the appearance of the area law for classical entropies in a variety of scenarios, including, for instance, non-Gaussian states and varying system sizes, and under typical experimental constraints, e.g., thermal fluctuations and finite sample size. Our simulation results are supported by an analytical model covering the early-time dynamics, which allows for a straightforward evaluation of all entropic quantities. With our findings, we pave the ground for an experimental observation of the area law in a continuous quantum many-body system without any prior assumptions on the underlying quantum state.

In this work, we present a detailed description of how the area law can be accessed from classical entropies, both from a theoretical and a practical perspective. In [64], we provide a concise summary of our findings and additionally discuss the process of local thermalization in terms of such entropies.

The remainder of this paper is organized as follows. In section II, we introduce the model system of our interest – a spin-1 Bose-Einstein condensate in the multi-well setup – with a special focus on the Hamiltonian and the spin observables. We proceed with mapping this system to a continuous-variable quantum system described by canonical commutation relations using the so-called undepleted pump approximation in section III. Therein, we also put forward phase-space representations of quantum states and show how these are related to the measurement distributions of the spin observables. The two approaches for simulating the dynamics, i.e. the truncated Wigner approximation for the full Hamiltonian and an approximate, analytically solvable model valid for early times are discussed in section IV. For the latter, we employ methods from Gaussian quantum information theory, which we describe in detail. Thereupon, in section V, we introduce the necessary background on the area law of quantum and classical entropies in phase space. Then, we describe the k NN machinery for estimating entropies from sampled

data and benchmark the method for several cases being relevant for its application to the system of our interest in section VI. Finally, we provide our main results in section VII. After a comparison of the analytical model with the TWA approach (section VII A) we show plenty of evidence for the emergence of the area law over time for a variety of classical entropies (section VII B), which is followed by systematic studies of its robustness. More precisely, we consider strongly non-Gaussian distributions (section VII C), a thermal initial state (section VII D), boundary effects (section VII E), varying system size (section VII F), varying sample size (section VII G) and different types of boundary conditions (section VII H). A comprehensive discussion of our results and some future prospects are given in section VIII.

Notation. We employ natural units $\hbar = k_B = 1$, denote quantum operators by bold letters, e.g. $\boldsymbol{\phi}$, and classical variables by normal letters, e.g. ϕ , (similarly for operations on operators / matrices, e.g. $\text{Tr}\{\boldsymbol{\rho}\} / \det\{\boldsymbol{\gamma}\}$) and equip vacuum expressions with a bar, e.g. \bar{Q} . Further, we use upper indices to refer to a specific well, e.g. Q^j for the j -th. well, or to a subsystem of a bipartition AB , e.g. Q^A , and put lower indices for the hyperfine levels, e.g. \boldsymbol{a}_0 , or relative modes, e.g. \boldsymbol{a}_\pm .

II. MULTI-WELL SPIN-1 BEC

We begin with a discussion of the multi-well setup of coupled spin-1 Bose-Einstein condensates. Then, we introduce the Hamiltonian governing the corresponding dynamics and the typical observables in such systems.

A. Setup

We consider an experimental setup consisting of a spin-1 BEC with ferromagnetic spin coupling, based on Lithium-7 [65, 66]. We assume the condensate’s geometry to be quasi one-dimensional, realizable via tight confinement along the radial direction, which can be implemented experimentally by ensuring that the radial trapping frequency is much larger than the longitudinal one, i.e. $\omega_r \gg \omega_l$. In the longitudinal direction, the BEC is further subjected to an optical lattice potential, which divides the overall BEC into N smaller-size BECs, which we refer to as wells. The depth of the optical lattice controls the potential barrier between neighboring wells. When tuned to sufficiently small energies, atoms can tunnel between neighboring wells, which establishes a nearest-neighbor interaction between the local degrees of freedom. This setup constitutes a discretized approximation to a continuous quantum field.

The internal degrees of freedom encoding the spin-1 system are the three hyperfine levels $m_F = 0, \pm 1$ of the $F = 1$ spin manifold of the electronic ground state. Initially, we consider all atoms to be prepared in the $m_F = 0$ mode. Thereafter, spin-changing collisions triggered by

off-resonant microwave dressing lead to the side modes being dynamically populated [65, 67, 68]. As this process is symmetric with respect to the side modes, the linear Zeeman shift is not relevant and hence, only the quadratic Zeeman shift is relevant for the dynamics.

We consider tunneling to only take place in the side modes and not in the initially macroscopically populated zero mode, such that information exchange between wells through tunneling only takes place after the generation of side mode population. Experimentally, this can be achieved, for example, by working in a state-dependent lattice, such that the side modes are weaker trapped than the zero mode. Another possibility is to tune the spin-changing collisions into resonance with excited states of the side modes and work in a deep optical lattice, such that atoms transferred to the side modes acquire enough kinetic energy to tunnel to neighboring sites of the deep optical lattice, whereas the kinetic energy of atoms in the zero mode remains insufficient for tunneling.

B. Modes and Hilbert space

We are interested in the dynamics of the internal degrees of freedom characterized by the three hyperfine levels $m_F \in \{-1, 0, 1\}$. To each hyperfine level m_F and well j we associate a set of independent bosonic creation and annihilation operators

$$[\mathbf{a}_{m_F}^j, \mathbf{a}_{m_F}^{j'\dagger}] = \delta^{jj'} \delta_{m_F m_F'}, \quad (1)$$

where the upper index labels the $N \in \mathbb{N} < \infty$ wells such that $j \in \{1, \dots, N\}$. For a single well j the underlying Hilbert space \mathcal{H}^j is a Fock space constructed from multi-particle states

$$|n_1^j, n_0^j, n_{-1}^j\rangle \equiv \frac{(\mathbf{a}_1^{j\dagger})^{n_1^j} (\mathbf{a}_0^{j\dagger})^{n_0^j} (\mathbf{a}_{-1}^{j\dagger})^{n_{-1}^j}}{\sqrt{n_1^j! n_0^j! n_{-1}^j!}} |0, 0, 0\rangle, \quad (2)$$

with the tensor product notation $|\cdot, \cdot, \cdot\rangle = |\cdot\rangle \otimes |\cdot\rangle \otimes |\cdot\rangle$ understood and the number of particles $n_{m_F}^j$ in mode j and level m_F being defined as the eigenvalue of the corresponding particle number operator

$$\mathbf{N}_{m_F}^j = \mathbf{a}_{m_F}^{j\dagger} \mathbf{a}_{m_F}^j. \quad (3)$$

Accordingly, the total number of particles in well j is measured by the operator

$$\mathbf{N}^j = \sum_{m_F=-1}^1 \mathbf{N}_{m_F}^j, \quad (4)$$

which has the eigenvalue $n^j = \sum_{m_F=-1}^1 n_{m_F}^j$. Then, the full Hilbert space is obtained by taking the tensor product with respect to all wells, i.e. $\mathcal{H} = \otimes_{j=1}^N \mathcal{H}^j$, such that a generic element of the full Fock basis reads

$$|n_1^1, n_0^1, n_{-1}^1; \dots; n_1^N, n_0^N, n_{-1}^N\rangle \equiv \bigotimes_{j=1}^N |n_1^j, n_0^j, n_{-1}^j\rangle. \quad (5)$$

C. Hamiltonian

The full Hamiltonian is composed of two main terms: a single-well Hamiltonian \mathbf{H}_{sw}^j describing the on-site dynamics of well j as well as a tunneling Hamiltonian \mathbf{H}_{t}^j encoding the coupling between neighboring wells j and $j+1$, such that in total we have

$$\mathbf{H} = \sum_{j=1}^N \mathbf{H}_{\text{sw}}^j + \sum_{j=1}^{N \text{ or } N-1} \mathbf{H}_{\text{t}}^j. \quad (6)$$

For the sake of generality, we keep the type of boundary conditions open at this point. More precisely, we allow for periodic or open boundary conditions, which are implemented by the second sum in (6) running up to N or $N-1$, i.e. coupling the first and the last well or not, respectively.

The single-well Hamiltonian for well j is given by

$$\begin{aligned} \mathbf{H}_{\text{sw}}^j &= c_0 \mathbf{N}^j (\mathbf{N}^j - \mathbb{1}) \\ &+ c_1 \left[\left(\mathbf{N}_0^j - (1/2)\mathbb{1} \right) \left(\mathbf{N}_1^j + \mathbf{N}_{-1}^j \right) \right. \\ &\quad \left. + \mathbf{a}_0^{j\dagger} \mathbf{a}_0^j \mathbf{a}_1^j \mathbf{a}_{-1}^j + \mathbf{a}_1^{j\dagger} \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_0^j \mathbf{a}_0^j \right] \\ &+ q \left(\mathbf{N}_1^j + \mathbf{N}_{-1}^j \right), \end{aligned} \quad (7)$$

and has been investigated in great detail, see e.g. [40, 41, 65, 67, 69, 70]. It contains three contributions: First, the on-site density-density interaction $c_0 > 0$ describing the repulsive interaction of atoms regardless of their hyperfine levels (petrol ellipse in Figure 1). Second, the spin-changing collision interaction $c_1 < 0$ encoding the generation of spin pairs in the $m_F = \pm 1$ hyperfine levels from the $m_F = 0$ level (red arrows). For clarity, we omitted the mean-field shifts in the second line of Eq. (7). Third, we included the quadratic $q > 0$ Zeemann shift (gray arrows, respectively).

The tunneling between equal hyperfine levels $m_F = \pm 1$ of neighboring wells j and $j+1$ is described by

$$\mathbf{H}_{\text{t}}^j = -J \sum_{m_F=\pm 1} \left(\mathbf{a}_{m_F}^{j\dagger} \mathbf{a}_{m_F}^{j+1} + \mathbf{a}_{m_F}^{(j+1)\dagger} \mathbf{a}_{m_F}^j \right), \quad (8)$$

with a non-negative tunnel rate $J \geq 0$ (blue arrows). Note again that the $m_F = 0$ mode does not couple to neighboring wells and also that in the limit $J \rightarrow 0$ the wells evolve independently.

D. Observables: Spin operators

The local observables in every well are the eight spin operators forming a representation of the Lie algebra $\mathfrak{su}(3)$ [65, 67]. They are constructed following the Jordan-Schwinger map [71, 72]: starting from an irreducible representation of the eight three-dimensional matrices G_α of $\mathfrak{su}(3)$ defined via

$$[G_\alpha, G_\beta] = f_{\alpha\beta\gamma} G_\gamma, \quad (9)$$

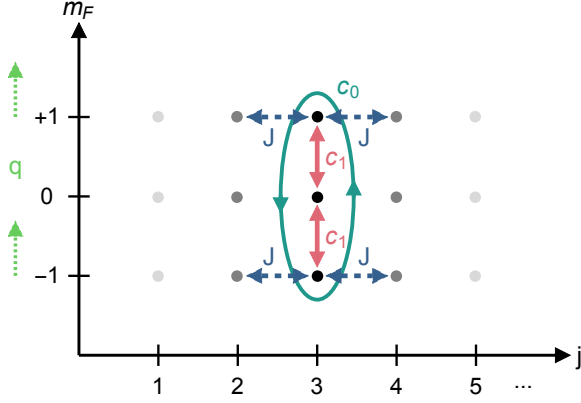


Figure 1: Illustration of the dynamics generated by the five contributions to the full Hamiltonian (6) for the $j = 3$ well. First-order (tunneling) and second-order (collisions) processes are indicated by dashed and solid arrows, respectively, while the energy shift due to the Zeemann effect is depicted by a dotted arrow.

where $f_{\alpha\beta\gamma}$ are the so-called structure constants (we omit their explicit form here) and $\alpha \in \{1, \dots, 8\}$ labels the matrices, the local quantum operators \mathbf{G}_α^j obeying the very same algebra are constructed following

$$\mathbf{G}_\alpha^j = \sum_{m_F, m_{F'}}^1 \mathbf{a}_{m_F}^{j\dagger} (G_\alpha)_{m_F m_{F'}} \mathbf{a}_{m_{F'}}^j, \quad (10)$$

with $(G_\alpha)_{m_F m_{F'}}$ denoting the $(m_F, m_{F'})$ -th. entry of the spin matrix G_α .

A common choice for the $\mathfrak{su}(3)$ matrices G_α is made by first constructing a $\mathfrak{su}(2)$ subspace (see [67] for explicit expressions for the corresponding matrices) via

$$[S_\alpha^j, S_\beta^j] = i\epsilon_{\alpha\beta\gamma} S_\gamma^j, \quad (11)$$

where $\alpha, \beta, \gamma \in \{x, y, z\}$, and thereupon deriving the three corresponding spin operators from (10), leading to

$$\begin{aligned} S_x^j &= \frac{1}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} (\mathbf{a}_1^j + \mathbf{a}_{-1}^j) + (\mathbf{a}_1^{j\dagger} + \mathbf{a}_{-1}^{j\dagger}) \mathbf{a}_0^j \right], \\ S_y^j &= \frac{i}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} (\mathbf{a}_1^j - \mathbf{a}_{-1}^j) - (\mathbf{a}_1^{j\dagger} - \mathbf{a}_{-1}^{j\dagger}) \mathbf{a}_0^j \right], \\ S_z^j &= \mathbf{a}_1^{j\dagger} \mathbf{a}_1^j - \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^j. \end{aligned} \quad (12)$$

The remaining five operators are the so-called quadrupole operators defined as

$$\mathbf{Q}_{\alpha\beta}^j = \{S_\alpha^j, S_\beta^j\} - \frac{4}{3} \delta_{\alpha\beta} \mathbb{1}, \quad (13)$$

with $\{.,.\}$ denoting the anticommutator. Here \mathbf{Q}_{yz}^j and \mathbf{Q}_{xz}^j are of special interest, which read

$$\begin{aligned} \mathbf{Q}_{yz}^j &= \frac{i}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} (\mathbf{a}_1^j + \mathbf{a}_{-1}^j) - (\mathbf{a}_1^{j\dagger} + \mathbf{a}_{-1}^{j\dagger}) \mathbf{a}_0^j \right], \\ \mathbf{Q}_{xz}^j &= \frac{1}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} (\mathbf{a}_1^j - \mathbf{a}_{-1}^j) + (\mathbf{a}_1^{j\dagger} - \mathbf{a}_{-1}^{j\dagger}) \mathbf{a}_0^j \right]. \end{aligned} \quad (14)$$

E. Readout schemes

We will analyze the system by gathering information about the measurement distributions of the variable pairs (S_x^j, Q_{yz}^j) over multiple wells. We discuss methods for reading out two types of such distributions in the following.

1. Separate detection

In BEC experiments the population of each mode, $n_{m_F}^j = \langle N_{m_F}^j \rangle$ is detected through absorption imaging after Stern-Gerlach separation of the different m_F components. From this the z -component of the local spin can be extracted as $S_z^j = n_1^j - n_{-1}^j$. Other spin components can be measured by applying a radio-frequency (rf) magnetic field prior to the absorption imaging. For a frequency matching the linear Zeeman shift, resonant Rabi oscillations are driven, described by the Hamiltonian (in the rotating wave approximation) [69, 70]

$$\mathbf{H}_{\text{rf}} = \sum_{j=1}^N \Omega_{\text{rf}} [\cos(\phi_{\text{rf}}) \mathbf{S}_y^j - \sin(\phi_{\text{rf}}) \mathbf{S}_x^j] \quad (15)$$

Here, Ω_{rf} is the Rabi frequency of the drive and ϕ_{rf} is a tunable phase. Applying the drive for a time $t = \pi/(2\Omega_{\text{rf}})$ thus allows one to map the spin along an arbitrary direction of the equatorial plane of the spin sphere onto \mathbf{S}_z^j . In the case where the $m_F = 0$ mode is a coherent state with population far larger than the side modes, this scheme is analogous to homodyne detection in optics, where the signal mode(s) are mixed with a local oscillator mode on a beam splitter to extract the field quadratures [73].

To toggle between measurements of spin operators and quadrupole operators one can add a time delay before the application of the rf-rotation. By tuning the quadratic Zeeman shift and the microwave dressing field in the single-well Hamiltonian Eq. (7) the $m_F = 0$ mode acquires a relative phase. For instance, measuring Q_{yz}^j instead of S_x^j requires an additional phase of $\pi/2$.

2. Simultaneous detection

Sophisticated methods to extract joint distributions over (S_x^j, Q_{yz}^j) , or equivalently (S_y^j, Q_{xz}^j) , in close analogy to the heterodyne detection protocol in quantum optics [29, 30, 74, 75] have been experimentally realized rather recently [40, 41, 70]. For this, one exploits the availability of additional, initially unoccupied, internal levels, in our example the $F = 2$ hyperfine manifold with five additional Zeeman levels. Using microwave fields resonantly coupling the levels of the $F = 1$ manifold with those of the $F = 2$ manifold, one can realize analogs of beam splitter operations between them. Splitting each mode equally and subsequently applying different rf-rotations

in the two manifolds realizes the simultaneous detection of, for example, \mathbf{S}_x^j and \mathbf{Q}_{yz}^j (see Ref. [40] for details). In the case of the $m_F = 0$ mode being macroscopically occupied, this corresponds to sampling from the Husimi Q -distribution in the phase space spanned by \mathbf{S}_x^j and \mathbf{Q}_{yz}^j (see section III for details).

III. FROM A SPIN-1 BEC TO A CONTINUOUS-VARIABLE QUANTUM SYSTEM

We point out the connection between the system of our interest, a multi-well spin-1 BEC, and a continuous-variable quantum system describing coupled bosonic oscillator modes. Thereupon, we introduce several kinds of phase-space descriptions.

A. Undepleted pump approximation

The undepleted pump regime captures the early-time dynamics of the Hamiltonian (6) when initially preparing the polar ground state of the spin-1 BEC in every well, i.e.

$$|\psi(0)\rangle = |0, \alpha, 0\rangle^{\otimes N} \quad (16)$$

where $|\alpha\rangle$ is the local coherent state of the zero mode with mean particle number $n = |\alpha|^2$. For large values n the $m_F = 0$ mode is macroscopically populated compared to the side modes $m_F = \pm 1$ in every well, i.e.

$$\langle N_0^j \rangle \gg \langle N_{\pm 1}^j \rangle, \quad (17)$$

for all j , which is equivalent to

$$n^j = \langle N^j \rangle \cong \langle N_0^j \rangle. \quad (18)$$

Under the above assumptions, we can approximate the $m_F = 0$ mode operators by their norms

$$\mathbf{a}_0^j = \mathbf{a}_0^{j\dagger} \cong \sqrt{n^j}, \quad (19)$$

such that $N_0^j \cong n^j$. Then, the full Hamiltonian (6) simplifies to

$$\begin{aligned} \mathbf{H}_{\text{up}} = & \sum_{j=1}^N \left[c_0 \left(\mathbf{a}_1^{j\dagger} \mathbf{a}_1^{j\dagger} \mathbf{a}_1^j \mathbf{a}_1^j + \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^j \mathbf{a}_{-1}^j + 2N_1^j N_{-1}^j \right) \right. \\ & + \tilde{c}_1^j \left(\mathbf{a}_1^j \mathbf{a}_{-1}^j + \mathbf{a}_1^{j\dagger} \mathbf{a}_{-1}^{j\dagger} \right) \\ & \left. + (\tilde{q}^j + c_0 n^j) \left(N_1^j + N_{-1}^j \right) \right] \\ & - J \sum_{j=1}^{N \text{ or } N-1} \left(\mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^{j+1} + \mathbf{a}_{-1}^{(j+1)\dagger} \mathbf{a}_{-1}^j \right. \\ & \quad \left. + \mathbf{a}_1^{j\dagger} \mathbf{a}_1^{j+1} + \mathbf{a}_1^{(j+1)\dagger} \mathbf{a}_1^j \right) \\ & + H_{\text{offset}}. \end{aligned} \quad (20)$$

Here, we defined the rescaled couplings

$$\tilde{c}_1^j = c_1 n^j, \quad \tilde{q}^j = c_1 \left(n^j - \frac{1}{2} \right) + q, \quad (21)$$

as well as a constant offset term

$$H_{\text{offset}} = c_0 \sum_{j=1}^N (n^j)^2, \quad (22)$$

which we will drop in the following.

B. Relative modes and canonically conjugate variables

In the undepleted pump regime, the relevant degrees of freedom and their Hamiltonian (20) can be mapped to a continuous-variable quantum system. To that end, we introduce the relative mode operators between the side modes

$$\mathbf{a}_{\pm}^j = \frac{1}{\sqrt{2}} \left(\mathbf{a}_1^j \pm \mathbf{a}_{-1}^j \right), \quad (23)$$

which also represent independent bosonic modes, since (1) implies

$$[\mathbf{a}_{\pm}^j, \mathbf{a}_{\pm}^{j'\dagger}] = \delta_{jj'}, \quad [\mathbf{a}_{\pm}^j, \mathbf{a}_{\mp}^{j'\dagger}] = 0. \quad (24)$$

Their associated canonical operators are defined as

$$\phi_{\pm}^j = \frac{1}{\sqrt{2}} \left(\mathbf{a}_{\pm}^{j\dagger} + \mathbf{a}_{\pm}^j \right), \quad \pi_{\pm}^j = \frac{i}{\sqrt{2}} \left(\mathbf{a}_{\pm}^{j\dagger} - \mathbf{a}_{\pm}^j \right), \quad (25)$$

which fulfill the canonical commutation relations

$$[\phi_{\pm}^j, \pi_{\pm}^{j'}] = i\delta^{jj'}, \quad [\phi_{\pm}^j, \pi_{\mp}^{j'}] = 0. \quad (26)$$

Interestingly, in the undepleted pump regime, the two pairs of canonical operators are equivalent to pairs of spin operators up to normalization, to wit

$$\begin{aligned} \mathbf{S}_x^j &= \sqrt{2n^j} \phi_+^j, & \mathbf{Q}_{yz}^j &= -\sqrt{2n^j} \pi_+^j, \\ \mathbf{S}_y^j &= \sqrt{2n^j} \phi_-^j, & \mathbf{Q}_{xz}^j &= -\sqrt{2n^j} \pi_-^j, \end{aligned} \quad (27)$$

which follows from using (19) in (12) and (14). This shows that, for short times, the relevant degrees of freedom mimic two pairs of canonical variables and hence the local Hilbert spaces \mathcal{H}^j decompose as $\mathcal{H}^j = \mathcal{H}_+^j \otimes \mathcal{H}_-^j$.

Next, we express the Hamiltonian (20) in terms of the newly defined operators (see Appendix A for details). We find the general decomposition

$$\mathbf{H}_{\text{up}} = \mathbf{H}_{\text{up}}^+ + \mathbf{H}_{\text{up}}^- + \mathbf{H}_{\text{up}}^{\text{mix}}, \quad (28)$$

where the form-equivalent \mathbf{H}_{up}^+ and \mathbf{H}_{up}^- contain all terms with only + or - modes, respectively, while $\mathbf{H}_{\text{up}}^{\text{mix}}$ contains all terms mixing the two relative modes \pm . In terms of

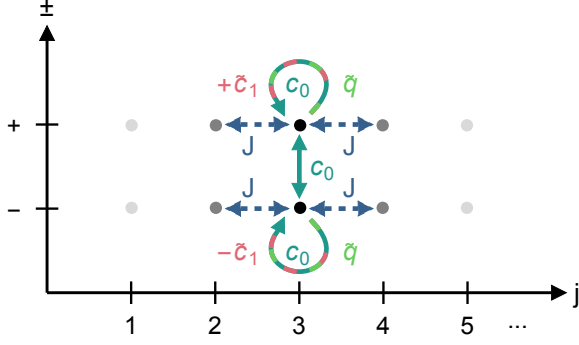


Figure 2: Sketch of the dynamics in the undepleted pump regime generated by (28) (in analogy to Figure 1).

relative mode operators, these two types of Hamiltonians read

$$\begin{aligned}
 \mathbf{H}_{\text{up}}^{\pm} = & \sum_{j=1}^N \left[c_0 \mathbf{a}_{\pm}^{j\dagger} \mathbf{a}_{\pm}^j \mathbf{a}_{\pm}^{j\dagger} \mathbf{a}_{\pm}^j \right. \\
 & + [\tilde{q}^j + c_0 (n^j - 1)] \mathbf{N}_{\pm}^j \\
 & \left. \pm \frac{\tilde{c}_1^j}{2} \left(\mathbf{a}_{\pm}^{j\dagger} \mathbf{a}_{\pm}^{j\dagger} + \mathbf{a}_{\pm}^j \mathbf{a}_{\pm}^j \right) \right] \\
 & - J \sum_{j=1}^{N \text{ or } N-1} \left(\mathbf{a}_{\pm}^{j\dagger} \mathbf{a}_{\pm}^{j+1} + \mathbf{a}_{\pm}^{(j+1)\dagger} \mathbf{a}_{\pm}^j \right), \\
 \mathbf{H}_{\text{up}}^{\text{mix}} = & 2c_0 \sum_{j=1}^N \mathbf{N}_+^j \mathbf{N}_-^j,
 \end{aligned} \tag{29}$$

whose dynamics are sketched in Figure 2. From the latter formulas, it becomes apparent that in the undepleted pump regime the relative mode operators, or equivalently the two pairs of canonical operators, constitute a complete set of observables for characterizing the early-time dynamics. Further, the dynamics within the \pm phase spaces differ only by a sign in the $\pm\tilde{c}_1$ term.

C. Canonical phase-space

As the observables of our interest now correspond to continuous variables, we can faithfully apply the powerful concepts of phase-space descriptions. The two sets of canonical operators $(\phi_{\pm}^j, \pi_{\pm}^j)$ span the two two-dimensional canonical phase spaces for well j , which are known to be isomorphic to the Euclidean plane \mathbb{R}^2 with measure [76, 77]

$$\int d\phi_{\pm}^j d\pi_{\pm}^j. \tag{30}$$

The field operators are conveniently combined into a single vector χ_{\pm}^j by defining $\chi_{\pm}^j = \phi_{\pm}^j$ for $j \in [1, N]$ and $\chi_{\pm}^j = \pi_{\pm}^{j-N}$ for $j \in [N+1, 2N]$, which we formally write

as [74, 75]

$$\chi_{\pm} = (\phi_{\pm}, \pi_{\pm})^T = (\phi_{\pm}^1, \dots, \phi_{\pm}^N, \pi_{\pm}^1, \dots, \pi_{\pm}^N)^T. \tag{31}$$

The canonical commutation relations (26) become

$$[\chi_{\pm}^j, \chi_{\pm}^{j'}] = i\Omega^{jj'} \mathbb{1}, \tag{32}$$

where

$$\Omega = (i\sigma_2) \otimes \mathbb{1}_N \tag{33}$$

denotes the symplectic metric revealing the symplectic structure of the canonical phase space and σ_2 is the second Pauli matrix.

For every well j we define the set of canonical coherent states associated with the relative modes \pm as displaced vacuum states [29, 30, 74, 75], i.e.

$$|\alpha_{\pm}^j\rangle = \mathbf{D}(\alpha_{\pm}^j) |0_{\pm}^j\rangle, \tag{34}$$

where $\mathbf{D}(\alpha_{\pm}^j)$ is the unitary displacement operator

$$\mathbf{D}(\alpha_{\pm}^j) = e^{\alpha_{\pm}^j \mathbf{a}_{\pm}^{j\dagger} - \alpha_{\pm}^{*j} \mathbf{a}_{\pm}^j}, \tag{35}$$

with the complex-valued phase fields being parameterized in terms of cartesian coordinates as

$$\alpha_{\pm}^j = \frac{1}{\sqrt{2}} \left(\phi_{\pm}^j + i\pi_{\pm}^j \right). \tag{36}$$

Importantly, the set of coherent states constitutes an overcomplete basis, i.e. the coherent states defined in (34) resolve the identity in the \pm subspaces

$$\mathbb{1} = \int \frac{d\phi_{\pm}^j d\pi_{\pm}^j}{2\pi} |\alpha_{\pm}^j\rangle \langle \alpha_{\pm}^j|, \tag{37}$$

but are not orthogonal to each other.

D. Phase space distributions

The phase-space picture enables the description of the system's state ρ^j in terms of classical phase-space distributions. In the following, we will consider the distributions associated with either the $+$ or the $-$ mode, whose corresponding density operators are obtained via the partial trace $\rho_{\pm}^j = \text{Tr}_{\mp} \{\rho^j\}$.

1. Wigner W -distribution

The arguably most prominent phase-space representation is the Wigner W -distribution [25], which is defined as the Fourier transform of the characteristic function, namely [74]

$$\begin{aligned}
 \mathcal{W}_{\pm}^j & \equiv \mathcal{W}_{\pm}^j(\phi_{\pm}^j, \pi_{\pm}^j) \\
 & = \int \frac{d\tilde{\phi}_{\pm}^j d\tilde{\pi}_{\pm}^j}{2\pi} e^{-i(\phi_{\pm}^j, \pi_{\pm}^j)\Omega(\tilde{\phi}_{\pm}^j, \tilde{\pi}_{\pm}^j)^T} \\
 & \quad \times \text{Tr} \left\{ \rho_{\pm}^j e^{i(\phi_{\pm}^j, \pi_{\pm}^j)\Omega(\tilde{\phi}_{\pm}^j, \tilde{\pi}_{\pm}^j)^T} \right\}.
 \end{aligned} \tag{38}$$

Here, we have chosen the normalization such that \mathcal{W}_{\pm}^j is normalized to unity with respect to the phase-space integral measure (30), i.e.

$$1 = \text{Tr} \left\{ \rho_{\pm}^j \right\} = \int d\phi_{\pm}^j d\pi_{\pm}^j \mathcal{W}_{\pm}^j. \quad (39)$$

It is well-known that the Wigner W -distribution can become negative when the underlying state is non-classical [78]. However, we note that both our simulation approaches are naturally based on positive Wigner W -distributions (see section IV).

2. Marginal distributions

Next, we introduce the measurement distributions over the canonical operators

$$\begin{aligned} f_{\pm}^j &\equiv f_{\pm}^j(\phi_{\pm}^j) = \text{Tr} \left\{ \rho_{\pm}^j |\phi_{\pm}^j\rangle \langle \phi_{\pm}^j| \right\}, \\ g_{\pm}^j &\equiv g_{\pm}^j(\pi_{\pm}^j) = \text{Tr} \left\{ \rho_{\pm}^j |\pi_{\pm}^j\rangle \langle \pi_{\pm}^j| \right\}, \end{aligned} \quad (40)$$

with the eigenvalue equations

$$\phi_{\pm}^j |\phi_{\pm}^j\rangle = \phi_{\pm}^j |\phi_{\pm}^j\rangle, \quad \pi_{\pm}^j |\pi_{\pm}^j\rangle = \pi_{\pm}^j |\pi_{\pm}^j\rangle, \quad (41)$$

understood. The distributions (40) are true probability density functions with normalizations

$$1 = \text{Tr} \left\{ \rho_{\pm}^j \right\} = \int d\phi_{\pm}^j f_{\pm}^j = \int d\pi_{\pm}^j g_{\pm}^j, \quad (42)$$

and can be identified with the marginals of the Wigner W -distributions

$$f_{\pm}^j = \int d\pi_{\pm}^j \mathcal{W}_{\pm}^j, \quad g_{\pm}^j = \int d\phi_{\pm}^j \mathcal{W}_{\pm}^j. \quad (43)$$

As such, they encode the full information of ϕ_{\pm}^j and π_{\pm}^j , but no information about their correlations. Hence, in the special case when ϕ_{\pm}^j and π_{\pm}^j are uncorrelated, the corresponding Wigner W -distribution decomposes into a product $W_{\pm}^j = f_{\pm}^j g_{\pm}^j$.

The separate detection scheme, or homodyne detection [79], described in section II E 1 approximately corresponds to sampling from the marginal distributions. Fundamentally, of course, this cannot generally be an exact correspondence for any finite atom number, as the detected quantities $n_{m_F}^j$ have a discrete spectrum (detected atom numbers), while the desired Wigner marginals are continuous. However, in the undepleted pump approximation, the Wigner marginals are generally well-approximated by the homodyne statistics [69, 70].

3. Husimi Q -distribution

Since the coherent state projectors $|\alpha_{\pm}^j\rangle \langle \alpha_{\pm}^j|$ are non-negative and resolve the identity, they constitute a posi-

tive operator-valued measure (POVM). The corresponding measurement distribution is the so-called Husimi Q -distribution [26]

$$\mathcal{Q}_{\pm}^j \equiv \mathcal{Q}_{\pm}^j(\phi_{\pm}^j, \pi_{\pm}^j) = \text{Tr} \left\{ \rho_{\pm}^j |\alpha_{\pm}^j\rangle \langle \alpha_{\pm}^j| \right\}. \quad (44)$$

While the Wigner W -distribution can be negative, the Husimi Q -distribution is *always* non-negative [27], since it stems from a POVM. More generally, it is bounded by $0 \leq \mathcal{Q}_{\pm}^j \leq 1$ [29, 30]. The normalization is induced by (37), leading to

$$1 = \text{Tr} \left\{ \rho_{\pm}^j \right\} = \int \frac{d\phi_{\pm}^j d\pi_{\pm}^j}{2\pi} \mathcal{Q}_{\pm}^j. \quad (45)$$

When operating in the regime $n_0^j \gg n_{\pm 1}^j$ the simultaneous detection scheme described in section II E 2 can be identified with the so-called eight-port-homodyne detection in quantum optics (also referred to as heterodyne detection). In this scheme, each measurement corresponds to drawing a sample from the Husimi- Q -distribution of the signal field [80].

IV. SIMULATING THE DYNAMICS

We simulate the dynamics of the spin-1 BEC in two different approximations. To capture most of the full Hamiltonian and to check the validity of the undepleted pump approximation, we perform a truncated Wigner simulation, see section IV A. Additionally, we derive an analytically solvable model based on (28) in section IV B.

A. Truncated Wigner Approximation

In the regime of high occupations, it is usually infeasible to exactly solve the exponentially complex dynamics in the full Hilbert space \mathcal{H} . Instead, as the populations are increased, the role of quantum fluctuations compared to the mean-field dynamics grows smaller and hence the predictions made by approximate techniques become more and more accurate, justifying a semiclassical approach to the problem. One such semiclassical technique is the truncated Wigner approximation (TWA). At its core, all mode operators (we drop all indices in this subsection for brevity) are replaced by complex numbers [81]

$$\mathbf{a}^{(\dagger)} \rightarrow \alpha^{(*)}, \quad (46)$$

which amounts to a lowest-order expansion of the Wigner-Weyl correspondence rules. Consequently, the Hamiltonian operator $\mathbf{H}(\mathbf{a}, \mathbf{a}^{\dagger})$ reduces to a classical Hamiltonian function $H(\alpha, \alpha^*)$ of the complex-valued phase-space coordinates (α, α^*) .

In general, the von Neumann equation

$$i\partial_t \mathbf{O}(t) = [\mathbf{O}(t), \mathbf{H}], \quad (47)$$

for some time-dependent operator \mathbf{O} , can be translated into an equation of motion for the operator's Weyl symbol O in phase space via the Moyal bracket

$$\{A, B\}_{\text{MB}} = \frac{A}{2} \sin \left[2 \left(\overleftarrow{\partial}_\alpha \overrightarrow{\partial}_{\alpha^*} - \overleftarrow{\partial}_{\alpha^*} \overrightarrow{\partial}_\alpha \right) \right] B, \quad (48)$$

which results in [81]

$$i\partial_t O(t) = \{O(t), H\}_{\text{MB}}. \quad (49)$$

In TWA, one is interested in the evolution of the elementary c -numbers α representing the mode operators to leading order, for which Eq. (49) simplifies to the classical Poisson bracket

$$i\partial_t \alpha(t) = \partial_{\alpha^*(t)} H, \quad (50)$$

since all higher-order derivatives vanish.

In the case when multiple modes are coupled as for the Hamiltonian of interest in Eq. (6), Eq. (50) defines a system of coupled differential equations that may be solved numerically using a suited integrator. Before being propagated in time, the initial set of coordinates $\alpha(0)$ of a given well j and mode m_F is obtained as Monte Carlo samples from the initial Wigner W -distribution, for which two cases are relevant. When considering the vacuum $|0\rangle$, the Wigner W -distribution takes the form of a Gaussian, which is centered at the origin and contains no correlations, i.e.

$$|\psi\rangle = |0\rangle \iff \alpha \sim \mathcal{N} \left[0, \frac{1}{2} \right] + i\mathcal{N} \left[0, \frac{1}{2} \right]. \quad (51)$$

The samples of coherent states $|\gamma\rangle$ have the same variance but are displaced by the square root of their mean particle number $\langle \mathbf{N} \rangle = |\gamma|^2$ from the origin

$$|\psi\rangle = |\gamma\rangle \iff \alpha \sim \mathcal{N} \left[\text{Re}(\gamma), \frac{1}{2} \right] + i\mathcal{N} \left[\text{Im}(\gamma), \frac{1}{2} \right], \quad (52)$$

since $|\gamma\rangle$ itself is obtained by acting with the displacement operator $\mathbf{D}(\gamma)$ on the vacuum, see (34).

In the presence of thermal fluctuations, the pure vacuum state $|0\rangle$ is replaced by the thermal ensemble $\rho \propto \exp(-\beta \mathbf{a}^\dagger \mathbf{a})$ with the inverse temperature $\beta = 1/T$. Accordingly, the standard deviations of the Wigner W -distribution are rescaled as

$$\frac{1}{2} \rightarrow \sqrt{1 + 2n_{\text{BE}}(\beta)} \frac{1}{2}, \quad (53)$$

which is equivalent to adding $n_{\text{BE}}(\beta)$ to both variances, where

$$n_{\text{BE}}(\beta) = \frac{1}{e^\beta - 1} \quad (54)$$

denotes the Bose-Einstein distribution.

Then, the expectation value of some observable \mathbf{O} is obtained as the stochastic average over all generated samples

$$\langle \mathbf{O} \rangle_{\text{TWA}} = \frac{1}{|\mathcal{S}|} \sum_{\alpha \in \mathcal{S}} O(\alpha, \alpha^*), \quad (55)$$

where \mathcal{S} denotes the set of all samples. Note here that we again rely on the correspondence between \mathbf{O} and O given in Eq. (46).

Having introduced all necessary tools in TWA, we want to briefly comment on its regimes of applicability. In the limit of high occupations $n \gg 1$ and initial states with minimal fluctuations, i.e. coherent states, the operator to c -number correspondence Eq. (46) is a justified simplification, since the relative fluctuations scale as $1/\sqrt{n}$. The fluctuations of the initial state are herein represented accurately, as they correspond to exact samples of the Wigner W -distribution of the initial state $|\psi\rangle$, while only those fluctuations that build up during the evolution due to the unitary evolution under \mathbf{H} are not captured. In the limit $\hbar \rightarrow 0$, the dynamics generated by TWA become exact. For a more thorough picture of this matter, see Ref. [82], in which the impossibility of quantum phase-space trajectories for generic (i.e. anharmonic) quantum systems is discussed. For further reading regarding TWA and other semiclassical techniques we refer the reader to [81] and [83].

B. Gaussian Model

Interestingly, the short-time dynamics can be described by a simple Gaussian model to a reasonable extent, which we discuss next.

1. Gaussian Hamiltonian

Starting from the undepleted pump Hamiltonian (28), we make two simplifying assumptions: First, we neglect the atomic collisions between the atoms, i.e. set $c_0 \approx 0$, which amounts to dropping all fourth- and one second-order contribution in (29). This results in a vanishing mixing Hamiltonian $\mathbf{H}_{\text{up}}^{\text{mix}} \approx 0$, which disentangles the \pm phase spaces and hence allows us to study their dynamics independently. Second, despite no explicit translational invariance, an approximate invariance exists in the center of the lattice. Since boundary effects are negligible for the subsystem of our interest (see section V A), we can assume all parameters to be equal across all wells, i.e. $\tilde{q}^j \approx \tilde{q}$, $\tilde{c}_1^j \approx \tilde{c}_1$ and $n^j \approx n$.

The resulting Hamiltonian is Gaussian, i.e. of second order in the relative mode operators, and reads for the $+$ mode (the $-$ mode can be treated on equal footing)

$$\begin{aligned} \mathbf{H}_{\text{up, Gauss}}^+ &= \sum_{j=1}^N \left[\tilde{q} \mathbf{N}_+^j + \frac{\tilde{c}_1}{2} \left(\mathbf{a}_+^j \mathbf{a}_+^j + \mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j\dagger} \right) \right] \\ &+ J \sum_{j=1}^{N \text{ or } N-1} \left(\mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j+1} + \mathbf{a}_+^{(j+1)\dagger} \mathbf{a}_+^j \right). \end{aligned} \quad (56)$$

The dynamics generated by the latter Hamiltonian is illustrated in Figure 3. As a result of its simple form,

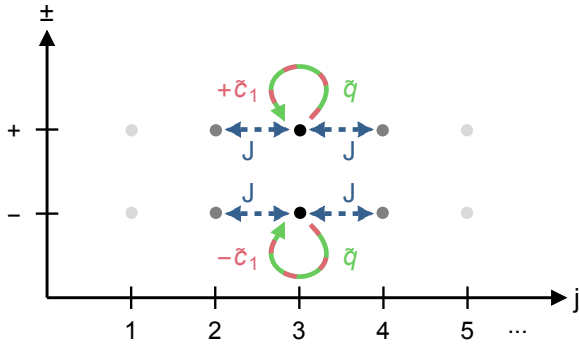


Figure 3: Sketch of the dynamics in the Gaussian regime generated by (56) (in analogy to Figure 1).

we can provide a more detailed description of how the different terms in (56) contribute to the overall dynamics (see Figure 4): The first term proportional to \tilde{q} causes a local rotation (green dashed circle in **d**), while the second term proportional to \tilde{c}_1 generates local squeezing (red dashed arrows in **d**). Both act locally on the j -th. $+$ phase space. The coupling term proportional to J builds up correlations between the wells (blue solid arrow in **i**) by stretching the local distributions proportionally to each other (see **g** versus **h**).

One may also express (56) through the canonical field operators (see Appendix A for details), leading to

$$\begin{aligned} H_{\text{up,Gauss}}^+ &= \frac{1}{2} \sum_{j=1}^N \left[\lambda^+ (\phi_+^j)^2 + \lambda^- (\pi_+^j)^2 \right] \\ &\quad - J \sum_{j=1}^{N \text{ or } N-1} (\phi_+^j \phi_+^{j+1} + \pi_+^j \pi_+^{j+1}) \quad (57) \\ &\equiv \frac{1}{2} \sum_{j,j'=1}^{2N} \chi_+^{jT} \Lambda^{jj'} \chi_+^{j'} \end{aligned}$$

where we introduced the two new couplings

$$\lambda^\pm = \tilde{q} \pm \tilde{c}_1 \quad (58)$$

for brevity¹ and the so-called Hamiltonian matrix Λ . The latter is a $2N \times 2N$, real and symmetric matrix of block-diagonal form

$$\Lambda = \begin{pmatrix} \Lambda^+ & 0 \\ 0 & \Lambda^- \end{pmatrix}, \quad (59)$$

where Λ^+ and Λ^- describe the dynamics of the fields ϕ_+^j and the momentum fields π_+^j , respectively. Both have the same form with non-vanishing entries only on the three

leading diagonals for open boundary conditions, i.e.

$$\Lambda^\pm = \begin{pmatrix} \lambda^\pm & -J & 0 & 0 \\ -J & \lambda^\pm & -J & 0 \\ 0 & -J & \lambda^\pm & \ddots \\ 0 & 0 & \ddots & \ddots \end{pmatrix}, \quad (60)$$

and additionally a contribution $-J$ in the first and last anti-diagonal entries for periodic boundary conditions. Hence, the diagonal entries in (60) describe single wells, while the next-to-leading diagonals (and possibly the outer two entries of the anti-diagonal) contain the coupling between neighboring wells. We also note that the Hamiltonian (57) is bounded from below by a positive number, i.e. has a positive-energy ground state, when $\lambda^\pm \geq 0$, which is fulfilled if $q \geq c_1/2$ and hence in particular for non-negative $q \geq 0$.

2. Mean and covariance matrix

It is well known that Gaussian Hamiltonians map Gaussian states to Gaussian states [74, 75]. Since the initial state $|0, \alpha, 0\rangle^{\otimes N}$ from Eq. (16) corresponds to an uncorrelated set of local vacua in the $+$ phase spaces, the initial state in the undepleted pump approximation is indeed Gaussian. Hence, the dynamics generated by (57) is fully encoded in the two lowest-order moments of the quantum state ρ_+ , i.e. its field expectation values

$$\chi_+^j = \text{Tr}\{\rho_+ \chi_+^j\} \quad (61)$$

and its covariance matrix [74, 75]

$$\gamma_+^{jj'} = \frac{1}{2} \text{Tr}\{\rho_+ \{\chi_+^j - \chi_+^j, \chi_+^{j'} - \chi_+^{j'}\}\}. \quad (62)$$

The latter is a real, symmetric, and non-negative $2N$ -dimensional matrix of block form

$$\gamma_+ = \begin{pmatrix} \mathcal{M}_+ & \mathcal{T}_+ \\ \mathcal{T}_+^T & \mathcal{N}_+ \end{pmatrix}, \quad (63)$$

and contains the two fundamental two-point correlation functions

$$\begin{aligned} \mathcal{M}_+^{jj'} &= \text{Tr}\{\rho_+ \phi_+^j \phi_+^{j'}\} - \phi_+^j \phi_+^{j'}, \\ \mathcal{N}_+^{jj'} &= \text{Tr}\{\rho_+ \pi_+^j \pi_+^{j'}\} - \pi_+^j \pi_+^{j'}, \end{aligned} \quad (64)$$

as well as the mixed correlator

$$\mathcal{T}_+^{jj'} = \text{Tr}\{\rho_+ \phi_+^j \pi_+^{j'}\} - \frac{i}{2} \delta_{jj'} - \phi_+^j \pi_+^{j'}. \quad (65)$$

Importantly, the covariance matrix is constrained by the Robertson-Schrödinger uncertainty relation in the form of a lower bound to its determinant [84–86]

$$\det \gamma_+ \geq \frac{1}{2^{2N}}, \quad (66)$$

¹ Note here that λ^\pm does *not* refer to the \pm phase spaces.

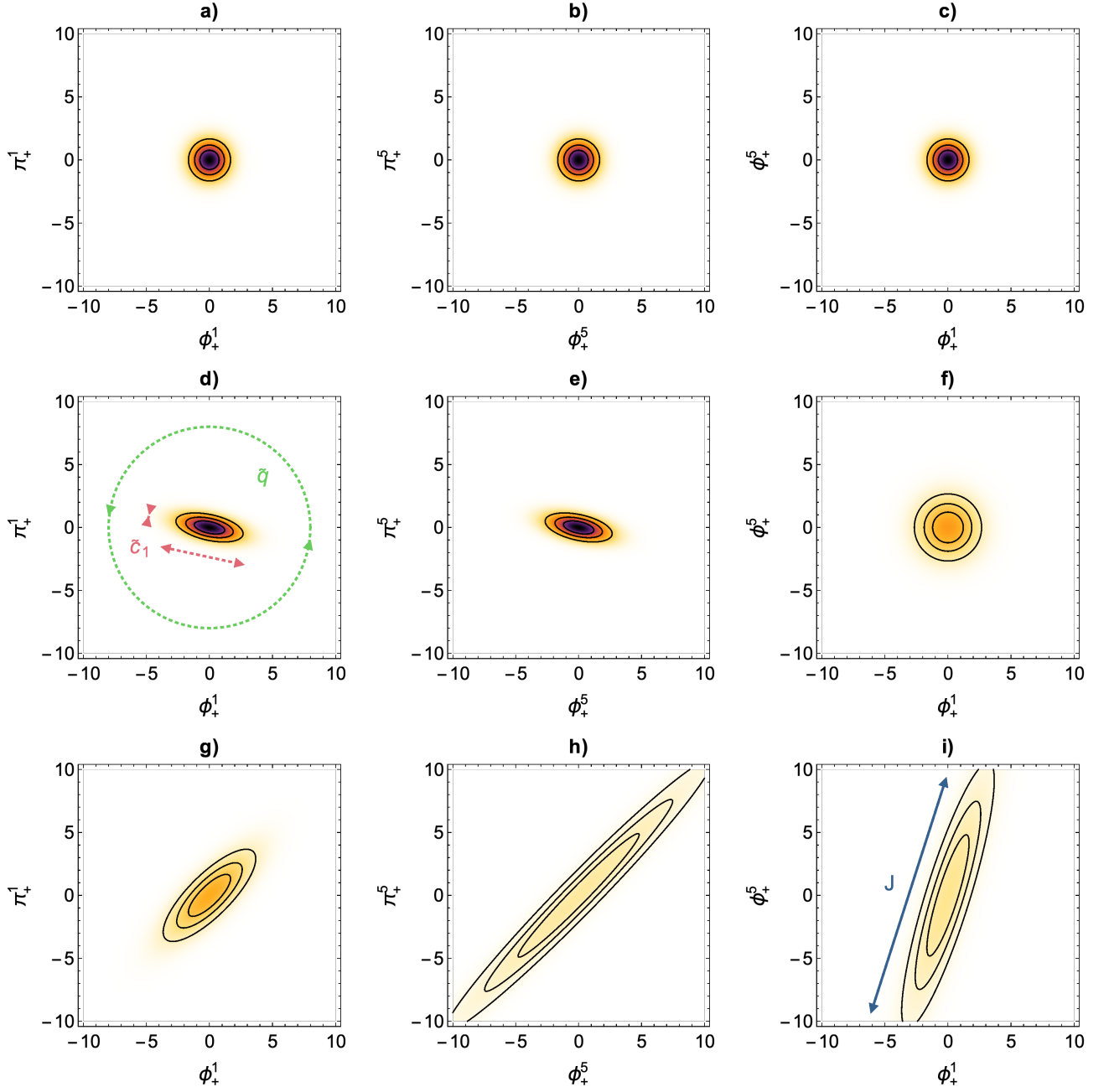


Figure 4: Illustrations of the effects of the three terms in the Gaussian Hamiltonian (56) on the $+$ phase-space distributions for $N = 9$ wells and open boundary conditions. The rows show the initial state, an evolved state for uncoupled wells, i.e. $J = 0$, and an evolved state for coupled wells, i.e. $J > 0$, respectively, while the columns depict the phase-space distributions of the first well, the fifth well in the center as well as the correlations in the fields between these two wells, respectively. For $t = 0$, see **a)**-**c)**, all wells are in the same state, and the wells are uncorrelated. When the uncoupled system has evolved, see **d)**-**f)**, all local distributions are equally squeezed (red dashed arrows) and rotated (green dashed arrow) by \tilde{c}_1 and \tilde{q} , respectively, and remain uncorrelated. For a non-zero tunnel rate $J > 0$, see **g)**-**i)**, the local distributions are stretched proportionally (note that the stretching is equal across the wells for periodic boundary conditions), and correlations between the wells built up (blue solid arrow).

with equality if and only if ρ_+ corresponds to a pure Gaussian state [74, 75]. Thus, for our analytical model, $\det \gamma_+ > 1/2^{2N}$ encodes mixedness of ρ_+ (see also [section VB 2](#)).

Importantly, χ_+ and γ_+ correspond to the first- and second-order moments of the Wigner W -distribution \mathcal{W}_+ , and the diagonal entries of γ_+ are encoded in f_+ and g_+ . While χ_+ is also the mean of the Husimi Q -distribution

\mathcal{Q}_+ , its covariance matrix V_+ acquires an additional contribution, which is a consequence of \mathcal{Q}_+ being a convolution of \mathcal{W}_+ , resulting in

$$V_+ = \gamma_+ + \frac{1}{2} \mathbb{1}. \quad (67)$$

3. Symplectic time evolution

The effect of the Hamiltonian (57) on the two lowest-order moments can be described by symplectic transformations in phase space [74, 75]. To that end, we recall a few basics of Gaussian quantum information theory. The Hamiltonian $\mathbf{H}_{\text{up,Gauss}}^+$ generates the time evolution of the initial state $\rho_+(0)$ at $t = 0$ via the unitary time evolution operator

$$\mathbf{U}_{\text{up,Gauss}}^+(t) = e^{-it\mathbf{H}_{\text{up,Gauss}}^+}, \quad (68)$$

in the sense that

$$\rho_+(t) = \mathbf{U}_{\text{up,Gauss}}^+(t) \rho_+(0) \mathbf{U}_{\text{up,Gauss}}^{+\dagger}(t). \quad (69)$$

In phase space, the unitary transformation $\mathbf{U}_{\text{up,Gauss}}^+(t)$ corresponds to a symplectic transformation $\mathcal{S}_{\text{up,Gauss}}^+(t)$, that is, a transformation which preserves the commutator in phase space (see Eq. (32)), i.e. leaves the symplectic form Ω invariant

$$\mathcal{S}_{\text{up,Gauss}}^+(t) \Omega \mathcal{S}_{\text{up,Gauss}}^{+T}(t) = \Omega. \quad (70)$$

When the Hamiltonian is written as a bilinear form as in the second line in (57), the symplectic time evolution matrix can be expressed through the Hamiltonian matrix Λ via [74, 75]

$$\mathcal{S}_{\text{up,Gauss}}^+(t) = e^{t\Omega\Lambda}. \quad (71)$$

This matrix can directly be applied to the two lowest-order moments of ρ_+ , yielding the simple time evolution equations [74, 75]

$$\begin{aligned} \chi_+(t) &= \mathcal{S}_{\text{up,Gauss}}^+(t) \chi_+(0), \\ \gamma_+(t) &= \mathcal{S}_{\text{up,Gauss}}^+(t) \gamma_+(0) \mathcal{S}_{\text{up,Gauss}}^{+T}(t). \end{aligned} \quad (72)$$

Since the initial state is the uncoupled vacuum state, we have $\chi_+(0) = 0$, and hence all phase-space distributions remain centered around the origin, i.e.

$$\chi_+(t) = \chi_+(0) = 0. \quad (73)$$

Instead, the initial covariance matrix

$$\gamma_+(0) = \frac{1}{2} \mathbb{1}, \quad (74)$$

evolves non-trivially following the second line of (72), which can be computed analytically for a given set of parameters.

Note that for a non-zero initial temperature, the initial covariance matrix is also proportional to the identity, but the variances increase due to thermal fluctuations, which results in

$$\gamma_+(0) = \left[\frac{1}{2} + n_{\text{BE}}(\beta) \right] \mathbb{1}. \quad (75)$$

Note also that adding half the identity to $\gamma_+(t)$ gives the time-evolved covariance matrix $V_+(t)$ of the Husimi Q -distribution. Since $\mathcal{S}\mathcal{S}^T = \mathbb{1}$ for all symplectic matrices \mathcal{S} one may equally apply the symplectic transformation directly to the defining equation of V_+ , i.e. (67).

4. Comment on field theory correspondence

At last, we comment on the field theory the Gaussian Hamiltonian (57) reproduces when taking the continuum limit. To that end, we introduce a lattice spacing ϵ and define the field operators

$$\phi_+(x) \equiv \frac{1}{\sqrt{\epsilon}} \phi_+^j, \quad \pi_+(x) \equiv \frac{1}{\sqrt{\epsilon}} \pi_+^j, \quad (76)$$

which fulfill the distribution-valued commutation relations

$$[\phi_+(x), \pi_+(x')] = i\delta(x - x'), \quad (77)$$

in the continuum limit $\epsilon \rightarrow 0$. Then, using (76) in (57) together with

$$\phi_+(x + \epsilon) = \phi_+(x) + \epsilon \partial_x \phi_+(x) + \mathcal{O}(\epsilon^2), \quad (78)$$

analogously for $\pi_+(x + \epsilon)$, results in

$$\mathbf{H}_{\text{up,Gauss}}^+ = \frac{1}{2} \int dx [\kappa^+(x) \phi_+^2(x) + \kappa^-(x) \pi_+^2(x)], \quad (79)$$

with the differential operators

$$\kappa^\pm(x) = \lambda^\pm + 1 + J\epsilon \partial_x, \quad (80)$$

to leading order in ϵ . Hence, a reasonable continuum limit requires the limit of infinite coupling $J \rightarrow \infty$ with the product $J\epsilon$ kept fixed, which effectively implements $\epsilon = 1/J \rightarrow 0$.

V. AREA LAW IN PHASE SPACE

We now discuss the scaling behaviors of quantum entropies associated with a subsystem. Then, we define *classical* entropies of the phase space distributions introduced in section III D and provide mathematical as well as heuristic arguments for why these entropies encode the area law.

A. Quantum entropies

We consider a subsystem of five wells from the N total wells, which can exchange energy and particles with its complement and therefore should be considered an open quantum system. We partition this subsystem into a subregion A consisting of the first $0 \leq M \leq 5$ wells and its complement B composed out of the remaining $5 - M$ wells (see Figure 5). The multi-well setup offers the possibility to study the discretized version of a continuous quantum field theory with the wells corresponding to lattice points. The local states associated with the subregions are, as usual, defined via the partial trace

$$\rho_+^A = \text{Tr}_B\{\rho_+\} \quad (81)$$

and analogous for ρ_+^B . The local mixedness of A is conveniently measured by the von Neumann entanglement entropy [87]

$$S(\rho_+^A) = -\text{Tr}\{\rho_+^A \ln \rho_+^A\}, \quad (82)$$

which serves as an entanglement measure if and only if the bipartite state ρ_+ is pure, in which case $S(\rho_+^A) > 0$ if and only if A and B are entangled and $S(\rho_+^A) = 0$ if not. However, when the bipartite state is mixed, for example, due to the presence of thermal fluctuations or coupling to the environment as considered here, one typically studies the quantum mutual information instead. It is defined as [88, 89]

$$I(\rho_+^A : \rho_+^B) = \text{Tr}\{\rho_+ (\ln \rho_+ - \ln \rho_+^A \otimes \rho_+^B)\}, \quad (83)$$

or as

$$I(\rho_+^A : \rho_+^B) = S(\rho_+^A) + S(\rho_+^B) - S(\rho_+), \quad (84)$$

if all three terms on the right are finite, and serves as the measure for the total correlations between A and B being zero if and only if $\rho_+ = \rho_+^A \otimes \rho_+^B$. Note that when the initial state is pure, the entanglement entropy and the quantum mutual information are proportional to each other, i.e. $I(\rho_+^A : \rho_+^B) = 2S(\rho_+^A)$.

Both quantities have been studied extensively for a large variety of quantum field theories and quantum many-body systems and it is well-known that for *typical* states of Hamiltonians with local interactions, they scale with the *area* of the entangling surface separating A and B , see e.g. [7, 8, 10, 11, 15] for reviews. In this context, states are called typical for example when they lie sufficiently close to the ground state of the Hamiltonian or when they are generated from an initial product state on short time scales after a quench [15]. In the following, we shall be concerned with the latter scenario.

For our setup, three competing length scales are relevant for the precise scaling with subsystem size M of the aforementioned quantum entropies: the system size N , the (initial) inverse temperature $\beta = 1/T$, and the subsystem size M itself. When the dominating scale is M , i.e. when

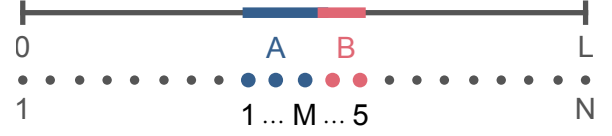


Figure 5: The open subsystem of interest (colored) is partitioned into A (blue) and B (red).

we consider the subinterval size being small compared to the system size and the inverse temperature $M \ll N, \beta$, then the entanglement entropy scales logarithmically (blue solid curve in Figure 6) [7, 8, 10, 11]

$$S(\rho_+^A) \sim \kappa_1 \ln(M + \kappa_2) + \kappa_3, \quad (85)$$

with κ_i being real constants. We normalize all entropic quantities such that they vanish at $M = 0$, resulting in the constraint

$$\kappa_2 = e^{-\kappa_3/\kappa_1}, \quad (86)$$

and hence only κ_1 and κ_3 are independent. Typically, κ_1 is universal, i.e. regularization independent, especially in conformal field theories where it corresponds to the central charge, while κ_3 is usually not.

When instead N is dominant, i.e. when the subinterval size is of the order of the system size $M \sim N$ and the temperature is still small $N \ll \beta$, the so-called finite-size area law holds (petrol dashed curve) [7, 8]

$$S(\rho_+^A) \sim \kappa_1 \ln \left[\frac{N}{\pi} \sin \left(\frac{\pi M}{N} \right) + \kappa_2 \right] + \kappa_3, \quad (87)$$

again with the constraint (86) understood. Since the quantum state of the whole system is close to being pure in this case, we have $S(\rho_+^A) \rightarrow 0$ when $M \rightarrow N$. Further, in the limit of small $M \ll N$, the latter equation reduces to (85) since $(N/\pi) \sin(\pi M/N) = M$ to first order in M .

If finite-temperature effects become relevant, i.e. for $\beta \sim N$, we instead expect a scaling of the form (red dotted curve) [7, 8]

$$S(\rho_+^A) \sim \kappa_1 \ln \left[\frac{\beta}{\pi} \sinh \left(\frac{\pi M}{\beta} \right) + \kappa_2 \right] + \kappa_3, \quad (88)$$

again with (86) understood. In this case, the entropy remains finite if computed for the whole system due to additional (classical) mixedness. While for small temperatures $\beta \gg N$ we obtain back (85), large temperatures $\beta \ll N$ lead to an extensive entropy obeying a volume law (green dot-dashed curve)

$$S(\rho_+^A) \sim \kappa_1 M, \quad (89)$$

in which case classical correlations dominate over the quantum ones.

Remarkably, the quantum mutual information follows the finite-size area law (87) in all of the four considered

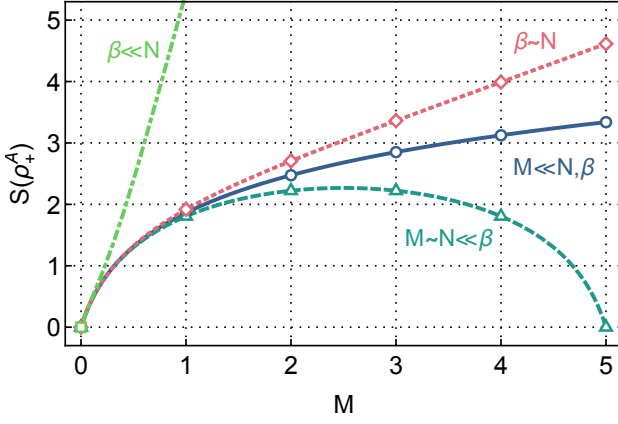


Figure 6: Scaling of the entanglement entropy $S(\rho_+^A)$ with subsystem size M for the four archetypal scenarios: large system size at zero temperature (blue curve), finite system size at zero temperature (petrol curve), finite temperature (red curve) and high temperature (green curve). For small M , the first three curves converge.

cases [18]. In particular, even if thermal fluctuations prevent the appearance of the area law in the entropies, the extensive contributions drop out when considering the quantum mutual information, which demonstrates its utility for observing an area law in actual experiments.

B. Classical entropies

We now introduce various entropic measures in the phase space associated with subsystem A (subsystem B and the full subsystem AB can be treated on equal footing). To that end, we shall group the set of corresponding local field variables according to

$$\chi_+^A = (\phi_+^A, \pi_+^A)^T = (\phi_+^1, \dots, \phi_+^M, \pi_+^1, \dots, \pi_+^M)^T. \quad (90)$$

Then, the distributions corresponding to A are obtained via integrating out all degrees of freedom belonging to subsystem B , for instance

$$\mathcal{W}_+^A = \int d\phi_+^B d\pi_+^B \mathcal{W}_+, \quad (91)$$

and similarly for the other two types of phase space distributions.

1. Standard entropies

Let us start with the Wigner entropy, which is defined as the differential entropy of the Wigner W -distribution, namely

$$S(\mathcal{W}_+^A) = - \int d\phi_+^A d\pi_+^A \mathcal{W}_+^A \ln \mathcal{W}_+^A. \quad (92)$$

It is well-defined, i.e. real, provided that $\mathcal{W}_+^A \geq 0$, otherwise it becomes complex-valued. A lower bound encoding the uncertainty principle has been conjectured in [90–92] and reads

$$S(\mathcal{W}_+^A) \geq S(\bar{\mathcal{W}}_+^A) = M(1 + \ln \pi), \quad (93)$$

with equality if and only if \mathcal{W}_+^A corresponds to a product of pure Gaussian states.

The marginal entropy of f_+^A is defined as

$$S(f_+^A) = - \int d\phi_+^A f_+^A \ln f_+^A, \quad (94)$$

and analogously for g_+^A . Since f_+^A and g_+^A are true probability density functions, their entropies are always well-defined. The corresponding entropic uncertainty relation has been put forward by Białynicki-Birula and Mycielski [93–96] (see also [97, 98] for reviews)

$$S(f_+^A) + S(g_+^A) \geq S(\bar{f}_+^A) + S(\bar{g}_+^A) = M(1 + \ln \pi), \quad (95)$$

which contains the same bound as the Wigner entropy conjecture (93). In fact, (95) would be a simple consequence of (93) when using the subadditivity of entropy, namely

$$S(\mathcal{W}_+^A) \leq S(f_+^A) + S(g_+^A), \quad (96)$$

with equality if and only if $\mathcal{W}_+^A = f_+^A g_+^A$. However, the bound in (95) is less tight since it is attained only for products of squeezed vacuum states.

At last, we introduce the Wehrl entropy [31, 32]

$$S(\mathcal{Q}_+^A) = - \int \frac{d\phi_+^A d\pi_+^A}{(2\pi)^M} \mathcal{Q}_+^A \ln \mathcal{Q}_+^A, \quad (97)$$

which is also always well-defined since $\mathcal{Q}_+^A \geq 0$. It is bounded from below by the Wehrl-Lieb inequality [31–34]

$$S(\mathcal{Q}_+^A) \geq S(\bar{\mathcal{Q}}_+^A) = M, \quad (98)$$

which is tight if and only if the state is a product of pure coherent states. It is also bounded by the Wigner and quantum entropies [37]

$$S(\mathcal{Q}_+^A) \geq S(\mathcal{W}_+^A) - M \ln \pi, S(\rho_+^A). \quad (99)$$

Note here that the additional constant $-M \ln \pi$ is a result of the different normalizations of the Wigner W - and Husimi Q -distributions. Therefore, (98) would also be a direct consequence of (93).

The three aforementioned types of entropies have in common that their lower bounds (93), (95) and (98) are attained for a product of pure vacuum states and scale with the volume of the subsystem A . In this sense, all these entropies are *classical* to leading order for arbitrary states with the main contribution coming from the vacuum, in contrast to the quantum entropies introduced in section V A which vanish for all pure states. In what follows, we will argue that quantum features such as the area law are still present in classical entropies, but hidden in the next-to-leading order terms.

2. Subtracted entropies

Recently, it has been argued that quantum features such as the area law of the entanglement entropy are also present in classical entropies when the leading order extensive contributions are subtracted [24]. Based on this idea, we define *subtracted* classical entropies by subtracting the extensive vacuum contributions appearing in the bounds of the corresponding entropic uncertainty relations, to wit

$$\begin{aligned}\Delta S(\mathcal{W}_+^A) &\equiv S(\mathcal{W}_+^A) - S(\bar{\mathcal{W}}_+^A), \\ \Delta S(f_+^A, g_+^A) &\equiv S(f_+^A) + S(g_+^A) - S(\bar{f}_+^A) - S(\bar{g}_+^A), \\ \Delta S(\mathcal{Q}_+^A) &\equiv S(\mathcal{Q}_+^A) - S(\bar{\mathcal{Q}}_+^A).\end{aligned}\quad (100)$$

Let us provide some simple arguments for why these entropies encode the very same features as their quantum analogs in the following.

First, we consider Gaussian quantum states, which provide a reasonable first-order approximation to the states generated by the Hamiltonian (6), see section IV B. A general Gaussian Wigner W -distribution is of the form

$$\mathcal{W}_+^A = \frac{1}{Z_+^A} e^{-\frac{1}{2}(x_+^A)^T (\gamma_+^A)^{-1} x_+^A}, \quad (101)$$

with $Z_+^A = (2\pi)^M \sqrt{\det \gamma_+^A}$ being a normalization constant. Computing the subtracted Wigner entropy for such a state gives [75, 99]

$$\Delta S(\mathcal{W}_+^A) = \frac{1}{2} \ln (2^{2M} \det \gamma_+^A) = S_2(\rho_+^A), \quad (102)$$

where

$$S_2(\rho_+^A) = -\ln \mathbf{Tr}\{(\rho_+^A)^2\} \quad (103)$$

denotes the quantum Rényi-2 entropy of the state ρ_+^A that corresponds to \mathcal{W}_+^A . Hence, the subtracted Wigner entropy *coincides* with a quantum entropy for Gaussian states. For completeness, let us also give the Gaussian expressions for the Husimi-based quantities, which read

$$\mathcal{Q}_+^A = \frac{1}{Z_+^A} e^{-\frac{1}{2}(x_+^A)^T (V_+^A)^{-1} x_+^A}, \quad (104)$$

with $Z_+^A = \sqrt{\det V_+^A}$ and

$$\Delta S(\mathcal{Q}_+^A) = \frac{1}{2} \ln \det V_+^A. \quad (105)$$

Note, however, that these entropies can not be related with quantum entropies in general.

For arbitrary states, including Wigner-negative states, a general relation between subtracted classical and quantum entropies can only be established for the subtracted Rényi-2 Wigner entropy [100, 101]

$$\Delta S_2(\mathcal{W}_+^A) = S_2(\rho_+^A). \quad (106)$$

Note here that the entropic order alters the lower bound in the corresponding entropic uncertainty relation and hence a different term compared to $\Delta S(\mathcal{W}_+^A)$ has to be subtracted. Although the relation (106) can *not* be generalized to other subtracted Rényi-Wigner entropies, it has been shown in [24] that the crucial feature of quantum entropies, i.e. the area law, is present for the entire family of subtracted Rényi-Wigner entropies beyond the Gaussian case, indicating that the area law may also appear for the classical entropies defined in (100), which we investigate in detail in section VII.

C. Classical mutual informations

Let us also introduce classical mutual informations as measures for correlations in phase space. For Wigner-positive states, we define the Wigner mutual information

$$I(\mathcal{W}_+^A : \mathcal{W}_+^B) = S(\mathcal{W}_+^A) + S(\mathcal{W}_+^B) - S(\mathcal{W}_+), \quad (107)$$

while for general states we define the marginal mutual informations

$$I(f_+^A : f_+^B) = S(f_+^A) + S(f_+^B) - S(f_+), \quad (108)$$

(similarly for g_+) and the Wehrl mutual information [37]

$$I(\mathcal{Q}_+^A : \mathcal{Q}_+^B) = S(\mathcal{Q}_+^A) + S(\mathcal{Q}_+^B) - S(\mathcal{Q}_+). \quad (109)$$

All of them are non-negative functionals being zero if and only if the two local distributions on A and B are uncorrelated.

It has been shown that every classical mutual information is a lower bound to the quantum mutual information [37, 102]. In particular, we have

$$I(\mathcal{W}_+^A : \mathcal{W}_+^B), I(\mathcal{Q}_+^A : \mathcal{Q}_+^B) \leq I(\rho_+^A : \rho_+^B), \quad (110)$$

indicating that classical mutual informations do *not* capture all correlations in general. However, the bound (110) is expected to be tighter than the standard second-order lower bounds on the quantum mutual information in terms of two-point correlation functions (which only faithfully describe Gaussian correlations).

In contrast to classical entropies, extensive contributions naturally cancel when considering classical mutual informations. For example, decomposing (107) in the sense of (84) shows that

$$I(\mathcal{W}_+^A : \mathcal{W}_+^B) = \Delta S(\mathcal{W}_+^A) + \Delta S(\mathcal{W}_+^B) - \Delta S(\mathcal{W}_+), \quad (111)$$

since the vacuum is uncorrelated, i.e. $S(\bar{\mathcal{W}}_+) = S(\bar{\mathcal{W}}_+^A) + S(\bar{\mathcal{W}}_+^B)$. For Gaussian states, we can find a relation analogous to (102), which reads [99]

$$\begin{aligned}I(\mathcal{W}_+^A : \mathcal{W}_+^B) &= \frac{1}{2} \ln \frac{\det \gamma_+^A \det \gamma_+^B}{\det \gamma_+} \\ &= I_2(\rho_+^A : \rho_+^B),\end{aligned}\quad (112)$$

with the Rényi-2 mutual information being defined as

$$I_2(\rho_+^A : \rho_+^B) = S_2(\rho_+^A) + S_2(\rho_+^B) - S_2(\rho_+). \quad (113)$$

Fortunately, the partly heuristic arguments for the appearance of the area law for classical entropies in [section VB2](#) can be made rigorous when considering classical mutual informations by adapting the argument in [18]. To that end, we first note that all three Hamiltonians (6), (28) and (56) can be decomposed according to

$$\mathbf{H} = \mathbf{H}^A + \mathbf{H}^B + \mathbf{H}^\partial, \quad (114)$$

where \mathbf{H}^A and \mathbf{H}^B denote the Hamiltonians of subsystems A and B , respectively, while \mathbf{H}^∂ contains the local interactions coupling A to B . Further, we consider the rather general class of thermal states for a given Hamiltonian \mathbf{H} , i.e.

$$\rho_+ = \frac{1}{Z} e^{-\beta \mathbf{H}}, \quad (115)$$

which are typically of non-Gaussian form. Then, every distribution \mathcal{O}_+ over measurement outcomes o_+ corresponding to some POVM \mathcal{O}_+ can be written as

$$\mathcal{O}_+ = \frac{1}{Z} e^{-\beta H}, \quad (116)$$

with the classical Hamiltonian H being defined implicitly via

$$\text{Tr}\{e^{-H} \mathcal{O}_+\} = e^{-H}. \quad (117)$$

For such a distribution, the classical entropy evaluates to

$$S(\mathcal{O}_+) = \ln Z + \beta \langle H \rangle_{\mathcal{O}_+}, \quad (118)$$

where

$$\langle H \rangle_{\mathcal{O}_+} = \int do_+ \mathcal{O}_+(o_+) H \quad (119)$$

is the classical energy expectation value with respect to the distribution $\mathcal{O}_+(o_+)$. Using (118), we find that the (classical) free energy is nothing but

$$F(\mathcal{O}_+) \equiv -\frac{\ln Z}{\beta} = \langle H \rangle_{\mathcal{O}_+} - \frac{S(\mathcal{O}_+)}{\beta}. \quad (120)$$

The latter is minimized by the thermal distribution \mathcal{O}_+ , a fact which can be derived, for instance, using the non-negativity of the classical relative entropy of any given distribution with respect to the thermal distribution \mathcal{O}_+ . As a special case, we can conclude that the thermal free energy is bounded from above by the free energy of the corresponding product distribution, i.e.

$$F(\mathcal{O}_+) \leq F(\mathcal{O}_+^A \mathcal{O}_+^B). \quad (121)$$

Expanding the latter inequality in the sense of (120) and using the additivity of classical entropies for product distributions yields

$$\begin{aligned} & S(\mathcal{O}_+^A) + S(\mathcal{O}_+^B) - S(\mathcal{O}_+) \\ & \leq \beta \left(\langle H \rangle_{\mathcal{O}_+^A \mathcal{O}_+^B} - \langle H \rangle_{\mathcal{O}_+} \right). \end{aligned} \quad (122)$$

The decomposition (114) implies

$$\langle H \rangle_{\mathcal{O}_+^A \mathcal{O}_+^B} - \langle H \rangle_{\mathcal{O}_+} = \langle H^\partial \rangle_{\mathcal{O}_+^A \mathcal{O}_+^B} - \langle H^\partial \rangle_{\mathcal{O}_+}, \quad (123)$$

which, together with the definition of the classical mutual information, finally leads to

$$I(\mathcal{O}_+^A : \mathcal{O}_+^B) \leq \beta \left(\langle H^\partial \rangle_{\mathcal{O}_+^A \mathcal{O}_+^B} - \langle H^\partial \rangle_{\mathcal{O}_+} \right). \quad (124)$$

The derived upper bound on the classical mutual information is a function of the classical boundary Hamiltonian H^∂ only, which proves the area law for all variants of classical mutual information, including the three of our interest, for local interactions and thermal states.

VI. ENTROPY ESTIMATION

We now turn to the challenging task of estimating classical entropies from finitely many samples. Albeit being a general problem, it arises especially for the system of our interest, since experimental runs are particularly costly. To that end, we rely on suitable k -nearest-neighbor estimators and benchmark their validity for various scenarios. For our implementation, we utilize the ‘Non-parametric Entropy Estimation Toolbox’ for Python publicly available at [103].

A. k -nearest-neighbor estimator

The integrals that give the Wigner or Wehrl entropy, Eqs. (92) and (97), respectively, are in general hard to estimate from samples. Unless the distribution is of a specific form, such as a Gaussian, that allows reexpressing the entropy through low-order correlators, one must use the available samples to approximate local densities of the distribution. A straightforward approach is to bin the samples onto a grid and use the relative frequencies in the histogram as approximations for the local densities. While this procedure presents an asymptotically unbiased estimator in the limit of infinite samples and small bins, it is problematic in practice, since the binning procedure presents an information bottleneck as it coarse grains the information about the precise sample position.

A more elaborate approach is to use the k -nearest-neighbor (k NN) statistics of the sample set to approximate local densities [49–63]. The intuition behind the nearest-neighbor statistics is that samples with short distances to their neighbors are located in regions of high probability density, while samples which are located in regions of low probability density have nearest neighbors that are far away. A formalization of this intuition allows to construct an asymptotically unbiased discretization-free estimator \hat{S} for differential entropies [49, 56]

$$\hat{S}(k, N_s) = g(k, N_s, d) + \frac{d}{N_s} \sum_{i=1}^{N_s} \log \epsilon^i(k) \quad (125)$$

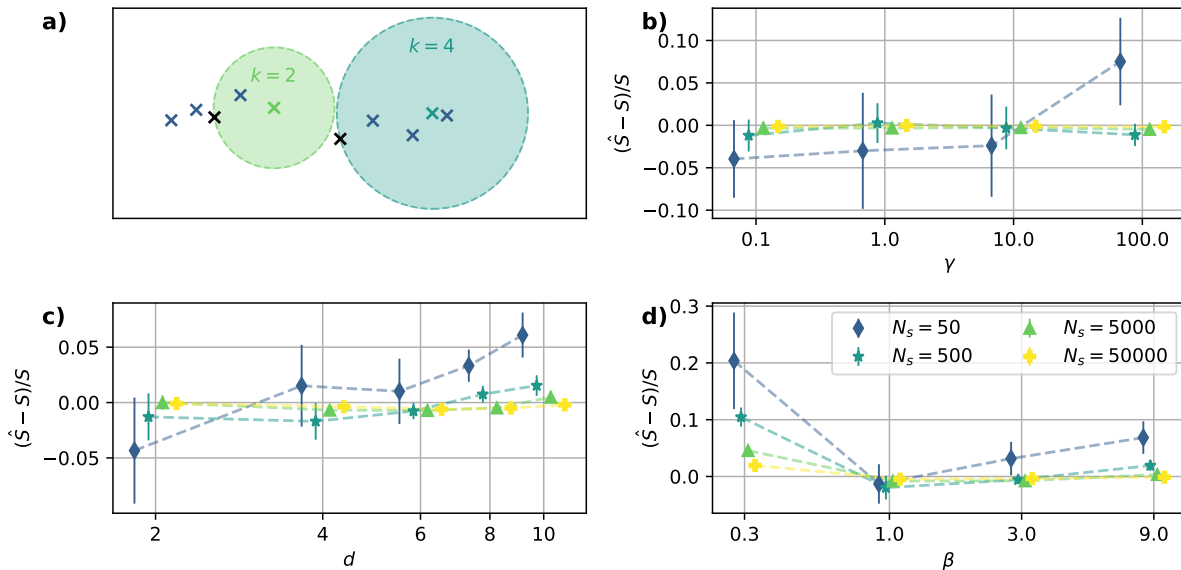


Figure 7: Benchmark of the entropy estimator \hat{S} for distributions that allow for an analytic comparison to S . In **a)**, we visualize the algorithm for a few Gaussian samples in $d = 2$ dimensions, with two random samples selected and two choices for the hyperparameter k , namely $k = 2$ and $k = 4$. The only assumption of the estimator is constant density in the light green and cyan circles respectively, an assumption less justified the larger k . Both k -th neighbors are shown in black. In **b)** we benchmark the entropy estimator on a squeezed covariance matrix in $d = 2$ dimensions with squeezing parameter γ , where large γ correspond to strong squeezing. The covariance matrix is given by $\Sigma = (\mathbf{1} + \gamma J) / \det(\mathbf{1} + \gamma J)^{1/d}$ where J is the matrix of all ones. All error bars are obtained from ten independent estimates and we use offsets on the x axis to aid in differentiating the different data points. In **c)** we let $\gamma = 0$ and vary the dimension d , observing decreasing performance with increasing dimension, as is expected. In **d)** we tune away from the Gaussian regime in $d = 4$ dimensions, using a generalized normal distribution [104] parameterized by β with scale α chosen such that the entropy S takes the value of a standard normal distribution in $d = 4$ dimensions. As can be seen from the density function in Eq. (127), small values of β correspond to heavy tails, $\beta = 2$ corresponds to a Gaussian shape, and the support of the distribution shrinks to the interval $[-1, 1]$ for $\beta \rightarrow \infty$. We denote the sample set size with N_s and choose the nearest neighbor parameter k to be three in all experiments.

where d is the dimension of the distribution, N_s is the number of samples and $\epsilon^i(k)$ is the distance of the i -th sample to its k -th neighbor. The term $g(k, N_s, d)$ is a constant that is independent of the samples and is given by [49, 56]

$$g(k, N_s, d) = -\psi(k) + \psi(N_s) + \log(c_d), \quad (126)$$

where ψ is the digamma function and c_d is the volume of the d -dimensional unit ball.

The asymptotical unbiasedness of the estimator means that one is guaranteed to obtain the true entropy in the limit of infinitely many samples, which directly stems from the observation that the sample distances must shrink to zero in this limit. This is in contrast to approaches that fit a probability density function of a specific form to the data, which may result in more accurate estimates in the regime of few samples, but will generally not converge to the true value when the amount of samples is increased.

We aim to give an intuitive picture of the k NN method in Figure 7 a), where the spheres, for which constant density is assumed, are drawn for $k = 2$ and $k = 4$ for two

different samples. This example reveals the only trade-off in the k NN algorithm, which is the so-called hyperparameter choice k . For small k , the assumption of constant density has stronger justification as the spheres are necessarily smaller compared to larger k . This, however, comes at the expense of stronger fluctuations, meaning larger statistical uncertainties. These can be systematically reduced by choosing larger k , as the volume of the spheres shows a power law in the dimension of the distribution. Typically, k is set to $k = 3$ or similar, depending on the situation at hand [103].

Since the estimation of differential entropies from samples is a task of fundamental interest [105] with applications ranging from statistics [106–108] to signal processing [109, 110] to machine learning and pattern recognition [111–114], various other paths beside the aforementioned one have been explored [115]. There exists a rich literature for non-parametric entropy estimators, that do not assume a specific form of the underlying probability density [49–55, 57–63]. These are either based on nearest neighbor statistics or arrive at an estimate for the

local density by using kernel density estimates. Recently, machine learning-inspired techniques have been explored. Giving up on the non-parametric property of the estimator results in the loss of the asymptotically unbiasedness at possibly increased performance. Numerous works have been put forward in this regard, with applications of estimating entropies[116] and mutual informations [117].

B. Benchmarks in low dimensions

We showcase the performance of the estimator for Gaussian distributions in [Figure 7 b](#)). The data points are obtained from 10 independent runs, meaning that a total of $10N_s$ data points were generated, from which we extracted means and standard deviations on the relative deviations $(\hat{S} - S)/S$. We study the performance of the algorithm in $d = 2$ dimensions as a function of the squeezing parameter γ , where $\gamma = 0$ corresponds to a normal distribution and large γ implies higher squeezing. Here, $S = 2.83$ for all values of γ . A central assumption in the derivation of the expression in [Eq. \(125\)](#) is uniform density inside the d -dimensional ball with radius $\epsilon^i(k)$ between sample i and its k -th neighbor. For larger squeezing, this assumption is violated more strongly, since the squeezing introduces a preferred direction at odds with the assumed isotropy, resulting in worse performance for larger γ .

C. Scaling with dimension

In [Figure 7 c](#)), we benchmark the estimator on a standard normal distribution for increasing dimension. The higher dimension implies that the assumption of constant density inside the d -dimensional spheres of radius $\epsilon^i(k)$ must hold for larger volumes if the number of samples is held constant. This will generally result in worse performance, as demonstrated.

D. Tuning away from the Gaussian regime

To test the estimator's performance away from the Gaussian regime, we benchmark it on a generalized normal distribution [104]. The latter has a probability density of the form

$$f_{\text{GND}}(x; \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left|\frac{x - \mu}{\alpha}\right|^\beta\right), \quad (127)$$

with mean μ , scale $\alpha > 0$, shape $\beta > 0$ and Γ denoting the gamma function. In [Figure 7 d](#)), we vary α and β in such a way that the true entropy S is kept constant at the value corresponding to the standard normal distribution in $d = 4$ dimensions, for which $\alpha = 1/2, \beta = 2$, i.e. $S = 5.67$. We observe the worst performance for small values of β , i.e. in the regime of heavy tails, which is expected as the central assumption of constant density of

the employed estimator needs to hold for larger regions in space. For large values of β and sufficiently many samples, the results improve (for $N_s = 5000$ (50000) the relative deviations for $\beta = 1$, $\beta = 3$ and $\beta = 9$ are -0.008, -0.007 and 0.003 (-0.004, -0.003 and -0.001)), in line with the intuition that the entropy of a uniform distribution should be easy to estimate. However, for $N_s = 50$ and $N_s = 500$ samples, this intuition seems to be misleading as the best performance is observed for $\beta = 1$ and $\beta = 3$, respectively, which resemble Gaussian-like distributions.

VII. RESULTS

We are now ready to present our main results: analytical and numerical observations of the area law for classical entropies. After a general comparison of the analytical model and TWA in terms of two-point correlators and entropic quantities, we systematically study typical effects, including non-Gaussian features, a thermal initial state, the influence of the subsystem's position within the total system, the total system size, the dependence of the estimated entropies on the number of samples and finally two types of boundary conditions.

Unless specified differently, we use the values $c_1 = -1/n$, $n = 1000$, such that our energy scale is given by $|nc_1| = 1$ (note that this renders *all* parameters and other quantities of our interest dimensionless). Further, we consider Lithium-7 for which $c_0 = 2c_1$, set $q = 2J$ with $J = 2$ and test $N = 20$ wells with open boundary conditions at the three different times $t = 0.5, 0.75, 1$. From the 20 wells, we study five wells located in the center of the lattice, i.e. wells 8, 9, 10, 11, 12, and neglect the remaining 15 wells from our examination, leaving us with the subsystem depicted in [Figure 5](#). We assume the initial temperature to be zero, i.e. $T = 0$, and base the entropy estimation on 10^4 samples. We systematically vary these parameters one by one and investigate their influences from [section VII C](#) to [section VII H](#). The plotted curves are either dashed or solid, corresponding to interpolations or fits of the (finite-size) area or the volume law.

A. Correlation matrices

As a first qualitative comparison between the analytical model and the truncated Wigner approximation that takes into account all interactions contained in [Eq. \(6\)](#), we show the correlation matrices between the phase-space variables ϕ_+ and π_+ in [Figure 8](#) for the five wells under scrutiny at the three different times. The correlation matrix is given by the normalized covariances

$$C_+^{jj'} = \frac{\gamma_+^{jj'}}{\sqrt{\gamma_+^{jj} \gamma_+^{j'j'}}}, \quad (128)$$

thereby rendering all entries to lie between -1 and 1. We want to emphasize that the analytical model is *fully* char-

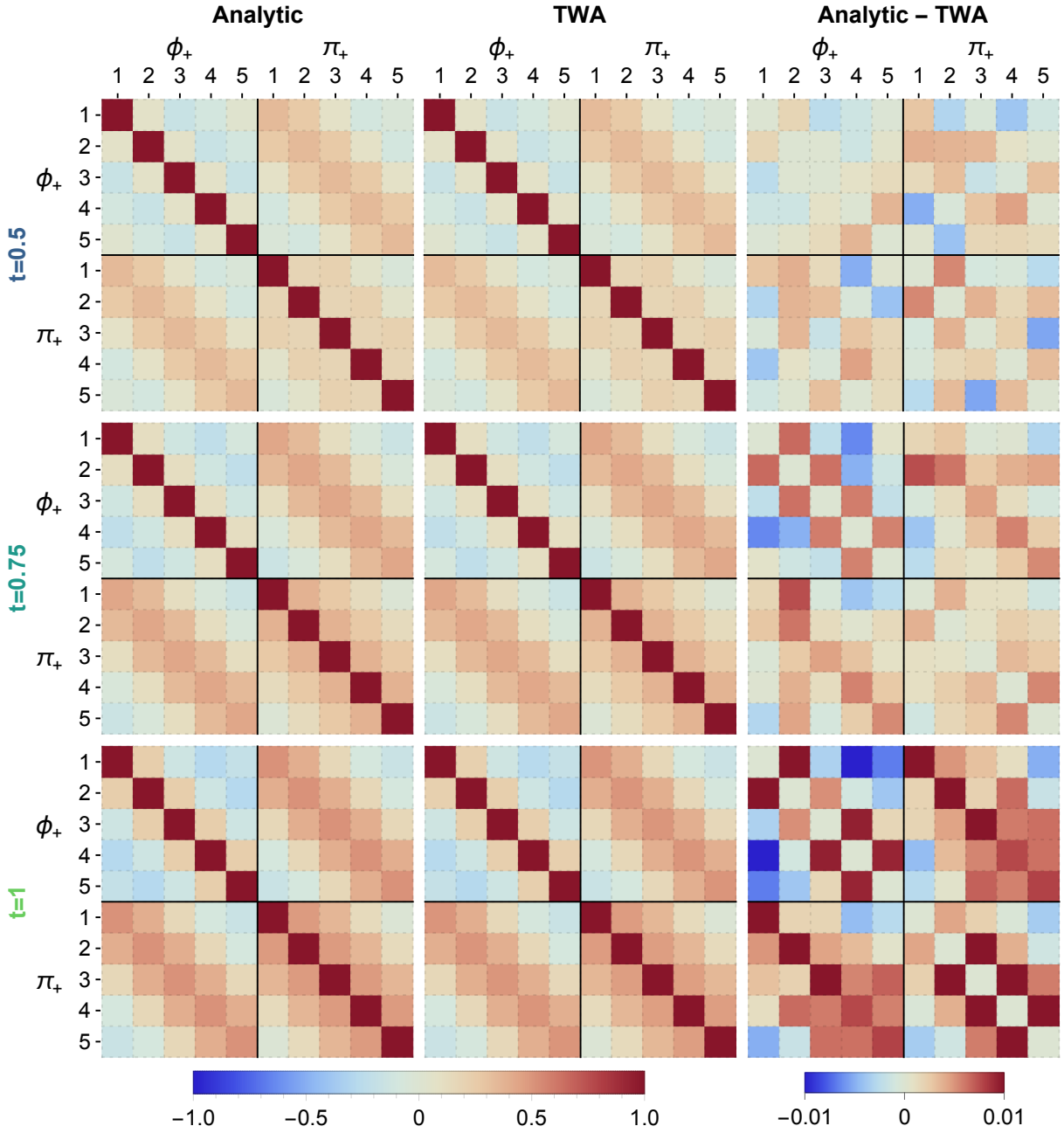


Figure 8: Time evolution of the correlation matrices (128) generated by the analytic Hamiltonian (57) (left column) and the TWA simulation (middle column) together with their difference (right column).

acterized by the observables in Eq. (128), reducing the complexity of extracting entropies to an extraction of its second moments. In contrast to that, the full Hamiltonian Eq. (6) contains more complex, i.e. higher-order terms, meaning that observables that contain higher-order fluctuations are required to fully characterize entropies.

Since the initial polar state ($t = 0$) in (16) is uncorrelated, its correlation matrix is the identity matrix. During the evolution, correlations between different wells are building up through the transport of atoms permitted by

the tunnel Hamiltonian given in Eq. (8). Simultaneously, correlations between ϕ_+^j and π_+^j within one well are generated through the internal dynamics in Eq. (7), leading to the complex structures visible in Figure 8. Remarkably, the analytical model captures most of the features visible in the normalized second-order correlators for the three times considered, with the absolute value of deviations not exceeding 1% (see right column).

Note, however, that this analysis does not permit any statements about the agreement of higher-order moments.

These can in principle deviate from a Gaussian prediction, due to the presence of fourth-order terms in Eq. (6). Crucially, we therefore do not assume the distribution to obey a specific functional form in our subsequent analysis of sample data generated by the means of TWA with regards to entropic quantities.

The good agreement between the two models leads us to conclude that the analytical model is a justified approximation in regimes where the contribution of fourth-order terms to the full Hamiltonian Eq. (6) is negligible. These regimes are characterized by small values of c_0 , i.e. Lithium-7, and large values of q , such that the undepleted pump approximation (see section III A) is justified. Furthermore, we expect the approximations made by the analytical model to be valid, especially for early times, which is why we restrict ourselves to times up to $t = 1$. An extended discussion for later times can be found in section VII C and in [64].

B. Observation of the area law

Having gained a qualitative understanding of the dynamics generated by the Hamiltonian (6), we now turn our attention to the dynamics of entropic quantities associated with the quantum state's phase space and measurement distributions. In this context, we want to discuss our main result Figure 9, which demonstrates the numerical observation of the area law from experimentally extractable quantities without assuming a specific functional form of the quantum state. Crucially, we only require 10^4 samples to estimate entropies of up to ten-dimensional distributions, which we deem experimentally feasible.

In the following, we consider the three distributions of interest, namely

- the Wigner W -distribution, introduced in Eq. (38) which has mainly theoretical interest due to its connection to quantum entropies in the early-time regime and its restricted accessibility,
- its marginal distributions, introduced in Eq. (40), motivated by the fact that they allow for direct experimental extraction as described in section II E 1,
- and finally the Husimi Q -distribution, introduced in Eq. (44) and directly measurable in experiments using the readout techniques explored in [40, 41] and described in section II E 2,

for subsystem A of varying size M , such that B contains $5 - M$ wells (see Figure 5). For the aforementioned distributions, we compute the absolute and subtracted entropies as well as the mutual information for the evolution times $t = 0.5$, $t = 0.75$, and $t = 1$. We compare the numerical TWA data (open markers) to the analytical data (solid markers) and fit (finite-size) area laws (solid lines).

We first discuss the absolute entropies shown in the first row of Figure 9, i.e. a), b), c). All entropies show similar

linear behavior and the area law is masked behind the leading order terms. These leading order contributions are subtracted in Figure 9 d), e) and f), as discussed in section VB 2, unveiling the area law in form of the typical logarithmic growth in all three types of distributions. Finally, the finite-size area law becomes apparent in all three classical mutual informations, see Figure 9 g) - i).

At this point, let us emphasize the challenges associated with the task of estimating the subtracted entropies: We are interested in the estimation of entropies of up to ten-dimensional distributions at an accuracy that is between one and two orders of magnitude higher than the signal itself (consider e.g. $\Delta S(W_+^A)/S(W_+^A) \approx 0.03$ for $M = 5$ wells at $t = 0.5$). Under this perspective, we consider the agreement between analytical and TWA results as remarkable.

This concludes the presentation of our main results. In the following, we carry out ablative studies to demonstrate the genericness of our experimental proposal. In particular, we explore later times (section VII C), where the distributions show stronger non-Gaussian features, consider imperfectly prepared, thermal initial states (section VII D) and investigate the influence of boundary effects by shifting the position of the five wells within the total system of $N = 20$ wells (section VII E). We also reduce the total system to $N = 5$ wells to hunt the finite-size area law for the subtracted entropies (section VII F). Finally, we aim to shed light on the sample complexity associated with these tasks when utilizing the k NN-estimator (section VII G) before considering the effect of periodic boundary conditions (section VII H).

C. Distributions: Gaussian versus non-Gaussian

To test the influence of non-Gaussian features of the distribution we consider the parameters $q = 2$ and $J = 0.5$ at later times, namely $t = 2$, $t = 3$, and $t = 4$. We expect stronger non-Gaussian features to emerge for smaller values of q , as we cross the polar to easy-plane ferromagnet phase transition [68], due to the (less detuned) squeezing generated by c_1 . Similar to Figure 8, we plot the discrepancies between the correlation matrices obtained with the analytical model and TWA for the three different times. Figure 10 shows the differences now become more pronounced, growing as large as 0.1, more than an order of magnitude larger compared to Figure 8. We particularly expect the analytical model to mainly give reliable results for the early time dynamics, so a disparity between the two is expected.

To gain more intuition about the characteristics of the distribution at those times, we show samples (ϕ_+^1, π_+^1) of the Wigner W -distribution of the first well in the second row of Figure 10, with all other degrees of freedom integrated out. While the local distribution at $t = 2$ still seems relatively Gaussian, an analysis of its 4th-moments reveals that Gaussianity is already strongly violated, with Isserlis' theorem [118] showing relative deviations as large

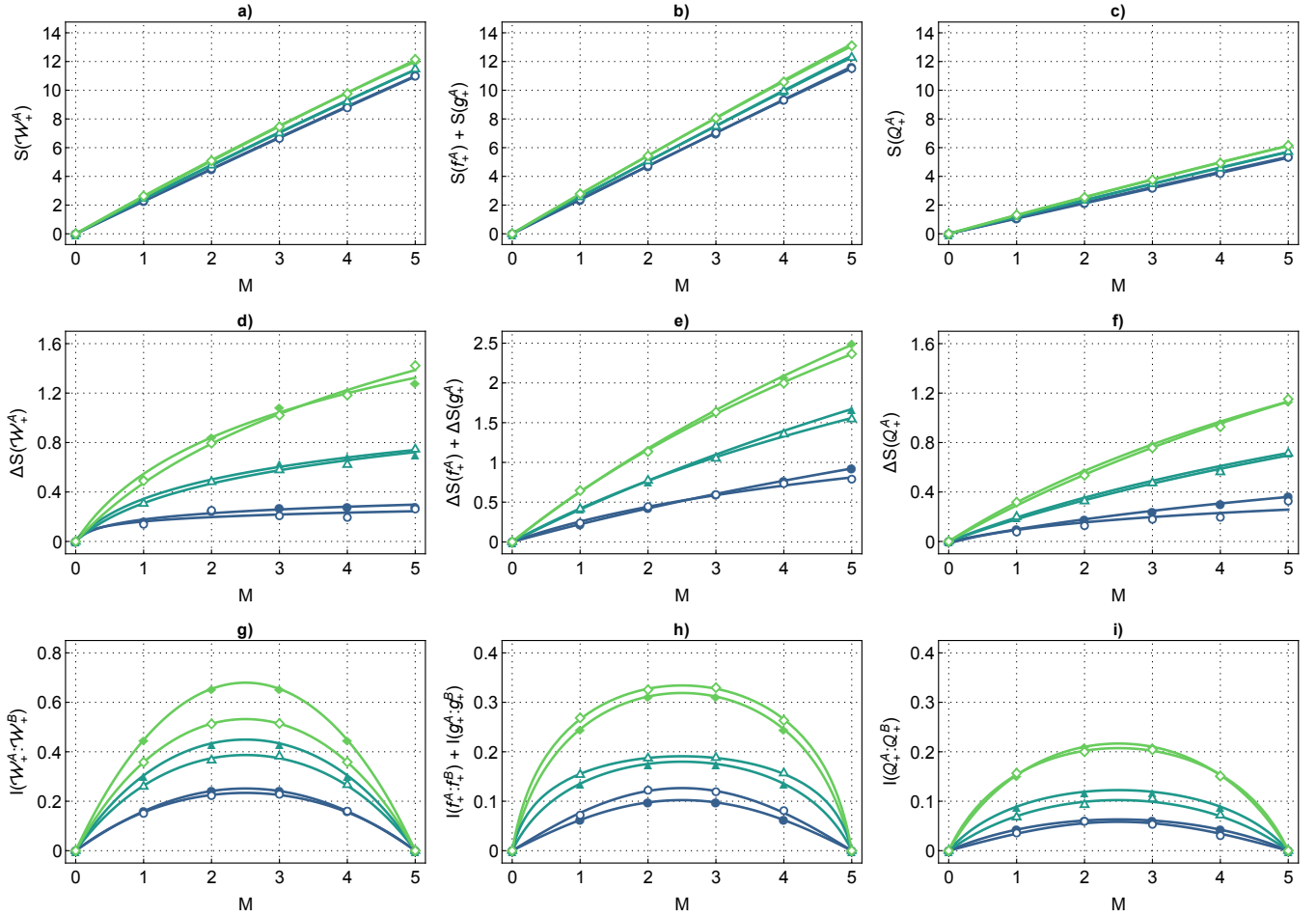


Figure 9: The upper, middle, and lower rows correspond to absolute entropies, subtracted entropies, and mutual information, respectively, while the left, middle, and right columns show Wigner, marginal, and Wehrl quantities. Solid (open) markers correspond to analytical (TWA) data, while solid lines depict fitted curves. The fit function for **d** - **f**) is the area law (85), with an additional extensive term for **a**) - **c**). For **g**) - **i**), we fit the finite-size area law (87).

as 4%. For $t = 3$ ($t = 4$) these deviations grow as large as 17% (60%). Hence, also the entropic quantities show stronger deviations. In the case of the Wigner W -distribution in Figure 10 **g**) substantial differences occur for more wells and late times, as the high dimensionality as well as the strong non-Gaussian features make it difficult for the estimator to reliably estimate the entropies using the 10^4 available samples. The mutual information shown in Figure 10 **h**) and **i**), show strong deviations between TWA and the analytical model, although a quantitative comparison does not seem meaningful given the strong non-Gaussian features shown in Figure 10 **d**) - **f**). A more informative observation is that the estimates of the mutual informations still change significantly upon increasing the sample size by one order of magnitude, albeit not changing their functional form. In the case of the Wigner mutual information, this underestimation is around 15% with respect to the currently given data, while it is around 7% for the Wehrl mutual information. We attribute this difference to the increased smoothness

of the Husimi Q -distribution compared to the Wigner W -distribution.

D. Initial state: vacuum versus thermal

Up to now, our discussion has not been concerned with experimental imperfections such as thermal noise. We here want to address this issue by considering thermal initial states that undergo the unitary evolution. These initial states are characterized by a temperature scale β that we choose to be on the order of a tenth of the system size, such that $\beta = 2$. The thermal fluctuations alter the computations of the dynamics by adding thermal noise to the quantum one-half, see Eqs. (53) and (75).

As expected, the subtracted Wigner entropy $\Delta S(W_+^A)$ rather features an extensive volume law than an area law, as can be seen in Figure 11 **a**). When computing both the Wigner and Wehrl mutual information in Figure 11 **b**) and **c**), we observe the finite-size area law in both

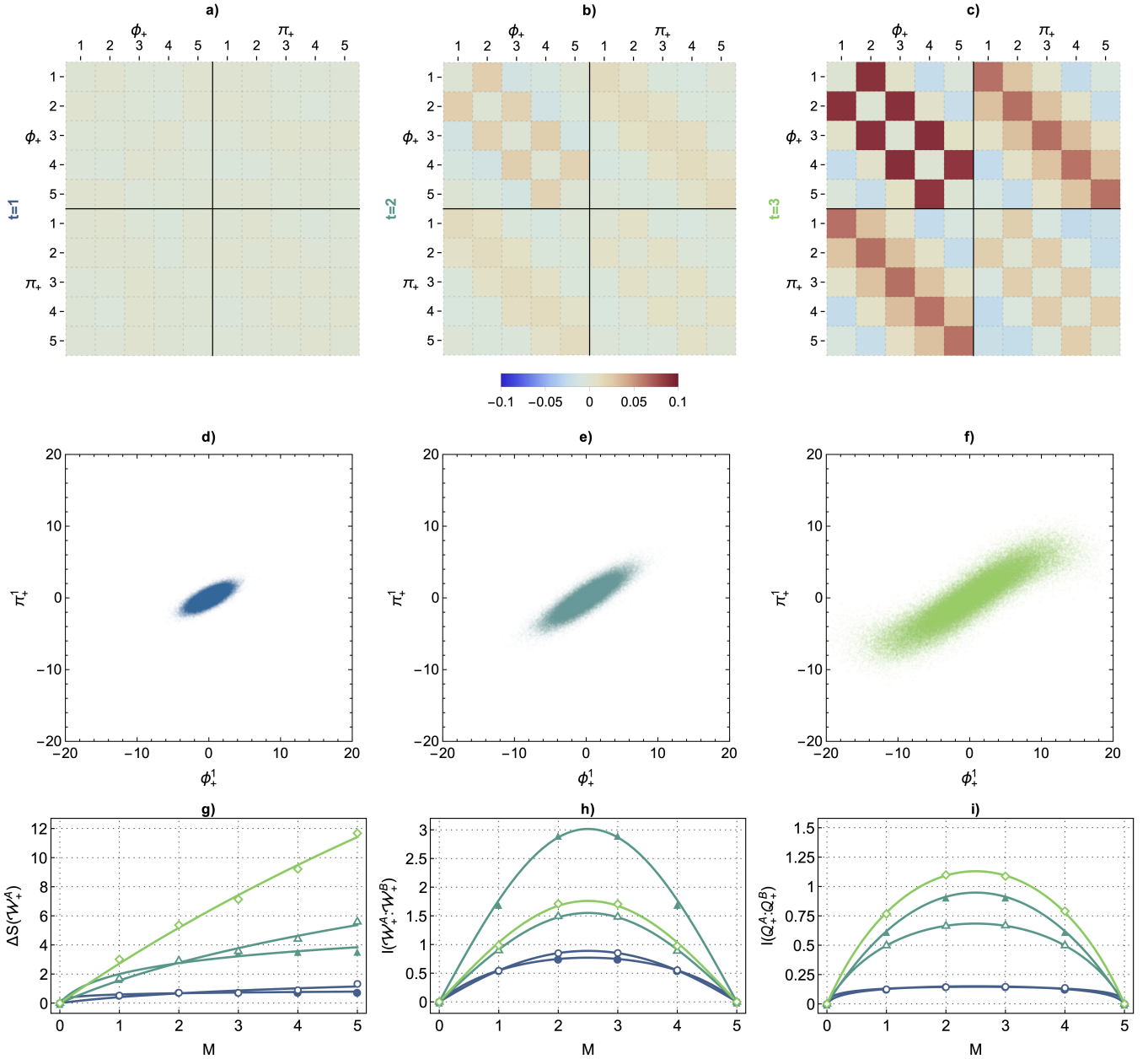


Figure 10: The differences between the analytical model and TWA in the correlation matrices are shown in **a)**, **b)** and **c)** for the three times $t = 1$, $t = 2$ and $t = 3$, respectively, and $q = 2$, $J = 0.5$. The discrepancies grow as large as 0.1, such that we can no longer meaningfully speak about agreement between the methods. The non-Gaussian form of the underlying phase-space distributions is illustrated in **d)** - **f)** in terms of sample points (ϕ_+^1, π_+^1) of the first well. The corresponding subtracted Wigner entropy, Wigner mutual information, and Wehrl mutual information are shown in **g)** - **i)**, for the analytical model (solid markers, only up to $t = 2$) as well as TWA (open markers). We observe the area law for the former and the finite-size area law for the two latter quantities, also in the non-Gaussian regime.

instances, just as in [Figure 9](#). The Wigner mutual information shows the same quantitative behavior observed in [Figure 9 g\)](#), since the contributions due to the initial thermal noise cancel out. In contrast, the Wehrl mutual information, is increased compared to [Figure 9 i\)](#). We attribute this to the inequality [\(110\)](#) becoming tight in the infinite temperature limit [\[102\]](#) (this is also evident

from [\(67\)](#): the additional term $(1/2)\mathbb{1}$ becomes irrelevant in this case). Importantly, the finite-size area law [\(87\)](#) describes the measurable and noisy data well, rendering the proposal robust against thermal noise. Hence, the considered measurement distributions indeed describe a suited setup to observe the area law experimentally.

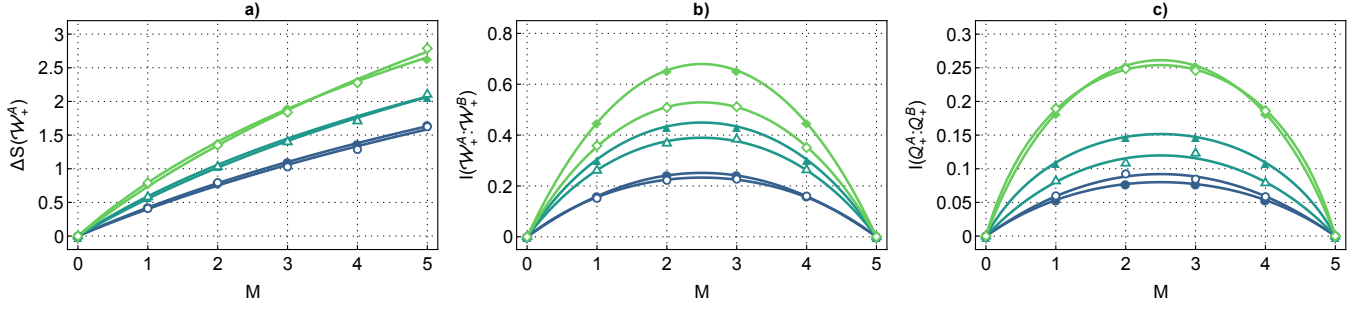


Figure 11: Entropic curves for a thermal initial state with $T = 1/2$. The fits correspond to the area law in **a)** and the finite-size area law in **b)** and **c)**, respectively.

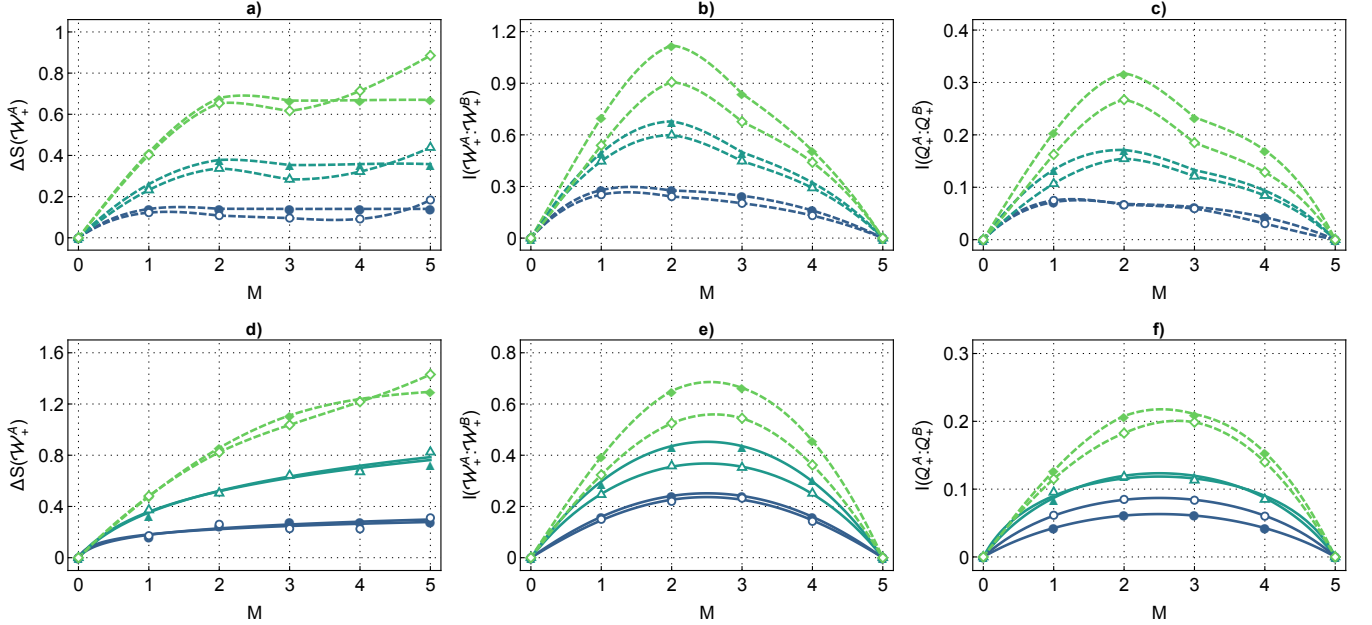


Figure 12: Entropic curves for the wells $M = 1 - 5$ (upper row) and $M = 3 - 7$ (lower row), i.e. when picking five wells out of $N = 20$ and partitioning these five wells into two subsystems as shown in [Figure 5](#). Near the boundary, the area laws are strongly distorted by boundary effects, illustrated by the interpolated curves in **a)** - **c)**. Further away from the boundary, boundary effects become only relevant when the boundary information had enough time to affect local quantities (note that all curves at $t = 1$ are interpolations), see e.g. the asymmetry of the mutual informations in **e)** and **f)**.

E. Subsystem position: center vs. outward

In this subsection, we will choose the position of the subsystem within the bigger system to be wells $M = 1 - 5$ and $M = 3 - 7$, to investigate how the boundary affects the area law.

After computing the subtracted Wigner entropy as well as the Wigner and Wehrl mutual informations for the two subsystem positions, we find the data to be well described by the area law for those instances where boundary effects are irrelevant. The upper row in [Figure 12](#) shows that the subsystem 1-5 is strongly affected by the boundary on the left, and no area law can be observed. In contrast, the data shown in the lower for the subsystem 3-7 is captured well

by the fitted curves, albeit the mutual informations are showing some slight asymmetry for $t = 1.0$, i.e. when the boundary information has had sufficient time to propagate to the considered subsystem.

F. System size: large vs. small

In contrast to the previous discussion, we now restrict the total system to $N = 5$ wells such that the evolution within the considered region is unitary. The small system size implies that the dynamics is dominated by boundary effects and the regular logarithmic area law is replaced by the finite-size area law also in the case of the subtracted classical entropies. More precisely, since the density ma-

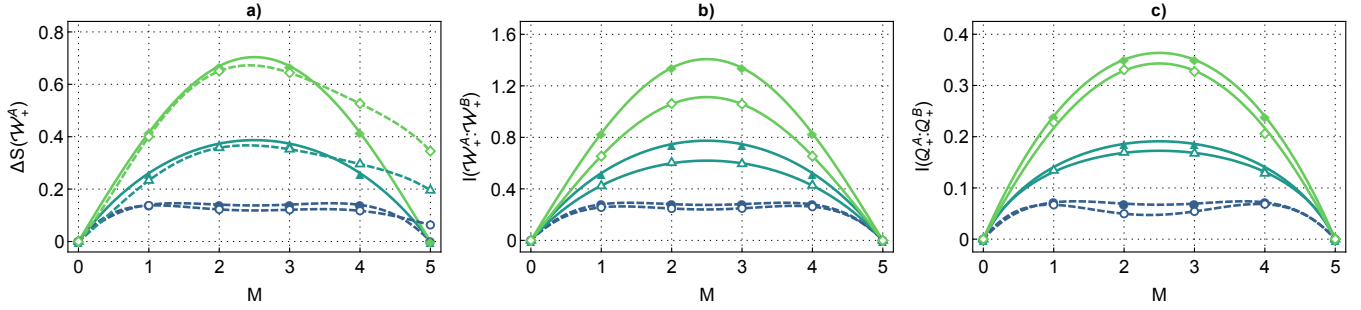


Figure 13: When the total system size equals the subsystem size, i.e. when $N = 5$, the area law for the subtracted classical entropies translates into a finite-size area law reflecting the purity of the global state. We remark the underperformance of the entropy estimator for the subtracted Wigner entropies in **a)** and that the build-up of the finite-size area law only occurs after a transient phase around $t = 0.5$. In both cases, the dashed curves correspond to interpolations, which we show for illustrative purposes.

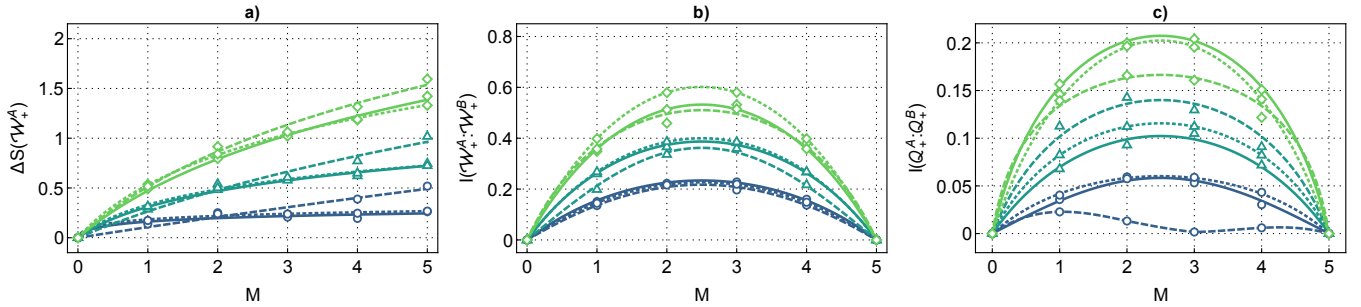


Figure 14: The feasibility of the entropy estimator is checked by varying sample size, where dashed, solid, and dotted lines correspond to 10^3 , 10^4 , and 10^5 samples. All curves are fits, the only exception being the 10^3 curve at $t = 0.5$ in **c)**. The convergence between 10^4 and 10^5 samples becomes visible especially for the subtracted Wigner entropy in **a)** and the Wehrl mutual information in **c)**.

trix of the five-well system remains pure throughout the evolution, the subtracted Wigner entropy returns to the value zero once the entire system is considered, as shown in [Figure 13 a\)](#). The reason why this is not captured with the TWA data is due to an insufficient amount of samples. Upon increasing the sample number to 10^5 , the subtracted Wigner entropy is much closer to zero for $M = 5$ wells (0.00, 0.06, and 0.18 instead of 0.06, 0.20, and 0.34 for $t = 0.50$, $t = 0.75$ and $t = 1.00$). The Wigner and Wehrl mutual information both feature the area law, but only starting at later times $t \geq 0.75$.

G. Sample size: small versus large

Thus far, we estimated entropies of the distributions generated through TWA using 10^4 samples, without questioning whether the k NN estimator had converged for the given sample set. Similar to the benchmarks presented in [section VI A](#), we want to better understand the sample complexity of the entropy estimation using the k NN-estimator. To this end, we carry out the estimations of the subtracted Wigner entropy, the Wigner mutual information and the Wehrl mutual information in [Fig-](#)

[ure 14 a\)](#), **b)** and **c)**, respectively, for 10^3 , 10^4 and 10^5 samples, indicated through dashed, solid and dotted lines respectively.

The subtracted Wigner entropy seems to be converged for all times when using 10^4 samples, as no significant differences exist to the data points that were obtained using 10^5 samples. The biggest visual differences exist for the Wigner mutual information for late times, which however mainly stem from the different scale between [Figure 14 a\)](#) and **b)**. For the experimentally relevant Wehrl mutual information, we observe convergence for 10^4 samples for almost all times, with an even further reduced scale compared to **b)**, rendering a potential experimental implementation efficient.

H. Boundary conditions: open versus periodic

While experimental setups will typically use open boundary conditions, a question of theoretical interest is whether the presented framework is sensitive to the type of boundaries that are employed. Therefore, we modify the tunnel Hamiltonian (8) to also allow atoms to jump from the last to the first well and show the same

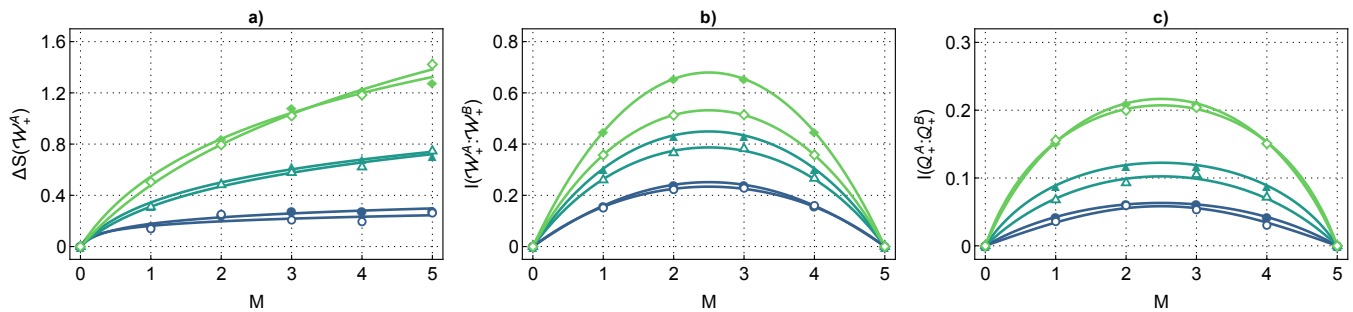


Figure 15: Entropic curves for periodic boundary conditions. As boundary effects are negligible for the subsystem under consideration, all curves agree with those in Figure 9 to a good extent.

quantities as previously in Figure 15. Also in this setting, we find that the area law persists. More precisely, the curves resemble those obtained for open boundary conditions (Figure 9), since the considered subsystem is sufficiently far away from the total system’s boundaries.

VIII. DISCUSSION

We have demonstrated the experimental feasibility of observing the area law in the subtracted classical entropies and mutual informations of measurable distributions in a multi-well spin-1 Bose-Einstein condensate using numerical techniques. As a testbed, we chose a system consisting of 20 wells that feature internal dynamics as well as tunneling among wells and read out two observables that obey a canonical commutation relation for early times. Focusing on a subregion of five wells, we find (finite-size) area laws in the Wigner distribution, its marginals as well as the Husimi-Q distribution, where the latter two lend themselves to experimental investigation using independent subsequent measurements [119] and a readout technique proposed in [40, 41], respectively. We have shown that all types of subtracted entropies as well as mutual informations are quantities worth studying in order to observe the area law and substantiated its robustness with respect to changing various parameters of the experimental setup from section VII C to section VIII H. The discussed area laws are observed dynamically, meaning that the underlying distributions are generated through quenches rather than ground-state preparations, simplifying the experimental requirements. The proposed procedure makes no assumptions regarding the functional form of the distribution under scrutiny, by estimating its differential entropy using a suited k -nearest-neighbor estimator to locally estimate the density at each sample point. In section VII G we demonstrated that for the estimator to converge we require on the order of 10^4 samples, which we deem experimentally feasible.

We simulated the system’s dynamics using an analytical model derived from the full Hamiltonian and TWA. While the analytical model only features quadratic terms corresponding to an integrable system, the dynamics predicted

by the TWA are more intricate and lead to late-time equilibration due to finite particle number and the non-integrable nature of the mean-field equations of motion. We accordingly found agreement between the two methods at early times and found pronounced differences at times $t \geq 2$. Particularly for parameter choices that feature stronger non-Gaussian features, such as a lower value of q violating the undepleted pump approximation more strongly, we observed greater disparities between TWA and the analytical model.

In the non-Gaussian regime, we found the k NN-estimator to require more samples to reach convergence, hinting at the difficulty of estimating entropies of ten-dimensional distributions while being agnostic to their functional form. As the k NN estimator is asymptotically unbiased, it is always possible to increase or decrease the sample size, to check whether the estimator has converged. We wish to state that while the k NN-estimator has a great appeal due to the guarantee of converging to the true entropy using sufficiently many samples, the estimation of differential entropies is a long-standing challenge and various ideas have been put forward to tackle the task more efficiently [49–63, 103, 105–117]. Of particular interest is the variational approach, which consists of adapting the parameters of a defined function by maximizing the likelihood of the observed samples. While this approach can work well, it is i) generally difficult to build a class of functions that includes the observed density distribution and ii) challenging to converge to the global minimum, rendering the approach uncontrolled. However, future research may aim to build functional forms that are directed at estimating the densities generated by specific classes of Hamiltonians, for which one may hope to arrive at more efficient estimators. Nevertheless, this would come at the expense of losing generality, which may be a significant drawback when applying the methodology to an experiment that suffers from undesired noise effects that one cannot account for analytically.

Having established the presence of quantum features in entropies of classical distributions, one can envision various ways forward. It will be interesting to further investigate which other platforms lend themselves to the herein-discussed approach and obey the area law. Par-

ticularly systems that are fully described by the set of observables being read out and systems with other types of degrees of freedom that are constrained by algebras different from canonical commutation relations are of interest. For instance, one may consider the hitherto discussed spin-1 BEC in the pseudo-spin 1/2 configuration, in which case three spin operators forming an $\mathfrak{su}(2)$ algebra constitute a complete representation of the system's state with the phase space distribution being represented on a sphere [76, 119]. But also non-atomic setups, such as integrated photonic waveguides may present promising candidate systems. We emphasize that the herein presented approach and analyses are applicable to any setup capable of preparing or dynamically generating an area law.

It will be equally interesting to see which other quantum phenomena can be accessed using classical entropies of measurement distributions. In this context, let us also highlight the two accompanying publications that study thermalization and the long time limit of the herein discussed system through the lens of classical entropies [64] and a work that is concerned with showing the generality of the area law for classical entropies [24].

ACKNOWLEDGEMENTS

The authors want to thank Thomas Gasenzer and Ido Sivovitz for helpful discussions during the development of the manuscript. T. H. is supported by the European

Union under project ShoQC within the ERA-NET Cofund in Quantum Technologies (QuantERA) program, as well as by the F.R.S.- FNRS under project CHEQS within the Excellence of Science (EOS) program. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster) and within the Collaborative Research Center SFB1225 (ISOQUANT). This work was partially financed by the Baden-Württemberg Stiftung gGmbH. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [120] at Jülich Supercomputing Centre (JSC).

Appendix A: Operator identities for the undepleted pump regime

We list a few identities to reexpress the Hamiltonian (20) in terms of the relative modes as well as the canonical variables. For the step from the local modes to the relative modes we invert (23), which yields

$$\mathbf{a}_1^j = \frac{1}{\sqrt{2}} (\mathbf{a}_+^j + \mathbf{a}_-^j), \quad \mathbf{a}_{-1}^j = \frac{1}{\sqrt{2}} (\mathbf{a}_+^j - \mathbf{a}_-^j), \quad (\text{A1})$$

leading to the identities

$$\begin{aligned} \mathbf{a}_1^{j\dagger} \mathbf{a}_1^{j\dagger} \mathbf{a}_1^j \mathbf{a}_1^j + \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^j \mathbf{a}_{-1}^j + 2N_1^j N_{-1}^j &= \mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j\dagger} \mathbf{a}_+^j \mathbf{a}_+^j + \mathbf{a}_-^{j\dagger} \mathbf{a}_-^{j\dagger} \mathbf{a}_-^j \mathbf{a}_-^j + 2N_+^j N_-^j, \\ \mathbf{a}_1^{j\dagger} \mathbf{a}_1^{j\dagger} + \mathbf{a}_1^j \mathbf{a}_1^j + \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^{j\dagger} + \mathbf{a}_{-1}^j \mathbf{a}_{-1}^j &= \mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j\dagger} + \mathbf{a}_+^j \mathbf{a}_+^j + \mathbf{a}_-^{j\dagger} \mathbf{a}_-^{j\dagger} + \mathbf{a}_-^j \mathbf{a}_-^j, \\ \mathbf{a}_1^j \mathbf{a}_{-1}^j + \mathbf{a}_1^{j\dagger} \mathbf{a}_{-1}^{j\dagger} &= \frac{1}{2} (\mathbf{a}_+^j \mathbf{a}_+^j + \mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j\dagger} - \mathbf{a}_-^j \mathbf{a}_-^j - \mathbf{a}_-^{j\dagger} \mathbf{a}_-^{j\dagger}), \\ N_1^j - N_{-1}^j &= \mathbf{a}_+^{j\dagger} \mathbf{a}_-^j + \mathbf{a}_-^{j\dagger} \mathbf{a}_+^j, \\ N_1^j + N_{-1}^j &= N_+^j + N_-^j, \\ \mathbf{a}_1^{j\dagger} \mathbf{a}_1^{j+1} + \mathbf{a}_1^{(j+1)\dagger} \mathbf{a}_1^j + \mathbf{a}_{-1}^{j\dagger} \mathbf{a}_{-1}^{j+1} + \mathbf{a}_{-1}^{(j+1)\dagger} \mathbf{a}_{-1}^j &= \mathbf{a}_+^{j\dagger} \mathbf{a}_+^{j+1} + \mathbf{a}_+^{(j+1)\dagger} \mathbf{a}_+^j + \mathbf{a}_-^{j\dagger} \mathbf{a}_-^{j+1} + \mathbf{a}_-^{(j+1)\dagger} \mathbf{a}_-^j. \end{aligned} \quad (\text{A2})$$

For the step from relative modes operators to canonical operators we invert (25), which gives

$$\mathbf{a}_\pm^j = \frac{1}{\sqrt{2}} (\phi_\pm^j + i\pi_\pm^j), \quad (\text{A3})$$

implying the identities

$$\begin{aligned} N_\pm^j &= \frac{1}{2} \left[(\phi_\pm^j)^2 + (\pi_\pm^j)^2 - 1 \right], \\ \mathbf{a}_\pm^{j\dagger} \mathbf{a}_\pm^{j\dagger} + \mathbf{a}_\pm^j \mathbf{a}_\pm^j &= (\phi_\pm^j)^2 - (\pi_\pm^j)^2, \\ \mathbf{a}_\pm^{j\dagger} \mathbf{a}_\pm^{j+1} + \mathbf{a}_\pm^{(j+1)\dagger} \mathbf{a}_\pm^j &= \phi_\pm^j \phi_\pm^{j+1} + \pi_\pm^j \pi_\pm^{j+1}. \end{aligned} \quad (\text{A4})$$

- [1] J. D. Bekenstein, Black holes and the second law, *Lett. Nuovo Cimento* **4**, 737 (1972).
 [2] J. D. Bekenstein, Black Holes and Entropy, *Phys. Rev.*

- D* **7**, 2333 (1973).
 [3] S. W. Hawking, Particle creation by black holes, *Comm. Math. Phys.* **43**, 199 (1975).

- [4] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, Quantum source of entropy for black holes, *Physical Review D* **34**, 373 (1986).
- [5] M. Srednicki, Entropy and area, *Physical Review Letters* **71**, 666 (1993).
- [6] C. Callan and F. Wilczek, On geometric entropy, *Physics Letters B* **333**, 55 (1994).
- [7] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, *J. Stat. Mech. Theo. Exp.* **2004**, P06002 (2004).
- [8] P. Calabrese and J. Cardy, Entanglement Entropy and Quantum Field Theory: A Non-Technical Introduction, *Int. J. Quantum Inf.* **04**, 429 (2006).
- [9] M. B. Hastings, An area law for one-dimensional quantum systems, *J. Stat. Mech.: Theo. Exp.* **2007**, P08024 (2007).
- [10] P. Calabrese and J. Cardy, Entanglement entropy and conformal field theory, *J. Phys. A Math. Theo.* **42**, 504005 (2009).
- [11] H. Casini and M. Huerta, Entanglement entropy in free quantum field theory, *J. Phys. A Math. Theo.* **42**, 504007 (2009).
- [12] S. Hollands and K. Sanders, *Entanglement Measures and their Properties in Quantum Field Theory* (Springer, 2018).
- [13] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, *Rev. Mod. Phys.* **80**, 517 (2008).
- [14] I. Peschel and V. Eisler, Reduced density matrices and entanglement entropy in free lattice models, *J. Phys. A: Math. Theo.* **42**, 504003 (2009).
- [15] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [16] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Entropy Scaling and Simulability by Matrix Product States, *Phys. Rev. Lett.* **100**, 030504 (2008).
- [17] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [18] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Area laws in quantum systems: Mutual information and correlations, *Phys. Rev. Lett.* **100**, 070502 (2008).
- [19] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, Measuring entanglement entropy in a quantum many-body system, *Nature* **528**, 77 (2015).
- [20] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, *Science* **353**, 794 (2016).
- [21] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, *Science* **364**, 260 (2019).
- [22] A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, and P. Zoller, Rényi Entropies from Random Quenches in Atomic Hubbard and Spin Models, *Phys. Rev. Lett.* **120**, 050406 (2018).
- [23] M. Tajik, I. Kukuljan, S. Sotiriadis, B. Rauer, T. Schweigler, F. Cataldini, J. Sabino, F. Møller, P. Schüttelkopf, S.-C. Ji, D. Sels, E. Demler, and J. Schmiedmayer, Verification of the area law of mutual information in a quantum field simulator, *Nat. Phys.* **19**, 1022 (2012).
- [24] T. Haas, Area Law from classical entropies, In preparation, 5 (2024).
- [25] E. P. Wigner, On the Quantum Correction For Thermodynamic Equilibrium, *Phys. Rev.* **40**, 749 (1932).
- [26] K. Husimi, Some formal properties of the density matrix, *Proc. Phys.-Math. Soc. Jap. 3rd Ser.* **22**, 264 (1940).
- [27] N. D. Cartwright, A non-negative Wigner-type distribution, *Phys. A* **83**, 210 (1976).
- [28] H.-W. Lee, Theory and application of the quantum phase-space distribution functions, *Phys. Rep.* **259**, 147 (1995).
- [29] W. P. Schleich, *Quantum Optics in Phase Space* (Wiley-VCH Verlag Berlin, 2001).
- [30] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, 2013).
- [31] A. Wehrl, General properties of entropy, *Rev. Mod. Phys.* **50**, 221 (1978).
- [32] A. Wehrl, On the relation between classical and quantum-mechanical entropy, *Rep. Math. Phys.* **16**, 353 (1979).
- [33] E. H. Lieb, Proof of an entropy conjecture of Wehrl, *Commun. Math. Phys.* **62**, 35 (1978).
- [34] E. H. Lieb and J. P. Solovej, Proof of an entropy conjecture for Bloch coherent spin states and its generalizations, *Acta Math.* **212**, 379 (2014).
- [35] S. Ditsch and T. Haas, Entropic distinguishability of quantum fields in phase space, [arXiv:2307.06128](https://arxiv.org/abs/2307.06128) (2023).
- [36] N. J. Cerf and T. Haas, Information and majorization theory for fermionic phase-space distributions, [arXiv:2401.08523](https://arxiv.org/abs/2401.08523) (2024).
- [37] S. Floerchinger, T. Haas, and H. Müller-Groeling, Wehrl entropy, entropic uncertainty relations, and entanglement, *Phys. Rev. A* **103**, 062222 (2021).
- [38] M. Gärttner, T. Haas, and J. Noll, Detecting continuous-variable entanglement in phase space with the Q distribution, *Phys. Rev. A* **108**, 042410 (2023).
- [39] M. Gärttner, T. Haas, and J. Noll, General Class of Continuous Variable Entanglement Criteria, *Phys. Rev. Lett.* **131**, 150201 (2023).
- [40] P. Kunkel, M. Prüfer, S. Lannig, R. Rosa-Medina, A. Bonnin, M. Gärttner, H. Strobel, and M. K. Oberthaler, Simultaneous Readout of Noncommuting Collective Spin Observables beyond the Standard Quantum Limit, *Phys. Rev. Lett.* **123**, 063603 (2019).
- [41] P. Kunkel, M. Prüfer, S. Lannig, R. Strohmaier, M. Gärttner, H. Strobel, and M. K. Oberthaler, Detecting Entanglement Structure in Continuous Many-Body Quantum Systems, *Phys. Rev. Lett.* **128**, 020402 (2022).
- [42] M. J. Collett, R. Loudon, and C. W. Gardiner, Quantum Theory of Optical Homodyne and Heterodyne Detection, *J. Mod. Opt.* **34**, 881 (1987).
- [43] J. W. Noh, A. Fougères, and L. Mandel, Measurement of the quantum phase by photon counting, *Phys. Rev. Lett.* **67**, 1426 (1991).
- [44] J. W. Noh, A. Fougères, and L. Mandel, Operational approach to the phase of a quantum field, *Phys. Rev. A* **45**, 424 (1992).
- [45] S. Stenholm, Simultaneous measurement of conjugate variables, *Ann. Phys.* **218**, 233 (1992).
- [46] U. Leonhardt and H. Paul, Phase measurement and Q function, *Phys. Rev. A* **47**, R2460 (1993).
- [47] C. R. Müller, C. Peuntinger, T. Dirmeier, I. Khan, U. Vogl, C. Marquardt, G. Leuchs, L. L. Sánchez-Soto, Y. S. Teo, Z. Hradil, and J. Řeháček, Evading Vacuum Noise: Wigner Projections or Husimi Samples?, *Phys.*

- Rev. Lett.* **117**, 070801 (2016).
- [48] O. Landon-Cardinal, L. C. G. Govia, and A. A. Clerk, Quantitative Tomography for Continuous Variable Quantum Systems, *Phys. Rev. Lett.* **120**, 090501 (2018).
- [49] L. F. Kozachenko and N. N. Leonenko, Sample Estimate of the Entropy of a Random Vector, *Probl. Peredachi Inf.* **23**, 9 (1987).
- [50] L. Györfi and E. C. van der Meulen, Density-free convergence properties of various estimators of entropy, *Comp. Stat. Data Anal.* **5**, 425 (1987).
- [51] H. Joe, Estimation of entropy and other functionals of a multivariate density, *Ann. Inst. Stat. Math.* **41**, 683 (1989).
- [52] P. Hall and S. C. Morton, On the estimation of entropy, *Ann. Inst. Stat. Math.* **45**, 69 (1993).
- [53] L. Birge and P. Massart, Estimation of Integral Functionals of a Density, *Ann. Stat.* **23**, 10.1214/aos/1176324452 (1995).
- [54] J. Beirlant, E. Dudewicz, L. Györfi, and E. Meulen, Nonparametric Entropy Estimation: An Overview, *Int. J. Math. Stat. Sci.* **6**, 17 (1997).
- [55] H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, Nearest Neighbor Estimates of Entropy, *Am. J. Math. Manag. Sci.* **23**, 301 (2003).
- [56] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, *Phys. Rev. E* **69**, 066138 (2004).
- [57] M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. N. Inverardi, A new class of random vector entropy estimators and its applications in testing statistical hypotheses, *J. Nonparametric Stat.* **17**, 277 (2005).
- [58] K. Sricharan, D. Wei, and A. O. Hero, Ensemble Estimators for Multivariate Entropy Estimation, *IEEE Trans. Inf. Theory* **59**, 4374 (2013).
- [59] K. Kandasamy, A. Krishnamurthy, B. Poczos, L. Wasserman, and J. M. Robins, Nonparametric von Mises Estimators for Entropies, Divergences and Mutual Informations, in *Adv. NIPS*, Vol. 28, edited by C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (2015).
- [60] S. Singh and B. Poczos, Finite-Sample Analysis of Fixed- k Nearest Neighbor Density Functional Estimators, in *Adv. NIPS*, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (2016).
- [61] T. B. Berrett, R. J. Samworth, and M. Yuan, Efficient multivariate entropy estimation via k -nearest neighbour distances, [arXiv:1606.00304](https://arxiv.org/abs/1606.00304) (2016).
- [62] W. Gao, S. Oh, and P. Viswanath, Demystifying Fixed k -Nearest Neighbor Information Estimators, *IEEE Trans. Inf. Theory* **64**, 5629 (2018).
- [63] C. Lu and J. Peltonen, Enhancing Nearest Neighbor Based Entropy Estimator for High Dimensional Distributions via Bootstrapping Local Ellipsoid, *Proc. AAAI* **34**, 5013 (2020).
- [64] Y. Deller, M. Gärttner, T. Haas, M. K. Oberthaler, M. Reh, and H. Strobel, Area laws and thermalization from classical entropies in a spin-1 Bose-Einstein condensate, Companion paper appearing in the same arXiv posting (2024).
- [65] Y. Kawaguchi and M. Ueda, Spinor Bose-Einstein condensates, *Phys. Rep.* **520**, 253 (2012).
- [66] S. Huh, K. Kim, K. Kwon, and J.-y. Choi, Observation of a strongly ferromagnetic spinor Bose-Einstein condensate, *Phys. Rev. Res.* **2**, 033471 (2020).
- [67] C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, and M. S. Chapman, Spin-nematic squeezed vacuum in a quantum gas, *Nat. Phys.* **8**, 305 (2012).
- [68] D. M. Stamper-Kurn and M. Ueda, Spinor Bose gases: Symmetries, magnetism, and quantum dynamics, *Rev. Mod. Phys.* **85**, 1191 (2013).
- [69] P. Kunkel, M. Prüfer, H. Strobel, D. Linnemann, A. Frölian, T. Gasenzer, M. Gärttner, and M. K. Oberthaler, Spatially distributed multipartite entanglement enables EPR steering of atomic clouds, *Science* **360**, 413 (2018).
- [70] P. Kunkel, *Splitting a Bose-Einstein condensate enables EPR steering and simultaneous readout of noncommuting observables*, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg (2019).
- [71] J. Schwinger, On Angular Momentum, Unpublished Report **Report Number NYO-3071** (1952).
- [72] S. K. Kim, Theorems on the Jordan-Schwinger representations of Lie algebras, *J. Math. Phys.* **28**, 2540 (1987).
- [73] C. Gross, H. Strobel, E. Nicklas, T. Zibold, N. Bar-Gill, G. Kurizki, and M. K. Oberthaler, Atomic homodyne detection of continuous-variable entangled twin-atom states, *Nature* **480**, 219 (2011).
- [74] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
- [75] A. Serafini, *Quantum Continuous Variables* (CRC Press, 2017).
- [76] W.-M. Zhang, D. H. Feng, and R. Gilmore, Coherent states: Theory and some applications, *Rev. Mod. Phys.* **62**, 867 (1990).
- [77] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States, 2nd ed.* (John Wiley and Sons, 2017).
- [78] A. Kenfack and K. Życzkowski, Negativity of the Wigner function as an indicator of non-classicality, *J. Opt., B Quantum Semiclass. Opt.* **6**, 396 (2004).
- [79] S. L. Braunstein, Homodyne statistics, *Phys. Rev. A* **42**, 474 (1990).
- [80] U. Leonhardt and H. Paul, Measuring the quantum state of light, *Prog. Quantum Electron.* **19**, 89 (1995).
- [81] A. Polkovnikov, Phase space representation of quantum dynamics, *Ann. Phys.* **325**, 1790 (2010).
- [82] M. Oliva, D. Kakofengitis, and O. Steuernagel, Anharmonic quantum mechanical systems do not feature phase space trajectories, *Phys. A: Stat. Mech. Appl.* **502**, 201 (2018).
- [83] P. Blakie, A. Bradley, M. Davis, R. Ballagh, and C. Gardiner, Dynamics and statistical mechanics of ultra-cold Bose gases using c-field techniques, *Adv. Phys.* **57**, 363 (2008).
- [84] H. P. Robertson, The Uncertainty Principle, *Phys. Rev.* **34**, 163 (1929).
- [85] H. P. Robertson, A general formulation of the uncertainty principle and its classical interpretation, *Phys. Rev.* **35**, 667 (1930).
- [86] E. Schrödinger, Zum Heisenbergschen Unschärfepinzipp, Sitzungsberichte der Preußischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse **14**, 296 (1930).
- [87] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, 1955).
- [88] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

- [89] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013).
- [90] Z. Van Herstraeten and N. J. Cerf, Quantum Wigner entropy, *Phys. Rev. A* **104**, 042211 (2021).
- [91] Z. Van Herstraeten, M. G. Jabbour, and N. J. Cerf, Continuous majorization in quantum phase space, *Quantum* **7**, 1021 (2023).
- [92] Z. Van Herstraeten, *Majorization theoretical approach to quantum uncertainty*, Ph.D. thesis, Université libre de Bruxelles (2021).
- [93] H. Everett, 'Relative State' Formulation of Quantum Mechanics, *Rev. Mod. Phys.* **29**, 454 (1957).
- [94] I. I. Hirschman, A Note on Entropy, *Am. J. Math.* **79**, 152 (1957).
- [95] W. Beckner, Inequalities in Fourier Analysis, *Ann. Math.* **102**, 159 (1975).
- [96] I. Białynicki-Birula and J. Mycielski, Uncertainty relations for information entropy in wave mechanics, *Commun. Math. Phys.* **44**, 129 (1975).
- [97] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, Entropic uncertainty relations and their applications, *Rev. Mod. Phys.* **89**, 015002 (2017).
- [98] A. Hertz and N. J. Cerf, Continuous-variable entropic uncertainty relations, *J. Phys. A Math. Theor.* **52**, 173001 (2019).
- [99] G. Adesso, D. Girolami, and A. Serafini, Measuring Gaussian Quantum Information and Correlations Using the Rényi Entropy of Order 2, *Phys. Rev. Lett.* **109**, 190502 (2012).
- [100] G. Manfredi and M. R. Feix, Entropy and Wigner functions, *Phys. Rev. E* **62**, 4665 (2000).
- [101] J. J. Włodarz, Entropy and Wigner Distribution Functions Revisited, *Int. J. Theor. Phys.* **42**, 1075 (2003).
- [102] E. H. Lieb and R. Seiringer, Stronger subadditivity of entropy, *Phys. Rev. A* **71**, 062329 (2005).
- [103] G. V. Steeg, Non-parametric Entropy Estimation Toolbox, <https://github.com/gregversteeg/NPEET> (2014).
- [104] S. Nadarajah, A generalized normal distribution, *J. Appl. Stat.* **32**, 685 (2005).
- [105] H. Hino, K. Koshijima, and N. Murata, Non-parametric entropy estimators based on simple linear regression, *Comput. Stat. Data Anal.* **89**, 72 (2015).
- [106] F. Tarasenko, On the evaluation of an unknown probability density function, the direct estimation of the entropy from independent observations of a continuous random variable, and the distribution-free entropy test of goodness-of-fit, *Proc. IEEE* **56**, 2052 (1968).
- [107] O. Vasicek, A Test for Normality Based on Sample Entropy, *J. R. Stat. Soc., B: Stat. Methodol.* **38**, 54 (1976).
- [108] H. Hino, K. Wakayama, and N. Murata, Entropy-based sliced inverse regression, *Comput. Stat. Data Anal.* **67**, 105 (2013).
- [109] E. G. Learned-Miller and J. W. F. III, ICA using spacings estimates of entropy, *J. Mach. Learn. Res.* **4**, 1271 (2003).
- [110] P. Comon, Independent component analysis, A new concept?, *Signal Process.* **36**, 287 (1994).
- [111] S. Mannor, D. Peleg, and R. Rubinfeld, The cross entropy method for classification, in *Proc. ICML '05* (2005).
- [112] R. Y. Rubinfeld and D. P. Kroese, *The Cross-Entropy Method* (Springer New York, 2004).
- [113] H. Hino and N. Murata, A Conditional Entropy Minimization Criterion for Dimensionality Reduction and Multiple Kernel Learning, *Neural Computat.* **22**, 2887 (2010).
- [114] H. Hino and N. Murata, Information estimators for weighted observations, *Neural Netw.* **46**, 260 (2013).
- [115] P. Czyż, F. Grabowski, J. E. Vogt, N. Beerenwinkel, and A. Marx, Beyond Normal: On the Evaluation of Mutual Information Estimators, [arXiv:2306.11078](https://arxiv.org/abs/2306.11078) (2023).
- [116] Z. Ao and J. Li, Entropy Estimation via Normalizing Flow, *Proc. AAAI* **36**, 9990 (2022).
- [117] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, Mutual Information Neural Estimation, in *Proceedings of the 35th International Conference on Machine Learning*, Proc. MLR, Vol. 80, edited by J. Dy and A. Krause (2018) pp. 531–540.
- [118] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition* (John Wiley and Sons, 2006).
- [119] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi, and M. K. Oberthaler, Fisher information and entanglement of non-Gaussian spin states, *Science* **345**, 424 (2014).
- [120] J. S. Centre, JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing, JLSRF (2019).

AREA LAWS AND THERMALIZATION FROM CLASSICAL ENTROPIES IN A SPIN-1 BOSE-EINSTEIN CONDENSATE

9.1 CONTRIBUTIONS

This chapter presents a finalized draft of the work in [E]. Tobias Haas originally proposed the project, and the proposal was refined in joint discussions between Tobias Haas, Martin Gärttner, Yannick Deller, Helmut Strobel, Markus Oberthaler, and me. Tobias Haas and I ran all simulations in close collaboration, with the analytical model implemented by Tobias Haas and the truncated Wigner simulation, as discussed in [Section 4.2](#), implemented by me. Tobias Haas and I wrote the initial draft of the manuscript and all other authors helped in refining it.

9.2 MOTIVATION

We take the preceding discussion in [Chapter 8](#) as motivation to demonstrate that the interplay between quantum thermalization and the transition from an area to a volume law is also maintained within the phase-space description. To this end, we explore the late-time regime where the local phase-space distributions become stationary. We again base our findings on classical simulations of quenches of a Spin-1 Bose-Einstein condensate, as discussed in [Section 2.7](#), and understand our contributions as an experimental proposal.

Area laws and thermalization from classical entropies in a spin-1 Bose-Einstein condensate

Yannick Deller,^{1,*} Martin Gärttner,^{2,†} Tobias Haas,^{3,‡} Markus
K. Oberthaler,^{1,§} Moritz Reh,^{1,¶} and Helmut Strobel^{1,**}

¹*Kirchhoff-Institut für Physik, Universität Heidelberg,
Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*

²*Institut für Festkörpertheorie und Optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany*

³*Centre for Quantum Information and Communication, École polytechnique de Bruxelles,
CP 165, Université libre de Bruxelles, 1050 Brussels, Belgium*

The scaling of local quantum entropies is of utmost interest for characterizing quantum fields, many-body systems, and gravity. Despite their importance, being nonlinear functionals of the underlying quantum state hinders their theoretical as well as experimental accessibility. Here, we show that suitably chosen classical entropies capture the very same features as their quantum analogs for an experimentally relevant setting. We describe the post-quench dynamics of a multi-well spin-1 Bose-Einstein condensate from an initial product state via measurement distributions of spin observables and estimate the corresponding entropies using the asymptotically unbiased k -nearest neighbor method. We observe the dynamical build-up of quantum correlations signaled by an area law, as well as local thermalization revealed by a transition to a volume law, both in regimes characterized by non-Gaussian quantum states. We emphasize that all relevant features can be observed at small sample numbers without assuming a specific functional form of the distributions, rendering our method directly applicable to a large variety of models and experimental platforms.

Introduction — The quantum entropy of a spatial subregion has proven to serve as a ubiquitous tool for studying the spatio-temporal structure of entanglement [1] and its role in various quantum phenomena, including local thermalization [2–5], quantum phase transitions [6], information scrambling [7–9] and black hole physics [10–13]. Arguably the most sought-after phenomenon in this context is the area law, which is signaled by a logarithmic growth of the local entropy for one-dimensional systems [14–22]. It appears in particular for the experimentally accessible scenario of a system prepared in a pure product state when the couplings of a local Hamiltonian are quenched before the system has thermalized locally [22]. In the latter case, the local entropy rather obeys a volume law, allowing for a macroscopic description using only a few thermodynamic quantities like temperature.

The main drawback of quantum entropic descriptions for many-body phenomena is their reliance on the knowledge of the full density matrix, which typically grows exponentially with the number of microscopic constituents. This especially has so far restricted the experimental access of quantum entropies to systems consisting of a few particles [23–25], as full tomography of the quantum state is, with no further assumptions, infeasible for larger systems approaching mesoscopic scales. For continuous systems, area laws have only been experimentally reported in a Gaussian scenario [26], while generally applicable methods have remained elusive.

Recently, the necessity of considering exclusively *quantum* entropies to probe quantum phenomena has been questioned. Suitably chosen *classical* entropies of (quasi-)probability distributions do also encode area and volume laws [27]. This ansatz naturally overcomes the need for

reconstructing the full quantum state – both for theoretical and experimental investigations. This method is specifically suitable for experimental platforms, which can directly sample from such distributions, see for example [28–36].

Here we show that area and volume laws are observable in state-of-the-art experiments with multi-well spin-1 Bose-Einstein condensates (BECs) [37, 38] by considering entropies of measurement distributions over spin observables. Starting from an initial product state, we find area laws being dynamically generated for intermediate evolution times following a quench, thereby confirming the growth of entanglement until the system thermalizes locally, where the same entropies exhibit volume law behavior. Importantly, we do so without assuming the functional form of the state and only rely on observables that are directly obtainable in standard experimental readouts [26, 31, 34–36], while reducing the sample complexity to a feasible level. We comprehensively discuss our method, including systematic checks for its validity and generality, in [39].

Notation — We use natural units $\hbar = k_B = 1$, write bold (normal) letters for quantum operators \mathbf{O} (classical variables O) as well as their traces and equip vacuum expressions with a bar, e.g. $\bar{\rho}$.

Multi-Well Spin-1 BEC — We consider a one-dimensional spin-1 BEC that extends over 20 wells, described by bosonic mode operators $[\mathbf{a}_{m_F}^j, \mathbf{a}_{m_F}^{j'\dagger}] = \delta^{jj'} \delta_{m_F m'_F}$ with $j \in \{1, \dots, N\}$ and $m_F \in \{0, \pm 1\}$. Starting from an initial product state with all zero modes ($m_F = 0$) being occupied coherently by $n = 10^3$ atoms,

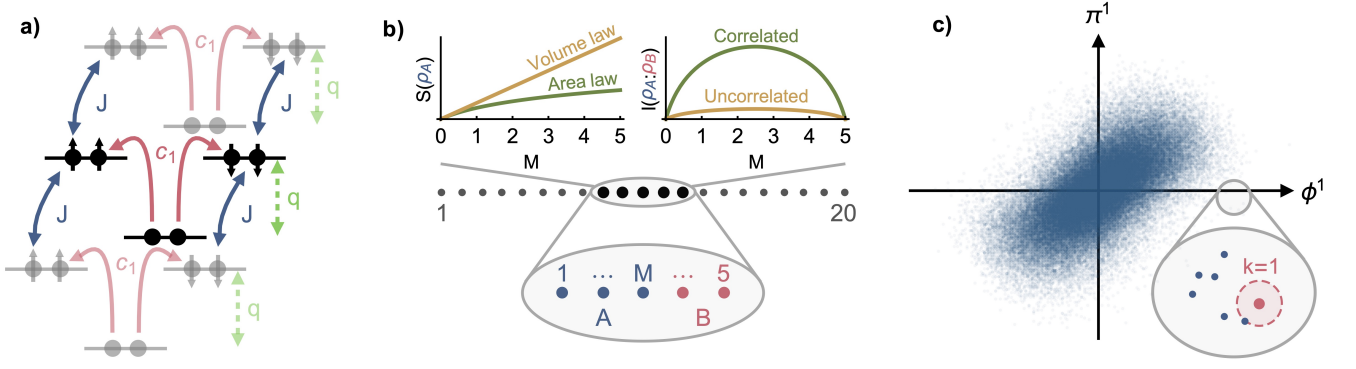


Figure 1: **a)** Illustration of relevant processes. The ± 1 modes of a spin-1 BEC are coupled to the 0 mode by spin-changing collisions with strength $c_1 < 0$ (red) and detuned by the quadratic Zeeman-shift $q > 0$ (green). The atoms in the ± 1 modes may hop to neighboring wells with strength $J > 0$ (blue). **b)** The full system consists of 20 wells, from which we exclusively analyze the five wells 8 – 12. We partition this subsystem into a left part A (blue) and a right part B (red) and study the scalings of information and correlation measures with A 's system size M . **c)** Samples of the Wigner W -distribution of the left-most well in subsystem A at time $t = 4$. The entropy is estimated from samples using the k NN-estimator by analyzing the distribution of distances to the k -th. neighbor for each sample, see magnified inset. Non-Gaussian features arise for higher-dimensional multi-well distributions, as measured by the relative entropy, see [40].

we consider a quench described by the Hamiltonian

$$\begin{aligned}
 H = & \sum_{j=1}^{20} q \left(N_1^j + N_{-1}^j \right) + c_0 N^j (N^j - 1) \\
 & + c_1 \left[\left(N_0^j - (1/2)\mathbb{1} \right) \left(N_1^j + N_{-1}^j \right) \right. \\
 & \quad \left. + \mathbf{a}_0^{j\dagger} \mathbf{a}_0^j \mathbf{a}_1^j \mathbf{a}_{-1}^j + \mathbf{a}_1^{j\dagger} \mathbf{a}_{-1}^j \mathbf{a}_0^j \mathbf{a}_0^j \right] \\
 & - J \sum_{j=1}^{19} \sum_{m_F=\pm 1} \left(\mathbf{a}_{m_F}^{j\dagger} \mathbf{a}_{m_F}^{j+1} + \mathbf{a}_{m_F}^{(j+1)\dagger} \mathbf{a}_{m_F}^j \right),
 \end{aligned} \tag{1}$$

featuring dynamics within single wells (first sum) as well as correlation build-up among wells (second sum), see Figure 1 a).

For early times, the zero mode is occupied macroscopically and the evolution is dominated by second-order fluctuations, such that (1) is well-approximated by an analytically solvable Gaussian model, which follows from treating the zero mode classically and dropping density-density interactions (see [39] for details)

$$\begin{aligned}
 H_{\text{up,Gauss}} = & \sum_{j=1}^{20} \left[\tilde{q} N^j + \frac{\tilde{c}_1}{2} \left(\mathbf{a}^j \mathbf{a}^j + \mathbf{a}^{j\dagger} \mathbf{a}^{j\dagger} \right) \right] \\
 & - J \sum_{j=1}^{19} \left(\mathbf{a}^{j\dagger} \mathbf{a}^{j+1} + \mathbf{a}^{(j+1)\dagger} \mathbf{a}^j \right).
 \end{aligned} \tag{2}$$

Here, we introduced the relative mode operators $\mathbf{a}^j = (\mathbf{a}_1^j + \mathbf{a}_{-1}^j)/\sqrt{2}$ as well as the rescaled couplings $\tilde{c}_1 = c_1 n$ and $\tilde{q} = c_1 \left(n - \frac{1}{2} \right) + q$.

Beyond this regime, the mesoscopic occupation justifies employing semi-classical approaches such as the truncated

Wigner approximation (TWA), in which the mode operators are demoted to c -numbers that obey an evolution dictated by classical mean field equations [41, 42]. The resulting model correctly captures the quantum fluctuations of the initial state, while neglecting higher-order corrections in \hbar for its evolution.

Measurement distributions — We analyze the information content of a subsystem of five wells (see Figure 1 b)) in terms of measurement distributions using phase-space methods. We focus on the two normalized spin-1 observables [37, 38]

$$\begin{aligned}
 \phi^j & \equiv \frac{S_x^j}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} \left(\mathbf{a}_1^j + \mathbf{a}_{-1}^j \right) + h.c. \right] / \sqrt{2n}, \\
 \pi^j & \equiv -\frac{Q_{yz}^j}{\sqrt{2n}} = \frac{-i}{\sqrt{2}} \left[\mathbf{a}_0^{j\dagger} \left(\mathbf{a}_1^j - \mathbf{a}_{-1}^j \right) - h.c. \right] / \sqrt{2n},
 \end{aligned} \tag{3}$$

which form a set of independent canonical operators $[\phi^j, \pi^{j'}] = i\delta^{jj'}\mathbb{1}$ with corresponding bosonic mode operators $\mathbf{a}^j, \mathbf{a}^{j\dagger}$ in the early-time regime [39]. Their Wigner W -distribution is defined via [43]

$$\begin{aligned}
 \mathcal{W}^j & \equiv \mathcal{W}^j(\phi^j, \pi^j) \\
 & = \int \frac{d\tilde{\phi}^j d\tilde{\pi}^j}{2\pi} e^{-i(\phi^j, \pi^j)\Omega(\tilde{\phi}^j, \tilde{\pi}^j)^T} \\
 & \quad \times \text{Tr} \left\{ \rho^j e^{i(\phi^j, \pi^j)\Omega(\tilde{\phi}^j, \tilde{\pi}^j)^T} \right\},
 \end{aligned} \tag{4}$$

with the symplectic form $\Omega = i\sigma_2$ and σ_2 being the second Pauli matrix. As \mathcal{W}^j is only accessible through costly Wigner tomography [44, 45], it is mainly of theoretical interest. Let us therefore also introduce more experimentally convenient distributions, namely

the Wigner marginals $f^j \equiv f^j(\phi^j) = \int d\pi^j \mathcal{W}^j$ and $g^j \equiv g^j(\pi^j) = \int d\phi^j \mathcal{W}^j$, accessible through homodyne measurements [28], as well as the Husimi Q -distribution, which is obtained by projecting onto the coherent states $|\alpha^j\rangle = \exp(\alpha^j \mathbf{a}^{j\dagger} - \alpha^{j*} \mathbf{a}^j) |0^j\rangle$ [31, 34–36, 46], with $\alpha^j = (\phi^j + i\pi^j)/\sqrt{2}$, leading to [47, 48]

$$\mathcal{Q}^j \equiv \mathcal{Q}^j(\phi^j, \pi^j) = \text{Tr} \{ \rho^j |\alpha^j\rangle \langle \alpha^j| \}. \quad (5)$$

Information and correlations from classical distributions — To analyze the local information content of the subsystem, we consider differential entropies of the classical distributions \mathcal{O}^A with respect to its left part A (see Figure 1 b))

$$S(\mathcal{O}^A) = - \int d\mu^A \mathcal{O}^A \ln \mathcal{O}^A, \quad (6)$$

where the integral measure $d\mu^A$ depends on the distribution under scrutiny [49]. We note that (6) is always well-defined for the non-negative marginal and Husimi Q -distributions, but is restricted to Wigner-positive states when applied to \mathcal{W}^A , which is an assumption implicitly made when working within TWA or Gaussian models.

Being measures of disorder, classical entropies over incompatible observables are bounded from below by their vacuum values via entropic uncertainty relations [50–55] (see [56, 57] for reviews). When considered for quantum many-body systems, these bounds scale with the number of modes, i.e. $S(\bar{\mathcal{O}}^A) \sim M$, showing that classical entropies are extensive to leading order as a result of vacuum contributions [58, 59]. However, as shown in [27], scalings induced by quantum phenomena, such as the area law, manifest themselves in the *next-to-leading* order terms. Thus, we define the so-called subtracted classical entropies as [27]

$$\Delta S(\mathcal{O}^A) \equiv S(\mathcal{O}^A) - S(\bar{\mathcal{O}}^A), \quad (7)$$

with the extensive vacuum contribution $S(\bar{\mathcal{O}}^A) \sim M$ being subtracted [60].

Let us further consider the classical version of the archetypical measure for correlations between the left and right parts of the subsystem, that is, the classical mutual information

$$I(\mathcal{O}^A : \mathcal{O}^B) = S(\mathcal{O}^A) + S(\mathcal{O}^B) - S(\mathcal{O}). \quad (8)$$

Being already defined via a relative entropic measure, no vacuum contributions have to be subtracted to reveal quantum features.

Connections to quantum information theory — In the context of the Gaussian model (2), the connection between subtracted classical and quantum entropies becomes a simple equality: in this case, we can establish $\Delta S(\mathcal{W}^A) = S_2(\rho^A)$, where $S_2(\rho^A)$ denotes the Rényi-2 entropy of the density matrix associated to \mathcal{W}^A [61]. Beyond Gaussianity, such simple relations can only be

established for the subtracted Rényi-2 entropy of \mathcal{W}^A [62]. However, in the following, we provide strong evidence that the scaling of the subtracted classical entropies (7) also extends to the non-Gaussian interacting case.

Furthermore, a connection to the quantum mutual information in the case of Gaussian states is straightforward and reads $I(\mathcal{W}^A : \mathcal{W}^B) = I_2(\rho^A : \rho^B)$ [61]. More generally, classical mutual informations constitute lower bounds to their quantum analogs by the uncertainty principle, i.e. [54, 63]

$$I(\mathcal{O}^A : \mathcal{O}^B) \leq I(\rho^A : \rho^B), \quad (9)$$

which are expected to be tighter than second-moment bounds beyond Gaussian states [64]. An immediate consequence of (9) is that the standard argument for the appearance of the area law for local interactions and thermal states presented in [65] applies also to any classical mutual information (see also [39]). Hence, classical mutual information, albeit typically not capturing all quantum correlations, shows the finite-size area law whenever its quantum analog does.

Methods — We generate 10^4 synthetic samples for the three distributions of our interest using TWA to simulate an experiment showcasing the feasibility of the proposed approach. In contrast to the estimation of low-order moments, extracting entropic quantities from a set of samples is more involved, since they are *functionals* of the underlying distributions. Given a set of samples, we employ the established k -nearest neighbor (k NN) method devised in [66–68] using information about the statistics of the nearest neighbors of each sample (see Figure 1 c)), to arrive at an estimate of its local density. These results are validated against the analytically solvable model (2) in the early-time regime. We give a more comprehensive validation of the k NN-estimator for our setup in [39]. Further, we define an energy scale by setting $nc_1 = -1$, consider Lithium-7 with $c_0 = -2c_1$ and set the quench parameters to $q = 4, J = 2$, for which non-Gaussian features arise around $t = 3$.

While the total system undergoes a unitary evolution dictated by the Hamiltonian in Eq. (1), the considered subsystem does not, as its entanglement with the rest of the system implies a mixed reduced density matrix [69]. In the following, we demonstrate the area law and local thermalization for the theoretically interesting, but experimentally difficult to access subtracted Wigner entropy, as well as for the experimentally amenable subtracted marginal entropy sum $\Delta S(f^A) + \Delta S(g^A)$, and the so-called Wehrl mutual information $I(\mathcal{Q}^A : \mathcal{Q}^B)$ (the remaining three quantities are presented in [40]).

Area law — We first study the early-time regime, that is, $t \leq 4$, in the upper row of Figure 2. At $t = 0$, the subsystem is in a pure product state, and all entropic measures evaluate to zero. Around $t = 1$, correlations among the wells start to build up, causing subsystem A to become entangled with its complement B . In this regime,

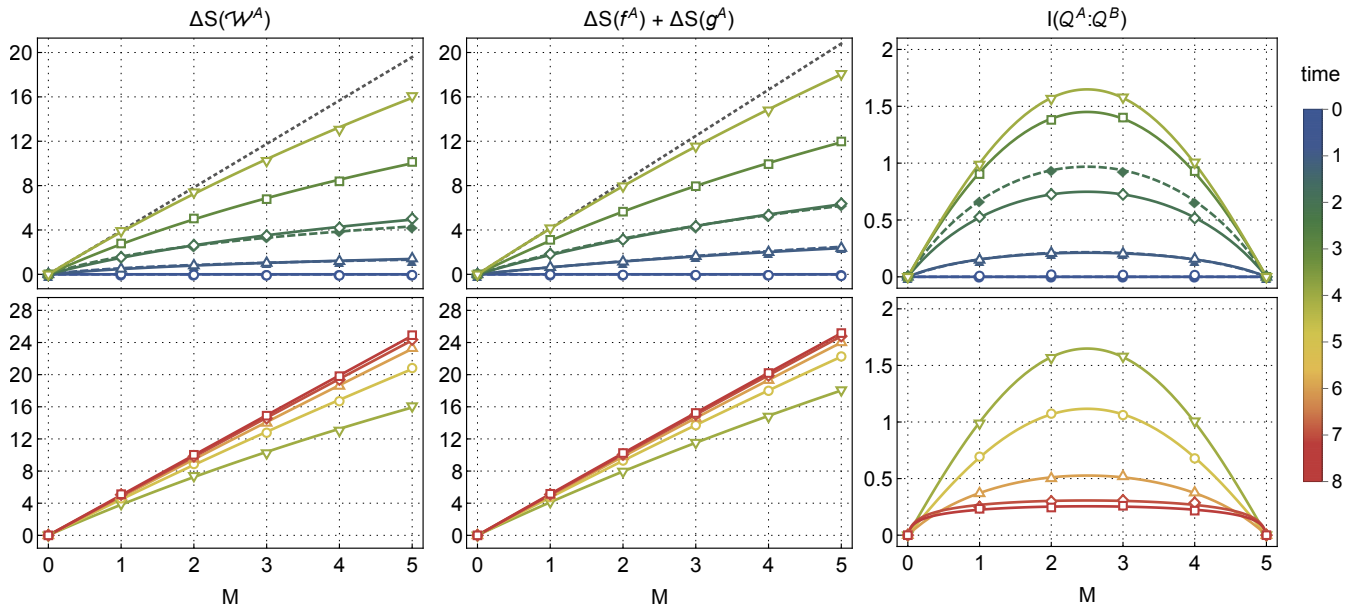


Figure 2: Analysis regarding the presence of area and volume laws at early times $t = 0, 1, 2, 3, 4$ (upper row) and late times $t = 4, 5, 6, 7, 8$ (lower row), respectively. Open (closed) plot markers denote TWA (analytic) results and the corresponding solid (dashed) curves are fits. In the early-time regime, we observe the subtracted classical entropies to fulfill a logarithmic growth with subsystem size in the sense of (10) (see [40] for the *standard* Wigner entropy). Their sublinear scaling is highlighted for $t = 4$ by straight lines (gray dotted), which are fitted to the first two data points. In accordance, we also find the finite-size area law (11) for the Wehrl mutual information. These findings hold true for both the TWA and the analytical approach, which agree in the Gaussian regime, i.e. up to $t = 3$ [40], thereby also validating the k NN estimator. For later times, the area law of the subtracted classical entropies tends into a stationary volume law (12), thereby demonstrating local thermalization. After the stationary point $t = 7$, the local temperature can be extracted via their inclines, which consistently yields $T \approx 5$. The appearance of local thermalization is further supported by the decreasing correlations between A and B towards zero as revealed by the evolution of the Wehrl mutual information.

subtracted classical entropies obey the area law, i.e. a logarithmic growth with system size M ,

$$\Delta S(\mathcal{O}^A) = \kappa_1 \ln(M + \kappa_2) + \kappa_3, \quad (10)$$

just as one would expect for the entanglement entropy [14–22]. The fit parameters κ_i are constrained by $\kappa_2 = e^{-\kappa_3/\kappa_1}$ to ensure $\Delta S(\mathcal{O}^A) = 0$ when $M = 0$. Around $t = 3$, the distributions begin to exhibit non-Gaussian features, which we quantify by the relative entropy with respect to the closest Gaussian distribution, see [40].

Similarly, the Wehrl mutual information signals the generation of correlations between A and B in terms of the finite-size area law [14]

$$I(\mathcal{O}^A : \mathcal{O}^B) = \kappa_1 \ln \left[\frac{5}{\pi} \sin \left(\frac{\pi M}{5} \right) + \kappa_2 \right] + \kappa_3, \quad (11)$$

which incorporates the reflection symmetry around $M = 2.5$. Again, the behavior coincides with what is expected for the quantum mutual information [65], with maximal correlations occurring at $t = 4$.

Local thermalization — For later times, i.e. in the regime $t \geq 4$ (lower row of Figure 2), the subtracted

classical entropies transition from an intermediate stage around $t = 5$ to an extensive growth with system size at $t = 7$. The latter remains stationary beyond $t = 7$, signaling that the system has thermalized locally in the considered degrees of freedom, with the remaining system serving as a heat bath. In this case, all entropies of our interest obey the volume law [3]

$$\Delta S(\mathcal{O}^A) = \beta M, \quad (12)$$

where $\beta = 1/T$ denotes the inverse local temperature. Indeed, both final entropic curves show an incline of $T \approx 5$, illustrating how the local temperature can be extracted from classical entropies by simple means. We have checked that this temperature depends only weakly on the quench parameters, as the dominating energy scale is set by the fourth-order term proportional to c_0 in (1).

While the classical entropies become extensive, the Wehrl mutual information still obeys the finite-size area law (11) for later times, which also highlights its robustness against thermal fluctuations. In contrast to the early-time dynamics, the correlations between A and B now decline monotonically towards local thermal equilibrium.

Discussion — We have demonstrated that quantum many-body phenomena can be probed with classical entropies by considering a concrete model system that can be readily realized experimentally. Specifically, we have shown that it is possible to observe the area law, that is, the characteristic logarithmic growth of the entanglement entropy, and the volume law, which indicates local thermalization, via subtracted classical entropies and mutual informations of experimentally accessible measurement distributions. Crucially, we have not assumed the state to obey a specific functional form and only relied on 10^4 samples which we deem experimentally feasible. Future work will address what other parallels between classical entropies and quantum entropies exist, especially for other degrees of freedom, and whether they also lend themselves as easily to experimental implementations as in the discussed work.

Acknowledgements — The authors thank Thomas Gasenzer and Ido Siovitz for helpful discussions during the development of the manuscript. T. H. is supported by the European Union under project ShoQC within the ERA-NET Cofund in Quantum Technologies (QuantERA) program, as well as by the F.R.S.-FNRS under project CHEQS within the Excellence of Science (EOS) program. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster) and within the Collaborative Research Center SFB1225 (ISO-QUANT). This work was partially financed by the Baden-Württemberg Stiftung gGmbH. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS [70] at Jülich Supercomputing Centre (JSC).

* yannick.deller@kip.uni-heidelberg.de

† martin.gaerttner@uni-jena.de

‡ tobias.haas@ulb.be

§ markus.oberthaler@kip.uni-heidelberg.de

¶ moritz.reh@kip.uni-heidelberg.de

** helmut.strobel@kip.uni-heidelberg.de

- [1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
- [2] S. Popescu, A. Short, and A. Winter, Entanglement and the foundations of statistical mechanics, *Nat. Phys.* **2**, 754 (2006).
- [3] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, Colloquium: Many-body localization, thermalization, and entanglement, *Rev. Mod. Phys.* **91**, 021001 (2019).
- [4] N. Dowling, S. Floerchinger, and T. Haas, Second law of thermodynamics for relativistic fluids formulated with relative entropy, *Phys. Rev. D* **102**, 105002 (2020).
- [5] S. Floerchinger and T. Haas, Thermodynamics from relative entropy, *Phys. Rev. E* **102**, 052117 (2020).
- [6] T. J. Osborne and M. A. Nielsen, Entanglement in a simple quantum phase transition, *Phys. Rev. A* **66**, 032110 (2002).
- [7] R. Jozsa and N. Linden, On the Role of Entanglement in Quantum-Computational Speed-Up, *Proc. Math. Phys. Eng. Sci.* **459**, 2011 (2003).
- [8] K. A. Landsman, C. Figgatt, T. Schuster, N. M. Linke, B. Yoshida, N. Y. Yao, and C. Monroe, Verified quantum information scrambling, *Nature* **567**, 61 (2019).
- [9] S. Xu and B. Swingle, Scrambling Dynamics and Out-of-Time-Ordered Correlators in Quantum Many-Body Systems, *PRX Quantum* **5**, 010201 (2024).
- [10] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, Quantum source of entropy for black holes, *Physical Review D* **34**, 373 (1986).
- [11] M. Srednicki, Entropy and area, *Physical Review Letters* **71**, 666 (1993).
- [12] C. Callan and F. Wilczek, On geometric entropy, *Physics Letters B* **333**, 55 (1994).
- [13] S. N. Solodukhin, Entanglement entropy in non-relativistic field theories, *JHEP* **2010** (4), 101.
- [14] P. Calabrese and J. Cardy, Entanglement entropy and quantum field theory, *J. Stat. Mech. Theo. Exp.* **2004**, P06002 (2004).
- [15] P. Calabrese and J. Cardy, Entanglement Entropy and Quantum Field Theory: A Non-Technical Introduction, *Int. J. Quantum Inf.* **04**, 429 (2006).
- [16] M. B. Plenio and S. Virmani, An Introduction to Entanglement Measures, *Quantum Inf. Comput.* **7**, 1–51 (2007).
- [17] M. B. Hastings, An area law for one-dimensional quantum systems, *J. Stat. Mech.: Theo. Exp.* **2007**, P08024 (2007).
- [18] L. Amico, R. Fazio, A. Osterloh, and V. Vedral, Entanglement in many-body systems, *Rev. Mod. Phys.* **80**, 517 (2008).
- [19] P. Calabrese and J. Cardy, Entanglement entropy and conformal field theory, *J. Phys. A Math. Theo.* **42**, 504005 (2009).
- [20] H. Casini and M. Huerta, Entanglement entropy in free quantum field theory, *J. Phys. A Math. Theo.* **42**, 504007 (2009).
- [21] I. Peschel and V. Eisler, Reduced density matrices and entanglement entropy in free lattice models, *J. Phys. A: Math. Theo.* **42**, 504003 (2009).
- [22] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, *Rev. Mod. Phys.* **82**, 277 (2010).
- [23] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, Measuring entanglement entropy in a quantum many-body system, *Nature* **528**, 77 (2015).
- [24] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, *Science* **353**, 794 (2016).
- [25] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, *Science* **364**, 260 (2019).
- [26] M. Tajik, I. Kukuljan, S. Sotiriadis, B. Rauer, T. Schweigler, F. Cataldini, J. Sabino, F. Møller, P. Schüttelkopf,

- S.-C. Ji, D. Sels, E. Demler, and J. Schmiedmayer, Verification of the area law of mutual information in a quantum field simulator, *Nat. Phys.* **19**, 1022 (2012).
- [27] T. Haas, Area Law from classical entropies, In preparation , 5 (2024).
- [28] U. Leonhardt and H. Paul, Measuring the quantum state of light, *Prog. Quantum. Electron.* **19**, 89 (1995).
- [29] G. Kirchmair, B. Vlastakis, Z. Leghtas, S. E. Nigg, H. Paik, E. Ginossar, M. Mirrahimi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Observation of quantum state collapse and revival due to the single-photon Kerr effect, *Nature* **495**, 205 (2013).
- [30] F. Haas, J. Volz, R. Gehr, J. Reichel, and J. Estève, Entangled states of more than 40 atoms in an optical fiber cavity, *Science* **344**, 180 (2014).
- [31] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi, and M. K. Oberthaler, Fisher information and entanglement of non-Gaussian spin states, *Science* **345**, 424 (2014).
- [32] G. Barontini, L. Hohmann, F. Haas, J. Estève, and J. Reichel, Deterministic generation of multiparticle entanglement by quantum Zeno dynamics, *Science* **349**, 1317 (2015).
- [33] C. Wang, Y. Y. Gao, P. Reinhold, R. W. Heeres, N. Ofek, K. Chou, C. Axline, M. Reagor, J. Blumoff, K. M. Sliwa, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, A Schrödinger cat living in two boxes, *Science* **352**, 1087 (2016).
- [34] P. Kunkel, M. Prüfer, H. Strobel, D. Linnemann, A. Frölian, T. Gasenzer, M. Gärtner, and M. K. Oberthaler, Spatially distributed multipartite entanglement enables EPR steering of atomic clouds, *Science* **360**, 413 (2018).
- [35] P. Kunkel, M. Prüfer, S. Lannig, R. Rosa-Medina, A. Bonnin, M. Gärtner, H. Strobel, and M. K. Oberthaler, Simultaneous Readout of Noncommuting Collective Spin Observables beyond the Standard Quantum Limit, *Phys. Rev. Lett.* **123**, 063603 (2019).
- [36] P. Kunkel, M. Prüfer, S. Lannig, R. Strohmaier, M. Gärtner, H. Strobel, and M. K. Oberthaler, Detecting Entanglement Structure in Continuous Many-Body Quantum Systems, *Phys. Rev. Lett.* **128**, 020402 (2022).
- [37] C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, and M. S. Chapman, Spin-nematic squeezed vacuum in a quantum gas, *Nat. Phys.* **8**, 305 (2012).
- [38] Y. Kawaguchi and M. Ueda, Spinor Bose–Einstein condensates, *Phys. Rep.* **520**, 253 (2012).
- [39] Y. Deller, M. Gärtner, T. Haas, M. K. Oberthaler, M. Reh, and H. Strobel, Area laws for classical entropies in a spin-1 Bose-Einstein condensate, Companion paper appearing in the same arXiv posting (2024).
- [40] See Supplemental Material at [URL will be inserted by publisher] for details.
- [41] A. Polkovnikov, Phase space representation of quantum dynamics, *Ann. Phys.* **325**, 1790 (2010).
- [42] P. Blakie, A. Bradley, M. Davis, R. Ballagh, and C. Gardiner, Dynamics and statistical mechanics of ultra-cold Bose gases using c-field techniques, *Adv. Phys.* **57**, 363 (2008).
- [43] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
- [44] W. P. Schleich, *Quantum Optics in Phase Space* (Wiley-VCH Verlag Berlin, 2001).
- [45] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, 2013).
- [46] P. Kunkel, *Splitting a Bose-Einstein condensate enables EPR steering and simultaneous readout of noncommuting observables*, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg (2019).
- [47] K. Husimi, Some formal properties of the density matrix, *Proc. Phys.-Math. Soc. Jap. 3rd Ser.* **22**, 264 (1940).
- [48] N. D. Cartwright, A non-negative Wigner-type distribution, *Phys. A* **83**, 210 (1976).
- [49] We have $d\mu^A = d\phi^A d\pi^A$ for $\mathcal{O}^A = \mathcal{W}^A$, $d\mu^A = d\phi^A (d\mu^A = d\pi^A)$ for $\mathcal{O}^A = f^A$ ($\mathcal{O}^A = g^A$) and $d\mu^A = d\phi^A \pi^A / (2\pi)^{\dim A}$ for $\mathcal{O}^A = \mathcal{Q}^A$.
- [50] I. Białyński-Birula and J. Mycielski, Uncertainty relations for information entropy in wave mechanics, *Commun. Math. Phys.* **44**, 129 (1975).
- [51] A. Wehrl, On the relation between classical and quantum-mechanical entropy, *Rep. Math. Phys.* **16**, 353 (1979).
- [52] E. H. Lieb, Proof of an entropy conjecture of Wehrl, *Commun. Math. Phys.* **62**, 35 (1978).
- [53] Z. Van Herstraeten and N. J. Cerf, Quantum Wigner entropy, *Phys. Rev. A* **104**, 042211 (2021).
- [54] S. Floerchinger, T. Haas, and H. Müller-Groeling, Wehrl entropy, entropic uncertainty relations, and entanglement, *Phys. Rev. A* **103**, 062222 (2021).
- [55] N. J. Cerf and T. Haas, Information and majorization theory for fermionic phase-space distributions, [arXiv:2401.08523](https://arxiv.org/abs/2401.08523) (2024).
- [56] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, Entropic uncertainty relations and their applications, *Rev. Mod. Phys.* **89**, 015002 (2017).
- [57] A. Hertz and N. J. Cerf, Continuous-variable entropic uncertainty relations, *J. Phys. A Math. Theor.* **52**, 173001 (2019).
- [58] S. Floerchinger, T. Haas, and M. Schröfl, Relative entropic uncertainty relation for scalar quantum fields, *SciPost Phys.* **12**, 089 (2022).
- [59] S. Ditsch and T. Haas, Entropic distinguishability of quantum fields in phase space, [arXiv:2307.06128](https://arxiv.org/abs/2307.06128) (2023).
- [60] The proportionality constant is $1 + \ln \pi$, $(1 + \ln \pi)/2$ and 1 for the Wigner W , the marginal and the Husimi Q -distribution, respectively.
- [61] G. Adesso, D. Girolami, and A. Serafini, Measuring Gaussian Quantum Information and Correlations Using the Rényi Entropy of Order 2, *Phys. Rev. Lett.* **109**, 190502 (2012).
- [62] A. Serafini, *Quantum Continuous Variables* (CRC Press, 2017).
- [63] E. H. Lieb and R. Seiringer, Stronger subadditivity of entropy, *Phys. Rev. A* **71**, 062329 (2005).
- [64] The upper bound $I(f^A : f^B) + I(g^A : g^B) \leq I(\rho^A : \rho^B)$ has been conjectured in [71].
- [65] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, Area laws in quantum systems: Mutual information and correlations, *Phys. Rev. Lett.* **100**, 070502 (2008).
- [66] L. F. Kozachenko and N. N. Leonenko, Sample Estimate of the Entropy of a Random Vector, *Probl. Peredachi Inf.* **23**, 9 (1987).
- [67] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, *Phys. Rev. E* **69**, 066138 (2004).
- [68] G. V. Steeg, Non-parametric Entropy Estimation Toolbox, <https://github.com/gregversteeg/NPEET> (2014).
- [69] P. Pechukas, Reduced Dynamics Need Not Be Completely Positive, *Phys. Rev. Lett.* **73**, 1060 (1994).

- [70] J. S. Centre, JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing, JLSRF (2019).
- [71] J. Schneeloch, C. J. Broadbent, and J. C. Howell, Uncertainty relation for mutual information, *Phys. Rev. A* **90**, 062119 (2014).

Supplementary material

I. CLASSICAL WIGNER ENTROPY

We illustrate the extensive growth of standard classical entropies by plotting the full classical Wigner entropy, i.e. without subtracting the vacuum contribution, in [Figure 1](#).

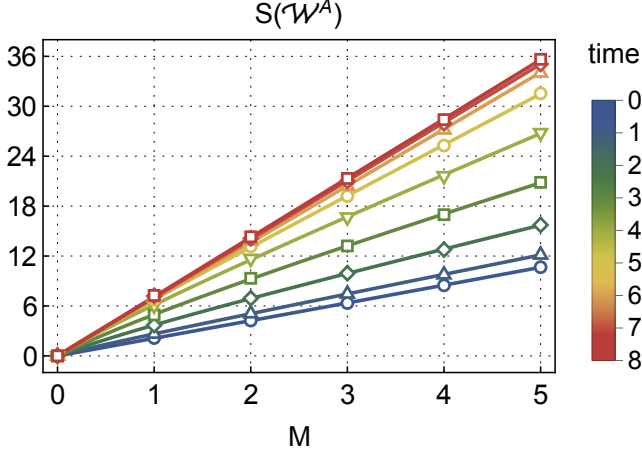


Figure 1. Time evolution of the full Wigner entropy $S(\mathcal{W}^A)$ for which the leading-order volume law is apparent for all times. The area law is barely visible on top of the extensive growth in the early-time regime, i.e. for $0 < t \leq 4$. Note that at $t = 0$ we have $S(\mathcal{W}^A) = S(\bar{\mathcal{W}}^A) = M(1 + \ln \pi) \approx 2.144M$.

II. NON-GAUSSIANITY

We consider a Gaussian model distribution

$$\mathcal{W}^{A,\text{Gauss}} = \frac{1}{Z^A} e^{-\frac{1}{2}(\chi^A)^T (\gamma^A)^{-1} \chi^A}, \quad (1)$$

where $\chi^A = (\phi^A, \pi^A)^T$ is a vector in phase space, $(\gamma^A)^{jj'} = \text{Tr}\{\rho^A\{\chi^j - \chi^j, \chi^{j'} - \chi^{j'}\}\}/2$ denotes the covariance matrix and $Z^A = (2\pi)^M \sqrt{\det \gamma^A}$ is a normalization constant. To assess the non-Gaussianity of a given distribution, \mathcal{W}^A , we introduce the Wigner relative entropy with respect to the nearest Gaussian distribution, i.e. the distribution with the same first- and second-order moments [\[1, 2\]](#)

$$S(\mathcal{W}^A \parallel \mathcal{W}^{A,\text{Gauss}}) = \int d\mu^A \mathcal{W}^A \ln \frac{\mathcal{W}^A}{\mathcal{W}^{A,\text{Gauss}}}. \quad (2)$$

Then, \mathcal{W}^A is (non-)Gaussian if and only if $S(\mathcal{W}^A \parallel \mathcal{W}^{A,\text{Gauss}})(>) = 0$. The non-negativity of the Wigner relative entropy translates into a Gaussian upper bound on the subtracted Wigner entropy, i.e. $\Delta S(\mathcal{W}^A) \leq \Delta S(\mathcal{W}^{A,\text{Gauss}})$, showing that resolving non-Gaussian features decreases the missing information about the underlying distribution. In this sense,

$S(\mathcal{W}^A \parallel \mathcal{W}^{A,\text{Gauss}})$ measures the additional information encoded in \mathcal{W}^A with respect to $\mathcal{W}^{A,\text{Gauss}}$.

To calculate the Wigner relative entropy [\(2\)](#) without reconstructing any distribution, we use [\(1\)](#) and perform a few straightforward simplifications, leading to

$$S(\mathcal{W}^A \parallel \mathcal{W}^{A,\text{Gauss}}) = \Delta S(\mathcal{W}^{A,\text{Gauss}}) - \Delta S(\mathcal{W}^A). \quad (3)$$

While $\Delta S(\mathcal{W}^A)$ is estimated using the k NN method, the subtracted Wigner entropy of the nearest Gaussian distribution is computed via

$$\Delta S(\mathcal{W}^{A,\text{Gauss}}) = \frac{1}{2} \ln \det (2\gamma^A), \quad (4)$$

such that only the covariance matrix has to be extracted from the TWA samples. We show the resulting relative entropy curves in [Figure 2](#) for all times discussed in the main text.

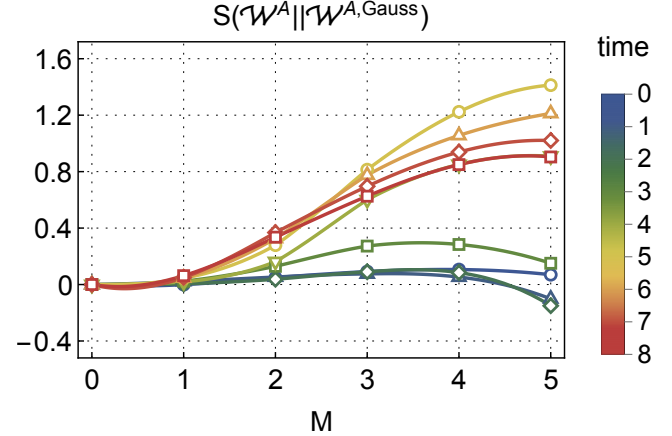


Figure 2. Time evolution of the non-Gaussianity measure $S(\mathcal{W}^A \parallel \mathcal{W}^{A,\text{Gauss}})$. Single-well distributions look rather Gaussian, while non-Gaussian features become apparent for larger subsystems. The non-Gaussianity is negligible for early times and peaks around $t \approx 5$, for which the relative information difference is $\sim 8\%$. We checked negative values at $M = 5$ for early times are caused by an insufficient number of samples, see [\[3\]](#) for details.

III. MODE OCCUPATIONS FOR LATE TIMES

A priori, it is unclear whether TWA gives meaningful results in the late-time limit where local thermalization occurs. As a semi-classical approximation, TWA is expected to hold whenever the momentum modes are occupied mesoscopically, that is, filled up to at least roughly one order of magnitude above the quantum one-half [\[4–7\]](#). In [Figure 3](#), we confirm that this condition is fulfilled for late times by plotting the momentum-mode occupations

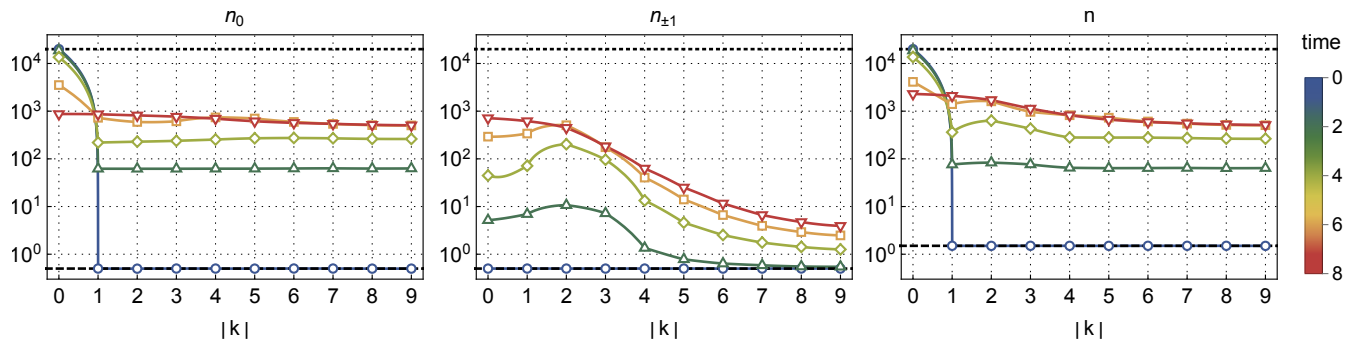


Figure 3. Dynamics of the momentum-mode occupations for the zero mode (left), the side modes (middle) and their sum (right). The atom number $n = 2 \times 10^4$ and the quantum one-half (three-half) are depicted by dotted and dashed lines, respectively.

$n_{m_F}(k) = \langle \mathbf{a}_{m_F}^{k\dagger} \mathbf{a}_{m_F}^k \rangle$ for the zero mode (left panel), the side modes (middle panel) and their sum (right panel).

IV. OTHER CLASSICAL INFORMATION-THEORETIC MEASURES

In analogy to Fig. 2 in the main text, we show the dynamics of the subtracted Wehrl entropy $\Delta S(\mathcal{Q}^A)$, the Wigner mutual information $I(\mathcal{W}^A : \mathcal{W}^B)$ and the marginal mutual information sum $I(f^A : f^B) + I(g^A : g^B)$ in Figure 4. All quantities behave as expected from Fig. 2.

-
- [1] S. Floerchinger, T. Haas, and M. Schröfl, Relative entropic uncertainty relation for scalar quantum fields, *SciPost Phys.* **12**, 089 (2022).
- [2] S. Ditsch and T. Haas, Entropic distinguishability of quantum fields in phase space, [arXiv:2307.06128](https://arxiv.org/abs/2307.06128) (2023).
- [3] Y. Deller, M. Gärttner, T. Haas, M. K. Oberthaler, M. Reh, and H. Strobel, Area laws for classical entropies in a spin-1 Bose-Einstein condensate, Companion paper appearing in the same arXiv posting (2024).
- [4] M. Werner and P. Drummond, Robust Algorithms for Solving Stochastic Partial Differential Equations, *Journal of Computational Physics* **132**, 312–326 (1997).
- [5] M. J. Steel, M. K. Olsen, L. I. Plimak, P. D. Drummond, S. M. Tan, M. J. Collett, D. F. Walls, and R. Graham, Dynamical quantum noise in trapped Bose-Einstein condensates, *Physical Review A* **58**, 4824–4835 (1998).
- [6] A. Sinatra, C. Lobo, and Y. Castin, Classical-Field Method for Time Dependent Bose-Einstein Condensed Gases, *Physical Review Letters* **87**, 210404 (2001).
- [7] A. Sinatra, C. Lobo, and Y. Castin, The truncated Wigner method for Bose-condensed gases: limits of validity and applications, *Journal of Physics B: Atomic, Molecular and Optical Physics* **35**, 3599–3631 (2002).

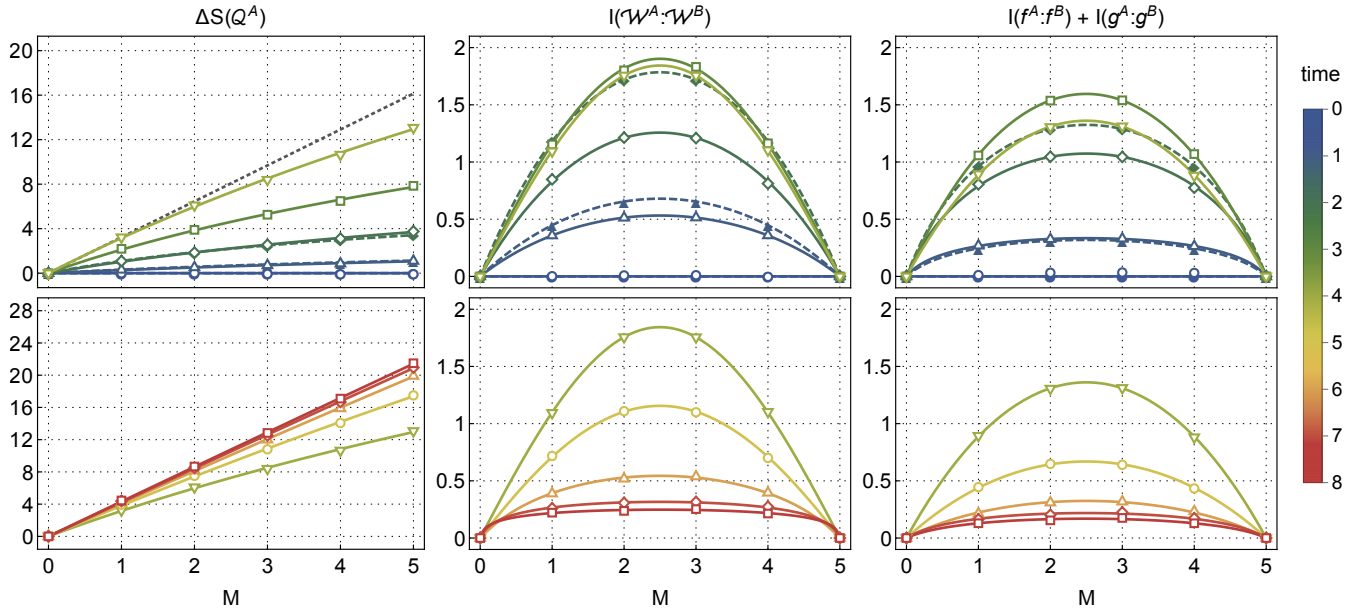


Figure 4. Same analysis as in Fig. 2 in the main text for the subtracted Wehrl entropy (left column), the Wigner mutual information (middle column) and the marginal mutual information sum (right column). All observed quantum features carry over to these three quantities as well. The local temperature $T \approx 5$ is also observed for the subtracted Wehrl entropy. Note here that the latter is based on the differently normalized Husimi Q -distribution, which we accounted for by subtracting $M \ln 2$.

CONCLUSION AND OUTLOOK

We have herein presented the results of various different projects that lie at the intersection of quantum and classical physics as well as semiclassical and variational simulation methods. The scope of the discussed topics is relatively broad and their only common theme is their numerical character. Let us thus give separate closing remarks for each of them.

10.1 NEURAL QUANTUM STATES

The rapid developments in machine learning caused profound consequences in all natural sciences. The arguably most famous example is the advent of AlphaFold [275], which allows to predict the three dimensional structures of proteins up to atomic accuracy based on their amino acid sequence. But also variational approaches in the context of quantum many-body physics, as those discussed in this thesis (see [Chapter 5](#) and [Chapter 7](#)), have profited tremendously from machine learning inspired techniques, opening up novel paths to solve the quantum many-body problem [121].

Today, neural quantum states give state-of-the-art results for challenging problems in two or more dimensions, such as for the real-time evolution of spin systems [154] as well as for ground states of frustrated spin models [98, 99], fermionic gases [170, 171] and molecules [172]. The projects that originated in the context of this thesis aim to push the range of scenarios applicable to treatment with neural quantum states and try to improve our understanding of their properties.

[Chapter 5](#) introduces a novel method to simulate the dynamics of open quantum systems, based on the framework of positive operator valued measures, thereby translating the problem of modeling a complex-valued density matrix to that of a discrete, real-valued probability distribution. Naturally, this opens up the door to employing various tools aimed at encoding probability distributions that have been developed within the machine learning community. The main drawback of this approach is the loss of a guarantee for the positivity of the density operator, since there exist probability distributions for which the eigenvalues can turn negative [207]. Nevertheless, we found this approach to be very competitive compared to other neural quantum state techniques. Future research may aim to use this method to explore phenomena of open quantum systems that so far eluded classical simulation schemes. Also, it would be interesting to benchmark the performance of network architectures beyond recurrent networks.

Particularly, convolutional networks seem to be a promising path forward, given their striking performance for unitary time evolutions [154].

Chapter 7 aims to shed light on the performance differences among different network architectures in ground state optimizations and their underlying reasons. As discussed in Section 4.1 the requirements for neural quantum states are slim; they must merely map basis configurations onto complex-valued coefficients in a differentiable fashion and thus a whole zoo of network architectures exists for the use as neural quantum states. We find that the ground states of stoquastic Hamiltonians, meaning those, that do not feature a sign rule, are easy to learn. Here, all network architectures give similar performances. However, when turning to frustrated models, pronounced differences arise. We identify the symmetrization procedure as a crucial switch towards learning the correct signs and show that autoregressive architectures are by design not optimally suited to handle such scenarios. Recent works investigate the limit of much larger network architectures while still employing second-order optimizers, relying on a reformulation of the stochastic reconfiguration method given in Eq. (4.12) [98, 99]. It would be highly interesting to investigate whether the observed performance differences continue to exist in those regimes.

Generally, neural quantum states have been demonstrated to be a highly promising technique for a wide range of applications. Nevertheless, their optimization is not without challenges, often relies on intuition, and can depend on seemingly unimportant design choices, such as the precise symmetrization scheme [C]. Thus, for neural quantum states to become mainstream applications that can be reliably used in new scenarios of interest, there exist numerous hurdles that future research must address. Subjectively, the most pressing question concerns stability during optimization which can be impaired due to many different factors, such as stochastic noise [159], initialization of the network parameters, infinities in the activation functions, particularly when using holomorphic networks, and even numerical precision [272] as well as whether to interpret the output as the coefficient $\psi(\vec{s})$ or its logarithm $\ln(\psi(\vec{s}))$ [99]. These problems are even more striking in the case of real-time evolution, as a deviation from the correct trajectory is bound to have a compounding effect for the following times. In the context of this thesis, we at one point aimed to model the dynamics of the fermionic $t - V$ model in two dimensions [276]. While early attempts on small 4×4 lattices gave promising results, scaling to larger 6×6 lattices proved challenging without an obvious cause.

10.2 GENERALIZATIONS TO CLASSICAL DYNAMICS

The project presented in [Chapter 6](#) adapts the techniques developed in [Chapter 5](#) and generalizes them to continuous probability densities, where the curse of dimensionality arises from discretizing continuous space in high dimensions. By using a suited ansatz function, such as an invertible neural network, we may reduce the computational complexity to a scaling that is subexponential. The price to pay is that the ability to store arbitrary densities on grids is replaced by only being able to represent those densities which can be encoded using the variational parameters of the ansatz function. We emphasize that a key advantage of normalizing flows lies in their ability to generate exact samples with normalized probabilities, allowing straightforward estimations of functionals of the distribution, such as its differential entropy.

The studied examples either allow for analytical or numerical solutions so that we have a clear benchmark available. For future applications, it would be interesting to probe those regimes and types of partial differential equations where other methods are bound to fail so that we can examine when the variational Monte-Carlo approach gives an edge over existing techniques.

10.3 AREA AND VOLUME LAWS IN BOSE-EINSTEIN CONDENSATES

In [Chapter 8](#) and [Chapter 9](#) we explore the presence of area- and volume-laws in the phase-space descriptions of spin-1 Bose-Einstein condensates following a quench from an initial product state. While area laws are typically observed for the quantum entropies of ground states, their presence in states obtained from a quench is also established [[190](#)]. What is novel here, is that these quantum features carry over to the classical, i.e. differential entropies of phase-space descriptions of quantum systems upon subtraction of the extensive vacuum contributions.

[Chapter 8](#) lays out the theoretical and numerical techniques that are employed in detail. In short, we synthetically generate samples of both the Wigner W - and Husimi Q -distribution (see Eqs. [\(2.60\)](#) and [\(2.64\)](#)) and estimate their entropy and mutual information using a general, asymptotically unbiased k -nearest neighbor estimator, while keeping the number of samples required for convergence at an experimentally feasible level. To ascertain the stability of the proposal with respect to measurement details such as quench parameters, readout time, thermal noise, subsystem position, system size, and boundary conditions we carry out ablative studies where we alter a single of the aforementioned settings. While we may safely confirm the presence of the area law in most of these cases, going to strongly non-Gaussian states proves challenging for the estimator. In [Chapter 9](#) we extend

our analysis to later times and observe the transition from an area- to a volume-law as the system thermalizes, thus allowing an information theoretical perspective on quantum thermalization from phase-space descriptions (also see [Section 2.8](#)).

As the presence of these information-theoretical phenomena in phase-space distributions is novel [277], various paths forward can be envisioned. For example, one could imagine carrying out a similar analysis for the ground state of the system at hand, although it may be expected that the experimental preparation of such states is considerably more challenging, as coherence during the adiabatic state preparation must be ensured. One might also extend the analysis at hand to two-dimensional setups, where area laws in the entanglement entropy for bosonic systems have been proven rigorously [278–281]. However, such considerations must always take into account the sample complexity to reach convergence when estimating the entropy. Thus, it may also be interesting to think about entropy estimation schemes that incorporate a physical prior, such as continuous normalizing flows that may be restricted to only encode such flows that are in agreement with Hamiltonian dynamics [282–284].

BIBLIOGRAPHY

- [1] A. Einstein, B. Podolsky, and N. Rosen, “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” *Phys. Rev.* **47**, 777–780 (1935).
- [2] M. A. Nielsen and I. L. Chuang, “Quantum Computation and Quantum Information,” *Cambridge University Press* (2012).
- [3] A. Einstein, “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt,” *Annalen der Physik* **322**, 132–148 (1905).
- [4] R. P. Feynman, “Simulating physics with computers,” *International Journal of Theoretical Physics* **21**, 467–488 (1982).
- [5] I. Buluta and F. Nori, “Quantum Simulators,” *Science* **326**, 108–111 (2009).
- [6] I. M. Georgescu, S. Ashhab, and F. Nori, “Quantum simulation,” *Rev. Mod. Phys.* **86**, 153–185 (2014).
- [7] E. Altman et al., “Quantum Simulators: Architectures and Opportunities,” *PRX Quantum* **2**, 017003 (2021).
- [8] A. Steane, “Quantum computing,” *Reports on Progress in Physics* **61**, 117–173 (1998).
- [9] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O’Brien, “Quantum computers,” *Nature* **464**, 45–53 (2010).
- [10] I. Bloch, J. Dalibard, and W. Zwerger, “Many-body physics with ultracold gases,” *Rev. Mod. Phys.* **80**, 885–964 (2008).
- [11] I. Bloch, J. Dalibard, and S. Nascimbène, “Quantum simulations with ultracold quantum gases,” *Nature Physics* **8**, 267–276 (2012).
- [12] M. Lewenstein, A. Sanpera, V. Ahufinger, B. Damski, A. Sen(De), and U. Sen, “Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond,” *Advances in Physics* **56**, 243–379 (2007).
- [13] J. L. O’Brien, A. Furusawa, and J. Vučković, “Photonic quantum technologies,” *Nature Photonics* **3**, 687–695 (2009).
- [14] J. W. Silverstone, D. Bonneau, J. L. O’Brien, and M. G. Thompson, “Silicon Quantum Photonics,” *IEEE Journal of Selected Topics in Quantum Electronics* **22**, 390–402 (2016).
- [15] R. Blatt and C. F. Roos, “Quantum simulations with trapped ions,” *Nature Physics* **8**, 277–284 (2012).

- [16] C. Schneider, D. Porras, and T. Schaetz, "Experimental quantum simulations of many-body physics with trapped ions," *Reports on Progress in Physics* **75**, 024401 (2012).
- [17] B. Reif, S. E. Ashbrook, L. Emsley, and M. Hong, "Solid-state NMR spectroscopy," *Nature Reviews Methods Primers* **1**, 2 (2021).
- [18] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch, "Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms," *Nature* **415**, 39–44 (2002).
- [19] L. Anderegg, L. W. Cheuk, Y. Bao, S. Burchesky, W. Ketterle, K.-K. Ni, and J. M. Doyle, "An optical tweezer array of ultracold molecules," *Science* **365**, 1156–1158 (2019).
- [20] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, "Measuring entanglement entropy in a quantum many-body system," *Nature* **528**, 77–83 (2015).
- [21] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, "Real-time dynamics of lattice gauge theories with a few-qubit quantum computer," *Nature* **534**, 516–519 (2016).
- [22] P. T. Brown, D. Mitra, E. Guardado-Sanchez, R. Nourafkan, A. Reymbaut, C.-D. Hébert, S. Bergeron, A.-M. S. Tremblay, J. Kokalj, D. A. Huse, P. Schauß, and W. S. Bakr, "Bad metallic transport in a cold atom Fermi-Hubbard system," *Science* **363**, 379–382 (2019).
- [23] M. A. Nichols, L. W. Cheuk, M. Okan, T. R. Hartke, E. Mendez, T. Senthil, E. Khatami, H. Zhang, and M. W. Zwierlein, "Spin transport in a Mott insulator of ultracold fermions," *Science* **363**, 383–387 (2019).
- [24] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, "String patterns in the doped Hubbard model," *Science* **365**, 251–256 (2019).
- [25] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature* **574**, 505–510 (2019).
- [26] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, "Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator," *Nature* **551**, 601–604 (2017).
- [27] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, "Probing many-body dynamics on a 51-atom quantum simulator," *Nature* **551**, 579–584 (2017).

- [28] K. V. Houcke, F. Werner, E. Kozik, N. Prokof'ev, B. Svistunov, M. J. H. Ku, A. T. Sommer, L. W. Cheuk, A. Schirotzek, and M. W. Zwierlein, "Feynman diagrams versus Fermi-gas Feynman emulator," *Nature Physics* **8**, 366–370 (2012).
- [29] M. J. H. Ku, A. T. Sommer, L. W. Cheuk, and M. W. Zwierlein, "Revealing the Superfluid Lambda Transition in the Universal Thermodynamics of a Unitary Fermi Gas," *Science* **335**, 563–567 (2012).
- [30] T. Ozawa and H. M. Price, "Topological quantum matter in synthetic dimensions," *Nature Reviews Physics* **1**, 349–357 (2019).
- [31] Y.-J. Lin, K. Jiménez-García, and I. B. Spielman, "Spin-orbit-coupled Bose–Einstein condensates," *Nature* **471**, 83–86 (2011).
- [32] A. J. Kollár, M. Fitzpatrick, and A. A. Houck, "Hyperbolic lattices in circuit quantum electrodynamics," *Nature* **571**, 45–50 (2019).
- [33] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch, "Observation of many-body localization of interacting fermions in a quasirandom optical lattice," *Science* **349**, 842–845 (2015).
- [34] C. E. Bradley, J. Randall, M. H. Abobeih, R. C. Berrevoets, M. J. Degen, M. A. Bakker, M. Markham, D. J. Twitchen, and T. H. Taminiau, "A Ten-Qubit Solid-State Spin Register with Quantum Memory up to One Minute," *Phys. Rev. X* **9**, 031045 (2019).
- [35] P. C. Humphreys, N. Kalb, J. P. J. Morits, R. N. Schouten, R. F. L. Vermeulen, D. J. Twitchen, M. Markham, and R. Hanson, "Deterministic delivery of remote entanglement on a quantum network," *Nature* **558**, 268–273 (2018).
- [36] E. Prati, K. Kumagai, M. Hori, and T. Shinada, "Band transport across a chain of dopant sites in silicon over micron distances and high temperatures," *Scientific Reports* **6**, 19704 (2016).
- [37] L. M. K. Vandersypen and M. A. Eriksson, "Quantum computing with semiconductor spins," *Physics Today* **72**, 38–45 (2019).
- [38] T. Hensgens, T. Fujita, L. Janssen, X. Li, C. J. V. Diepen, C. Reichl, W. Wegscheider, S. D. Sarma, and L. M. K. Vandersypen, "Quantum simulation of a Fermi–Hubbard model using a semiconductor quantum dot array," *Nature* **548**, 70–73 (2017).
- [39] A. González-Tudela, C.-L. Hung, D. E. Chang, J. I. Cirac, and H. J. Kimble, "Subwavelength vacuum lattices and atom–atom interactions in two-dimensional photonic crystals," *Nature Photonics* **9**, 320–325 (2015).

- [40] M. Hartmann, F. Brandão, and M. Plenio, “Quantum many-body phenomena in coupled cavity arrays,” *Laser & Photonics Reviews* **2**, 527–556 (2008).
- [41] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, “Experimental boson sampling,” *Nature Photonics* **7**, 540–544 (2013).
- [42] V. D. Vaidya, Y. Guo, R. M. Kroeze, K. E. Ballantine, A. J. Kollár, J. Keeling, and B. L. Lev, “Tunable-Range, Photon-Mediated Atomic Interactions in Multimode Cavity QED,” *Phys. Rev. X* **8**, 011002 (2018).
- [43] M. A. Norcia, R. J. Lewis-Swan, J. R. K. Cline, B. Zhu, A. M. Rey, and J. K. Thompson, “Cavity-mediated collective spin-exchange interactions in a strontium superradiant laser,” *Science* **361**, 259–262 (2018).
- [44] E. J. Davis, G. Bentsen, L. Homeier, T. Li, and M. H. Schleier-Smith, “Photon-Mediated Spin-Exchange Dynamics of Spin-1 Atoms,” *Phys. Rev. Lett.* **122**, 010405 (2019).
- [45] M. J. Hartmann, “Quantum simulation with interacting photons,” *Journal of Optics* **18**, 104005 (2016).
- [46] L. W. Clark, N. Schine, C. Baum, N. Jia, and J. Simon, “Observation of Laughlin states made of light,” *Nature* **582**, 41–45 (2020).
- [47] I. D. Leroux, M. H. Schleier-Smith, and V. Vuletić, “Implementation of Cavity Squeezing of a Collective Atomic Spin,” *Phys. Rev. Lett.* **104**, 073602 (2010).
- [48] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, “Quantum metrology with nonclassical states of atomic ensembles,” *Rev. Mod. Phys.* **90**, 035005 (2018).
- [49] G. Barontini, L. Hohmann, F. Haas, J. Estève, and J. Reichel, “Deterministic generation of multiparticle entanglement by quantum Zeno dynamics,” *Science* **349**, 1317–1321 (2015).
- [50] S. Welte, B. Hacker, S. Daiss, S. Ritter, and G. Rempe, “Photon-Mediated Quantum Gate between Two Neutral Atoms in an Optical Cavity,” *Phys. Rev. X* **8**, 011018 (2018).
- [51] S. de Léséleuc, V. Lienhard, P. Scholl, D. Barredo, S. Weber, N. Lang, H. P. Büchler, T. Lahaye, and A. Browaeys, “Observation of a symmetry-protected topological phase of interacting bosons with Rydberg atoms,” *Science* **365**, 775–780 (2019).
- [52] M. Ganzhorn, D. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, “Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer,” *Phys. Rev. Appl.* **11**, 044092 (2019).

- [53] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature* **549**, 242–246 (2017).
- [54] P. J. J. O'Malley et al., "Scalable Quantum Simulation of Molecular Energies," *Phys. Rev. X* **6**, 031007 (2016).
- [55] R. Ma, B. Saxberg, C. Owens, N. Leung, Y. Lu, J. Simon, and D. I. Schuster, "A dissipatively stabilized Mott insulator of photons," *Nature* **566**, 51–57 (2019).
- [56] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, "Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator," *Phys. Rev. X* **8**, 031022 (2018).
- [57] J. W. Britton, B. C. Sawyer, A. C. Keith, C.-C. J. Wang, J. K. Freericks, H. Uys, M. J. Biercuk, and J. J. Bollinger, "Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins," *Nature* **484**, 489–492 (2012).
- [58] J. G. Bohnet, B. C. Sawyer, J. W. Britton, M. L. Wall, A. M. Rey, M. Foss-Feig, and J. J. Bollinger, "Quantum spin dynamics and entanglement generation with hundreds of trapped ions," *Science* **352**, 1297–1301 (2016).
- [59] D. Jaksch and P. Zoller, "The cold atom Hubbard toolbox," *Annals of Physics* **315**, 52–79 (2005).
- [60] L. W. Cheuk, M. A. Nichols, M. Okan, T. Gersdorf, V. V. Ramasesh, W. S. Bakr, T. Lompe, and M. W. Zwierlein, "Quantum-Gas Microscope for Fermionic Atoms," *Phys. Rev. Lett.* **114**, 193001 (2015).
- [61] T. Langen, S. Erne, R. Geiger, B. Rauer, T. Schweigler, M. Kuhnert, W. Rohringer, I. E. Mazets, T. Gasenzer, and J. Schmiedmayer, "Experimental observation of a generalized Gibbs ensemble," *Science* **348**, 207–211 (2015).
- [62] M. J. Martin, M. Bishof, M. D. Swallows, X. Zhang, C. Benko, J. von-Stecher, A. V. Gorshkov, A. M. Rey, and J. Ye, "A Quantum Many-Body Spin System in an Optical Lattice Clock," *Science* **341**, 632–636 (2013).
- [63] S. Kolkowitz, S. L. Bromley, T. Bothwell, M. L. Wall, G. E. Marti, A. P. Koller, X. Zhang, A. M. Rey, and J. Ye, "Spin-orbit-coupled fermions in an optical lattice clock," *Nature* **542**, 66–70 (2016).
- [64] C. Viermann, M. Sparn, N. Liebster, M. Hans, E. Kath, Á. Parra-López, M. Tolosa-Simeón, N. Sánchez-Kuntz, T. Haas, H. Strobel, S. Floerchinger, and M. K. Oberthaler, "Quantum field simulator for dynamics in curved spacetime," *Nature* **611**, 260–264 (2022).

- [65] M. Prüfer, P. Kunkel, H. Strobel, S. Lannig, D. Linnemann, C.-M. Schmied, J. Berges, T. Gasenzer, and M. K. Oberthaler, "Observation of universal dynamics in a spinor Bose gas far from equilibrium," *Nature* **563**, 217–220 (2018).
- [66] L. D. Marco, G. Valtolina, K. Matsuda, W. G. Tobias, J. P. Covey, and J. Ye, "A degenerate Fermi gas of polar molecules," *Science* **363**, 853–856 (2019).
- [67] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, "Quantum Algorithms for Quantum Chemistry and Quantum Materials Science," *Chemical Reviews* **120**, 12685–12717 (2020).
- [68] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, "Simulating Chemistry Using Quantum Computers," *Annual Review of Physical Chemistry* **62**, 185–207 (2011).
- [69] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz," *Quantum Science and Technology* **4**, 014008 (2018).
- [70] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, "Quantum Chemistry in the Age of Quantum Computing," *Chemical Reviews* **119**, 10856–10915 (2019).
- [71] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry," *Rev. Mod. Phys.* **92**, 015003 (2020).
- [72] M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, "Boson localization and the superfluid-insulator transition," *Phys. Rev. B* **40**, 546–570 (1989).
- [73] K. Sheshadri, H. R. Krishnamurthy, R. Pandit, and T. V. Ramakrishnan, "Superfluid and Insulating Phases in an Interacting-Boson Model: Mean-Field Theory and the RPA," *Europhysics Letters (EPL)* **22**, 257–263 (1993).
- [74] J. K. Freericks and H. Monien, "Phase diagram of the Bose-Hubbard Model," *Europhysics Letters (EPL)* **26**, 545–550 (1994).
- [75] D. van Oosten, P. van der Straten, and H. T. C. Stoof, "Quantum phases in an optical lattice," *Phys. Rev. A* **63**, 053601 (2001).
- [76] 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).
- [77] R. N. Palmer, C. Moura Alves, and D. Jaksch, "Detection and characterization of multipartite entanglement in optical lattices," *Phys. Rev. A* **72**, 042335 (2005).

- [78] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, “Quantum thermalization through entanglement in an isolated many-body system,” *Science* **353**, 794–800 (2016).
- [79] A. Friedenauer, H. Schmitz, J. T. Glueckert, D. Porras, and T. Schaetz, “Simulating a quantum magnet with trapped ions,” *Nature Physics* **4**, 757–761 (2008).
- [80] R. Nandkishore and D. A. Huse, “Many-Body Localization and Thermalization in Quantum Statistical Mechanics,” *Annual Review of Condensed Matter Physics* **6**, 15–38 (2015).
- [81] G. B. Mbeng, A. Russomanno, and G. E. Santoro, *The quantum Ising chain for beginners*, 2020.
- [82] S. Suzuki, J.-i. Inoue, and B. K. Chakrabarti, “Quantum Ising Phases and Transitions in Transverse Ising Models,” *Springer Berlin Heidelberg* (2013).
- [83] L. Savary and L. Balents, “Quantum spin liquids: a review,” *Reports on Progress in Physics* **80**, 016502 (2016).
- [84] Y. Zhou, K. Kanoda, and T.-K. Ng, “Quantum spin liquid states,” *Rev. Mod. Phys.* **89**, 025003 (2017).
- [85] C. Broholm, R. J. Cava, S. A. Kivelson, D. G. Nocera, M. R. Norman, and T. Senthil, “Quantum spin liquids,” *Science* **367**, eaayo668 (2020).
- [86] J. Eisert, M. Friesdorf, and C. Gogolin, “Quantum many-body systems out of equilibrium,” *Nature Physics* **11**, 124–130 (2015).
- [87] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, “Evidence for the utility of quantum computing before fault tolerance,” *Nature* **618**, 500–505 (2023).
- [88] T. Begušić and G. K.-L. Chan, “Fast classical simulation of evidence for the utility of quantum computing before fault tolerance,” *arXiv:2306.16372* (2023).
- [89] J. Tindall, M. Fishman, M. Stoudenmire, and D. Sels, “Efficient tensor network simulation of IBM’s Eagle kicked Ising experiment,” *arXiv:2306.14887* (2023).
- [90] E. G. D. Torre and M. M. Roses, “Dissipative mean-field theory of IBM utility experiment,” *arXiv:2308.01339* (2023).
- [91] S. Patra, S. S. Jahromi, S. Singh, and R. Orus, “Efficient tensor network simulation of IBM’s largest quantum processors,” *arXiv:2309.15642* (2023).
- [92] H.-J. Liao, K. Wang, Z.-S. Zhou, P. Zhang, and T. Xiang, “Simulation of IBM’s kicked Ising experiment with Projected Entangled Pair Operator,” *arXiv:2308.03082* (2023).

- [93] F. Becca and S. Sorella, "Quantum Monte Carlo Approaches for Correlated Systems," *Cambridge University Press* (2017).
- [94] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, "Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo," *New Journal of Physics* **11**, 083026 (2009).
- [95] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," *Annals of Physics* **326**, 96–192 (2011).
- [96] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of Physics* **349**, 117–158 (2014).
- [97] D. Wu et al., "Variational Benchmarks for Quantum Many-Body Problems," *arXiv:2302.04919* (2023).
- [98] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, "A simple linear algebra identity to optimize Large-Scale Neural Network Quantum States," *arXiv:2310.05715* (2023).
- [99] A. Chen and M. Heyl, "Efficient optimization of deep neural quantum states toward machine precision," *arXiv:2302.01941* (2023).
- [100] W.-Y. Liu, S.-S. Gong, Y.-B. Li, D. Poilblanc, W.-Q. Chen, and Z.-C. Gu, "Gapless quantum spin liquid and global phase diagram of the spin-1/2 $J_1 - J_2$ square antiferromagnetic Heisenberg model," *Science Bulletin* **67**, 1034–1041 (2022).
- [101] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks* **4**, 251–257 (1991).
- [102] B. Csáji, "Approximation with Artificial Neural Networks," *M.Sc. Thesis* (2001).
- [103] A. Kratsios and E. Bilokopytov, "Non-Euclidean Universal Approximation," *arXiv:2006.02341* (2020).
- [104] D.-X. Zhou, "Universality of deep convolutional neural networks," *Applied and Computational Harmonic Analysis* **48**, 787–794 (2020).
- [105] A. Heinecke, J. Ho, and W.-L. Hwang, "Refinement and Universal Approximation via Sparsely Connected ReLU Convolution Nets," *IEEE Signal Processing Letters* **27**, 1175–1179 (2020).
- [106] J. Park and I. W. Sandberg, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation* **3**, 246–257 (1991).
- [107] D. Yarotsky, "Universal Approximations of Invariant Maps by Neural Networks," *Constructive Approximation* **55**, 407–474 (2021).

- [108] M. Zakwan, M. d'Angelo, and G. Ferrari-Trecate, "Universal Approximation Property of Hamiltonian Deep Neural Networks," *IEEE Control Systems Letters*, 1–1 (2023).
- [109] S. Sonoda and N. Murata, "Neural network with unbounded activation functions is universal approximator," *Applied and Computational Harmonic Analysis* **43**, 233–268 (2017).
- [110] S. Park, C. Yun, J. Lee, and J. Shin, "Minimum Width for Universal Approximation," *arXiv:2006.08859* (2020).
- [111] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems* **2**, 303–314 (1989).
- [112] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks* **6**, 861–867 (1993).
- [113] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numerica* **8**, 143–195 (1999).
- [114] Z. Shen, H. Yang, and S. Zhang, "Optimal approximation rate of ReLU networks in terms of width and depth," *Journal de Mathématiques Pures et Appliquées* **157**, 101–135 (2022).
- [115] G. Gripenberg, "Approximation by neural networks with a bounded number of nodes at each level," *Journal of Approximation Theory* **122**, 260–266 (2003).
- [116] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, "The Expressive Power of Neural Networks: A View from the Width," *arXiv:1709.02540* (2017).
- [117] B. Hanin and M. Sellke, "Approximating Continuous Functions by ReLU Nets of Minimal Width," *arXiv:1710.11278* (2017).
- [118] P. Kidger and T. Lyons, "Universal Approximation with Deep Narrow Networks," *arXiv:1905.08539* (2019).
- [119] A. Kratsios and L. Papon, "Universal Approximation Theorems for Differentiable Geometric Deep Learning," *arXiv:2101.05390* (2021).
- [120] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems* **25** (2012).
- [121] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science* **355**, 602–606 (2017).
- [122] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, "Neural-network quantum state tomography," *Nature Physics* **14**, 447–450 (2018).

- [123] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, “Efficient quantum state tomography,” *Nature Communications* **1**, 149 (2010).
- [124] T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, “Scalable Reconstruction of Density Matrices,” *Phys. Rev. Lett.* **111**, 020401 (2013).
- [125] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, “Efficient tomography of a quantum many-body system,” *Nature Physics* **13**, 1158–1162 (2017).
- [126] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, “Quantum State Tomography via Compressed Sensing,” *Phys. Rev. Lett.* **105**, 150401 (2010).
- [127] C. Schwemmer, G. Tóth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, “Experimental Comparison of Efficient Tomography Schemes for a Six-Qubit State,” *Phys. Rev. Lett.* **113**, 040503 (2014).
- [128] C. A. Riofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, “Experimental quantum compressed sensing for a seven-qubit system,” *Nature Communications* **8**, 15305 (2017).
- [129] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, “Permutationally Invariant Quantum Tomography,” *Phys. Rev. Lett.* **105**, 250403 (2010).
- [130] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter, “Permutationally invariant state reconstruction,” *New Journal of Physics* **14**, 105001 (2012).
- [131] 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).
- [132] S. Czischek, M. S. Moss, M. Radzihovsky, E. Merali, and R. G. Melko, “Data-enhanced variational Monte Carlo simulations for Rydberg atom arrays,” *Phys. Rev. B* **105**, 205108 (2022).
- [133] M. S. Moss, S. Ebadi, T. T. Wang, G. Semeghini, A. Bohrdt, M. D. Lukin, and R. G. Melko, “Enhancing variational Monte Carlo using a programmable quantum simulator,” [arXiv:2308.02647](https://arxiv.org/abs/2308.02647) (2023).
- [134] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, “Recurrent neural network wave functions,” *Phys. Rev. Res.* **2**, 023358 (2020).

- [135] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, “Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems,” *Phys. Rev. Lett.* **124**, 020503 (2020).
- [136] K. Choo, G. Carleo, N. Regnault, and T. Neupert, “Symmetries and Many-Body Excitations with Neural-Network Quantum States,” *Phys. Rev. Lett.* **121**, 167204 (2018).
- [137] K. Choo, T. Neupert, and G. Carleo, “Two-dimensional frustrated J_1 – J_2 model studied with neural network quantum states,” *Phys. Rev. B* **100**, 125124 (2019).
- [138] X. Liang, M. Li, Q. Xiao, H. An, L. He, X. Zhao, J. Chen, C. Yang, F. Wang, H. Qian, L. Shen, D. Jia, Y. Gu, X. Liu, and Z. Wei, “ 2^{1296} Exponentially Complex Quantum Many-Body Simulation via Scalable Deep Learning Method,” [arXiv:2204.07816](https://arxiv.org/abs/2204.07816) (2022).
- [139] M. Li, J. Chen, Q. Xiao, F. Wang, Q. Jiang, X. Zhao, R. Lin, H. An, X. Liang, and L. He, “Bridging the Gap between Deep Learning and Frustrated Quantum Spin System for Extreme-Scale Simulations on New Generation of Sunway Supercomputer,” *IEEE Transactions on Parallel and Distributed Systems* **33**, 2846–2859 (2022).
- [140] H. Chen, D. Hendry, P. Weinberg, and A. E. Feiguin, “Systematic improvement of neural network quantum states using a Lanczos recursion,” [arXiv:2206.14307](https://arxiv.org/abs/2206.14307) (2022).
- [141] Y. Nomura and M. Imada, “Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy,” *Phys. Rev. X* **11**, 031034 (2021).
- [142] F. Ferrari, F. Becca, and J. Carrasquilla, “Neural Gutzwiller-projected variational wave functions,” *Phys. Rev. B* **100**, 125131 (2019).
- [143] J.-Q. Wang, R.-Q. He, and Z.-Y. Lu, “Variational optimization of the amplitude of neural-network quantum many-body ground states,” [arXiv:2308.09664](https://arxiv.org/abs/2308.09664) (2023).
- [144] X. Liang, S.-J. Dong, and L. He, “Hybrid convolutional neural network and projected entangled pair states wave functions for quantum many-particle states,” *Phys. Rev. B* **103**, 035138 (2021).
- [145] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, “Solving frustrated quantum many-particle models with convolutional neural networks,” *Phys. Rev. B* **98**, 104426 (2018).
- [146] A. Szabó and C. Castelnovo, “Neural network wave functions and the sign problem,” *Phys. Rev. Res.* **2**, 033075 (2020).

- [147] E. Ledinauskas and E. Anisimovas, “Scalable Imaginary Time Evolution with Neural Network Quantum States,” [arXiv:2307.15521](#) (2023).
- [148] C. Roth, A. Szabó, and A. H. MacDonald, “High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks,” *Physical Review B* **108**, 054410 (2023).
- [149] C. Roth and A. H. MacDonald, “Group Convolutional Neural Networks Improve Quantum State Accuracy,” [arXiv:2104.05085](#) (2021).
- [150] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, “Broken-Symmetry Ground States of the Heisenberg Model on the Pyrochlore Lattice,” *Phys. Rev. X* **11**, 041021 (2021).
- [151] Y. Nomura, “Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry,” *Journal of Physics: Condensed Matter* **33**, 174003 (2021).
- [152] L. Yang, Z. Leng, G. Yu, A. Patel, W.-J. Hu, and H. Pu, “Deep learning-enhanced variational Monte Carlo method for quantum many-body physics,” *Phys. Rev. Res.* **2**, 012039 (2020).
- [153] A. Chen, K. Choo, N. Astrakhantsev, and T. Neupert, “Neural network evolution strategy for solving quantum sign structures,” *Phys. Rev. Res.* **4**, L022026 (2022).
- [154] M. Schmitt and M. Heyl, “Quantum Many-Body Dynamics in Two Dimensions with Artificial Neural Networks,” *Phys. Rev. Lett.* **125**, 100503 (2020).
- [155] M. Schmitt, M. M. Rams, J. Dziarmaga, M. Heyl, and W. H. Zurek, “Quantum phase transition dynamics in the two-dimensional transverse-field Ising model,” *Science Advances* **8**, eabl6850 (2022).
- [156] I. L. Gutiérrez and C. B. Mendl, “Real time evolution with neural-network quantum states,” *Quantum* **6**, 627 (2022).
- [157] S. Czischek, M. Gärttner, and T. Gasenzer, “Quenches near Ising quantum criticality as a challenge for artificial neural networks,” *Phys. Rev. B* **98**, 024311 (2018).
- [158] K. Donatella, Z. Denis, A. Le Boité, and C. Ciuti, “Dynamics with autoregressive neural quantum states: Application to critical quench dynamics,” *Phys. Rev. A* **108**, 022210 (2023).
- [159] D. Hofmann, G. Fabiani, J. H. Mentink, G. Carleo, and M. A. Sentef, “Role of stochastic noise and generalization error in the time propagation of neural-network quantum states,” *SciPost Phys.* **12**, 165 (2022).

- [160] S.-H. Lin and F. Pollmann, “Scaling of Neural-Network Quantum States for Time Evolution,” *physica status solidi (b)* **259** (2022).
- [161] B. Jónsson, B. Bauer, and G. Carleo, “Neural-network states for the classical simulation of quantum computing,” [arXiv:1808.05232](https://arxiv.org/abs/1808.05232) (2018).
- [162] M. J. Hartmann and G. Carleo, “Neural-Network Approach to Dissipative Quantum Many-Body Dynamics,” *Phys. Rev. Lett.* **122**, 250502 (2019).
- [163] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, “Variational Neural-Network Ansatz for Steady States in Open Quantum Systems,” *Phys. Rev. Lett.* **122**, 250503 (2019).
- [164] N. Yoshioka and R. Hamazaki, “Constructing neural stationary states for open quantum many-body systems,” *Phys. Rev. B* **99**, 214306 (2019).
- [165] A. Nagy and V. Savona, “Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems,” *Phys. Rev. Lett.* **122**, 250501 (2019).
- [166] D. Luo, Z. Chen, J. Carrasquilla, and B. K. Clark, “Autoregressive Neural Network for Simulating Open Quantum Systems via a Probabilistic Formulation,” *Phys. Rev. Lett.* **128**, 090501 (2022).
- [167] D. Eeltink, F. Vicentini, and V. Savona, “Variational dynamics of open quantum systems in phase space,” [arXiv:2307.07429](https://arxiv.org/abs/2307.07429) (2023).
- [168] O. Dugan, P. Y. Lu, R. Dangovski, D. Luo, and M. Soljačić, “Q-Flow: Generative Modeling for Differential Equations of Open Quantum Dynamics with Normalizing Flows,” [arXiv:2302.12235](https://arxiv.org/abs/2302.12235) (2023).
- [169] K. Choo, A. Mezzacapo, and G. Carleo, “Fermionic neural-network states for ab-initio electronic structure,” *Nature Communications* **11**, 2368 (2020).
- [170] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, “Neural-network quantum states for ultra-cold Fermi gases,” [arXiv:2305.08831](https://arxiv.org/abs/2305.08831) (2023).
- [171] G. Pescia, J. Nys, J. Kim, A. Lovato, and G. Carleo, “Message-Passing Neural Quantum States for the Homogeneous Electron Gas,” [arXiv:2305.07240](https://arxiv.org/abs/2305.07240) (2023).
- [172] I. von Glehn, J. S. Spencer, and D. Pfau, “A Self-Attention Ansatz for Ab-initio Quantum Chemistry,” [arXiv:2211.13672](https://arxiv.org/abs/2211.13672) (2022).

- [173] J. Hermann, Z. Schätzle, and F. Noé, “Deep-neural-network solution of the electronic Schrödinger equation,” *Nature Chemistry* **12**, 891–897 (2020).
- [174] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, “Ab initio solution of the many-electron Schrödinger equation with deep neural networks,” *Phys. Rev. Res.* **2**, 033429 (2020).
- [175] H. Lange, F. Döschl, J. Carrasquilla, and A. Bohrdt, “Neural network approach to quasiparticle dispersions in doped anti-ferromagnets,” *arXiv:2310.08578* (2023).
- [176] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, “Fermionic wave functions from neural-network constrained hidden states,” *Proceedings of the National Academy of Sciences* **119**, e2122059119 (2022).
- [177] G. Cassella, H. Sutterud, S. Azadi, N. D. Drummond, D. Pfau, J. S. Spencer, and W. M. C. Foulkes, “Discovering Quantum Phase Transitions with Fermionic Neural Networks,” *Phys. Rev. Lett.* **130**, 036401 (2023).
- [178] J. Stokes, J. R. Moreno, E. A. Pnevmatikakis, and G. Carleo, “Phases of two-dimensional spinless lattice fermions with first-quantized deep neural-network quantum states,” *Phys. Rev. B* **102**, 205122 (2020).
- [179] D. Luo and B. K. Clark, “Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions,” *Phys. Rev. Lett.* **122**, 226401 (2019).
- [180] T. Zhao, J. Stokes, and S. Veerapaneni, “Scalable neural quantum states architecture for quantum chemistry,” *Machine Learning: Science and Technology* **4**, 025034 (2023).
- [181] A. Lovato, C. Adams, G. Carleo, and N. Rocco, “Hidden-nucleons neural-network quantum states for the nuclear many-body problem,” *Phys. Rev. Res.* **4**, 043178 (2022).
- [182] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, “Neural-network quantum states for periodic systems in continuous space,” *Phys. Rev. Res.* **4**, 023138 (2022).
- [183] J. M. Martyn, K. Najafi, and D. Luo, “Variational Neural-Network Ansatz for Continuum Quantum Field Theory,” *Phys. Rev. Lett.* **131**, 081601 (2023).
- [184] P. F. Bedaque, H. Kumar, and A. Sheng, “Neural Network Solutions of Bosonic Quantum Systems in One Dimension,” *arXiv:2309.02352* (2023).
- [185] O. Kaestle and A. Carmele, “Efficient bit encoding of neural networks for Fock states,” *Phys. Rev. A* **103**, 062409 (2021).

- [186] D.-L. Deng, X. Li, and S. Das Sarma, “Quantum Entanglement in Neural Network States,” *Phys. Rev. X* **7**, 021021 (2017).
- [187] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, “Quantum Entanglement in Deep Learning Architectures,” *Phys. Rev. Lett.* **122**, 065301 (2019).
- [188] X.-Q. Sun, T. Nebabu, X. Han, M. O. Flynn, and X.-L. Qi, “Entanglement features of random neural network quantum states,” *Phys. Rev. B* **106**, 115138 (2022).
- [189] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, “Quantum entanglement,” *Rev. Mod. Phys.* **81**, 865–942 (2009).
- [190] J. Eisert, M. Cramer, and M. B. Plenio, “Colloquium: Area laws for the entanglement entropy,” *Rev. Mod. Phys.* **82**, 277–306 (2010).
- [191] D. N. Page, “Average entropy of a subsystem,” *Phys. Rev. Lett.* **71**, 1291–1294 (1993).
- [192] S. K. Foong and S. Kanno, “Proof of Page’s conjecture on the average entropy of a subsystem,” *Phys. Rev. Lett.* **72**, 1148–1151 (1994).
- [193] S. Sen, “Average Entropy of a Quantum Subsystem,” *Phys. Rev. Lett.* **77**, 1–3 (1996).
- [194] S. Sachdev and J. Ye, “Gapless spin-fluid ground state in a random quantum Heisenberg magnet,” *Phys. Rev. Lett.* **70**, 3339–3342 (1993).
- [195] J. Maldacena and D. Stanford, “Remarks on the Sachdev-Ye-Kitaev model,” *Phys. Rev. D* **94**, 106002 (2016).
- [196] V. Bettaque and B. Swingle, “NoRA: A Tensor Network Ansatz for Volume-Law Entangled Equilibrium States of Highly Connected Hamiltonians,” *arXiv:2303.16946* (2023).
- [197] G. Passetti, D. Hofmann, P. Neitemeier, L. Grunwald, M. A. Sentef, and D. M. Kennes, “Can Neural Quantum States Learn Volume-Law Ground States?” *Phys. Rev. Lett.* **131**, 036502 (2023).
- [198] H.-P. Breuer and F. Petruccione, “The Theory of Open Quantum Systems,” *Oxford University Press* (2007).
- [199] A. Rivas and S. F. Huelga, “Open Quantum Systems: An Introduction,” *Springer Berlin Heidelberg* (2012).
- [200] I. Rotter and J. P. Bird, “A review of progress in the physics of open quantum systems: theory and experiment,” *Reports on Progress in Physics* **78**, 114001 (2015).
- [201] D. Manzano, “A short introduction to the Lindblad master equation,” *AIP Advances* **10**, 025106 (2020).

- [202] D. Nigro, "On the uniqueness of the steady-state solution of the Lindblad–Gorini–Kossakowski–Sudarshan equation," *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 043202 (2019).
- [203] V. V. Albert, "Lindbladians with multiple steady states: theory and applications," *arXiv:1802.00010* (2018).
- [204] J. Dalibard, Y. Castin, and K. Mølmer, "Wave-function approach to dissipative processes in quantum optics," *Phys. Rev. Lett.* **68**, 580–583 (1992).
- [205] K. Mølmer, Y. Castin, and J. Dalibard, "Monte Carlo wave-function method in quantum optics," *Journal of the Optical Society of America B* **10**, 524 (1993).
- [206] C. Fuchs, M. Hoang, and B. Stacey, "The SIC Question: History and State of Play," *Axioms* **6**, 21 (2017).
- [207] C. A. Fuchs and R. Schack, "Quantum-Bayesian coherence," *Rev. Mod. Phys.* **85**, 1693–1715 (2013).
- [208] W. P. Schleich, *Quantum Optics in Phase Space* (Wiley, Feb. 2001).
- [209] A. Polkovnikov, "Phase space representation of quantum dynamics," *Annals of Physics* **325**, 1790–1852 (2010).
- [210] I. Białynicki-Birula and J. Mycielski, "Uncertainty relations for information entropy in wave mechanics," *Communications in Mathematical Physics* **44**, 129–132 (1975).
- [211] T. Haas, "An Entropic Perspective on Equilibrium, Uncertainty and Entanglement," *Heidelberg University Library* (2022).
- [212] S. N. Bose, "Plancks Gesetz und Lichtquantenhypothese," *Zeitschrift für Physik* **26**, 178–181 (1924).
- [213] A. Einstein, "Quantentheorie des einatomigen idealen Gases," *John Wiley & Sons, Ltd* (1924).
- [214] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, "Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor," *Science* **269**, 198–201 (1995).
- [215] K. B. Davis, M. - Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, "Bose-Einstein Condensation in a Gas of Sodium Atoms," *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
- [216] P. Kunkel, "Splitting a Bose-Einstein condensate enables EPR steering and simultaneous readout of noncommuting observables," *Heidelberg University Library* (2019).
- [217] P. Kunkel, M. Prüfer, S. Lannig, R. Rosa-Medina, A. Bonnin, M. Gärttner, H. Strobel, and M. K. Oberthaler, "Simultaneous Readout of Noncommuting Collective Spin Observables beyond the Standard Quantum Limit," *Phys. Rev. Lett.* **123**, 063603 (2019).

- [218] T. Mori, T. N. Ikeda, E. Kaminishi, and M. Ueda, “Thermalization and prethermalization in isolated quantum systems: a theoretical overview,” *Journal of Physics B: Atomic, Molecular and Optical Physics* **51**, 112001 (2018).
- [219] D. A. Abanin, E. Altman, I. Bloch, and M. Serbyn, “Colloquium: Many-body localization, thermalization, and entanglement,” *Rev. Mod. Phys.* **91**, 021001 (2019).
- [220] M. Ueda, “Quantum equilibration, thermalization and prethermalization in ultracold atoms,” *Nature Reviews Physics* **2**, 669–681 (2020).
- [221] J. M. Deutsch, “Eigenstate thermalization hypothesis,” *Reports on Progress in Physics* **81**, 082001 (2018).
- [222] P. W. Anderson, “Absence of Diffusion in Certain Random Lattices,” *Phys. Rev.* **109**, 1492–1505 (1958).
- [223] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, “Scaling Theory of Localization: Absence of Quantum Diffusion in Two Dimensions,” *Phys. Rev. Lett.* **42**, 673–676 (1979).
- [224] G. Roati, C. D’Errico, L. Fallani, M. Fattori, C. Fort, M. Zaccanti, G. Modugno, M. Modugno, and M. Inguscio, “Anderson localization of a non-interacting Bose–Einstein condensate,” *Nature* **453**, 895–898 (2008).
- [225] D. S. Wiersma, P. Bartolini, A. Lagendijk, and R. Righini, “Localization of light in a disordered medium,” *Nature* **390**, 671–673 (1997).
- [226] Y. Lahini, A. Avidan, F. Pozzi, M. Sorel, R. Morandotti, D. N. Christodoulides, and Y. Silberberg, “Anderson Localization and Nonlinearity in One-Dimensional Disordered Photonic Lattices,” *Phys. Rev. Lett.* **100**, 013906 (2008).
- [227] J. Billy, V. Josse, Z. Zuo, A. Bernard, B. Hambrecht, P. Lugan, D. Clément, L. Sanchez-Palencia, P. Bouyer, and A. Aspect, “Direct observation of Anderson localization of matter waves in a controlled disorder,” *Nature* **453**, 891–894 (2008).
- [228] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv:2010.11929* (2020).
- [229] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, “Big Transfer (BiT): General Visual Representation Learning,” *arXiv:1912.11370* (2020).

- [230] X. Chen, C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, Y. Lu, and Q. V. Le, "Symbolic Discovery of Optimization Algorithms," [arXiv:2302.06675](#) (2023).
- [231] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," [arXiv:1706.03762](#) (2017).
- [232] T. B. Brown et al., "Language Models are Few-Shot Learners," [arXiv:2005.14165](#) (2020).
- [233] H. Touvron et al., "Llama 2: Open Foundation and Fine-Tuned Chat Models," [arXiv:2307.09288](#) (2023).
- [234] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," *Nature* **529**, 484–489 (2016).
- [235] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," *Rev. Mod. Phys.* **91**, 045002 (2019).
- [236] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges," [arXiv:2104.13478](#) (2021).
- [237] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science* **313**, 504–507 (2006).
- [238] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," *Proceedings of the 25th international conference on Machine learning*, 536–543 (2008).
- [239] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," *Proceedings of the 24th international conference on Machine learning*, 791–798 (2007).
- [240] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," *Applied Optics* **29**, 4790 (1990).
- [241] C. Mouton, J. C. Myburgh, and M. H. Davel, "Stride and Translation Invariance in CNNs," *Springer International Publishing* (2020).
- [242] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," [arXiv:1311.2901](#) (2013).
- [243] J. L. Elman, "Finding Structure in Time," *Cognitive Science* **14**, 179–211 (1990).
- [244] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Computing and Informatics Journal* **3**, 334–340 (2018).

- [245] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current status and future directions," *International Journal of Forecasting* **37**, 388–427 (2021).
- [246] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543 (2014).
- [247] K. S. Tai, R. Socher, and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks," *arXiv:1503.00075* (2015).
- [248] S. Merity, N. S. Keskar, and R. Socher, "Regularizing and Optimizing LSTM Language Models," *arXiv:1708.02182* (2017).
- [249] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *MIT Press* (1986).
- [250] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation* **9**, 1735–1780 (1997).
- [251] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *arXiv:1406.1078* (2014).
- [252] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," *arXiv:1710.10903* (2017).
- [253] J. H. Ferziger and M. Perić, "Computational Methods for Fluid Dynamics," *Springer Berlin Heidelberg* (2002).
- [254] D. Freedman, "Brownian Motion and Diffusion," *Springer New York* (1983).
- [255] C. Reisinger, "Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben," *Heidelberg University Library* (2004).
- [256] M. S. Albergó, G. Kanwar, S. Racanière, D. J. Rezende, J. M. Urban, D. Boyda, K. Cranmer, D. C. Hackett, and P. E. Shanahan, "Flow-based sampling for fermionic lattice field theories," *Phys. Rev. D* **104**, 114507 (2021).
- [257] M. S. Albergó, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, "Flow-based sampling in the lattice Schwinger model at criticality," *Phys. Rev. D* **106**, 014514 (2022).

- [258] R. Abbott, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, B. Tian, and J. M. Urban, “Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions,” *Phys. Rev. D* **106**, 074506 (2022).
- [259] O. Dugan, P. Y. Lu, R. Dangovski, D. Luo, and M. Soljačić, “Q-Flow: Generative Modeling for Differential Equations of Open Quantum Dynamics with Normalizing Flows,” [arXiv:2302.12235](https://arxiv.org/abs/2302.12235) (2023).
- [260] G. Deco and W. Brauer, “Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures,” *Neural Networks* **8**, 525–535 (1995).
- [261] A. Hyvärinen and P. Pajunen, “Nonlinear independent component analysis: Existence and uniqueness results,” *Neural Networks* **12**, 429–439 (1999).
- [262] E. G. Tabak and E. Vanden-Eijnden, “Density estimation by dual ascent of the log-likelihood,” *Communications in Mathematical Sciences* **8**, 217–233 (2010).
- [263] E. G. Tabak and C. V. Turner, “A Family of Nonparametric Density Estimation Algorithms,” *Communications on Pure and Applied Mathematics* **66**, 145–164 (2012).
- [264] L. Dinh, D. Krueger, and Y. Bengio, “NICE: Non-linear Independent Components Estimation,” [arXiv:1410.8516](https://arxiv.org/abs/1410.8516) (2014).
- [265] L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using Real NVP,” [arXiv:1605.08803](https://arxiv.org/abs/1605.08803) (2016).
- [266] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe, “Analyzing Inverse Problems with Invertible Neural Networks,” [arXiv:1808.04730](https://arxiv.org/abs/1808.04730) (2018).
- [267] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, “Neural Ordinary Differential Equations,” [arXiv:1806.07366](https://arxiv.org/abs/1806.07366) (2018).
- [268] F. Verstraete and J. I. Cirac, “Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions,” [arXiv:cond-mat/0407066](https://arxiv.org/abs/cond-mat/0407066) (2004).
- [269] M. Levin and C. P. Nave, “Tensor Renormalization Group Approach to Two-Dimensional Classical Lattice Models,” *Phys. Rev. Lett.* **99**, 120601 (2007).
- [270] Z. Y. Xie, H. C. Jiang, Q. N. Chen, Z. Y. Weng, and T. Xiang, “Second Renormalization of Tensor-Network States,” *Phys. Rev. Lett.* **103**, 160601 (2009).

- [271] M. Hibat-Allah, E. M. Inack, R. Wiersema, R. G. Melko, and J. Carrasquilla, “Variational neural annealing,” *Nature Machine Intelligence* **3**, 952–961 (2021).
- [272] M. Bukov, M. Schmitt, and M. Dupont, “Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape,” *SciPost Phys.* **10**, 147 (2021).
- [273] M. Oliva, D. Kakofengitis, and O. Steuernagel, “Anharmonic quantum mechanical systems do not feature phase space trajectories,” *Phys. A: Stat. Mech. Appl.* **502**, 201–210 (2018).
- [274] P. Blakie, A. Bradley, M. Davis, R. Ballagh, and C. Gardiner, “Dynamics and statistical mechanics of ultra-cold Bose gases using c-field techniques,” *Adv. Phys.* **57**, 363–455 (2008).
- [275] J. Jumper et al., “Highly accurate protein structure prediction with AlphaFold,” *Nature* **596**, 583–589 (2021).
- [276] J. Nys and G. Carleo, “Variational solutions to fermion-to-qubit mappings in two spatial dimensions,” *Quantum* **6**, 833 (2022).
- [277] T. Haas, “Area Law from classical entropies,” In preparation (2024).
- [278] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, “Quantum source of entropy for black holes,” *Phys. Rev. D* **34**, 373–383 (1986).
- [279] M. Srednicki, “Entropy and area,” *Phys. Rev. Lett.* **71**, 666–669 (1993).
- [280] M. Cramer, J. Eisert, M. B. Plenio, and J. Dreißig, “Entanglement-area law for general bosonic harmonic lattice systems,” *Phys. Rev. A* **73**, 012309 (2006).
- [281] M. Cramer and J. Eisert, “Correlations, spectral gap and entanglement in harmonic quantum systems on generic lattices,” *New Journal of Physics* **8**, 71 (2006).
- [282] C.-W. Huang, L. Dinh, and A. Courville, “Augmented Normalizing Flows: Bridging the Gap Between Generative Flows and Latent Variable Models,” *arXiv:2002.07101* (2020).
- [283] D. J. Rezende, S. Racanière, I. Higgins, and P. Toth, “Equivariant Hamiltonian Flows,” *arXiv:1909.13739* (2019).
- [284] Y. D. Zhong, B. Dey, and A. Chakraborty, “Symplectic ODE-Net: Learning Hamiltonian Dynamics with Control,” *arXiv:1909.12077* (2019).

ACKNOWLEDGMENTS

First and foremost, I want to thank Prof. Dr. Martin Gärttner for giving me the opportunity to work at the fascinating intersection of quantum theory, numerical modeling, and machine learning. I've been truly fortunate to have had such an engaged and kind advisor who taught me so much on both topical matters and beyond, ever since attending the 'Quantum and Neural Networks' master seminar four years ago. I greatly enjoyed bouncing ideas back and forth over lunch, celebrating the small and big successes and simply working in such a positive and supportive environment, which crucially persisted even when we would hit a wall. I am delighted to see the group live on in a new place and wish you all the best for this new chapter.

I want to express my gratitude towards Prof. Dr. Björn Malte Schäfer, who agreed to act as a referee for this thesis even though the topics of my PhD are quite removed from his personal research interests.

Importantly, this thesis would not have been possible without the support of Dr. Markus Schmitt, whose expertise in GPU computing and design of the jVMC library enabled quick iterations and thus many of the contained results.

I am grateful to all former and current group members for making coming to work so enjoyable, including both the Heidelberg generation of Tobias Schmale, Oliver Stockdale, Adrian Aasen, Niklas Euler, Adrian Brämer, and Javad Vahedi as well as the new members Elena Callus, Martina Jung and Mauro D'Achille. I am particularly thankful for being given the opportunity to supervise the exceptionally talented bachelor students Moritz Epping, Tim Rein, and Maximilian Rieger.

Furthermore, I want to thank all the great people that are part of the SynQS collaboration. I am grateful to Prof. Dr. Markus Oberthaler and Dr. Helmut Strobel for the shared insights regarding cold atoms and Prof. Dr. Thomas Gasenzer for an open door concerning questions about the truncated Wigner approximation. I am particularly glad to have collaborated closely with the BEC experiment of Yannick Deller, Alexander Schmutz and David Feiz. I greatly enjoyed the company of Nikolas Liebster, Marius Sparr, Elinor Kath, Ido Siovitz, Niklas Rasch, Philipp Heinen, Lasse Kundy, Verena Oehmke, and all other members during Friday breakfasts and certain Wednesday nights.

I have learned a lot from Tobi Haas, whose persistence and attention to detail is remarkable. I will cherish the memories of a few extremely productive and fun weeks marked by hour-long zoom calls, in which I came to appreciate the value of *slowly* increasing complexity.

I also want to express my gratitude towards Stefanie Czischek and Juan Carrasquilla, who agreed to host me in Canada for a three-month research stay, which gave me a unique opportunity to see how research is conducted elsewhere. I want to also thank the friends I made along the way there, namely Faith Oyedemi, Andrew Jreissaty, Schuyler Moss, and Roeland Wiersema. Additionally, I would like to thank the Baden-Württemberg Stiftung for financing the stay.

I am happy to have met Rouven Koch during my PhD journey at a conference in Warsaw and I am glad that we stayed in touch.

I am immeasurably fortunate to have some truly amazing friends who accompanied me on this journey. This goes out to Lea Rau, Şamil Güler, Chi-Ho Wai, Yannick Grimm, Benjamin Glogauer, and Bianca Benner. Also, I have had the best time living together with Annika Jungk, Tim Ahrens, Hannah von Ammon, Lorenz Paetzke, Maike Gnutzmann, Lorenzo Biasi, and Firat & Mira Terzi.

I want to thank my family for their continuous support ever since I decided to study physics more than eight years ago. It has been a great time and I am glad that you were all part of it. I am deeply saddened that my dad is no longer around to witness this final chapter of my academic journey.

Finally, I want to express my deepest gratitude to Antonia, my partner and best friend. Coming home to you never ceases to put a smile on my face. I am *so* looking forward to all the time ahead of us :)