PHASE CLASSIFICATION BY TENSORIAL KERNEL SUPPORT VECTOR MACHINE

NICOLAS SADOUNE

München, 2024

PHASE CLASSIFICATION BY TENSORIAL KERNEL SUPPORT VECTOR MACHINE

Dissertation

an der Fakultät für Physik
der Ludwig-Maximilians-Universität München

vorgelegt von

NICOLAS SADOUNE

aus München

München, den 16. Juli 2024

Dedicated to my love, Aurela

ABSTRACT

The determination of the ground state phase diagram of strongly interacting many-body system is an inherently hard problem. With exception of few models, analytical as well as numerical solutions are very challenging. Instead of searching for a brute-force solution, the focus may be shifted to the investigation of key physical properties. Avoiding a full description of the many-body system, the machine learning model of tensorial kernel support vector machine (TKSVM) allows to detect phase transitions and to describe each phase in terms of local observables. In this thesis, TKSVM, which was initially developed for studying classical systems, is generalized to the analysis of quantum systems. The generalization is based on informationally complete measurements, shadow tomography, and modifications of the tensorial kernel. Crucially, the *unsupervised* and *interpretable* traits of the learning model are preserved.

The abilities of TKSVM are demonstrated in three applications. In the first application, training data is obtained from trapped-ion quantum computers without error correction. Two families of quantum states, which both admit state preparation through shallow circuits, are investigated. In both cases, the phase transition from a symmetry-protected topological phase to a trivial phase is successfully learned and the respective string and trivial order parameters are determined. This application emphasizes the utility of TKSVM in the context of the current noisy intermediate-scale quantum (NISQ) era. The second and third applications rely on synthetic quantum data, which allows to benchmark the learning model in the absence of experimental limitations such as system size and quantum noise effects. For a one-dimensional cluster model with two free parameters, the phase diagram is successfully constructed and the respective order parameters are detected. Moreover, the toric code model subject to external fields is investigated and the explicit stabilizer operators of the topological phase are identified.

Subsequent to the three quantum applications, a classical spin model defined on the pyrochlore lattice is finally investigated. In the low temperature phase of the model, thermalization issues inhibit Monte Carlo simulations and lead to training data of poor quality. Nonetheless, by combined iterations of machine learning and human interpretation, a subsystem symmetry as well as an intricate hybrid dipolar-quadrupolar order are discovered.

ZUSAMMENFASSUNG

Die Bestimmung des Grundzustandsphasendiagramms eines stark wechselwirkenden Vielteilchensystems ist ein von Natur aus schwieriges Problem. Mit Ausnahme einiger Modelle sind analytische sowie numerische Lösungen sehr anspruchsvoll. Anstatt nach einer expliziten Lösung zu suchen, ist es daher oft sinnvoll die charakteristischen physikalischen Eigenschaften des Systems zu untersuchen. Das maschinelle Lernmodell TKSVM vermeidet eine vollständige Beschreibung des Vielteilchensystems und ermöglicht es Phasenübergänge zu erkennen sowie jede Phase anhand lokaler Observablen zu beschreiben. In dieser Arbeit wird TKSVM, welches ursprünglich zur Untersuchung klassischer Systeme entwickelt wurde, zur Analyse von Quantensystemen erweitert. Die Verallgemeinerung basiert auf informationell vollständigen Messungen, Schattentomographie, und Modifikationen des Tensor-Kernels. Entscheidend ist, dass die unbeaufsichtigten und interpretierbaren Merkmale des Lernmodells erhalten bleiben.

Das Lernmodell wird in mehreren Anwendungen demonstriert. In der ersten Anwendung werden die Trainingsdaten von Quantencomputern basierend auf gefangenen Ionen ohne Fehlerkorrektur generiert. Es werden zwei Familien von Quantenzuständen untersucht, welche beide eine Zustandspräparation mit kurzen Schaltkreisen erlauben. In beiden Fällen wird der Phasenübergang von einer symmetriegeschützten topologischen Phase zu einer trivialen Phase erfolgreich erlernt und die jeweiligen String- und trivialen Ordnungsparameter bestimmt. Diese Anwendung hebt den Nutzen von TKSVM in der aktuellen NISQ Ära hervor. Die zweite sowie die dritte Anwendung basiert auf synthetischen Quantendaten, die es ermöglichen, das Lernmodell ohne experimentelle Einschränkungen wie Systemgröße und Quantenrauscheffekte zu benchmarken. Das Phasendiagramm eines eindimensionalen Clustermodells mit zwei freien Parametern wird erfolgreich konstruiert und die jeweiligen Ordnungsparameter ermittelt. Darüber hinaus wird das *toric code* Modell unter Einwirkung externer Felder untersucht und die expliziten Stabilisatoroperatoren der topologischen Phase identifiziert.

Abschließend wird ein klassisches Spinmodell auf dem Pyrochlorgitter untersucht. In der Niedertemperaturphase des Modells führen Thermalisierungsprobleme während der Monte-Carlo-Simulation zu Trainingsdaten von minderer Qualität. Durch kombinierte Iterationen von maschinellem Lernen und Interpretation durch den Nutzer werden dennoch eine Subsystemsymmetrie sowie eine komplizierte hybride Dipol-Quadrupol-Ordnung festgestellt.

PUBLICATIONS

The contents of publication [1], preprint [2], and of a manuscript that is currently in preparation [3] are presented in this thesis, and therefore contain partial textual overlap. These three manuscripts are the result of collaborations with various authors, all of which contributed substantially to the scientific work and writing. In partial fulfilment of the requirements for the degree of PhD, another scientific work [4] not presented in this thesis has been published.

[1] Nicolas Sadoune, Giuliano Giudici, Ke Liu, and Lode Pollet. "Unsupervised interpretable learning of phases from many-qubit systems." In: *Phys. Rev. Res.* 5 (1 2023), p. 013082. DOI: 10.1103/PhysRevResearch.5.013082. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.5.013082.

[2] Nicolas Sadoune, Ke Liu, Han Yan, Ludovic D. C. Jaubert, Nic Shannon, and Lode Pollet. *Human-machine collaboration: ordering mechanism of rank−2 spin liquid on breathing pyrochlore lattice.* 2024. arXiv: 2402.10658 [cond-mat.str-el]. URL: https://arxiv.org/abs/2402.10658.

[3] Nicolas Sadoune, Ivan Pogorelov, Claire Edmunds, Giuliano Giudici, Giacomo Giudice, Christian D. Marciniak, Martin Ringbauer, Thomas Monz, and Lode Pollet. "Learning symmetry-protected topological order from trapped-ion experiments." In: *To be submitted* (2024).

[4] Nicolas Sadoune and Lode Pollet. "Efficient and scalable path integral Monte Carlo simulations with worm-type updates for Bose-Hubbard and XXZ models." In: *SciPost Phys. Codebases* (2022), p. 9. DOI: 10.21468/SciPostPhysCodeb.9. URL: https://scipost.org/10.21468/SciPostPhysCodeb.9.

ACKNOWLEDGMENTS

CONTENTS

LIST OF TABLES

ACRONYMS

AKLT   Affleck-Kennedy-Lieb-Tasaki

MPS   matrix product state

MUB   mutually unbiased bases

NISQ   noisy intermediate-scale quantum

POVM   positive operator-valued measure

SIC-POVM   symmetric informationally complete POVM

SPT   symmetry-protected topological

SVM   support vector machine

TKSVM   tensorial kernel support vector machine

INTRODUCTION

At present quantum computing is without a doubt one of the most active research areas of physics and computer science. Its inception dates back to 1980, when the idea of a quantum computer was first formulated by several key scientist. One of them was Manin, who mentioned the core idea of quantum computing in his book *Computable and Non-Computable* [140]. Therein he broadly described a computing device exploiting superposition and entanglement of quantum states. Another well known pioneer of quantum computing was Benioff. Viewing computers as a physical systems, he derived the first quantum mechanical description of a computer and the computation process. The specific goal was to find the Hamiltonian, which guides the time evolution of a Turing machine, i. e. the universal representative of all classical computation, in some initial state [7]. The most famous originator of quantum computing is however Feynman, whose idea it was to map physical laws onto a computer [36]. In the classical limit, this idea implies the simulation of local differential equations, but not all processes in nature have a useful classical approximation. Therefore, the computer is supposed to mimic quantum mechanical processes exactly as they would happen in nature or a quantum mechanical model of nature. An important restriction is that the number of computation resources required to simulate the quantum mechanical system do not scale exponentially with the size of the physical system. Feynman showed that the Turing machine is not up to the task.

Thus for simulating quantum physics, a new kind of computation model was required, the quantum Turing machine. Analogous to its classical counterpart, this machine was sought to be the universal representative of all quantum computation. The Church-Turing-hypothesis states that any function is computable in a deterministic way in finite number of steps if and only if it can be produced by a Turing machine. In a publication in 1985 [31], Deutsch argued that this hypothesis implies a physical principle. Namely, that the Turing machine suffices to simulate any finitely realizible physical system. He further reasoned that because classical physics is continuous and the Turing machine discrete, the principle is violated. To restore the validity of this principle, Deutsch formalized the quantum generalization of the Turing machine, and showed that the theory of quantum mechanics and the quantum Turing machine are again compatible with the physical principle implied by the Church-Turing-hypothesis. Specifically, the generalization of the Turing machine consisted in

extending the state-space by allowing superpositions of classical states. During a computation step, the quantum Turing machine transitions from one superposition to another in a unitary way [81].

The quantum Turing machine, even though conceptually achieving the goal of simulating quantum physics, is not the default computation model used today, but instead it is the computation model called *quantum circuits*. Today's quantum circuits originate from yet another contribution of Deutsch in 1989. Namely, from the quantum computational networks [30], that he designed to generalize the theory of boolean circuits used in classical computing. Introducing quantum wires and quantum gates as the analogue of the classical logical gate resulted in the first version of the quantum circuit. The proof that quantum circuits are equivalent to the quantum Turing machine in terms of computational power and complexity followed by Yao a few years later [25]. Because of practical limitations regarding the efficient implementation of quantum algorithms on a quantum Turing machine, the computation model of quantum circuits soon replaced the quantum Turing machine entirely and became the standard [81].

Besides the simulation of quantum physical processes, this new platform of computation called for the development and exploration of an entire field of research known as *quantum algorithms*. Naturally, the existing theory of computational complexity expanded by a whole array of quantum complexity classes [133]. The objective of quantum algorithms research is solving specific tasks with significant advantage over the best achievable solution on a classical computer. Quantum advantage may be with respect to memory requirements or runtime, and in the in the best case the task at hand would even have practical relevance for modern day technology. In 1994, Shor indeed developed a prime factorization algorithm [120] that not only exhibits exponential speedup compared to the best known classical algorithm, but also lends itself to practical application. Unsurprisingly, the combinations of these two properties sparked a lot of interest in quantum computation. As a practical example, some public-key cryptography methods are based on the hardness of factorization. With Shor's algorithm factorization becomes efficient, rendering these cryptography methods useless. However, reliable quantum computers at the scale required to really break these kind of cryptography methods are not yet available. In anticipation of their availability, quantum-secure cryptography methods became an active field of research. The theoretical success of Shor's algorithm spawned hope that similarly efficient algorithms could be developed for other and more general tasks than factorization. But as it turned out, quantum algorithms solving practical tasks with an exponential speedup are very rare. One famous early quantum algorithm exemplifying this is Grovers algorithm [48], solving a task with a much wider range of applications. The task is the identification of one specific target state (e. g. represented by a bit-string) out of

a set of N possible states, by querying an oracle the least possible amount of times. It is equivalent to the task of searching for an entry in a completely unstructured database. Due to the lack of structure, there is classically no better way than a linear search with a runtime $\mathcal{O}(N)$ (big O notation). Employing a technique now known as *amplitude amplification*, Grover's algorithm achieves a runtime of $\mathcal{O}(\sqrt{N})$. It thus hints that exponential advantage might only be possible for very specific tasks. Although the advantage is not exponential but only quadratic, it is nonetheless impressive and motivates further research into similar general purpose tasks.

In regards to the task of physical simulation, there is of course one very obvious exponential advantage. This exponential advantage is in computation memory, and is in principle what defines a quantum computer. While N classical bits have a discrete state space of size 2^N, N qubits have a continuous state space consisting of linear combinations of the classical states with 2^N independent complex coefficients (minus 1 if the quantum states are normalized). Representing a general quantum state thus requires exponentially more classical than quantum bits. Efficiently preparing a general quantum state on a quantum computer is however a whole topic of its own, which will not be covered in this thesis. Instead, assuming an efficient way for state preparation is known, the considered task will be to gain as much information as possible about the state. The assumption does not imply any kind of knowledge about the prepared state itself. Although the quantum computer has the capacity to store large amounts of information, the amount of information that can be accessed efficiently is limited. This limitation is commonly referred to as *accessibility problem*. Of course all information can eventually be extracted, but it might require a number of identical copies of the state scaling exponentially with the number of qubits storing the state. Common methods for extracting all information, known as *full quantum state tomography* typically require $\mathcal{O}(2^{2N})$ copies of a N-qubit quantum state [49, 90]. The accessibility problem is also manifest in the *quantum Fourier transform*, which is the key ingredient of Shor's algorithm, and the source of its exponential advantage. Computing the discrete Fourier transform of 2^N complex amplitudes requires $\mathcal{O}(N2^N)$ computation steps classically but only $\mathcal{O}(N^2)$ steps on a quantum computer. However, in case of the quantum Fourier transform, the transformed amplitudes are not easy to read out after they have been computed. This is not an issue in Shor's algorithm, because the transformed amplitudes are merely an intermediate result used for further computation and are never explicitly read out. The final output, i.e. two prime factors, contains no information about Fourier transformed amplitudes at all.

Fortunately, it is usually sufficient and satisfactory to understand the characteristic physical properties of an unknown state, thereby circumventing the accessibility problem of a full description of the

state. This endeavour can be extended to a parametrized set of unknown states, rather than a single state. Analyzing which subsets of states share the same physical properties is then synonym to learning the phase diagram with respect to the set's parameters. The major task considered in this thesis is the determination of the ground state phase diagram of a parametrized quantum many-body Hamiltonian and the characterization of each phase. This is achieved with a hybrid quantum-classical scheme, rather than pure quantum computation. In this hybrid approach, the quantum computer is used to realize a quantum ground state for fixed Hamiltonian parameters, and to produce data by sampling the prepared state. Crucial to this approach is that the number of required identical copies of the state is independent of system size. Repeating the state preparation and sampling procedure for different values of Hamiltonian parameters produces one dataset per parameter choice, results in large amounts of data. And this is where the classical part of the hybrid approach comes into play. Processing and extracting useful information from large amounts of data is a task perfectly suitable for *machine learning* models.

Machine learning is nowadays a valuable tool for data analysis in most areas of technology and research, all the while itself being subject of active research. Specifically in the context of quantum many-body physics, machine learning models are frequently used for the representation of a wavefunction or density matrix, improving existing computational methods such as Monte Carlo, quantum state tomography, phase classification and inferring phase transitions like in this thesis, and more [17, 78]. When speaking of machine learning, the *neural network* is the first type of model that comes to mind. Inspired by the mammalian brain, neural networks consist of connected *neurons* as basic computation units. Similar to its biological counterpart, neural networks learn by establishing connections between the neurons. Connections between neurons are modelled by numbers and constitute (among others) the internal parameters of the neural network. During the learning stage, these internal parameters are tuned to achieve a desired result. Overall, a neural network can be seen as a directed weighted graph whose *architecture* plays a crucial role for what kind of task it is suited for. Arguably the most popular architectures within and outside of physics applications include the *restricted Boltzmann machine* [79], the *convolutional neural network* [71, 73], the *generative adversarial network* [4] and the *variational auto-encoder* [61]. What most neural network applications have in common is their extensive amount of internal parameters. Regarding physics applications, this is in principle unproblematic as long as their number does not scale exponentially with the physical system size. However, there can be an issue concerning the learning model's *interpretability*. Admittedly, the generative and predictive power of neural networks are more than impressive, but it is of little worth if no new theoretical insight

is gained. While for industrial applications understanding the decision making of the learning model is often of no relevance at all, e. g. self-driving cars, the situation is entirely different for physics applications, where the purpose of the learning model is ultimately better understanding of the underlying problem or data. This circumstance was the motivation of Ponte and Melko, when they first employed a support vector machine (SVM) as machine learning tool to analyze condensed matter models in 2017 [97]. They demonstrated the strong interpretability of SVM by considering several prototypical Ising-type models and exhibiting that the internal parameters directly reflect the order parameter of the respective symmetry broken phases.

Building on the work of Ponte and Melko, Greitemann *et al.* extended the range of applicability of the SVM to general classical O(3)-breaking multipolar orders [44, 45]. This was achieved by introducing a *tensorial kernel* with the ability to capture even the most intricate multipolar orders [87]. Shortly after, Liu *et al.* discovered a special significance of one of the intrinsic parameters of the SVM [75]. Namely the *bias parameter*, acting as an indicator for the existence of phase transitions. The SVM is technically a *supervised* machine learning model, meaning that the training data must be labelled according to a priori knowledge on the physics of the underlying model or states. Exploiting the criterion imposed by the bias parameter, however, supervision of the learning task can be delegated entirely to the machine, dismissing the need for any physical information about the input. TKSVM thus represents an unsupervised and interpretable learning model, capable of classifying and characterizing physical phases of classical many-body systems. The main contribution of this thesis lies in the generalization and adaptation of the TKSVM learning model to quantum many-body problems, see Figure 1.1 for an overview of the workflow.

Classical data is comparably simple to produce. Any classical O(3)-spin model may in principle be simulated by means of Monte Carlo simulation. Surely even for classical models difficulties such as thermalization issues due to a complex energy landscape may arise. But quantum data is at present undoubtedly much more difficult to obtain given that only few quantum many-body models can even be simulated in a somewhat reliable way. It could be argued that this is mainly due to the maturity of classical computers contrasted with quantum computers that are still a theoretical and especially practical work in progress.

On the theoretical side, efficient ground state preparation of an arbitrary Hamiltonian is still an open question [72]. Existing theoretical techniques to prepare ground states on a quantum device usually rely on a specific structure of the Hamiltonian, specialized circuit layouts, the use of ancilla qubits or heuristics and often only approximate the true ground state [33, 41, 74, 95]. The experimental realization
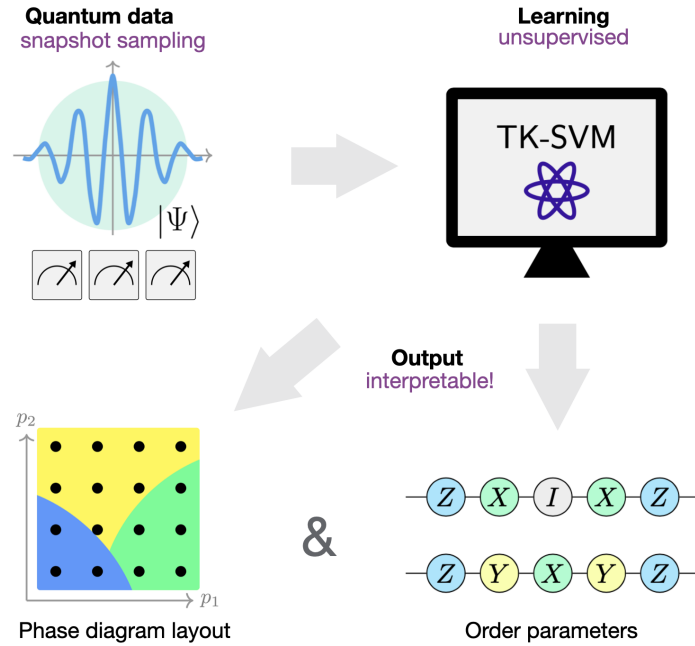
Figure 1.1: Workflow of TKSVM. Input data is produced by sampling quantum states. After training, the phase diagram topology as well as local order parameters are extracted from the internal parameters of the learning model.

of quantum computing seems even more challenging, due to the necessity to shield the quantum system from external disturbances to avoid decoherence, while at the same time manipulating it in order to carry out calculations. Since the inception of quantum computing many different, physical implementations have been proposed and implemented. The technologies fundamentally differ in what kind of physical two-level system is used to represent a qubit. Naturally, the choice of the qubit realization governs the way in which intialization, gate operations and readout are effected. There are several key quantities such as gate fidelity, coherence time, gate execution times, qubit control and connectivity, and scalability, that are indicative of the quality of a quantum computer. However, their sheer numbers can be misleading and therefore different methods such as volumetric benchmarking [8] have been proposed to compare different platforms. Another possible metric used in the past is how well a quantum computer outperforms a classical one for a given generic task. This is again sub-optimal since classical computer are also constantly scaled up in capacity and are thus not a fixed point of reference. Besides, such a metric might also strongly depend on the chosen task. Given the current rate of experimental progress and innovation, efforts to establish a ranking among the most promising quantum computing platforms would be in vain anyway. At present, the platforms with the most impressive achievements or with the most promising potential

include trapped ions [55, 98], super-conducting qubits [2, 60], silicon qubits [83], and neutral atoms [9, 21].

The scale of computations that are reliably performed on any of these platforms is limited by *quantum noise*, originating from imperfect initialization or quantum gates, faulty measurements or general decoherence due to interaction with the environment. To overcome the current NISQ era [100] continuous effort is being made to reduce noise, e. g. by improving hardware, but fundamentally some degree of imperfections will always remain. Therefore, in addition to noise reduction, methods to detect and correct computation errors caused by the noise during runtime are being developed and demonstrated [13, 98, 121]. These methods, called *error correction codes*, prescribe the way in which several physical qubits represent a single protected logical qubit. Moreover, they describe how gate operations are effected on a multicomponent logical qubits and how errors are detected and corrected in the redundant qubit representation. Each code has an intrinsic error threshold that must not be crossed in order for the code to be effective, thus even with error correction the noise must be minimized by improving gate fidelities etc. Since protected qubits are comprised of several physical ones, scaling up their number becomes even more challenging, and the time it will take to achieve fully fault-tolerant quantum computation is hard to estimate. Nonetheless, past and current progress admit some optimism.

The premise of this thesis is that reliable ground state preparation will become possible for many interesting if not arbitrary quantum many-body Hamiltonians in the future. Because of the current limitation due to quantum noise, the focus lies on the methodology and benchmarking of TKSVM by application to well understood many-body models, rather than on the investigation of models whose characteristic physics are unknown. The capabilities of TKSVM to gain physical insight are demonstrated, and the problem of investigating poorly understood models is accordingly inhibited only by state preparation.

The thesis is structured as follows: Chapter 2 is dedicated to the properties of the special type of measurements that are required for quantum applications of TKSVM. Several sampling techniques to generate training data are covered. Next, the notion of *shadow tomography* and *classical shadows* is reviewed in Section 3.1. The concept is employed to compute estimators of local observables from snapshot data, which is relevant for the definition of the tensorial kernel. In Chapter 4, the basics of the SVM are reviewed and the tensorial kernel is introduced. Both the classical and the quantum variant of the TKSVM are discussed and their key differences highlighted.

Finally, in Part ii, four different applications of TKSVM are showcased. The first and most important application is centered around two parametrized families of states, both admitting state preparation with shallow quantum circuits. Quantum data is obtained from state-of-the-

art trapped-ion devices without error correction in collaboration with the University of Innsbruck. For both families, the phase transition from a symmetry-protected topological phase to a trivial phase is successfully learned and the respective string and trivial order parameters are determined. It is worth mentioning that some of the datasets were produced on the world's first *compact* (two 19-inch server racks) quantum computer [96] meeting DiVincenzo's criteria for quantum computation [32].

In the second application, a one-dimensional cluster model is considered. Relying on synthetic data, the phase diagram of the model is successfully constructed and the order parameters of its phases, especially the string order parameters, are detected. The model is closely related to one of the families of states considered in the first application, but goes beyond experimental limitations in terms of system size.

As third application, the toric code model subject to external magnetic fields is investigated. Input data is obtained from *exact diagonalization* of the Hamiltonian. Since the topological phase of the model has *global* character, it can't be fully characterized by TKSVM, which is based on local quantities. Nevertheless, the explicit stabilizer operators of the topological phase are obtained, which may be seen as local traits of global order.

In contrast to the first three applications, the fourth and last application treats a classical model, meaning that spin degrees of freedom are approximated as three-component real vectors. The model is defined by Heisenberg interactions on the pyrochlore lattice, i. e. a three-dimensional network of corner-sharing tetrahedra, with additional Dzyaloshinskii–Moriya interaction on a subset of tetrahedra. Rather than determining the already known phase diagram of the model, the focus lies on characterizing one of its phases. By employing TKSVM, a subsystem symmetry as well as an intricate hybrid dipolar-quadrupolar order are discovered.

Part I

THEORY

QUANTUM DATA ACQUISITION

The input for TKSVM is always a set of *snapshots*, i.e. local measurements repeatedly taken at every site of the quantum many-body system under investigation. While many machine learning algorithms rely on projective measurements (also called von Neumann measurements) to produce snapshots, this is not sufficient for the purpose of TKSVM. When taking projective measurements, e.g. in the computational basis, some amount of information about the original quantum state is unavoidably lost. This loss of information can be remedied by resorting to generalized measurements, historically called positive operator-valued measure (POVM), with special properties. In the subsequent section, a summary about quantum measurements and their required properties will be given, illustrated by the specific examples used in Part ii. The remainder of this chapter treats several methods for efficiently sampling these quantum measurements from the many-body system to produce input data.

2.1 GENERALIZED MEASUREMENTS

Any quantum measurement is mathematically defined by a collection of measurement operators $M = \{M_l\}$, fulfilling the identity and non-negativity conditions [86]

$$\sum_l M_l = I \text{ and } \langle\psi|M_l|\psi\rangle \geqslant 0 \; \forall \, |\psi\rangle. \tag{2.1}$$

Both these conditions are necessary to interpret the expectation values of the measurement operators as probabilities. Assume a quantum system is in a state $|\psi\rangle$ prior to measuring, then the probability that the outcome M_l occurs is given by $\langle\psi|M_l|\psi\rangle$. Projective measurements only qualify as such, if furthermore the measurement operators are mutually orthogonal projectors $M_k M_l = \delta_{kl} M_k$. By allowing the measurement operators to be non-orthogonal, one obtains the wider class of generalized measurements. Thus projective measurements are merely a special case of generalized measurements. In contrast to projective measurement, generalized measurements can be *informationally complete*. This is the case when the measurement operators suffice to express any density matrix $\rho \in \mathbb{C}^{d \times d}$ as a linear combination $\rho = \sum_l c_l M_l$ with $c_l \in \mathbb{C}$. This can only be achieved if the measurement has a minimum size of $|M| \geqslant d^2$. A projective measurement can never be informationally complete, given that a set of more than d projectors can never be mutually orthogonal.

Subsequently, two important types of generalized measurements, mutually unbiased bases (MUB) and symmetric informationally complete POVM (SIC-POVM), will be discussed. Moreover, the Haar measure is used to construct a generalized measurement and a crucial relation with the other two kinds of POVM is established.

MUTUALLY UNBIASED BASES   The idea behind MUB is simple: since measurements in the computational basis are not informationally complete, add more orthogonal bases to the measurement until it becomes informationally complete. The term *mutually unbiased* refers to the relation between the different orthogonal bases that are part of the measurement. Two orthogonal bases $\{\,|v_k\rangle\,\}$ and $\{\,|w_l\rangle\,\}$ are mutually unbiased, if the inner product of any pair of basis states always has the same magnitude $|\langle v_k|w_l\rangle| = 1/\sqrt{d} \;\forall\; k, l$. A natural question is about the existence and the maximal number of MUB in a given dimension d. The maximal number of MUB is generally at most $d + 1$. But when d is the power of a prime, exactly $d + 1$ MUB do exist, and explicit constructions thereof are known [64, 135].

A set of MUB states can directly be converted into a generalized measurement by defining projectors $|v_l\rangle\langle v_l|$ onto each of the basis states $|v_l\rangle$. The projectors have to be sub-normalized as $M_l := |v_l\rangle\langle v_l|/(d + 1)$ in order to satisfy the identity condition in Equation 2.1. Informational completeness can only be achieved if there are at least $(d + 1)$ MUB, meaning the set is maximal in case of prime power dimensions. Since in this work all POVM are necessarily informationally complete, the term MUB should always be understood as a *maximal* set of MUB from here on.

Certainly the best known example of MUB is the set of eigenstates of the Pauli matrices

$$\left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}, \frac{1}{\sqrt{2}}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\}, \frac{1}{\sqrt{2}}\left\{\begin{pmatrix}1\\i\end{pmatrix}, \begin{pmatrix}1\\-i\end{pmatrix}\right\}. \quad (2.2)$$

In higher dimensions MUB are less commonly known. Applying the procedure from [135], a maximal set of MUB for $d = 3$ can be computed as

$$\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}, \frac{1}{\sqrt{3}}\left\{\begin{pmatrix}q_-\\1\\1\end{pmatrix}, \begin{pmatrix}1\\q_-\\1\end{pmatrix}, \begin{pmatrix}q_+\\q_+\\1\end{pmatrix}\right\},$$
$$\frac{1}{\sqrt{3}}\left\{\begin{pmatrix}q_+\\1\\1\end{pmatrix}, \begin{pmatrix}1\\q_+\\1\end{pmatrix}, \begin{pmatrix}q_-\\q_-\\1\end{pmatrix}\right\}, \frac{1}{\sqrt{3}}\left\{\begin{pmatrix}q_+\\q_-\\1\end{pmatrix}, \begin{pmatrix}q_-\\q_+\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right\} \quad (2.3)$$

with $q_\pm = e^{\pm 2\pi i/3}$.

Generalized measurements constructed from MUB have an important advantage over other types of POVM. Namely, they are a collection

of multiple projective measurements, where no projective measurement is favored over another. This can be exploited by randomly choosing a basis in which to measure, and then performing a projective measurement in that basis, as shown later in Section 2.3. Compared to implementing the POVM directly, this approach requires less resources.

SIC-POVM   The second important class of generalized measurements used in this work are SIC-POVM, which are derived from a set of d^2 states, such that the magnitude of the inner product is perfectly uniform through all possible pairs of states. More explicitly, the magnitude of the inner-product equals $|\langle v_k|v_l\rangle| = 1/\sqrt{d+1}$ for $k \neq l$. In contrast to MUB, none of states are orthogonal. SIC-POVM are known in dimensions up to $d = 151$ either analytically or numerically [40], but it remains an open question if they exist in all dimensions. Furthermore SIC-POVM are minimal, in the sense that they saturate the size requirement $|M| \geqslant d^2$ necessary for informational completeness. Depending on the use case, this can yield computational advantages compared to larger sized POVM. In $d = 2$ dimensions only two SIC-POVM exist, and

(a) UNBIASED BASES   (b) TETRAHEDRAL
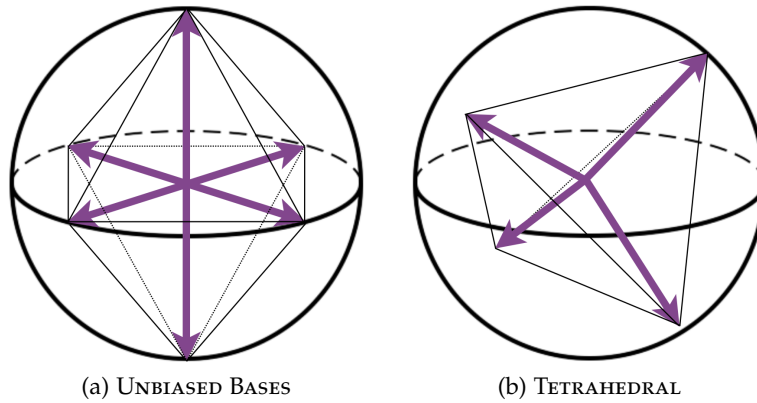
Figure 2.1: Two common POVM for $d = 2$ visualized in the Bloch sphere. (a) MUB vectors spanning a regular octahedron. (b) State vectors defining a SIC-POVM. They span a regular tetrahedron.

they are related by inversion symmetry. Following the construction from [108], an explicit example of a SIC-POVM is defined by the states

$$\left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}, \begin{pmatrix} \beta \\ i\alpha \end{pmatrix}, \begin{pmatrix} \beta \\ -i\alpha \end{pmatrix} \right\} \tag{2.4}$$

with $\alpha, \beta = \sqrt{(3 \pm \sqrt{3})/6}$. A global rotation does not result in a qualitatively different measurement since it merely represents a change of the coordinate system. Meanwhile in $d = 3$ dimensions, there exist

infinitely many qualitatively distinct SIC-POVM. One of them expressed in terms of its defining state vectors is given by

$$
\frac{1}{\sqrt{2}}\left\{ \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -\omega \\ 0 \end{pmatrix}, \begin{pmatrix} \omega^2 \\ 0 \\ -\omega \end{pmatrix}, \right.
$$
$$
\left. \begin{pmatrix} 0 \\ \omega^2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -\omega^2 \\ 0 \end{pmatrix}, \begin{pmatrix} -\omega \\ 0 \\ \omega^2 \end{pmatrix}, \begin{pmatrix} 0 \\ \omega \\ -1 \end{pmatrix} \right\} \quad (2.5)
$$

where $\omega = e^{i\pi/3}$. Explicit examples are given for dimensions $d = 2$ and $d = 3$ because these are the local Hilbert space dimensions of the models that will be discussed in Part ii. For dimension $d = 2$, quantum states admit an intuitive visualization and are therefore displayed in Figure 2.1 to emphasize their symmetry.

HAAR MEASURE   Sets of random projectors or equivalently sets of random pure states can also be used to construct informationally complete POVM. Usually the corresponding measurement operators are not specified explicitly, but instead a probability density function over them is considered. Measurement operators are then generated by sampling according to the probability density [80]. In many applications, a uniform probability density over pure states is desired, and this density is known as the Haar measure.

Consider the set of pure states in d dimensions, i. e. the sphere S^{d-1} of normalized vectors in the complex vector space $\mathbb{C}^d$. Global phases are not observable, hence two states $|v\rangle$ and $|w\rangle$ are equivalent $|v\rangle \equiv |w\rangle$, if they merely differ by a phase $|v\rangle = e^{i\theta}|w\rangle$. Taking this into account, define the quotient space with respect to the equivalence relation as $\mathbb{C}S^{d-1} := S^{d-1}/\equiv$. An informationally complete infinite size POVM is then immediately defined by the set of projectors onto the states $|v\rangle \in \mathbb{C}S^{d-1}$ distributed according to the Haar measure μ on this space. In order to satisfy the continuous equivalent of the identity condition

$$
\int_{\mathbb{C}S^{d-1}} |v\rangle \langle v| \, d\mu(v) = I \quad (2.6)
$$

(cf. Equation 2.1), the Haar measure must be normalized as $\mu(\mathbb{C}S^{d-1}) = \int_{\mathbb{C}S^{d-1}} d\mu(v) = d$. The resulting generalized measurement is rarely used for practical purposes, because its implementation is less resource efficient than for other kinds of POVM and the additional overhead needed for bookkeeping of many parameters make it rather unwieldy, especially in higher dimensions. There is, however, an important property of the Haar measure that is very desirable in the context of *shadow tomography*, which is the subject of next chapter. Namely, the quantum

channel by which the density matrix is transformed when taking the expectation value over measurement outcomes, has a simple analytical form. Formally that quantum channel is defined as

$$Q(\rho) = \mathbb{E}\big(|v\rangle\langle v|\big) = \int_{CS^{d-1}} |v\rangle\langle v| \operatorname{Tr}\big[|v\rangle\langle v|\rho\big] d\mu(v). \qquad (2.7)$$

Evaluating this integral explicitly [47] yields the concise formula

$$Q(\rho) = \frac{\rho + \operatorname{Tr}(\rho)I}{d+1}. \qquad (2.8)$$

Technically this formula holds not only for density matrices, but for general Hermitian matrices. Note that by substituting the identity matrix I for ρ in Equations 2.7 and 2.8, the identity condition Equation 2.6 is recovered.

COMPLEX PROJECTIVE DESIGNS   In principle, the quantum channel could be computed for any POVM. This might however be tedious for a general POVM in lack of additional structure or symmetry. To save effort, one might employ a class of finite-sized and informationally complete POVM possessing the same quantum channel as the Haar measure. Specifically, for a finite generalized measurement $M = \{M_l\}$, the requirement that the respective quantum channels match is expressed as

$$\frac{|M|}{d}\sum_l M_l \operatorname{Tr}\big[M_l\rho\big] = \int_{CS^{d-1}} |v\rangle\langle v| \operatorname{Tr}\big[|v\rangle\langle v|\rho\big] d\mu(v). \qquad (2.9)$$

Rewriting the POVM elements in terms of their defining states $M_l = \frac{d}{|M|}|v_l\rangle\langle v_l|$ and dividing both sides by $\mu(CS^{d-1})$, which equals d, yields

For SIC-POVM $|M| = d^2$, while for MUB $|M| = d(d+1)$.

$$\frac{1}{|M|}\sum_l f\big(|v_l\rangle\langle v_l|\big) = \frac{1}{\mu(CS^{d-1})}\int_{CS^{d-1}} f\big(|v\rangle\langle v|\big) d\mu(v) \qquad (2.10)$$

where $f(x) = x\operatorname{Tr}[x\rho]$. The class of POVM fulfilling this equation for even more general functions f are called *complex projective designs*, because the domain of the right-hand side, CS^{d-1}, is isomorphic to the complex projective space CP^{d-1}. To be more precise, any finite POVM M is called a complex projective t-design, if Equation 2.10 holds for any polynomial f in d variables and of degree t or less.

All SIC-POVM as well as all MUB are at least $t = 2$-designs [64]. In summary, the usage of SIC-POVM or MUB is advantageous due to their symmetry properties, implying the same quantum channel as for the Haar measure while allowing for more resource efficient implementation (e. g. shallow quantum circuits). Furthermore, in the context of quantum state reconstruction, which is implicitly used in TKSVM, they minimize the statistical noise originating from the finiteness of data sets [115].

2.2 MATRIX PRODUCT STATE SAMPLING

The generalized measurements discussed so far are defined on a Hilbert space of dimension d. From here on this shall refer to the dimension of a single site (*local* Hilbert space) of an interacting many-body quantum state. To produce snapshots to feed to the machine learning model, the generalized measurements are successively taken on each site of the system. The way this is achieved depends on the platform that the quantum state to be sampled is represented in. With the goal of benchmarking the capabilities of TKSVM in mind, an efficient and reliable classical representation of the quantum states is very desirable. One such representation is the matrix product state (MPS) representation, which will be briefly reviewed before describing an algorithm for sampling POVM.

MATRIX PRODUCT STATES   The discussion of MPS in this paragraph covers the minimal aspects needed in order to understand the sampling procedure. For a profound introduction to the matter, the reader may refer to [93] or [114].

Consider an arbitrary pure quantum state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes N}$ characterizing a N-site system, with d-dimensional local state spaces $\{\sigma_i\}$ at each site $i \in \{1, \ldots, N\}$. The state is most generally expressed as

$$|\psi\rangle = \sum_{\sigma_1, \ldots, \sigma_N} c_{\sigma_1 \ldots \sigma_N} |\sigma_1 \ldots \sigma_N\rangle \tag{2.11}$$

with exponentially many coefficients $c_{\sigma_1 \ldots \sigma_N}$. By subsequently reshaping and decomposing the object $c_{\sigma_1 \ldots \sigma_N}$, any such state can in principle be represented as a MPS

$$|\psi\rangle = \sum_{\sigma_1, \ldots, \sigma_N} A^{\sigma_1} A^{\sigma_2} \cdots A^{\sigma_{N-1}} A^{\sigma_N} |\sigma_1 \ldots \sigma_N\rangle \tag{2.12}$$

where the object A^{σ_i} is a matrix of a certain *bond dimension*. In general, the bond dimension scales exponentially with the number of sites, rendering the transformation numerically intractable. This is why in practice, the bond dimension is truncated to a fixed maximal value, resulting in an approximate representation of the original state. Whether or not this approximation is appropriate, depends on the nature of the state. In fact, there exists a special class of quantum states whose bond dimension does not grow with the number of sites N. For these states, a suitable truncation only discards singular values that vanish anyway, and thus the state remains *exact*. Often times the term MPS is used for *exact* MPS in the literature. Furthermore, the term MPS usually refers to systems in one spatial dimension, e. g. a chain of sites. Even though technically MPS can be formulated for systems in higher spatial dimensions, they are not numerically manageable and hence rarely, if ever, discussed.

Truncation happens during the Singular-Value-Decomposition, which is heavily used in the context of MPS. It refers to discarding a subset of singular values.

Note that as written in Equation 2.12, the system has open boundary conditions, and rather than matrices, the edge-objects A^{σ_1} and A^{σ_N} represent a row- and a column-vector, respectively. The alternative formulation for periodic boundary conditions requires taking the trace over all A-objects $\mathrm{Tr}(A^{\sigma_1} A^{\sigma_2} \cdots A^{\sigma_{N-1}} A^{\sigma_N})$, where the edges are proper matrices and not vectors.

RIGHT CANONICAL MPS   The MPS representation of a state, Equation 2.12, is not unique. For example, by choosing matrices C_i and G_i fulfilling $C_i G_i = I$ for $i \in \{1, \dots, N\}$, and inserting them in between each product of A-matrices, the representation of the state is transformed to

$$|\psi\rangle = \sum_{\sigma_1, \dots, \sigma_N} B^{\sigma_1} B^{\sigma_2} \cdots B^{\sigma_{N-1}} B^{\sigma_N} |\sigma_1 \dots \sigma_N\rangle \qquad (2.13)$$

with $B^{\sigma_1} = A^{\sigma_1} C_1$, $B^{\sigma_i} = G_{i-1} A^{\sigma_i} C_i$ for $1 < i < N$ and $B^{\sigma_N} = G_{N-1} A^{\sigma_N}$. This circumstance is generally exploited to facilitate MPS computations and reduce their numerical cost. For instance, transforming the MPS into a special form prior to sampling from it, saves the computational cost of many matrix multiplications. The *right canonical* form of a MPS is defined by the following condition

$$\sum_{\sigma_i} B^{\sigma_i} B^{\sigma_i \dagger} = I \quad \text{for all } i \in \{1, \dots, N\} \qquad (2.14)$$

which is represented graphically in Figure 2.2. One may equiva-
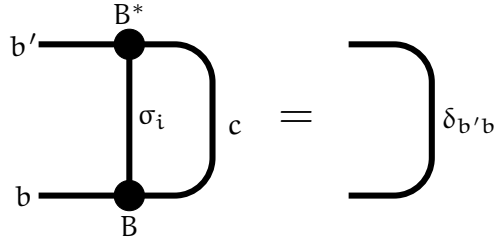
Figure 2.2: Right normalization condition for the B-matrices. The summation runs over the physical index σ_i as well as the right bond index c. Explicitly, the normalization condition reads $\sum_{\sigma_i c} B^{\sigma_i}_{bc} B^{\sigma_i *}_{cb'} = \delta_{b'b}$. In case of open boundary conditions, B^{σ_N} is a vector with a single (left) bond index, and thus the sum only runs over the physical index σ_N. If the normalization condition is fulfilled for all sites i, the MPS is right canonical.

lently transform the MPS into left canonical form, which is defined by Equation 2.14 with swapped order of B-matrices, $B^{\sigma_i} B^{\sigma_i \dagger} \to B^{\sigma_i \dagger} B^{\sigma_i}$. To achieve the aforementioned computational advantage, the site-sequential sampling algorithm must start from the left-most site if the MPS is in right canonical form, and from the right-most site if the MPS is in left canonical form.

SAMPLING ALGORITHM   Given a right canonical MPS with open boundary conditions, the aim is to sequentially sample a finite POVM $M = \{M_l\}$ at each site. Recall that each POVM element is associated with a normalized local state $|v_l\rangle$, $M_l = \frac{d}{|M|} |v_l\rangle \langle v_l|$. The algorithm splits into two steps, that are repeated at each site i. Starting from the left-most site $i = 1$:

- STEP 1: Compute the discrete probability density $\{p_l\}$ to obtain the measurement outcome M_l on site i. The outcome M_l is treated as a random variable in this context. Numerically sample from the probability density, to generate a concrete realization of M_l, denoted M_{l_i}, and save its index l_i as part of the final measurement string.

- STEP 2: Collapse site i to the measured state $|v_{l_i}\rangle$, and perform the appropriate contractions to reduced the size, i.e. the number of sites, of the MPS. More specifically, apply the state $|v_{l_i}\rangle$ to the MPS, prior to summing over the physical index σ_i and the right bond index of the B-vectors on site i. This effectively reduces the B-matrices on site $i + 1$ to vectors and prepares them for sampling in the next iteration.

- PROCEED to site $i + 1$.

The final output of the algorithm is a measurement string $l_1 l_2 \cdots l_N \in \{1, \ldots, |M|\}^N$, which represents a single snapshot. The algorithm is summarized graphically in Figure 2.3.

The updated $\tilde{B}$-vectors on site $i + 1$, obtained from contracting the sampled local state $|v_{l_i}\rangle$ with the $\tilde{B}$-vectors of the previous site i and the B-matrices of site $i + 1$, can be expressed in a recursive formula

$$\tilde{B}_c^{\sigma_{i+1}} = \sum_{\sigma_i, b} v_{l_i}^{\sigma_i} \tilde{B}_b^{\sigma_i} B_{bc}^{\sigma_{i+1}} \tag{2.15}$$

with the initial condition $\tilde{B}^{\sigma_1} = B^{\sigma_1}$. At the last site N, the original B-objects are vectors with only one (left) bond index. Therefore the updated $\tilde{B}$-objects are scalars and the index c drops out of Equation 2.15. Assuming the bond dimension and the physical dimension d are the same at all sites, generating a single snapshot scales quadratically with the bond dimension and the physical dimension, and linearly with the number of sites and the POVM-size $|M|$.

2.3   QUANTUM CIRCUIT SAMPLING

Unfortunately not all quantum states admit a MPS representation, but only those with a favorable scaling of the bond dimension as a function of system size. Quantum computers, on the other hand, are in theory suitable for efficiently representing any quantum state. Since the origin of the snapshot-like input data does not matter, TKSVM can

$$\langle l_1 \cdots l_{i-1}; \psi |$$

$$\to \{ p_l \}$$

$$| l_1 \cdots l_{i-1}; \psi \rangle$$

(a) STEP 1

$$\langle v_{l_i} | l_1 \cdots l_{i-1}; \psi \rangle \qquad\qquad | l_1 \cdots l_i; \psi \rangle$$

(b) STEP 2
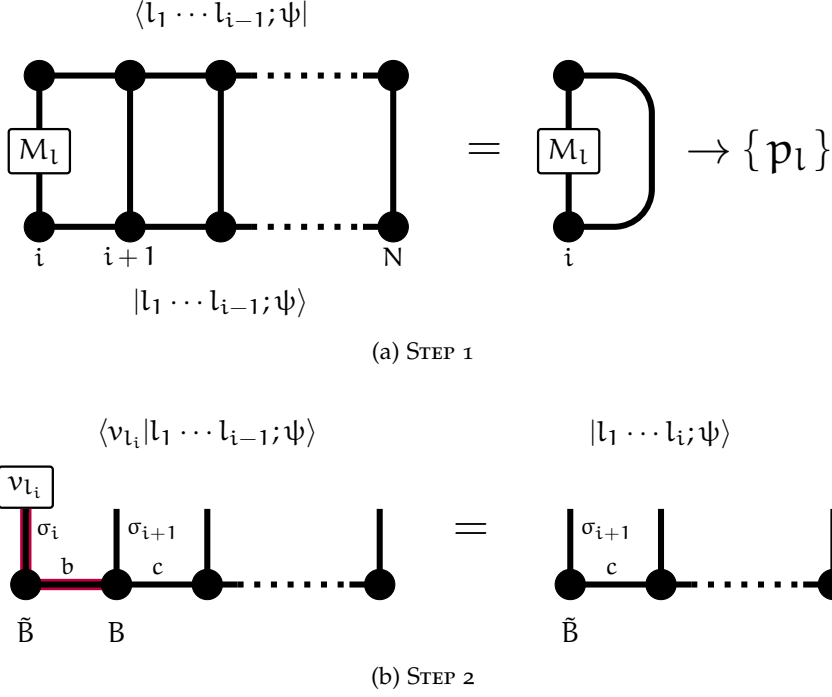
Figure 2.3: Algorithm for site-sequential sampling of a local POVM from a MPS. The first $i-1$ sites have already been sampled and the respective outcomes are stored in the partial measurement string $l_1 \cdots l_{i-1}$. Further the original N-site state $|\psi\rangle$ has been collapsed to a $(N-i+1)$-site state denoted as $|l_1 \cdots l_{i-1}; \psi\rangle := \left(\langle v_{l_1} | \otimes \cdots \otimes \langle v_{l_{i-1}} | \right) |\psi\rangle$. (a) Computation of the probability density for POVM outcomes M_l on site i. All contractions to the right of site i reduce to the identity, due to the right canonical form of $|\psi\rangle$. (b) Collapse of site i to the state $|v_{l_i}\rangle$, followed by contraction of one physical σ_i and one bond index b, highlighted in purple. The B-matrices at site $i+1$ become vectors $B_{bc}^{\sigma_{i+1}} \to \tilde{B}_c^{\sigma_{i+1}}$, and the number of sites is reduced by one.

be employed in combination with a universal quantum computer to investigate arbitrary states. Setting aside the issue of reliable state preparation, this section treats POVM sampling methods from states encoded as quantum circuits. Because the measurements are local, they do not significantly impact the overall circuit depth.

NAIMARK'S DILATION THEOREM   For practical reasons, the direct experimental implementation of generalized measurements is usually circumvented by converting it to a simple projective measurement on an extended space. The physical system that is to be measured is extended by introducing an ancillary system. With unitary evolution of the composite system, the POVM outcomes acting on the physical system are mapped to projective outcomes acting on the composite system. The existence of such a map is guaranteed by *Naimark's dilation theorem* [91]. For simplicity, the following discussion is restricted to

the measurement of a single site. In case of a multi-site system, the same technique is applied to each site simultaneously, contrary to the sequential MPS sampling.

Consider a POVM M and a state $|\psi\rangle$, both defined on a d-dimensional Hilbert space. Additionally, define projective measurements on the composite system as $P = \{P_1, \cdots, P_{|M|}\}$, and associate each POVM element with one of the projective elements, $M_l \leftrightarrow P_l$. Further, assume the ancillary system is in the initial state $|A\rangle$. Due to Naimark's theorem, it is always possible to find a unitary U acting on the composite system $|A\rangle \otimes |\psi\rangle$, such that the probability of measuring P_l on the composite system matches the probability of measuring outcome M_l on the physical system

$$\langle A| \otimes \langle \psi| \, U^\dagger P_l U \, |A\rangle \otimes |\psi\rangle = \langle \psi|M_l|\psi\rangle \,. \tag{2.16}$$

The arguably simplest and most generic way to implement Naimark's theorem is measuring the ancillary system only, as demonstrated in [86]. Towards this end, the ancillary system's dimension must equal $|M|$ and the orthogonal projectors are given by $P_l = |l\rangle \langle l| \otimes I$, where $\{|l\rangle\}$ forms an orthonormal basis for the ancillary space. Then, the unitary is specified by its action onto the composite system of dimension $d_{comp} = d \cdot |M|$,

$$U |A\rangle \otimes |\psi\rangle = \sum_l |l\rangle \otimes \sqrt{M_l} |\psi\rangle \,. \tag{2.17}$$

It is important to note that this equation defines the action of U only for physical states coupled to one specific ancillary state $|A\rangle$. Equation 2.17 doesn't specify how U acts onto any ancillary state other than its initial state $|A\rangle$. Therefore, as specified above, U is technically an *isometry* from a d-dimensional into an d_{comp}-dimensional space, rather than a proper unitary. Specifically, writing $U = \sum_l |l\rangle \langle A| \otimes \sqrt{M_l}$, trivially fulfilling Equation 2.17, is *not* unitary. Quantum circuits are however based on unitary evolution, implying that the isometry must be extended to a full unitary for application in a circuit. In matrix language, the isometry can be seen as a set of d columns, that are a subset of the d_{comp} columns of a full unitary. Thus the full unitary is obtained by determining the remaining $d_{comp} - d$ columns such that the resulting $(d_{comp} \times d_{comp})$-matrix is unitary. There are several methods to achieve this, the most common of which is likely the

Gram-Schmidt process. Equation 2.16 for the action of U as specified in Equation 2.17 is confirmed by

$$\langle A | \otimes \langle \psi | U^\dagger P_l U | A \rangle \otimes | \psi \rangle$$

$$= \sum_{l',l''} \langle l' | \otimes \langle \psi | \left(| l \rangle \langle l | \otimes \sqrt{M_{l'}}^\dagger \sqrt{M_{l''}} \right) | l'' \rangle \otimes | \psi \rangle$$

$$= \sum_{l',l''} \langle l' | l \rangle \langle l | l'' \rangle \langle \psi | \sqrt{M_{l'}}^\dagger \sqrt{M_{l''}} | \psi \rangle \tag{2.18}$$

$$= \langle \psi | M_l | \psi \rangle \,.$$

While in this instance the composite system has dimension $d_{comp} = d \cdot |M|$, there exists another method yielding a lower dimension $d_{comp} = |M|$ [29, 92]. To achieve a lower dimension, first each POVM element is associated with a local sub-normalized state $| v_l \rangle$, $M_l = | v_l \rangle \langle v_l |$, with $| \langle v_l | v_l \rangle | = d/|M|$. These states define the isometry

$$\hat{V} = \begin{pmatrix} \mathbf{v}_1^\dagger \\ \mathbf{v}_2^\dagger \\ \vdots \\ \mathbf{v}_{|M|}^\dagger \end{pmatrix} \in \mathbb{C}^{|M| \times d} \tag{2.19}$$

where $\mathbf{v}_l$ denotes to the column vector representation of the state $| v_l \rangle$; $\mathbf{v}_l^\dagger$ is thus the row vector with conjugated elements. By definition, any isometry $\hat{V}$ has orthonormal columns, allowing to extend it into a full unitary without altering the existing columns. Orthonormality of the existing columns is guaranteed since M satisfies the identity condition $\sum_l | v_l \rangle \langle v_l | = I$. Hence any matrix $\hat{V}$ constructed this way is indeed an isometry. Next, unitarize $\hat{V}$ by extending the states, i. e. determine the remaining columns to obtain

Notation: the symbol V denotes an operator, whereas the symbol $\hat{V}$ denotes the corresponding matrix in the standard computational basis.

$$\hat{U} = \begin{pmatrix} \mathbf{v}_1^\dagger & \mathbf{w}_1^\dagger \\ \mathbf{v}_2^\dagger & \mathbf{w}_2^\dagger \\ \vdots & \vdots \\ \mathbf{v}_{|M|}^\dagger & \mathbf{w}_{|M|}^\dagger \end{pmatrix} \in \mathbb{C}^{|M| \times |M|}, \quad \hat{U}^\dagger \hat{U} = I \tag{2.20}$$

thereby implicitly defining orthonormal extended states $| u_l \rangle = | v_l \rangle \oplus | w_l \rangle$. To confirm that U indeed properly maps the POVM to projective measurements $P_l = | l \rangle \langle l |$ on the extended space, Equation 2.16 must be verified. For simplicity, assume that the initial state of the ancillary system, $| A \rangle$, can be chosen such that the initial composite state becomes

$$| A \rangle \otimes | \psi \rangle = | \psi \rangle \oplus | 0^{|M| - d} \rangle \equiv \begin{pmatrix} \psi \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{C}^{|M|} \tag{2.21}$$

where ψ is the column vector representation of $|\psi\rangle$. Then, one can immediately confirm Equation 2.16

$$\langle A| \otimes \langle\psi| U^\dagger P_l U |A\rangle \otimes |\psi\rangle$$
$$= \langle\psi| \oplus \langle 0^{|M|-d}| \; U^\dagger |l\rangle \langle l| U \; |\psi\rangle \oplus |0^{|M|-d}\rangle$$
$$= \langle\psi| V^\dagger |l\rangle \langle l| V |\psi\rangle \tag{2.22}$$
$$= \langle\psi|v_l\rangle \langle v_l|\psi\rangle$$
$$= \langle\psi|M_l|\psi\rangle$$

exploiting the fact that $V^\dagger$ maps the computational basis state $|l\rangle$ onto $|v_l\rangle$. The assumption in Equation 2.21 that there exists an ancilla state $|A\rangle$ such that $|A\rangle \otimes |\psi\rangle = |\psi\rangle \oplus |0^{|M|-d}\rangle$, cannot always be made. It is only possible, if the dimension of the ancillary system d_A can be chosen so that $d_{comp} = d \cdot d_A$ equals $|M|$. In that case $|A\rangle$ is simply the computational basis state $(1,0,\ldots,0)^T \in \mathbb{C}^{d_A}$. However, in general the POVM size $|M|$ isn't necessarily compatible with d_{comp}, e. g. if $d = 2$ and $|M| = 3$, there is no valid dimension of the ancillary system d_A such that $d \cdot d_A = |M|$. To remedy the situation, one must choose $d_{comp} > |M|$. Then there are d_{comp} projective measurement outcomes for only $|M|$ POVM outcomes. The unitary U connecting them can, however, be conveniently chosen such that the probability of the excess $d_{comp} - |M|$ projective outcomes vanishes.

Depending on the endianness used to represent states in the computational basis, either $|A\rangle \otimes |\psi\rangle$ or the reversed order $|\psi\rangle \otimes |A\rangle$ produce the correct state vector $(\psi^\dagger, 0,\ldots,0)^\dagger$. Here the big endian convention is used to represent operators and states.

EXPLICIT EXAMPLES   The approach from Equation 2.17 applied to the tetrahedral SIC-POVM in $d = 2$ (Expression 2.4) yields a unitary with a subset of unspecified columns (marked by a $\times$-symbol):

$$\hat{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha^2 & \alpha\beta & \times & \times & \times & \times & \times & \times \\ \alpha\beta & \beta^2 & \times & \times & \times & \times & \times & \times \\ \alpha^2 & -\alpha\beta & \times & \times & \times & \times & \times & \times \\ -\alpha\beta & \beta^2 & \times & \times & \times & \times & \times & \times \\ \beta^2 & -i\alpha\beta & \times & \times & \times & \times & \times & \times \\ i\alpha\beta & \alpha^2 & \times & \times & \times & \times & \times & \times \\ \beta^2 & i\alpha\beta & \times & \times & \times & \times & \times & \times \\ -i\alpha\beta & \alpha^2 & \times & \times & \times & \times & \times & \times \end{pmatrix} \tag{2.23}$$

with $\alpha, \beta = \sqrt{(3 \pm \sqrt{3})/6}$. The dimension of the ancillary system is $d_A = 4$, to accommodate all four outcomes, and the total dimension of the composite system is $d_{comp} = 8$.
Meanwhile, applying the more efficient approach from Equation 2.20 allows the realization of Naimark's theorem with a smaller dimension

of the composite system $d_{comp} = 4$. One explicit example for the unitarization of the isometry $\hat{V}$ is given by [29]

$$\hat{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha & \beta \\ \alpha & -\beta \\ \beta & -i\alpha \\ \beta & i\alpha \end{pmatrix} \longrightarrow \hat{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} \alpha & \beta & \beta & \alpha \\ \alpha & -\beta & \beta & -\alpha \\ \beta & -i\alpha & -\alpha & i\beta \\ \beta & i\alpha & -\alpha & -i\beta \end{pmatrix} . \quad (2.24)$$

The resource-optimal implementation of generalized measurements based on MUB in dimension d requires an ancillary systems with dimensions $d_A = d + 1$. On the other hand, optimally realizing SIC-POVM in dimension d, requires an ancillary system merely the same size as the physical, $d_A = d$.

RANDOMIZED IMPLEMENTATION OF MUB   In the context of quantum computation, implementing a POVM through Naimark's theorem requires not only ancillary qudits but also entangling gates acting on the composite system. Per se, any POVM acting on a low-dimensional Hilbert space can be implemented with high fidelity, since the resulting circuits are shallow. But considering that the measurement portion of the quantum circuit is preceded by state preparation, any additional depth diminishing data quality should be prevented. Because the entangling gates are typically more prone to error than the single qudit gates, the number of entangling gates should be minimized before minimizing the number of single qudit gates. As an alternative to Naimark's theorem, MUB may be implemented in a *randomized* fashion to avoid the use of an ancillary system and thus reduce the number of entangling gates to zero. The idea behind the randomized implementation is to exploit a structural property of MUB. Specifically, MUB can be seen as a collectiton of $d + 1$ distinct projective measurements. None of the projective measurements is favored over another, permitting to choose from them randomly with uniform probability $1/(d+1)$. For illustration, consider the generalized measurement constructed from the MUB in dimension $d = 2$ (cf. Expression 2.2)

$$M = \frac{1}{3} \big\{ |z_\pm\rangle \langle z_\pm| , |x_\pm\rangle \langle x_\pm| , |y_\pm\rangle \langle y_\pm| \big\}. \quad (2.25)$$

Independently of the quantum state ρ, the probability of the measurement outcome being in basis a is uniform

$$\frac{1}{3}\mathrm{Tr}(|a_+\rangle \langle a_+| \rho) + \frac{1}{3}\mathrm{Tr}(|a_-\rangle \langle a_-| \rho) = \frac{1}{3} \quad \forall a \in \{x, y, z\}. \quad (2.26)$$

Instead of directly implementing projective measurements in bases other than the computational, it can be more efficient to rotate the state followed by a projective measurement in the computational basis. Figure 2.4 shows a comparison between the implementation of MUB based on Naimark's theorem and the randomized implementation.

(a) NAIMARK'S THEOREM

(b) RANDOMIZED
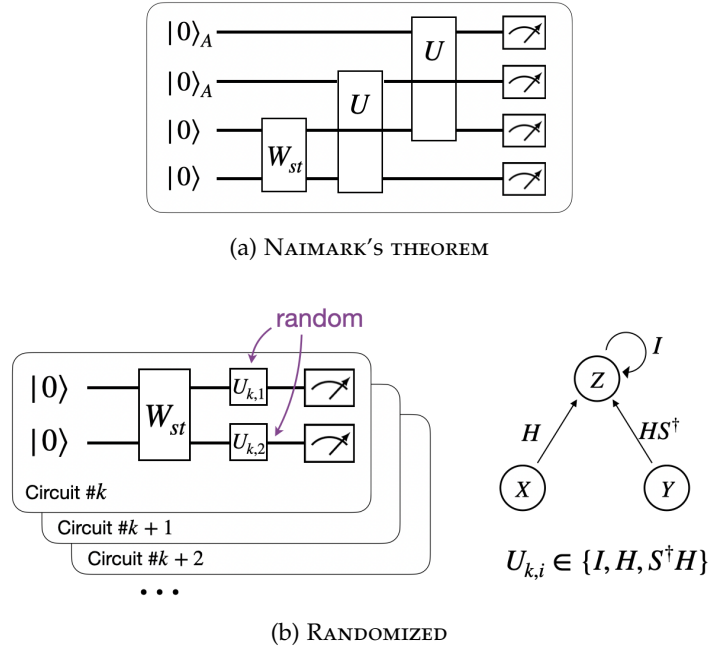
Figure 2.4: Implementation of a MUB-POVM for a physical system consisting of two sites. The unitary W_{st} prepares the state to be measured. (a) Deterministic circuit resulting from Naimark's theorem. For each physical site, one additional ancillary site is needed. The unitary U directly links the projective measurement outcomes of each physical–ancillary site pair to the POVM outcomes acting on the physical site only. (b) Non-deterministic circuit resulting from randomized implementation of MUB. No ancillary sites are needed. Instead, the random unitaries $U_{k,i}$ rotate from one of the unbiased bases into the computational basis in each circuit instance k. The operations I, H, S denote identity, Hadamard, and phase gate, respectively.

The drawback of randomized implementation is that the single site unitaries have to be randomly generated after each snapshot. Such a frequent reconfiguration of the quantum circuit can be unpractical when the required number of snapshots is large.

It is noteworthy, that randomized implementation of MUB can also be applied to the MPS sampling algorithm, discussed in the previous section. Rather than computing expectation values of all $|M| = d(d+1)$ measurement operators in STEP 1, one only needs to compute the expectation values of the d measurement operators from a single randomly chosen orthogonal basis. The computational cost is thus reduced by a factor $d+1$, which can have significant impact in practise, especially for large bond dimensions.

ESTIMATION OF LOCAL OBSERVABLES

An integral part of TKSVM is the statistical estimation of sets of local observables. The method used to compute these estimators from POVM snapshots will be the subject of the first section of this chapter. In the second section, the type of observables and the way they affect the learning model's expressibility will be covered.

3.1 SHADOW TOMOGRAPHY

The method of shadow tomography introduced by Aaronson [1], decisively differs from other types of tomography methods regarding its goal. While usually the term tomography refers to methods aimed at the reconstruction of an unknown quantum state ρ, shadow tomography aims at the estimation of target functions $\text{Tr}(O_\mu \rho)$ defined by a set of observables $\{O_\mu\}$. It is motivated by the argument that full tomography is excessive for specialized tasks, in the sense that understanding the physical properties of a model does not necessarily require precise knowledge of the full density matrix. Moreover, the curse of dimensionality limits the applicability of full tomography to systems with few degrees of freedom. For a system of N qubits for example, the number of measurements (copies) of ρ required for efficient state tomography schemes scales as $\mathcal{O}(2^{2N})$ [49, 90].

Both these issues are circumvented in shadow tomography, for which the number of measurements required to estimate a set of target functions up to arbitrary cumulative error scales as $\mathcal{O}(\log^4(m)\log(D))$, where m is the number of target functions and D the dimension of the state, e.g. $D = 2^N$ for N qubits. Implementing the procedure as proposed by Aaronson has substantial drawbacks with regard to hardware demands. It requires *collective* measurements on multiple copies to be taken *gently*, meaning that the measurement only slightly disturbs the state rather than destroying it. The collective aspect of the measurements requires simultaneous preparation of $\mathcal{O}(D)$ copies of ρ and the gentle aspect requires exponentially deep quantum circuits.

Introducing the notion of *classical shadows*, which is a purely classical description of the quantum state ρ, Huang *et al.* modified the procedure such that logarithmic scaling in the number of target functions is retained while the necessity for gentle and collective measurements is removed, thereby rendering the process feasible for experimental implementation [53]. Their modifications come at the price of reintroducing some exponential scaling, but only with respect to the locality of the observables and not with respect to system size. In the

context of many-body systems, the *locality* of an observable refers to the number of sites on which the observable acts non-trivially. The number of required POVM samples to estimate a set of m different observables up to cumulative error ϵ scales as $\mathcal{O}(\log(m) \max_\mu \|O_\mu\|_{\text{shadow}}^2 / \epsilon^2)$, where the norm $\|\cdot\|_{\text{shadow}}$ depends on the POVM. For the measurements described in the previous chapter, this norm is bounded by the operator norm as $\|O_\mu\|_{\text{shadow}}^2 \leqslant c^r \|O_\mu\|_\infty^2$ with c a constant and r the locality of O_μ. When the POVM snapshots are tensor products of MUB outcomes with local dimension $d = 2$, one obtains $c = 3$ for instance.

The original definition of classical shadows by Huang *et al.* for qubit systems is based on unitary ensembles combined with projective measurement. In the following discussion, the immediate generalization to quantum systems with arbitrary local Hilbert space dimension is derived instead. As another generalization, the derivation below relies on POVM in lieu of the unitary ensemble. The unitary ensemble employed by Huang *et al.*, i.e. tensor products of random single qubit clifford gates, is recovered by the equivalent POVM based on MUB in $d = 2$.

Consider a set of N_s independent POVM samples $\{M_1^{(l)}, M_2^{(l)}, \dots, M_N^{(l)}\}_{l=1}^{N_s}$ of a N-site quantum state ρ with local Hilbert space dimension d. Each sample l is a tensor products of single site POVM outcomes. For notational convenience, the outcomes $M_j^{(l)}$ shall denote properly normalized rather than sub-normalized projectors in this section. The expectation value of the POVM samples represents a N-site quantum channel

$$Q_N(\rho) = \mathbb{E}\left(\bigotimes_{j=1}^N M_j\right) = \lim_{N_s \to \infty} \frac{1}{N_s} \sum_{l=1}^{N_s} \left[\bigotimes_{j=1}^N M_j^{(l)}\right]. \tag{3.1}$$

The goal is to invert the quantum channel to obtain a representation of ρ in terms of POVM data. This representation is exact in the limit $N_s \to \infty$, but approximate for a finite number of samples. Intuitively, the quantum channel is invertible if and only if the N-site POVM (tensor product) is informationally complete. This is true if the local POVM are informationally complete on the d dimensional Hilbert space. The approximate expression for the quantum state reads

$$\rho \approx \frac{1}{N_s} \sum_{l=1}^{N_s} \hat{\rho}^{(l)} \tag{3.2}$$

$$\hat{\rho}^{(l)} = Q_N^{-1}\left(\bigotimes_{j=1}^N M_j^{(l)}\right) = \bigotimes_{j=1}^N Q^{-1}(M_j^{(l)}). \tag{3.3}$$

The set $\{\hat{\rho}^{(l)}\}$ is called the *classical shadow* of ρ, and constitutes the base for the estimation of target observables. Due to the tensor product structure of the samples, the N-site quantum channel as well as its

inverse factorize into a product of single site quantum channels, as expressed in the last equality. If the local POVM is a complex projective design (introduced in Section 2.1), the single site quantum channel is given by Equation 2.8 which is readily invertible as

$$Q^{-1}(A) = (d+1)A - \text{Tr}(A)I \tag{3.4}$$

where A denotes an arbitrary Hermitian matrix. Naively, the target functions would be estimated as

$$\text{Tr}(O_\mu \rho) \approx \frac{1}{N_s} \sum_l \text{Tr}(O_\mu \hat{\rho}^{(l)}). \tag{3.5}$$

But rather than averaging over the classical shadow directly, Huang *et al.* actually employ a *median of means* estimation of the target functions. In this method of statistical outlier elimination, the classical shadow is partitioned into equally sized bins wherein the average is computed individually, and the median of the set of averages is taken as the final estimate. Explicitly, for a number K of bins

$$\text{Tr}(O_\mu \rho) \approx \text{median} \left\{ \text{Tr}(O_\mu \hat{\rho}^1), \dots, \text{Tr}(O_\mu \hat{\rho}^K) \right\} \tag{3.6}$$

$$\hat{\rho}^k = \frac{1}{\lfloor N_s/K \rfloor} \sum_{l=s(k)}^{k \lfloor N_s/K \rfloor} \hat{\rho}^{(l)} \tag{3.7}$$

with $s(k) = (k-1)\lfloor N_s/K \rfloor + 1$.

The scaling $\mathcal{O}(\log(m) \max_\mu \|O_\mu\|_{\text{shadow}}^2 / \epsilon^2)$ mentioned above was rigorously derived only in conjunction with the median of means method, and doesn't apply in case of the direct average of Equation 3.5.

As opposed to shadow tomography, the objective of TKSVM is not the computation of target functions with high precision. Considering two quantum states ρ and ρ' the goal is rather to determine a hyperplane separating the two states in the space spanned by the target functions. For this specific task, it is more suitable not to take the median to eliminate outliers, but to use the full set of K bin averages as input data. This is justified because TKSVM itself appropriately treats outliers in the input data by assigning them a lower weight compared to non-outliers during the optimization process. The optimization problem underlying TKSVM is covered in detail in Chapter 4.

Assume the N-site quantum state under investigation possesses translational invariance with respect to a cluster of sites. In this context, a cluster shall be defined as a collection of n neighboring sites with a fixed geometric shape, e. g. six sites forming a hexagon in the honeycomb lattice. Further assume that all r-local target functions act on sites within a single cluster, implying $r \leqslant n$. Then the system may be partitioned into a set of site clusters respecting the translational symmetry. And due to the translational invariance, each cluster can be

Sample Average
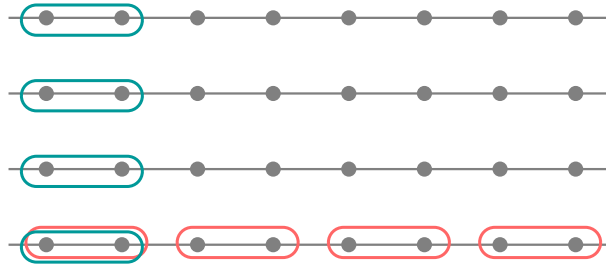
Cluster Average

Figure 3.1: Cluster averaging and sample averaging techniques for a cluster size $n = 2$ on the chain lattice. Each chain represents one N-site sample with one POVM outcome at each site. Taking the *sample average* implies averaging computed features within a cluster at a fixed location over different samples, while taking the *cluster average* implies averaging over all clusters within a fixed sample.

seen as an independent POVM sample of a reduced n-site state, with $n \ll N$ in typical applications. Computing target functions over clusters shall be referred to as *cluster average* technique. It can optionally be combined with the standard *sample average* technique described above. When combined, however, only one technique should be performed with binning. Usually, the sample average is performed with binning and the cluster average without binning. The difference between these averaging techniques is visualized in Figure 3.1. Overall, the purpose of the cluster average technique is to reduce the number of required N-site samples for adequate target function estimation by exploiting the translational symmetry of the state.

Are the sample average and the cluster average techniques equivalent? In case of the sample average all samples are always completely uncorrelated because each sample is taken independently. The quantum state is destroyed after measurement, requiring to prepare the state anew to take the next sample. On the other hand, the measurements from different cluster within the same sample are *approximately* uncorrelated only if the correlation length of the state is small compared to the system size. In practice it is therefore usually safe to assume that sample average and cluster average are equivalent if the system size is large enough.

The required number of samples for the learning model of TKSVM to be successful depends most notably on the locality of the underlying target function, the system size, and quality of the data (quantum noise). Therefore it is investigated not in generality, but numerically for the specific applications in Chapters 5 and 6.

Target functions computed as input for TKSVM are always r-local tensor products constructed from a predetermined operator basis, which is a set of Hermitian operators acting on a single site. The operator basis can be seen as the language in which TKSVM expresses what it has learned. Therefore, its size directly impacts the *expressibility*, which in this context signifies the ability to yield analytical expressions of a physical observable characterizing a specific quantum state. The choice of the operator basis is completely independent of the POVM that is used for sampling quantum states, which is possible because informationally complete measurements allow the reconstruction of any observables. In practice, different operator bases can hence be tried out without the need to re-sample the quantum state.

Physical observables are represented by Hermitian operators. Therefore, the operator basis must span the space of all Hermitian matrices in order to achieve the highest possible expressibility. This implies that the size of the operator basis must at least equal d^2. A straightforward choice of a complete Hermitian operator basis would be to use one operator for each independent element of a Hermitian matrix. Specifically, this results in d operators with a single 1 somewhere on the diagonal and 0's elsewhere, $(d^2 - d)/2$ symmetric operators with a single pair of 1's in the off-diagonal part and 0's elsewhere, as well as $(d^2 - d)/2$ Hermitian operators with a single pair of $i, -i$ entries in the off-diagonal part and again 0's elsewhere. However, if the type of the interacting sites is known, a more suitable choice can be made. E. g. for interacting spin degrees of freedom, the operators measuring spin along the three spatial axes are a good starting point. To render the resulting incomplete operator basis complete, it may be extended by considering products and/or linear combinations of the spin operators. In case of spin-1/2 systems adding a single operator proportional to the identity matrix to the Pauli matrices completes the operator basis, while for spin-1 systems six more operators are needed. The spin-1 operators are defined as

$$\tau^x = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tau^y = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \tau^z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$(3.8)$$

and a complete operator basis is for example given by

$$B = \{\tau^x, \tau^y, \tau^z, (\tau^x)^2, (\tau^y)^2, (\tau^z)^2, [\tau^y, \tau^z]_+, [\tau^x, \tau^z]_+, [\tau^x, \tau^y]_+\} \quad (3.9)$$

where $[\cdot, \cdot]_+$ denotes the anti-commutator.

Besides expressibility, also the runtime of TKSVM is affected by the operator basis. A larger operator basis increases the total number of

target functions proportional to $|B|^r$ (shown in the next chapter), which in turn implies longer runtime for both target function estimation and the TKSVM optimization process. Therefore it may be advantageous to rely on a incomplete operator basis, if it is still sufficiently expressive for the task at hand. In practise, finding a good balance between expressibility and runtime is achieved by starting with a small operator basis and extending it systematically until success.

With an operator basis given by $B = \{O^a\}$, a target function of locality or *rank* r is estimated as

$$\mathrm{Tr}\left(O_{j_1}^{a_1}O_{j_2}^{a_2}\cdots O_{j_r}^{a_r}\hat{\rho}^k\right) = \frac{1}{\lfloor N_s/K\rfloor}\sum_{l=s(k)}^{k\lfloor N_s/K\rfloor}\mathrm{Tr}\left[O_{j_1}^{a_1}Q^{-1}(M_{j_1}^{(l)})\right]\times$$

$$\mathrm{Tr}\left[O_{j_2}^{a_2}Q^{-1}(M_{j_2}^{(l)})\right]\times\cdots\times\mathrm{Tr}\left[O_{j_r}^{a_r}Q^{-1}(M_{j_r}^{(l)})\right] \quad (3.10)$$

within the k-th bin, cf. Equation 3.7. Outside of the support of the observable all factors equal unity, because the projectors $M_j^{(l)}$ have unit trace which is conserved by the inverse quantum channel

$$\mathrm{Tr}\left[Q^{-1}(M_j^{(l)})\right] =$$
$$(d+1)\mathrm{Tr}(M_j^{(l)}) - \mathrm{Tr}(M_j^{(l)})\mathrm{Tr}(I) = \mathrm{Tr}(M_j^{(l)}) = 1. \quad (3.11)$$

Because there is only a fixed amount of possible combinations of the operator O_j^a and the measurement outcome $M_j^{(l)}$ at each site j, the factors $\mathrm{Tr}\left[O_j^aQ^{-1}(M_j^{(l)})\right]$ can pre-computed and stored in a lookup table of size $|B| \times |M|$ to save computational effort.

Whenever the operator basis is large enough to produce an operator proportional to the identity matrix as a linear combination, which is necessarily the case for a complete operator basis, then at a given rank r all lower rank target functions are computed implicitly along the way. For illustration, consider the spin-1 operator basis $B = \{\tau^x, \tau^y, \tau^z, (\tau^x)^2, (\tau^y)^2, (\tau^z)^2\}$ which produces the identity like $(\tau^x)^2 + (\tau^y)^2 + (\tau^z)^2 = 2I$. Any rank 1 observable $\tau_1 \in \mathrm{span}(B)$ is implicitly computed at rank 2, disguised as a sum of target functions $\mathrm{Tr}[\tau_1(\tau_2^x)^2\rho] + \mathrm{Tr}[\tau_1(\tau_2^y)^2\rho] + \mathrm{Tr}[\tau_1(\tau_2^z)^2\rho] = 2\,\mathrm{Tr}[\tau_1\rho]$. Such redundancies must be masked in order to facilitate interpretation of TKSVM results.

TENSORIAL KERNEL SUPPORT VECTOR MACHINE

A SVM [12, 128] in its most basic form is a binary classifier approximating an unknown indicator function. The approximation is constructed from a *training set* of labeled data points, which can be regarded as samples of the indicator function. In essence, the SVM's objective is to find a hyperplane which separates the two classes. Since there are generally infinitely many different hyperplanes separating the classes, the aim is to find the hyperplane that is optimal with respect to a specific constraint.

Consider a set of data points $\mathbf{x}_k \in \mathbb{R}^d$, each labelled $y_k \in \{-1, 1\}$. Data points with the same label represent a class. In the simplest case where the two classes are linearly separable, there exists a hyperplane defined by a vector $\mathbf{w}$ and a constant b, such that the signed distance from the plane correctly predicts the class label of each data point,

$$\text{sgn}(\mathbf{w} \cdot \mathbf{x}_k - b) = y_k \ \forall k. \tag{4.1}$$

The vector $\mathbf{w}$ represents the normal to the hyperplane and the constant b, called the bias parameter, a shift thereof. In Figure 4.1a some possible hyperplanes are displayed for a linearly separable problem. The optimization goal of the SVM is to find the hyperplane $\mathbf{w} \cdot \mathbf{x} - b = 0$, called decision boundary, which maximizes the margin between data points of opposite classes. Upper and lower boundaries of the margin are formally defined as $\mathbf{w} \cdot \mathbf{x} - b = \pm 1$. This sets the width of the margin equal $2/\|\mathbf{w}\|$, as shown in Figure 4.1b. Typically and for mere convenience, instead of maximizing the margin width, the square of its inverse is minimized, and the optimization problem reformulated as

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 \text{ with respect to } \mathbf{w}, b$$
$$\text{subject to } y_k(\mathbf{w} \cdot \mathbf{x}_k - b) \geqslant 1 \ \forall k. \tag{4.2}$$

This problem is solved by means of Lagrange multipliers λ_k enforcing the constraints. The Lagrangian for the constrained optimization problem reads [12]

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_k \lambda_k(1 - y_k(\mathbf{w} \cdot \mathbf{x}_k - b)). \tag{4.3}$$

Setting the gradient of $\mathcal{L}$ with respect to the optimization parameters equal zero, one obtains

$$\mathbf{w} = \sum_k \lambda_k y_k \mathbf{x}_k \quad \text{and} \quad \sum_k \lambda_k y_k = 0. \tag{4.4}$$

(a) NON-OPTIMAL

(b) MARGIN WIDTH

(c) HARD MARGIN

(d) SOFT MARGIN
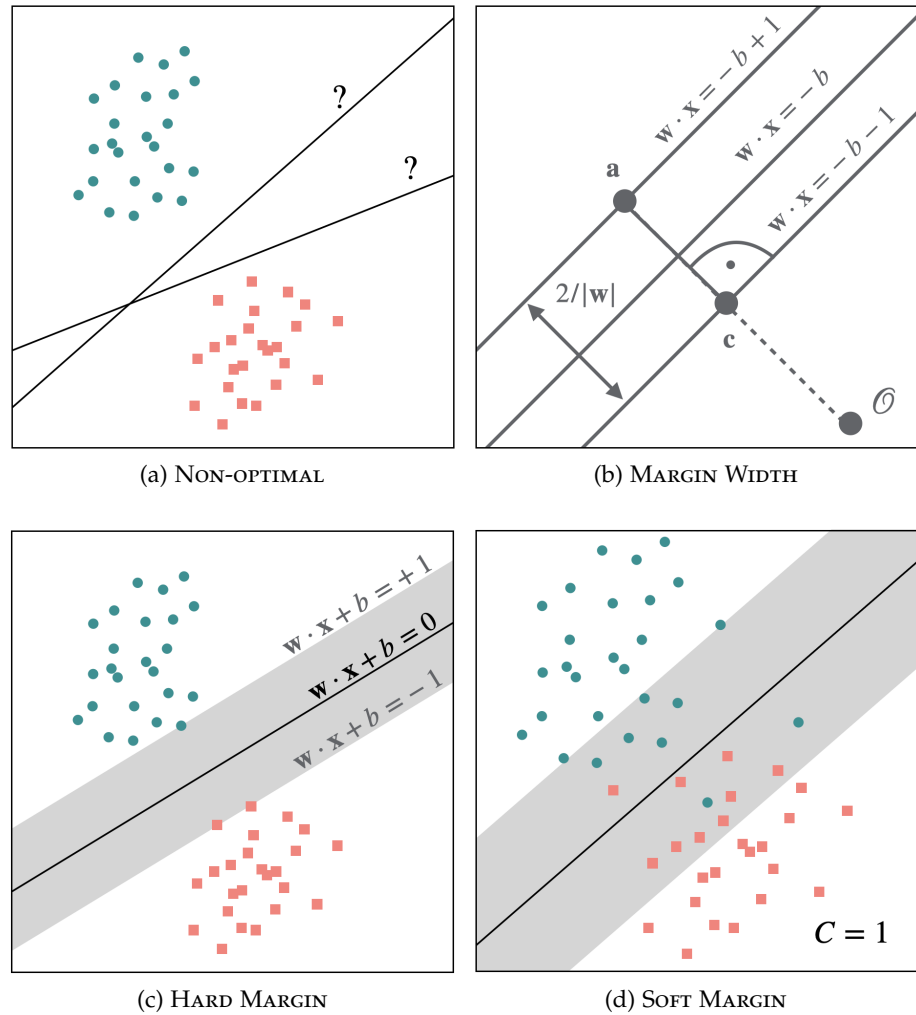
Figure 4.1: Linear classification problems solvable by SVM in dimension $d = 2$. (a) Linear separable problem with several possible hyperplanes. (b) Computation of the margin width. With $\mathcal{O}$ the origin, the margin width is given by $\|\mathbf{a} - \mathbf{c}\| = |\|\mathbf{a}\| - \|\mathbf{c}\|| = 1/\|\mathbf{w}\||-b + 1 + b + 1| = 2/\|\mathbf{w}\|$. (c) Data separated by the optimal hyperplane, maximizing the margin width. Data points are not allowed to penetrate the hard margin. (d) Non-separable data with the optimal soft margin at a given penalty parameter $C = 1$.

Substituting these expression back into the Lagrangian yields the dual problem which is expressed in terms of λ_k only. The Lagrangian for the dual problem is given by

$$\mathcal{L} = \sum_k \lambda_k - \frac{1}{2} \sum_{j,k} \lambda_j \lambda_k y_j y_k \mathbf{x}_j \cdot \mathbf{x}_k \tag{4.5}$$

and the corresponding dual optimization problem becomes

$$\text{Maximize } \sum_k \lambda_k - \frac{1}{2} \sum_{j,k} \lambda_j \lambda_k y_j y_k \mathbf{x}_j \cdot \mathbf{x}_k \text{ with respect to } \{\lambda_k\}$$
$$\text{subject to } \lambda_k \geqslant 0, \ \sum_k \lambda_k y_k = 0 \ \forall k. \tag{4.6}$$

In optimization theory, minimizing the primal problem is equivalent to maximizing the dual problem.

Whenever the problem is not perfectly separable by a hyperplane, i. e. Equation 4.1 cannot be satisfied, the optimization problem needs to be relaxed. This is achieved by allowing data points to penetrate the margin at a cost $C > 0$. The optimization problem with softened constraint is expressed as

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_k \xi_k \text{ with respect to } \mathbf{w}, b, \{\xi_k\}$$
$$\text{subject to } y_k(\mathbf{w} \cdot \mathbf{x}_k - b) \geqslant 1 - \xi_k, \ \xi_k \geqslant 0 \ \forall k \tag{4.7}$$

where ξ_k denote so-called slack variables. Misclassified datapoints lying on the wrong side of the decision boundary have slack variables $\xi_k > 1$ whereas correctly classified points within and outside of the margin have $\xi_k < 1$ and $\xi_k = 0$, respectively. The dual relaxed problem can be obtained from the primal relaxed problem in the same fashion as the dual strict problem is obtained from the primal strict problem. It merely differs from the dual strict problem by the additional constraint that the Lagrange multipliers are bounded from above by the penetration cost $C \geqslant \lambda_k \geqslant 0$. Importantly, all slack variables ξ_k are eliminated in the dual problem.

The motivation behind considering the dual problem is that in contrast to the primal problem, it is solvable by *quadratic programming* methods [20, 88]. These numerical methods guarantee convergence to the global maximum and polynomial runtime with respect to the number of datapoints. Once the optimization is completed, the class label of a test sample $\mathbf{x}$ is predicted by the sign of the decision function, defined as

$$D(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} - b = \sum_k \lambda_k y_k \mathbf{x}_k \cdot \mathbf{x} - b. \tag{4.8}$$

All training samples with non-zero coefficients λ_k, intuitively all those lying within or at the boundary of the margin, are called *support vectors*. Note that in case of a soft margin, the penetration cost C is not subject to optimization but rather an input parameter set by the user. Finite

positive values of C represent a soft and therefore broader margin, while the original hard and therefore narrow margin is recovered in the limit $C \to \infty$, see Figures 4.1c and 4.1d. Thus, the number of support vectors is indirectly controlled by C. In practice, the choice of C is difficult since it is not bounded. Without any knowledge of the data, one is left to benchmark the trained model against independent test sets for different orders of magnitude of C. Fortunately, there exists an alternative formulation eliminating the parameter C in favor of a new bounded regularization parameter $\nu \in [0, 1)$ [19, 113]. This parameter directly controls the number of support vectors and is easier to choose appropriately, at the cost of a more involved theoretical derivation, which is omitted here.

4.2   THE KERNEL TRICK

In many cases the two classes are not at all separable by a hyperplane in the original space $\mathfrak{X} = \mathbb{R}^d$ of the training samples. Fortunately, it is in principle always possible to find a map into a higher dimensional space $\mathbf{g} : \mathfrak{X} \to \mathfrak{G}$, where the data becomes separable again. Once all training samples are mapped into the higher-dimensional space, linear SVM can be applied to solve the optimization problem in the new space $\mathfrak{G}$. Specifically, in all equations and definitions above, simply substitute $\mathbf{x} \mapsto \mathbf{g}(\mathbf{x})$. This approach is however not always applicable, because performing calculations in $\mathfrak{G}$ might not be numerically feasible. In the worst case, the dimension of $\mathfrak{G}$ might even be infinite.

The *kernel trick* solves this problem by avoiding explicit computations of $\mathbf{g}$ as well as all subsequent computations in $\mathfrak{G}$. To appreciate the kernel trick, note that the Lagrangian and the decision function in the space $\mathfrak{G}$

$$\mathcal{L} = \sum_k \lambda_k - \frac{1}{2} \sum_{j,k} \lambda_j \lambda_k y_j y_k \mathbf{g}(\mathbf{x}_j) \cdot \mathbf{g}(\mathbf{x}_k) \tag{4.9}$$

$$D(\mathbf{x}) = \sum_k \lambda_k y_k \mathbf{g}(\mathbf{x}_k) \cdot \mathbf{g}(\mathbf{x}) - b \tag{4.10}$$

still only rely on the evaluation of inner products, cf. Equations 4.5 and 4.8. This circumstance allows for the definition of a kernel function $\mathcal{K} : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$, which mimics the inner product of two elements of $\mathfrak{G}$ without ever directly mapping into $\mathfrak{G}$, see Figure 4.2a. In fact, if the kernel $\mathcal{K}$ is positive-definite, known as *Mercer's condition*, it is always guaranteed to mimic an inner product in some auxiliary space $\mathfrak{G}$. This space does not even need to be known, much less the map from $\mathfrak{X}$ into it. To illustrate the idea, consider the quadratic kernel

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^2 = x_1^2 y_1^2 + 2 x_1 x_2 y_1 y_2 + x_2^2 y_2^2 \tag{4.11}$$
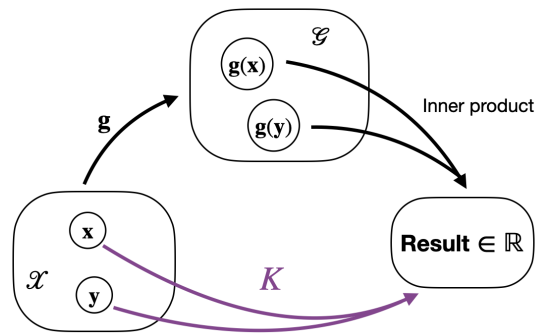
(a) KERNEL SHORTCUT
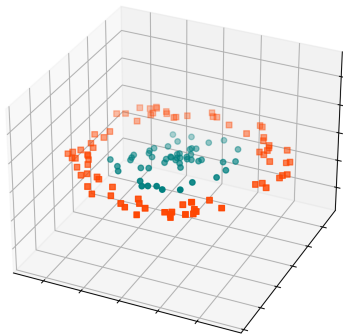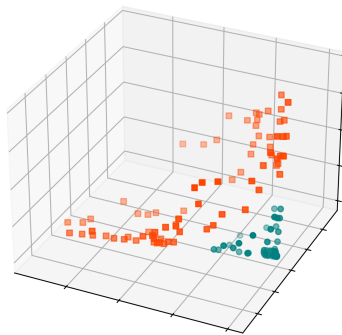
(b) ORIGINAL SPACE $\mathcal{X}$ (c) AUXILIARY SPACE $\mathcal{G}$

Figure 4.2: Illustration of the kernel trick. (a) Shortcut to the result of the inner product taken in $\mathcal{G}$, avoiding explicit computations into and from the auxiliary space. (b) Original two-dimensional data $\mathbf{x} = (x_1, x_2, 0)^\top$, which is not separable by linear SVM. (c) Explicitly mapped samples $\mathbf{g}(\mathbf{x}) = (x_1^2, \sqrt{2}x_1 x_2, x_2^2)^\top$. In the auxiliary space $\mathcal{G}$, the data becomes separable by linear SVM.

where for simplicity two-dimensional input data $(\dim(\mathcal{X}) = 2)$ is considered. It is easily verified, that this kernel mimics the inner product $\mathbf{g}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{y})$ for

$$
\mathbf{g} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1 x_2 \\ x_2^2 \end{pmatrix}. \tag{4.12}
$$

Figures 4.2b and 4.2c visualize the action of this kernel. In case of arbitrary dimension of the input data $\dim(\mathcal{X}) = d$, the quadratic kernel mimics the inner product in an auxiliary space of dimension $\dim(\mathcal{G}) = d(d+1)/2$. Taking the inner product in $\mathcal{G}$ thus implies $\mathcal{O}(d^2)$ floating point operations, whereas evaluating the quadratic kernel directly implies only $\mathcal{O}(d)$. This advantage becomes substantial when input data is high-dimensional, as it is the case in the context of TKSVM where d ranges from hundreds to thousands. More complicated kernels can imply even higher dimensions of $\mathcal{G}$, e. g. simply adding a constant to the quadratic kernel $\mathcal{K}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + c)^2$ already implies an auxiliary space of dimension $\dim(\mathcal{G}) = d(d+1)/2 + d + 1$.

In addition to saving computation cost, finding an appropriate kernel to solve a specific problem is usually easier than finding an appropriate map $\mathbf{g}$. Consider for example, the commonly used *radial basis function kernel*, $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \exp(-\epsilon \|\mathbf{x} - \mathbf{y}\|^2)$. Although Mercer's condition guarantees the existence of $\mathbf{g}$, finding an expression for it is not as simple as for the quadratic kernel. Obviously there is no single choice of a kernel suitable for every classification task, and some kind of intuition or knowledge about the distribution of the data is therefore needed in order to choose a suitable kernel.

4.3   BINARY PHASE CLASSIFICATION

Distinguishing two physical phases of a many-body model can be viewed as a binary classification problem. For this purpose, samples are collected and labelled according to their phase. The decision function assigns any test sample to one of the phases based on physical quantities manifest in the training data. Unsurprisingly, the success of this approach relies on the choice of a suitable kernel. In the earliest application of SVM for phase classification [97], a quadratic kernel is used to classify the ordered against the disordered phase of several Ising-type models. Since the decision boundary represents the phase transition, it contains the same information as the order parameter. Indeed, the *analytical expression* for the order parameter is directly encoded in the decision function. Beyond linear orders such as the magnetization of Ising-type models, a quadratic kernel is however no longer appropriate and classification fails due to inseparability. This motivated the introduction of a *tensorial kernel*, suitable to capture any orientational order breaking the rotational symmetries of

isotropic space, $O(3)$, to one of its many subgroups [87]. Besides order parameters, the tensorial kernel also enables the SVM to learn ground state constraints, characteristic of more exotic phases such as classical spin liquids. The resulting TKSVM framework for classical spin models has been developed and demonstrated in [45, 75] and successfully applied to various models in [43, 76, 103, 104]. Generalizing TKSVM to quantum models, which is the main contribution of this thesis, implies modifications to the tensorial kernel compared to the classical case. The key properties, definition, and modifications of TKSVM are the subject of this section.

INTERPRETABILITY   One key feature of TKSVM is its high degree of *interpretability*. In the context of machine learning, interpretability is the degree to which a human can consistently predict the model's result [59]. In other words, if the decision of the trained model on the class of a test sample can be easily understood by a human, then the interpretability of the model is high. In contrast to many other learning models with sometimes thousands of internal parameters lacking any physical meaning, the decision function of TKSVM always encapsulates physical quantities that can be measured in experiment or simulation.

TENSORIAL KERNEL   The tensorial kernel is generally defined in terms of a local *operator basis* (cf. Section 3.2), which directly impacts the expressibility of TKSVM. In the classical limit of a spin model, $S \to \infty$, the spin operator is usually replaced by a normalized real three-component vector and the site-local operator basis is given by the three classical spin components $\{S^x, S^y, S^z\}$. Meanwhile, for a quantum model of spin-1/2 degrees of freedom, a sensible choice would be the Pauli operators $\{X, Y, Z\}$. After choosing a site-local operator basis $B = \{O^a\}$ of size $|B|$, a map $\boldsymbol{\phi}$ is introduced to construct feature vectors whose components comprise estimators of all possible monomials of a *rank* r within a given *cluster* of size n

$$\boldsymbol{\phi} = \left\{ \left\langle O_{j_1}^{a_1} O_{j_2}^{a_2} \cdots O_{j_r}^{a_r} \right\rangle \right\} \tag{4.13}$$

where $j_1, \ldots, j_r \in \{1, \ldots, n\}$ and $a_1, \ldots, a_r \in \{1, \ldots, |B|\}$. Recall from Chapter 3 that the cluster is a subset of neighboring sites in the lattice. Its size and shape, as well as the rank of the monomials are hyperparameters. Since besides the lattice geometry, no knowledge about the physical model under investigation is assumed, it is generally not clear how to choose the rank and cluster to successfully classify the underlying data. While for the cluster, a guiding principle can be the unit cell of the lattice, or multiples thereof, the optimal rank is typically determined on trial and error basis, starting at rank 1 and increasing systematically.
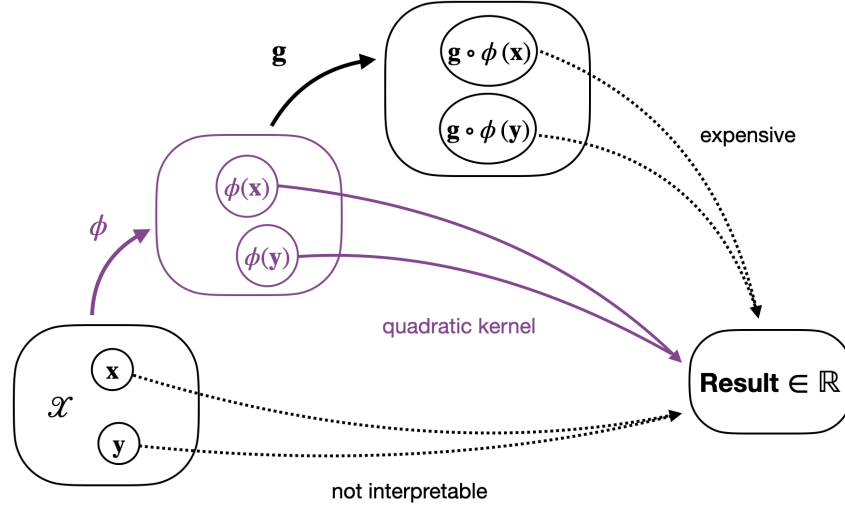
Figure 4.3: Different computational routes to achieve the same result during SVM optimization. With an appropriate choice of rank and cluster, $\boldsymbol{\phi}$ maps the physical samples into a space where the problem becomes separable using a quadratic kernel. The function $\mathbf{g}$ maps the quadratically separable into a linearly separable problem. A possible kernel shortcut from the original space $\mathcal{X}$ to the result, mimicking the inner product in the codomain of $\boldsymbol{\phi}$, would yield a decision function that is not or just weakly interpretable. The route highlighted in purple represents the tensorial kernel and is a good tradeoff between interpretability and computational cost.

The brackets $\langle \cdot \rangle$ in Equation 4.13 represent two very different estimation methods for classical and quantum data. In the classical case, the raw input samples are system configurations obtained from classical Monte Carlo simulations and have the format $\mathbf{x} = \{\mathbf{S}_i\} = \{S_i^a\}$ with $a = x, y, z$ and $i = 1, \ldots, N$ for a N-site system. A single feature vector $\boldsymbol{\phi}$ is computed from a single snapshot $\mathbf{x}$ by the cluster averaging technique (cf. Figure 3.1) without binning. As an example, one specific component $\langle S_1^z \rangle$ of a classical rank 1 feature vector is computed as

$$\langle S_1^z \rangle = \frac{1}{\lfloor N/n \rfloor} \sum_{\text{clusters}} S_1^z. \tag{4.14}$$

In case of a quantum model, the raw input is in the form of POVM samples and a single feature vector is computed by means of shadow tomography, as explained in detail in the previous chapter. For instance, the expression $\langle X_1 \rangle$ represents an estimator of $\text{Tr}(X_1 \rho)$.

The symbols $\mathbf{x}$ *and* $\mathbf{y}$ *may refer to either a single raw sample or a set of raw samples, depending on which averaging technique is used.*

In general, the tensorial kernel is defined as

$$\mathcal{K}\big(\boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{y})\big) = \big[\boldsymbol{\phi}(\mathbf{x}) \cdot \boldsymbol{\phi}(\mathbf{y})\big]^2 \tag{4.15}$$

which is technically a quadratic kernel with respect to the feature vectors $\boldsymbol{\phi}(\mathbf{x})$ and $\boldsymbol{\phi}(\mathbf{y})$. The explicit computation of $\boldsymbol{\phi}$ is necessary to ensure interpretability, see Figure 4.3.

Denote the training data as $\{\mathbf{x}_k\}$ and introduce the component-wise notation $\boldsymbol{\phi}(\mathbf{x}) = \{\phi_\mu(\mathbf{x})\}$ for the feature vector. To retrieve the order parameter, the decision function is reshaped such that terms concerning the training data $\{\mathbf{x}_k\}$ are separated from terms concerning the test data $\mathbf{x}$ to be classified,

$$
\begin{aligned}
D(\mathbf{x}) &= \sum_k \lambda_k y_k \big[\boldsymbol{\phi}(\mathbf{x}_k) \cdot \boldsymbol{\phi}(\mathbf{x})\big]^2 - b \\
&= \sum_k \lambda_k y_k \Big[\sum_\mu \phi_\mu(\mathbf{x}_k)\phi_\mu(\mathbf{x})\Big]^2 - b \\
&= \sum_{\mu\nu} C_{\mu\nu}\phi_\mu(\mathbf{x})\phi_\nu(\mathbf{x}) - b
\end{aligned}
\tag{4.16}
$$

$$
C_{\mu\nu} = \sum_k \lambda_k y_k \phi_\mu(\mathbf{x}_k)\phi_\nu(\mathbf{x}_k).
\tag{4.17}
$$

The resulting *coefficient matrix* $C_{\mu\nu}$ depends only on the training data weighted by the SVM optimization parameters. Its non-vanishing components correspond to the physical features manifest in the training data, based on which the machine predicts the class of the test data $\mathbf{x}$. Therefore, the interpretation of the coefficient matrix directly yields the analytical expression of the underlying order parameter.

DIMENSIONAL REDUCTION   The total number of components of the feature vector from Equation 4.13 equals $(n|B|)^r$. However, some components are redundant and can be eliminated, thereby reducing the dimension of the input data and consequently the runtime of TKSVM. Not only are there less target functions to be computed, but the evaluation of inner products during SVM optimization also benefits from the lower dimension. Since reordering of a given monomial leaves its expectation value unchanged, e. g. $\langle O_1^3 O_2^2 \rangle_x = \langle O_2^2 O_1^3 \rangle_x$, imposing an ordering on the composite indices $(a_1, j_1) \leqslant (a_2, j_2) \leqslant \cdots \leqslant (a_r, j_r)$ during feature vector construction eliminates all redundancies. The resulting feature vector dimension is given by

$$
\dim(\boldsymbol{\phi}(\mathbf{x})) = \binom{n|B| + r - 1}{r} = \frac{(n|B| + r - 1)!}{r!(n|B| - 1)!}.
\tag{4.18}
$$

For classical systems the ordering is not strict, meaning that two index tuples are allowed to be equal. Contractions of spin components on the same site j, such as $\langle S_j^x S_j^y \rangle$, are meaningful in a classical setting. For quantum systems, on the other hand, multiplication of multiple site operators acting on the same site always reduce to another Hermitian operator, $O_j^a O_j^b = O_j'$. Therefore a strict ordering $j_1 < j_2 < \cdots < j_r$ must be imposed on the site indices. The feature vector dimension then becomes

$$
\dim(\boldsymbol{\phi}(\mathbf{x})) = \binom{n}{r} \cdot |B|^r = \frac{n!|B|^r}{r!(n-r)!}
\tag{4.19}
$$

RANK	1	2	3	4	5	6
$(3n)^r$	18	324	5832	104976	$\sim 10^6$	$\sim 10^7$
classical	18	171	1140	5985	26334	100947
quantum	18	135	540	1215	1458	729

Table 4.1: Comparison of feature vector dimensions before and after removing redundancies, using a $n = 6$-site cluster and assuming an operator basis of size $|B| = 3$. Due to the binomial coefficient, the dimension decreases in the quantum case when the rank approaches n.

implying that the rank r cannot be chosen larger than the number of sites in the cluster n. Table 4.1 shows a comparison of the feature vector dimension for different ranks.

CLASSICAL EXAMPLE   Consider the classification problem of a classical anti-ferromagnetic phase against a trivial featureless phase for a one dimensional system of N Heisenberg spins. System configurations $\mathbf{x} = \{\mathbf{S}_i\} = \{S_i^a\}$ with $i = 1, \ldots, N$ and $a = x, y, z$ are obtained from classical Monte Carlo simulations and serve as training data. The order parameter of the anti-ferromagnetic phase is given by the staggered magnetization $O^{st} = \langle \mathbf{S}_{2i-1} - \mathbf{S}_{2i} \rangle$. A $n = 2$-site cluster of neighboring spins and rank $r = 1$ are an appropriate choice of hyperparameters to captured this order parameter. With this choice, the feature vectors are computed as

$$\boldsymbol{\phi}(\mathbf{x}) = \{\langle S_j^a \rangle\} = \left\{ \frac{1}{\lfloor N/n \rfloor} \sum_{\text{clusters}} S_j^a \right\} \tag{4.20}$$

in the first step of TKSVM. Here the index $j \in \{1, 2\}$ labels the two sites within the chosen cluster. As mentioned earlier, the expectation value in the classical case is computed by taking the average over all non-overlapping 2-site clusters within a single sample of the N-site system. The final output of TKSVM is the coefficient matrix displayed in Figure 4.4, which must be interpreted by a human. To this end, the non-vanishing components of the coefficient matrix are substituted into the decision function (Equation 4.16)

$$\begin{aligned} D(\mathbf{x}) &= \left[\sum_{j,a} \langle S_j^a \rangle^2 - 2 \sum_a \langle S_1^a \rangle \langle S_2^a \rangle \right] - b \\ &= \sum_a \left[\langle S_1^a \rangle - \langle S_2^a \rangle \right]^2 - b \\ &= \langle \| \mathbf{S}_1 - \mathbf{S}_2 \| \rangle^2 - b \\ &= \langle \| O^{st} \| \rangle^2 - b \end{aligned} \tag{4.21}$$
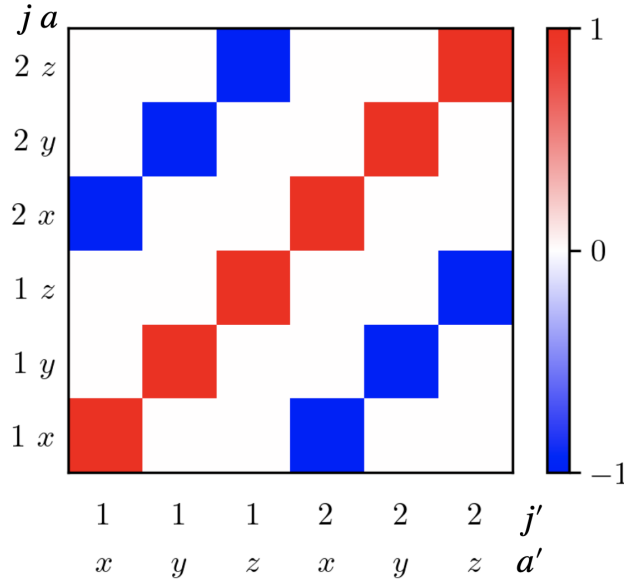
Figure 4.4: Rank 1 coefficient matrix $C_{\mu\nu}$ of the classical antiferromagnetic phase classified against the trivial phase. The axes are labelled by the composite indices $\mu = (a, j)$ and $\nu = (a', j')$ and each pixel represents a contraction $\langle S_j^a \rangle \langle S_{j'}^{a'} \rangle$ weighted by the SVM. The matrix is normalized such that the maximum modulus of its elements equals one.

revealing that the decision function encodes the squared magnitude of the staggered magnetization.

In this example, the importance of the quadratic kernel becomes apparent. Instead of attempting to learn the order parameter O^{st} directly, the quadratic kernel permits TKSVM to learn the squared magnitude $\|O^{st}\|^2$. This seems sensible since the staggered magnetization is not invariant under $O(3)$ rotations, leading to a quadratically but not linearly separable distribution of samples with different orientations. Meanwhile, the squared magnitude is invariant, thus avoiding cancellation. In the absence of such symmetries, however, a linear kernel would also work, and lead to a coefficient vector instead of a matrix.

QUANTUM EXAMPLE   Consider the problem of classifying a spin-1/2 valence bond solid against a trivial featureless phase. On a chain lattice with an even number of sites N, the wave function of the valence bond solid is given by the tensor product of singlets

$$|VBS\rangle = \bigotimes_{i=1}^{N/2} \left[|\uparrow\rangle_{2i-1} |\downarrow\rangle_{2i} - |\downarrow\rangle_{2i-1} |\uparrow\rangle_{2i} \right]. \tag{4.22}$$

At rank 1, classification is impossible due to inseparability of the two phases. In this case the coefficient matrix has random entries and cannot be interpreted. The minimal choice of hyperparameters is therefore $r = 2$ and a cluster of size $n = 2$, resulting in the coefficient
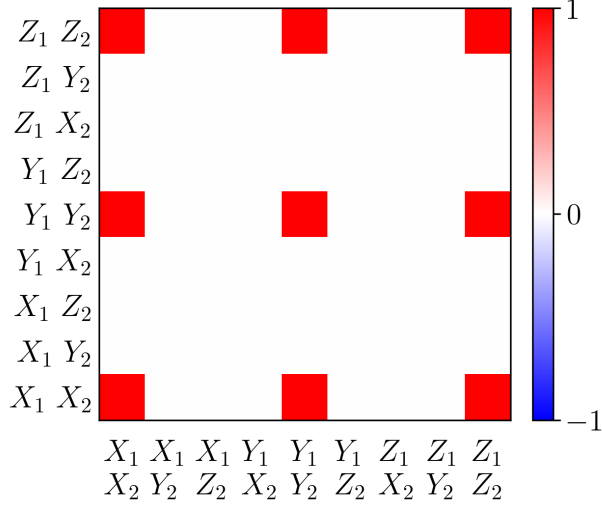
Figure 4.5: Rank 2 coefficient matrix $C_{\mu\nu}$ of a valence bond solid phase classified against a trivial phase, using the Pauli matrices as operator basis. Each pixel represents rank 2 contractions, such as $\langle X_1 X_2 \rangle \langle Y_1 Y_2 \rangle$, weighted by the SVM.

matrix displayed in Figure 4.5. Analogous to the classical example, the non-vanishing coefficients are substituted into the expression for the decision function (Equation 4.16)

$$
\begin{aligned}
D(\mathbf{x}) &= \sum_{\sigma,\sigma'} \langle \sigma_1 \sigma_2 \rangle \langle \sigma'_1 \sigma'_2 \rangle - b \\
&= \langle \|\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2\| \rangle^2 - b \\
&= \langle \|O^{\mathrm{VBS}}\| \rangle^2 - b
\end{aligned}
\tag{4.23}
$$

where $\sigma, \sigma' \in \{X, Y, Z\}$ and $\boldsymbol{\sigma} = (X, Y, Z)^{\mathsf{T}}$, to retrieve the underlying correlations.

4.4 MULTIPLE PHASE CLASSIFICATION

All the classification problems discussed so far are solved in a *supervised* manner. Two datasets are respectively collected from two different physical phases, labelled accordingly, and used to train the learning model. The trained learning model is then interpreted to retrieve physical quantities discriminating the two phases. Supervision of the learning task is equivalent to knowledge of a phase transition *before* training. TKSVM however, is an *unsupervised* learning model. During the training stage, the two datasets are assigned distinct labels *without* knowing whether they belong to different physical phases or not. Rather than the user informing the learning model about the relation between the two training sets, the machine informs the user via the bias parameter b.

THE BIAS CRITERION   In order to illustrate the importance of the bias parameter b of the decision function, the example of the classical anti-ferromagnetic phase classified against a disordered featureless phase shall be revisited. In the resulting decision function, Equation 4.21, the bias parameter was left undetermined. The fact that the samples from the disordered phase lack any kind of features, i.e. all monomials have a vanishing expectation, implicitly fixes the value of b. Specifically, any sample $\mathbf{x}_{\text{dis}}$ from the disordered phase has a vanishing staggered magnetization $\langle \|O^{\text{st}}\| \rangle = 0$, and thus the decision function consists of the bias term only, $D(\mathbf{x}_{\text{dis}}) = -b$. Assuming that the class label of the disordered phase is -1, the ideal support vectors of the disordered phase define the lower margin boundary $D(\mathbf{x}) = -1$, implying $b = 1$. In fact, the sign of the bias relative to the sign of the class label indicates the orientation of the phase transition. Assuming that samples from two phases A and B are labelled $+1$ and -1, respectively, the bias b equals $+1$, if A is ordered and B disordered, and vice versa when $b = -1$. As a consequence, the bias serves as a criterion for the presence of a phase transition. In the presence of a transition it will take a value close to unity, while in the absence of a transition significant violations $b \neq 1$ are expected, see Figure 4.6.

While the discussion here is based on a specific example, there exists a more general argument to support the heuristic bias criterion in [42]. Although the argument still only applies for the scenario of a single symmetry breaking phase transition taking place between A and B, several extensions of the bias criterion to more complex situations were observed and verified empirically. Those extensions are, according to Greitemann [42]:

> First, if one phase possesses two or more orders, while a subset of them vanishes when entering the disordered phase and the remaining ones only diminish in magnitude, $|b|$ will typically be slightly larger than unity, owing to a contribution from the difference in magnitude of the persevering orders. Such behavior can occur when dealing with vestigial orders and partial symmetry breaking.
> Second, if the two sets of samples originate from the same ordered phase and, hence, are characterized in the same way, b can dramatically exceed unity, $|b| \gg 1$. Nevertheless, in those cases, the sign of b retains its physical meaning: A positive $b_{A,B}$ indicates that A is relatively deeper in the ordered phase.
> Third, b can also differ significantly from ± 1 but fall into the interval $(-1, 1)$. This may happen when both sample sets originate from nontrivial phases featuring different characteristics. In that case, even though $C_{\mu\nu}$ can capture the characteristics of both phases, the sign of b will lose its above interpretation. Namely, the TK-SVM can still iden-

(a) Separable

(b) Inseparable
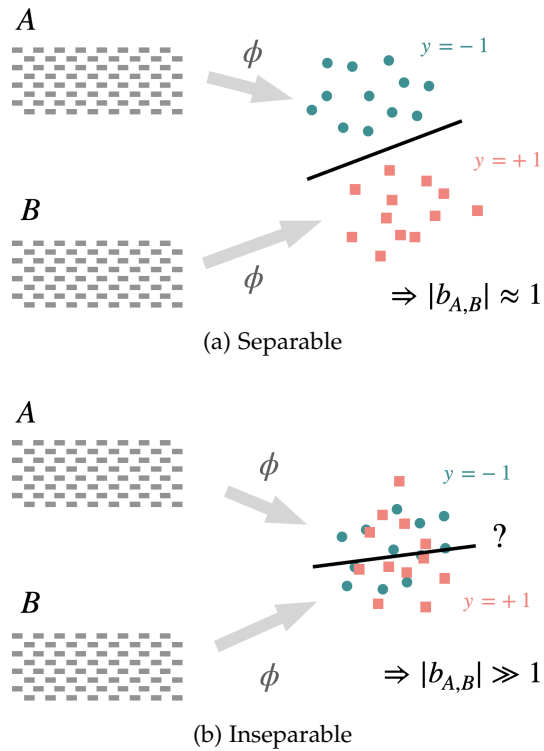
Figure 4.6: Two datasets of POVM samples labeled A and B are mapped to two sets of feature vectors with class labels $y = -1$ and $y = +1$, respectively. (a) If A and B are in different phases, the SVM successfully determines the hyperplane separating the classes in feature space, indicated by the absolute bias being close to unity. The decision function encodes the underlying order parameter. (b) If on the other hand, A and B are part of the same phase, the sets of feature vectors become inseparable and the absolute bias drastically exceeds unity. Because of this inseparability, the hyperplane has no physical meaning.

tify them as distinct phases, but one cannot interpret their relation in terms of a simple order-disorder transition. Lastly, this bias criterion is also applicable for crossovers between phases of different cooperative behavior. Rather than learning the order parameter associated with the broken symmetry of a particular phase transition, the decision function will encode the constraint that governs the cooperative behavior of the spins. Note that in this case the "ordered phase" in the above argument is the one which is more constrained, even though the spins are still disordered in the sense that no symmetry is broken. Again, a bias in $(-1, 1)$ is indicative of a situation where no statement on the relative level of disorder can be made, i. e. both phases exhibit constraints which are mutually incompatible.

In conclusion, the bias parameter is a useful indicator for the relationship between two phase points A and B with class labels $+1$ and -1, respectively:

$$
b_{A,B} \begin{cases} \approx 1, & \text{A in the ordered phase} \\ \approx -1, & \text{B in the ordered phase} \\ \gg 1 \text{ or } \ll -1, & \text{A, B in the same phase} \\ \in (-1, 1) & \text{not directly comparable.} \end{cases} \tag{4.24}
$$

The bias criterion as discussed here applies to classical models, and has never been benchmarked on quantum models. Considering that the working principle of TKSVM, aside from data acquisition and feature construction methods, is identical for classical and quantum models, the bias criterion can be expected to remain valid. In Chapters 5 and 6 the validity of the bias criterion will be confirmed for some specific quantum models. Although this is not a proof, it is an important step towards the empirical verification similar to the classical case.

PHASE DIAGRAM TOPOLOGY   Consider a physical model with free parameters such as interaction strength, external field strength, or temperature. Furthermore assume training data has been sampled over a fixed grid of points in the space of model parameters, i. e. over a grid of phase points. A naive application of TKSVM would entail binary classification between randomly selected pairs of phase points, thereby learning whether or not the pair of points belong to the same physical phase and if they don't, what physical quantity discriminates between them. This would lead to a multitude of redundant results, because usually the number of distinct phases is much lower than the number of sampled phase points. Fortunately, due to the bias criterion,

it is possible to avoid these redundancies by systematically locating approximate phase boundaries through the grid of phase points. With the rough phase boundaries determined, a single representative phase point may be selected in each phase, and subsequently used for interpretation of the coefficient matrix by the human user, thereby saving the effort of interpreting similar coefficient matrices many times over. As an alternative to the selection of a single representative phase point per phase, the whole set of phase points identified as a single phase may be merged and treated as one single dataset.

The determination of approximate phase boundaries is achieved by analyzing a weighted undirected graph constructed using the bias criterion. Vertices of the graph correspond to phase points whereas the weighted edges are given by the normalized biases resulting from binary classification between each pair of phase points. Since there is no necessity for the edges to be directed, it is sufficient to consider the absolute of the bias, and the criterion from Equation 4.24 simplifies to

$$
|b_{A,B}| \begin{cases} \gg 1 & \text{if } A, B \text{ in the same phase,} \\ \approx 1 & \text{if } A, B \text{ in different phases.} \end{cases} \tag{4.25}
$$

This immediately implies that regions of the graph that are highly connected, meaning the edges have a large weight, represent the same physical phase. Meanwhile, at the boundary of two phases, the vertices are weakly connected. Therefore, partitioning the graph yields the approximate layout of the phase diagram, see Figure 4.7a.

For technical reasons, when assigning the graph's edge weights, the absolute biases are normalized using a Lorentzian function

$$
w(b) = 1 - \frac{b_c^2}{(|b| - 1)^2 + b_c^2} \in [0, 1) \tag{4.26}
$$

where the parameter b_c sets a characteristic scale quantifying the condition "$\gg 1$" in Equation 4.25. Typically the choice of b_c is uncritical, and the partition of the graph is robust against varying b_c over several orders of magnitude [43].

SPECTRAL GRAPH PARTITIONING   To conclude this chapter, the method used to partition the graph shall be discussed. There are many standard methods to partition a graph, but TKSVM has the strict restriction that the the final number of phases, i.e. parts of the graph, is not known beforehand. The method employed is the *spectral graph partitioning* technique introduced by Fiedler [37, 38].

A weighted graph $G = (V, E, w)$ is defined as the tuple of a set of vertices V, a set of edges E and a set of normalized weights w on the edges. All graphs produced by TKSVM are of a restricted class of undirected graphs with no edges that connect a vertex to itself and no multiple edges connecting the same two vertices. The goal is to find a two-way partition $V = V_1 \cup V_2$ such that $V_1 \cap V_2 = \emptyset$ and $|V_1| \approx |V_2|$.
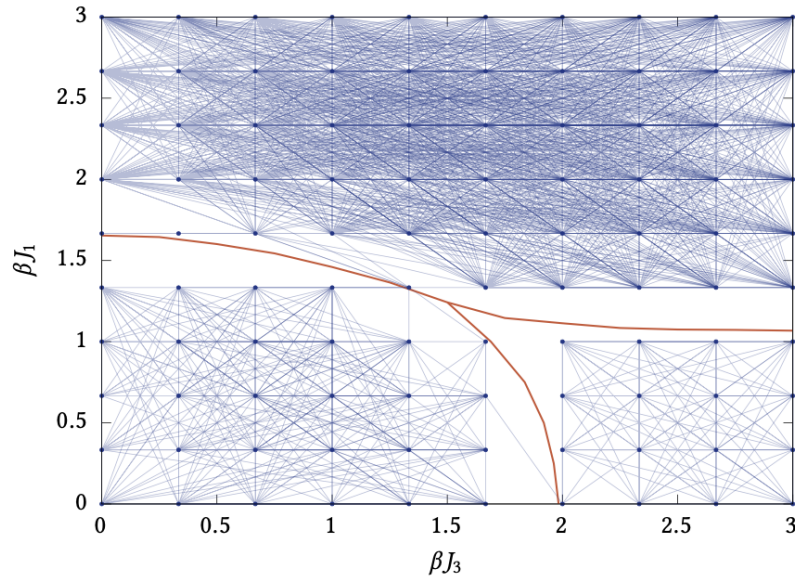
(a) GRAPH CONSTRUCTION
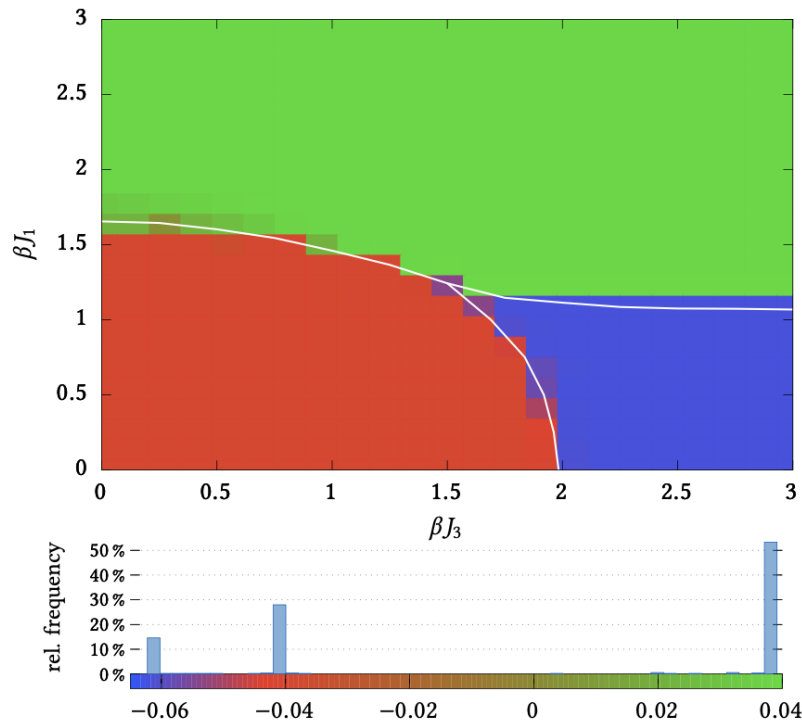
(b) GRAPH PARTITION

Figure 4.7: Exemplary graph construction and partition for phase classification of a classical spin model with two parameters. Figure and caption are adapted from [42]. (a) Graph representation of the relation between points on a 10×10 grid as inferred from the bias b of the corresponding SVM classifiers. The phase boundaries are intersected by only few graph edges and this happens mostly where grid points are very close to the phase transition. (b) Upper panel: Elements of the Fiedler vector corresponding to the graph obtained for a 23×23 grid. Each element is rendered as a pixel at the location of its corresponding grid point in parameter space and color-coded according to its value. Lower panel: A histogram of the elements' values is shown next to the color scale used to encode them in the upper panel.

Additionally, the summed weight of all edges connecting V_1 and V_2, i.e. the connectivity, shall be minimal. Fiedler's theory makes use of the Laplace matrix defined as

$$\Lambda_{i \neq j} = -w_{ij}, \quad \Lambda_{ii} = \sum_j w_{ij} \tag{4.27}$$

where $w_{ij} = 0$ if $(i,j) \notin E$. Further, define a partition vector as $\mathbf{x} = \{x_i = \pm 1\}$ where a positive (negative) sign indicates that the vertex i belongs to V_1 (V_2). For simplicity, assume that the graph is connected, since the treatment of multi-component graphs can always be reduced to separate treatment of the individual disconnected components.

Given an arbitrary partition vector $\mathbf{x}$, the quadratic form $\mathbf{x}^\mathsf{T} \Lambda \mathbf{x}$ equals four times the connectivity

$$\mathbf{x}^\mathsf{T} \Lambda \mathbf{x} = \sum_{(i,j) \in E} w_{ij} (x_i - x_j)^2 = 4 \sum_{(i,j) \in E_{con}} w_{ij} \tag{4.28}$$

where $E_{con} \subset E$ is the set of edges that connect V_1 and V_2. The ojective is to find a partition vector that minimizes the quadratic form while maintaining approximately equally sized parts. By changing from discrete $x_i \in \{-1, 1\}$ to continuous variables $z_i \in [-1, 1]$ and diagonalizing Λ, it can be shown that

$$\min(\mathbf{z} \Lambda \mathbf{z}) = |V| \, \lambda_2 \tag{4.29}$$

Any Laplace matrix has non-negative eigenvalues.

with λ_2 the second smallest eigenvalue of Λ. Thus the argument minimizing the quadratic form is simply the corresponding eigenvector $\mathbf{z}_{(2)}$ of Λ. To recover a discrete solution from the optimal continuous partition vector $\mathbf{z}_{(2)}$, the sign function is applied to each component

$$\mathbf{x}_{min} = \text{sgn}(\mathbf{z}_{(2)}). \tag{4.30}$$

For a general multi- rather than two-way partition of the graph, in place of the sign function applied to the Fiedler vector, clustered appearances of similar entries can indicate more than two parts. Because in applications of TKSVM the final number of parts of the graph is unknown, a multi-way partition and hence a continuous representation of the Fiedler vector is required, see Figure 4.7b.

Part II

APPLICATIONS

LEARNING FROM TRAPPED-ION EXPERIMENTS

Present-day NISQ computers [100] have shown tremendous progress in the coherent control of several qubits via high-fidelity quantum operations and efficient readout. Nonetheless, there are still substantial limitations on the achievable size of the quantum system as well as on the quality of the quantum operations that can be performed. Individual quantum computation platforms exhibit different trade-offs between benchmark metrics such as gate fidelity, coherence time, gate execution times, qubit control and connectivity, and scalability. For computations with few high-quality qubits, trapped-ion quantum computers are well-suited as they exhibit high gate fidelities [51] and universal connectivity [26, 125]. With the fault-tolerant era still years ahead, it is of fundamental interest to identify use-cases of NISQ devices which lead to practical applications in non-tailored problems.

A non-trivial question in the context of phase classification is posed by the discrimination between symmetry-protected topological (SPT) phases with string orders and topologically trivial phases. Addressing this problem on a trapped-ion quantum computer requires preparing such states and performing measurements within current state-of-the-art limits. In this application of TKSVM, quantum data for two distinct families of states is obtained from trapped-ion devices based on qubits [96] and on qutrits [109]. The main purpose is to show that the task is achievable in the NISQ era. In the conducted experiments, the primary limiting factor is the number of gates that can be applied, because the fidelity rapidly degrades when it is increased. Owing to these limitations, this application is centered around MPS models with low bond dimension, because these states can be prepared with a number of gates scaling linearly with system size. This allows to reliably prepare states on the qubit and qutrit platforms with up to eight qubits and five qutrits, respectively.

5.1 INVESTIGATED QUANTUM STATES

Each of the investigated families of quantum states corresponds to a parametrized closed set of states that have a MPS representation with bond dimension 2. One family comprises a subset of ground states of a spin-1/2 cluster model, whereas the other family consists of spin-1 states that are derived from the first family.

FAMILY OF SPIN-1/2 STATES   Consider a set of translation invariant MPS of the form

$$|\psi(g)\rangle \propto \sum_{s_1\ldots s_L} \mathrm{Tr}\left[B_{s_1}(g)\ldots B_{s_L}(g)\right] |s_1 \ldots s_L\rangle \tag{5.1}$$

where L is the number of lattice sites, and g is a real parameter to be defined below. For a spin-1/2 system, $s_j = 0, 1$ labels the spin-1/2 or qubit computational basis states. Explicitly, the MPS tensors for this model [134] are given by

$$B_0 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & g \\ 0 & 0 \end{pmatrix}. \tag{5.2}$$

The MPS interpolates between the cluster state at $g = 1$ and a trivial product state at $g = -1$, without breaking the $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry of the cluster model generated by $X_1 X_3 X_5 \ldots$ and $X_2 X_4 X_6 \ldots$. A phase transition occurs at $g = 0$, separating a SPT phase from a trivial phase [134]. The SPT phase can be characterized by the following string order parameter [94]

Like before X, Y, Z denote the Pauli matrices.

$$\lim_{\ell \to \infty} \left\langle Z_1 Y_2 \left(\prod_{j=3}^{\ell-2} X_j\right) Y_{\ell-1} Z_\ell \right\rangle = \begin{cases} \frac{-4g}{(1-g)^2} & \text{if } g < 0 \\ 0 & \text{if } g \geqslant 0. \end{cases} \tag{5.3}$$

The string order parameter therefore has a non-zero expectation value in the SPT phase ($g < 0$). The MPS state defined by the tensor in Equation 5.2 is the ground state of the Hamiltonian

$$H = \sum_j g_{zxz} Z_j X_{j+1} Z_{j+2} - g_{zz} Z_j Z_{j+1} - g_x X_j \tag{5.4}$$

where the sum runs over lattice sites, and the relation between the MPS parameter $g \in [-1, 1]$ and the Hamiltonian coupling constants is $g_{zxz} = (g-1)^2, g_{zz} = 2(1-g^2), g_x = (1+g)^2$. This Hamiltonian exhibits three phases: A topological phase protected by a $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry, a trivial paramagnetic phase, and a symmetry-broken anti-ferromagnetic phase. The MPS line traverses the former two phases, and its phase transition point corresponds to the tri-critical point where all three phases meet, see Figure 5.1. The simplicity of the MPS parent Hamiltonian is not instrumental to the analysis that follows. In fact, the only reason why states along the MPS line will be considered exclusively, is that they can be prepared via shallow circuits.

FAMILY OF SPIN-1 STATES   In order to define a family of spin-1 quantum states with the same properties as the spin-1/2 model in Equation 5.1, consider an isometry P, which maps two consecutive spin-1/2 degrees of freedom to a single spin-1 degree of freedom

$$\frac{|X_+ X_-\rangle \pm i |X_- X_+\rangle}{\sqrt{2}} \to |\pm\rangle$$

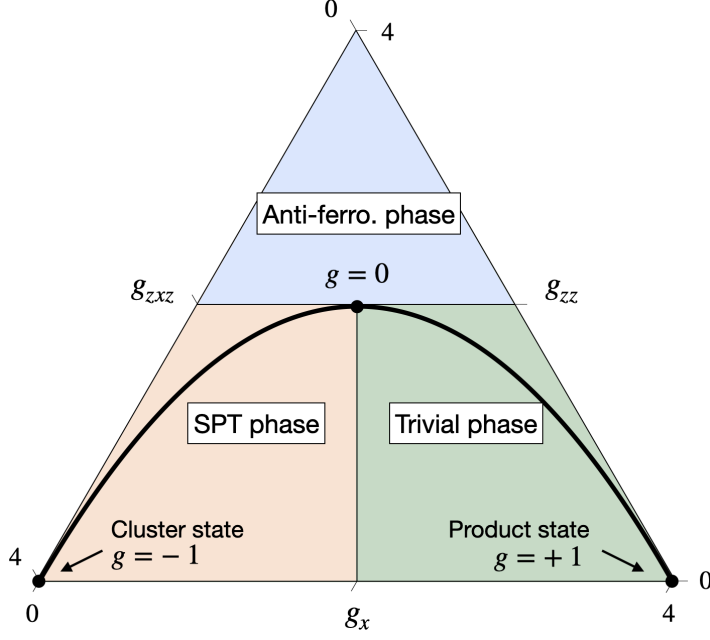$$|X_+ X_+\rangle \to |\circ\rangle \tag{5.5}$$

Figure 5.1: Phase diagram of the three-parameter one dimensional cluster model with a path from the SPT into the trivial para-magnetic phase and through the tri-critical point. Along this path the ground state is an exact MPS in the thermodynamic limit. Reproduced from Smith *et al.* [122].

where $+, \circ, -$ label the spin-1 or qutrit computational basis states, and $|X_\pm\rangle$ are the eigenstates of X. When applied to the cluster state ($g = -1$), this transformation yields the Affleck-Kennedy-Lieb-Tasaki (AKLT) state [3, 130], which is the prototypical example of the same SPT order encoded in the cluster state but realized on a qutrit system. Because of the invariance under SO(3) transformations, its parent Hamiltonian has the simple form

$$H = \sum_j \mathbf{S}_j \cdot \mathbf{S}_{j+1} + \frac{1}{3}(\mathbf{S}_j \cdot \mathbf{S}_{j+1})^2 \tag{5.6}$$

where $\mathbf{S}_j$ denotes the spin-1 vector acting on site j. The map P in Equation 5.5 relates the symmetries of the cluster state to the symmetries of the AKLT state

$$P(X \otimes I) = -e^{i\pi\tau^x}P$$
$$P(I \otimes X) = -e^{i\pi\tau^y}P \tag{5.7}$$

with the spin-1 matrices defined as

$$\tau^x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tau^y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \tau^z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{5.8}$$

This mapping can be extended to the whole MPS line parametrized by Equation 5.2 to obtain a family of spin-1 states with the following explicit MPS tensor

$$
B_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1-g}{2} + i\frac{1+g}{2} & ig \\ -i & \frac{g-1}{2} - i\frac{1+g}{2} \end{pmatrix}
$$
$$
B_\circ = \begin{pmatrix} \frac{1+g}{2} & g \\ 1 & \frac{1+g}{2} \end{pmatrix}
$$
$$
B_- = B_+^*.
$$

(5.9)

For $g = -1$, the standard MPS representation of the AKLT state [114] is recovered via the gauge transformation

$$
\frac{1}{\sqrt{2}} \begin{pmatrix} i & -1 \\ i & 1 \end{pmatrix}.
$$

(5.10)

For $g = 1$, the MPS represents the product state $\bigotimes |\circ\rangle$. For all values of g, the transfer matrix can be diagonalized analytically, and all local quantities can be expressed in closed form. For instance, the string order parameter characterizing the SPT phase reads

$$
\lim_{\ell \to \infty} \left\langle \tau_1^z \prod_{j=2}^{\ell-1} e^{i\pi\tau_j^z} \tau_\ell^z \right\rangle = \begin{cases} -\frac{16}{9} \frac{g}{(1-g)^2} & \text{if } g < 0 \\ 0 & \text{if } g \geqslant 0. \end{cases}
$$

(5.11)

Analogous to the spin-1/2 MPS in Equation 5.2, the string order parameter is non-zero for $g < 0$ and vanishes at $g = 0$, where a phase transition to a trivial phase occurs. For $g \neq -1$ the MPS in Equation 5.9 is not SO(3)-invariant in contrast to the AKLT state. For this reason, its parent Hamiltonian is more intricate than Equation 5.6 and its derivation omitted.

5.2   STATE PREPARATION

The MPS representation allows for easy and resource efficient implementation on gate based quantum computing platforms. Any infinite size translational invariant MPS can be directly transformed into a finite size quantum circuit. This is achieved through a straight-forward procedure pointed out by Barratt *et al.* [6], and schematically depicted in Figure 5.2. The first step consists of transforming the MPS into right-canonical form and interpreting the right-normalized tensor as an isometry from the left virtual space (green leg in Figure 5.2) to the product of physical and right virtual spaces (black and orange legs in Figure 5.2). Subsequently, the isometry is extended to a unitary matrix by addition of an extra leg with the same dimension as the physical space. This leg is equivalent to a dummy index for the MPS tensor when applied to the state $|0\rangle$ in the spin-1/2 case and $|00\rangle$ or $|\circ\rangle$ in the
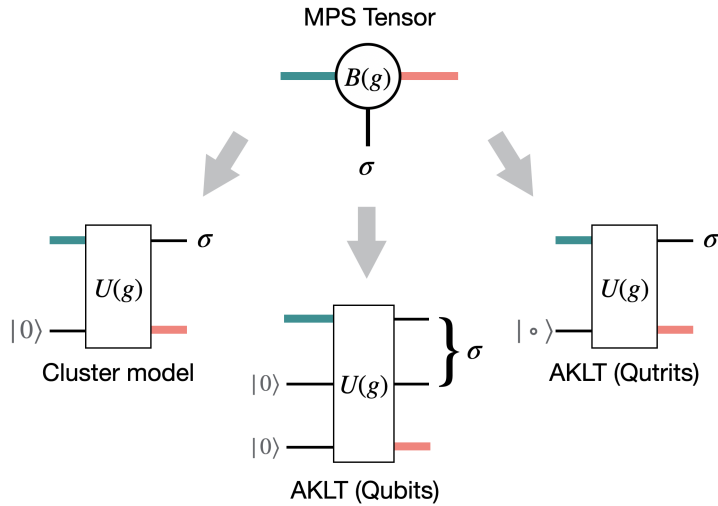
Figure 5.2: The MPS tensor is converted to a 2-qubit unitary in case of the spin-1/2 states. Meanwhile the MPS tensor of the spin-1 states can be translated into either a 3-qubit unitary, or into a 2-qutrit unitary. In the first case, the physical Hilbert space (labelled σ) is too large, while in the second case, the virtual Hilbert space (green and orange legs) is too large.

spin-1 case. But its action on the remaining physical states is undetermined and adjustable to make the gate unitary. The *unitarization* of the isometry is not unique, and can therefore be optimized with respect to the number of native gates necessary to implement the resulting unitary matrix on the quantum computer. Several approaches to solve this task sometimes referred to as *multi-state preparation* exist [56, 126] and have been implemented in publicly available software tools such as *UniversalQCompiler* [57] and *BQSKit* [139]. Both are used in this work to unitarize MPS tensors. To mimic an infinite system, the uppermost unitary U_1 in Figure 5.3 is constructed by unitarization of the non-trivial left eigenvector of the MPS transfer matrix.

For the spin-1/2 family of states, a quantum circuit realization based on qubits is optimal, since the physical dimension of the MPS tensor equals its virtual dimension. For the spin-1 states in Equation 5.9, however, there is a mismatch between physical and virtual dimensions. As a consequence, there are two possible choices to implement the spin-1 states on a quantum computer. Namely, utilizing either qubits or qutrits as computation units, neither of which is optimal. On a qubit platform, the physical Hilbert space of a single spin-1 is emulated by a pair of qubits. The basis states $|00\rangle, |01\rangle, |10\rangle$ are associated with the spin-1 basis states $|\circ\rangle, |+\rangle, |-\rangle$, respectively, while the extra basis state $|11\rangle$ is neglected. On a qutrit platform, on the other hand, the physical dimension of the MPS tensor matches the dimension of a qutrit, but the virtual Hilbert space dimension does not. Therefore the virtual Hilbert space is embedded into the subspace spanned by the qutrit basis states $|+\rangle, |-\rangle$. The dimensions of these different implementations are listed

Dimension	physical	virtual
Spin 1/2 MPS tensor	2	2
Spin 1/2 Qubit unitary	2	2
Spin 1 MPS tensor	3	2
Spin 1 qubit unitary	4	2
Spin 1 qutrit unitary	3	3

Table 5.1: Hilbert space dimensions for the MPS tensor and corresponding implementations as unitaries.

in Table 5.1.

To illustrate the unitarization procedure, consider the phase point $g = -1$ in the spin-1 family of states, i.e. the AKLT state. After right-normalization, the transformed MPS tensors read

$$\tilde{B}_+ = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \ \tilde{B}_\circ = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ \tilde{B}_- = \tilde{B}_+^* \tag{5.12}$$

and the corresponding three-qubit and two-qutrit isometries acting on quantum states are given by

$$\frac{1}{\sqrt{6}} \begin{pmatrix} 0 & \sqrt{2} \\ 1 & i \\ 1 & -i \\ 0 & 0 \\ -\sqrt{2} & 0 \\ i & -1 \\ -i & -1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{6}} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \sqrt{2} \\ 1 & i \\ 1 & -i \\ -\sqrt{2} & 0 \\ i & -1 \\ -i & -1 \end{pmatrix} \tag{5.13}$$

respectively. Note that different permutations of rows are possible, depending on the bit/trit ordering convention used. In the convention chosen here, the first and second row of $\tilde{B}_\circ$ become the first and fifth row of the qubit unitary, and the fourth and seventh row of the qutrit unitary. Vanishing rows indicate the excess dimension of the physical space in case of the qubit isometry, where the state $|11\rangle$ has no physical meaning. On the other hand, the vanishing rows of the qutrit isometry arise because of the excess dimension of the virtual space. Decomposing the three-qubit unitary in terms of the standard universal gate set consisting of general three parameter single qubit rotations and the CNOT gate by means of the *column-by-column* decomposition scheme [56] yields the circuit depicted in

(a) Spin-1/2 family

(b) Spin-1 family (Qubits)

(c) Spin-1 family (Qutrits)

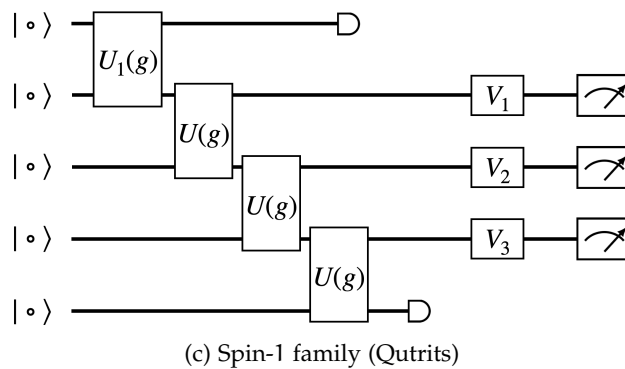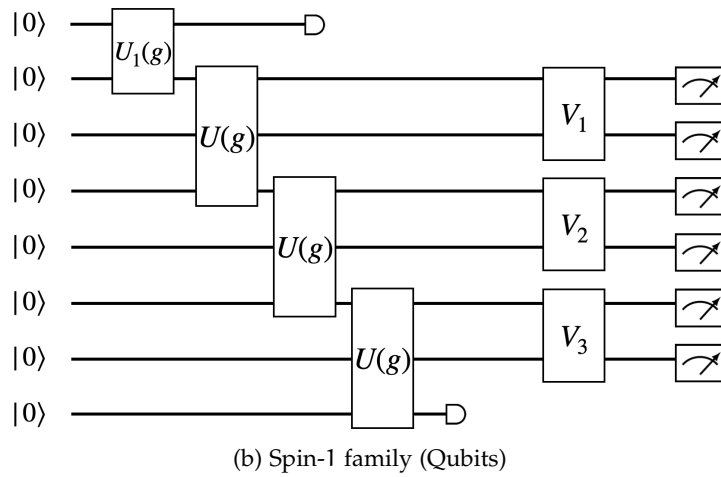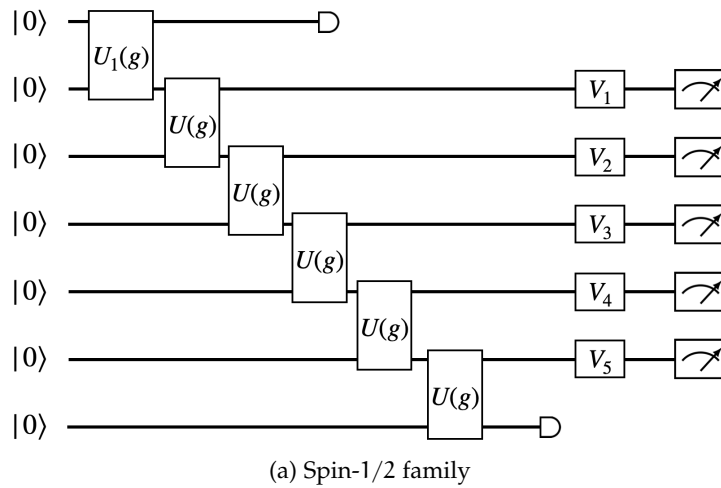Figure 5.3: Quantum circuits with randomized measurements for trapped-ion experiments. Infinite size MPS are represented as finite size quantum circuits at the cost of two ancilla qubits/qutrits that are not measured. The unitaries V_i are chosen randomly and rotate the physical degree of freedom into one of the MUB, hence each configuration of the V-unitaries represents a distinct quantum circuit.
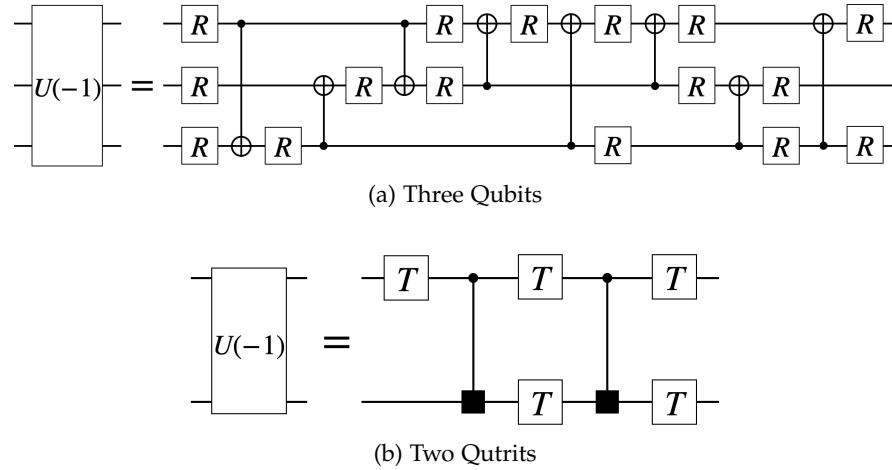
(a) Three Qubits

(b) Two Qutrits

Figure 5.4: Circuit layout resulting from the conversion of an isometry to a full unitary with near-optimal gate count. The gates R and T represent general single qubit and single qutrit rotations, respectively. The entangling two-qutrit CINC gate is represented by a black square on one qutrit and control on another.

Figure 5.4a. Decomposing the three-qutrit unitary naturally requires a different gate set acting on qutrits. Consider the universal gate set consisting of general single qutrit rotations with eight parameters and the entangling CINC gate, which can be seen as a generalization of the CNOT gate to higher dimensions. In dimension $d = 3$ and using the notation $+, \circ, - \equiv 0, 1, 2$, its action is defined as

$$\text{CINC}: \begin{cases} |j, k\rangle \mapsto |j, k\rangle, & \text{if } j < 2 \\ |j, k\rangle \mapsto |j, (k + 1) \bmod 3\rangle, & \text{if } j = 2 \end{cases} \tag{5.14}$$

incrementing the state of the target qutrit, if and only if the control qutrit is in the state $|2\rangle$. For this gate set, the two-qutrit isometry decomposition is obtained by a numerical optimization technique based on *instantiation* of parametrized circuit templates [126]. The resulting optimized circuit is displayed in Figure 5.4b.

MEASUREMENT   In the measurement portion of the circuits succeeding state preparation, informationally complete generalized measurements based on MUB are implemented in a *randomized* manner as discussed in Section 2.3. Recall that generalized measurements constructed from MUB have an important advantage over other types of generalized measurements. Namely, they are a collection of multiple projective measurements, where no projective measurement is favored over another. This is exploited by randomly choosing a basis in which to measure, and then perform a projective measurement in that basis, resulting in a randomized quantum circuit, see Figure 5.3 and compare to Figure 2.4b.

GATE COUNT   Besides the isometries preparing the quantum states, the measurement unitaries V_i also contribute to the total gate count of the circuits. They consist of merely a single general qudit rotation in case of the spin-1/2 states and also in case of the qutrit implementation of the spin-1 states, but require three entangling CNOT gates and several single qubit rotations in case of the qubit implementation of the spin-1 states. Of course V_i can also be the identity, when the measurement is performed in the computational basis, not requiring any gates. For different values of g the gate count can vary. For instance, the environment unitary U_1 for the spin-1/2 family requires a single CNOT gate for $g \leqslant 0$ but two CNOT gates otherwise. The bulk isometry U has two CNOT gates for all values of g. For a system size of 5 physical and 2 ancillary qubits, this amounts to a total of $11-12$ CNOT gates. Meanwhile, the qubit implementation of the spin-1 states has a single CNOT for the environment unitary, 9 CNOT gates for the bulk unitary, and 3 for any measurement unitary that is not the identity, yielding a total of 37 entangling gates for the circuits of Figure 5.3b, if $V_i \neq I \; \forall i$. Only at the exact AKLT point $g = -1$, the bulk unitary can be converted with 8 instead of 9 CNOT gates. Finally, the qutrit realization of the spin-1 family of states yields circuits with a single CINC gate for the environment unitary and $2-3$ CINC gates for the bulk unitary, resulting in a total of either 7 or 10 entangling gates, depending on the value of g. In the experiment, rather than CNOT and CINC gates, the *Møller-Sørensen* entangling gates [124] are used. This implies another compilation step which is achieved by the algorithm described by Maslov [77].

5.3   EXPERIMENT DETAILS

The experiment is performed on two trapped-ion devices. Both devices operate with $^{40}Ca^+$ ions confined in macroscopic, linear Paul (blade) traps. Information is encoded in Zeeman sub-levels $4S_{1/2}$ and $3D_{5/2}$ of the ions, which are connected via an optical quadrupole transition at 729 nm. The optical systems of the respective traps provide single-ion addressing capabilities allowing for arbitrary single-qubit and arbitrary-pair two-qubit gate operations. The native gate set of the setup consists of single-qubit rotations around an axis in the equatorial plane of the Bloch sphere $R(\theta, \phi)$, implemented via laser pulses resonant with the optical transition, single-qubit Z-rotations virtually implemented in software, and maximally-entangling two-qubit gates $XX(\pm\pi/2) = \exp(\mp i\frac{\pi}{4}X \otimes X)$ implemented via the Mølmer-Sørensen interaction [124]. The $XX(\pm\pi/2)$ operation is equivalent to the CNOT gate up to local rotations [77].

Qubit setup: For the qubit experiment, the quantum states are encoded in the magnetic Zeeman sub-levels as $|0\rangle = |4S_{1/2}, m_J = -1/2\rangle$

and $|1\rangle = |3D_{5/2}, m_J = -1/2\rangle$. A detailed setup description can be found in [96, 98]. Measurements are submitted as Qiskit [58] circuits that undergo a custom transpilation procedure. Compared to the qutrit setup, the qubit setup has the benefit of hardware upgrades enabling overall higher-performing operations.

Qutrit setup: For the qutrit experiment, the quantum states are encoded in the magnetic Zeeman sub-levels as $|+\rangle = |4S_{1/2}, m_J = -3/2\rangle$, $|\circ\rangle = |4S_{1/2}, m_J = -1/2\rangle$, and $|-\rangle = |3D_{5/2}, m_J = -1/2\rangle$. The experimental setup used for the qutrit measurements has lower overall gate fidelities and shorter coherence times when compared to the system used for the qubit experiments [96]. Due to geometric constraints, the optical single-ion addressing also results in larger crosstalk between ions and lower coupling strengths to the resonant transitions. Typical gate fidelities can be found in [109]. The qutrit circuits were created using the compiler infrastructure in the python package BQSKit [139].

5.4   PHASE CLASSIFICATION AND CHARACTERIZATION

Before running TKSVM, an operator basis used to express the learned local quantities must be fixed. Throughout this application, the Pauli matrices are used in the spin-1/2 case and the set $\{\tau^x, \tau^y, \tau^z, (\tau^x)^2, (\tau^y)^2, (\tau^z)^2\}$ in the spin-1 case. Neither is complete, but both are sufficient for successful phase classification and characterization of the underlying orders. In this application, TKSVM is used with a linear rather than quadratic kernel, implying that a coefficient vector rather than a matrix is extracted from the decision function.

For each family of states, we consider 21 values of g picked equidistantly in the interval $g \in [-1, 1]$. The path through the cluster model is realized on five physical and two ancillary sites, resulting in $3^5 = 243$ possible measurement configurations, i. e. choices of the V-unitaries in Figure 5.3. In each measurement configuration, 500 projective samples are collected. This results in a total of $121\,500$ samples per value of g. Meanwhile, the spin-1 family of states is realized for three physical and two ancillary spin-1 sites with a physical local Hilbert space dimension $d = 3$, implying $4^3 = 64$ possible measurement configurations. In case of the qubit implementation, where each spin-1 site is emulated by a qubit pair, $2\,000$ projective samples per measurement configuration are collected. The total number of samples is then $128\,000$ per value of g. Due to the imperfect realization of this state in the experiment, however, a fraction of the samples contain the projective qubit pair outcome $|11\rangle$. These samples have no valid spin-1 basis state correspondence and are discarded. The fraction of discarded samples ranges from 14% to 22% depending on g, resulting in an average of about $106\,000$ usable samples per value of g. In case of the the qutrit implementation, on the other hand, $3\,000$ samples per measurement configuration are collected, none of which must be
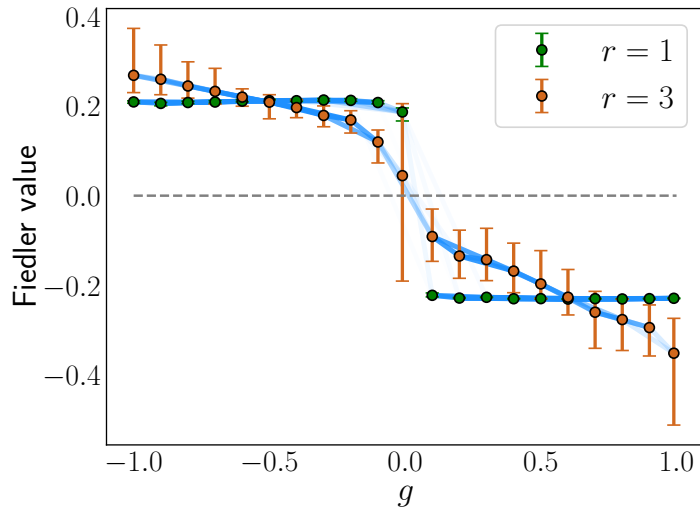
(a) Spin-1/2 family
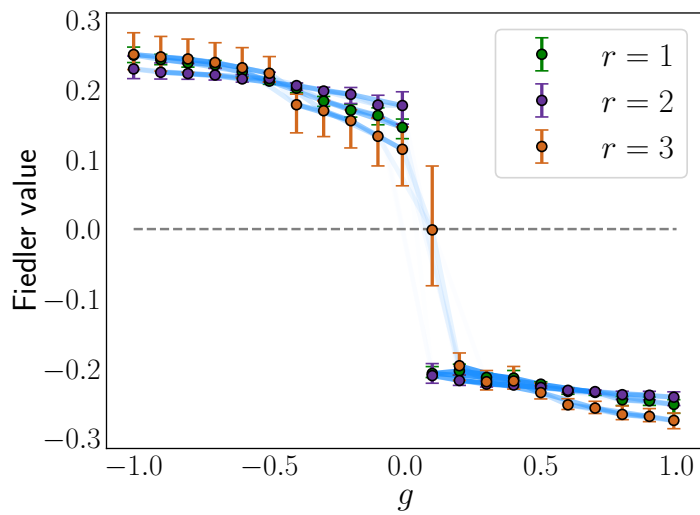
(b) Spin-1 family (qubit platform)

Figure 5.5: Phase classification for MPS paths realized in trapped ions with qubit circuits. Graphs consisting of phase points as vertices and bias parameters as weighted undirected edges are partitioned, resulting in a bi-partition in all cases. The normalized weight of the edges is represented by their intensity. Each vertex has a Fiedler value, whose sign indicates which component of the graph, i.e. which phase, it belongs to. Error bars are obtained by Jackknife resampling [34]. The phase classification result of the spin-1 family realized on the qutrit platform is displayed in Section 5.5
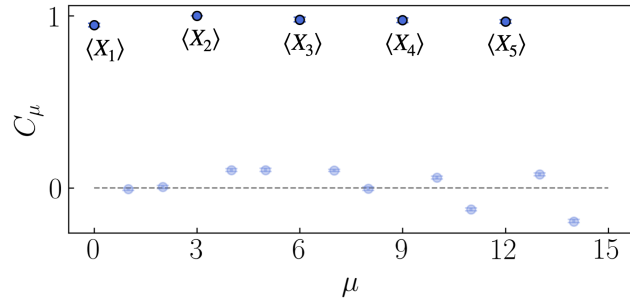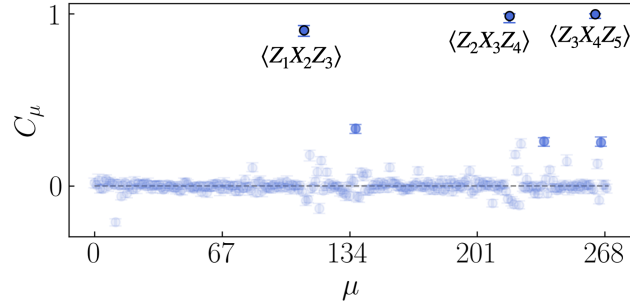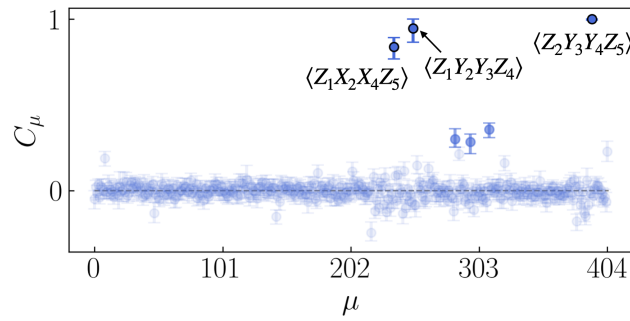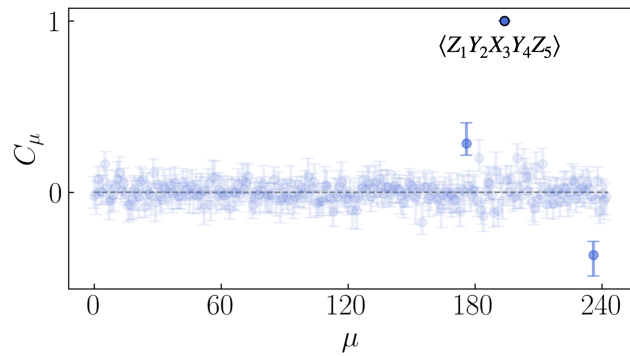
(a) g > 0 rank 1

(b) g < 0 rank 3

(c) g < 0 rank 4

(d) g < 0 rank 5

Figure 5.6: Coefficient vector C_μ for different merged datasets at different ranks. The cluster size is always fixed at $n = 5$ matching the whole system size. Entries with the largest magnitude are highlighted and labeled with the corresponding analytical expression.

discarded. This amounts to $192\,000$ MUB samples per value of g. The resulting machine-learned phase diagrams are displayed in Figure 5.5. As will be shown in the subsequent section, the qubit platform yields slightly more consistent results regarding phase classification of the spin-1 family of states.

The phase classification algorithm, relying on the bias of the decision functions and graph partitioning theory, correctly identifies two different phases, with a transition at $g = 0$. For each model, all datasets which have been identified as part of the same phase are merged, thereby obtaining one dataset for $g < 0$ and one dataset for $g > 0$. This is done to ensure that the features of the phases as a whole are learned, rather than the features of a single point within the phase. Finally, TKSVM is run several times more, classifying the merged datasets against a sets of *random samples*. The random samples consist of uniformly generated projectors onto the MUB states, therefore lacking any kind of order at any rank. Extracting and interpreting the coefficient vector C_μ then yields analytical expressions of local observables, able to discriminate the two phases.

The resulting order parameters discriminating the phases of the spin-1/2 family of states are discussed first. In Figure 5.6a, the corresponding rank 1 coefficient vector for the $g > 0$ merged dataset (trivial paramagnetic phase) is displayed. Within the cluster of size $n = 5$ the same prominent local observable $\langle X \rangle$ is found at each site. Hence, the minimal unit-cell capturing the order is just a single site and choosing the hyper-parameter $n = 1$ would have been sufficient. At higher ranks, the manifest features are just redundant products of the features already observed at rank 1, e.g. $\langle X_1 X_2 \rangle$ or $\langle X_3 X_4 X_5 \rangle$, and therefore not shown. Up to rank 5, no significant features appear that can't be reduced to products of rank 1 features. Therefore the underlying local order characterizing the phase is given by

$$O_{g>0} = X_j. \tag{5.15}$$

Meanwhile, the $r = 1$ decision function for the merged dataset $g < 0$ (SPT phase) has an absolute bias drastically exceeding unity, implying that TKSVM detects no difference to the random samples and classifies the datasets as the same phase. The resulting coefficient vector has no meaning and is therefore not displayed. The same is true at $r = 2$. Increasing the rank further, the bias parameter approaches unity and distinct features become apparent in the coefficient vector, see Figures 5.6b, 5.6c, and 5.6d. The higher rank features can all be understood as non-trivial *overlapping* products of the rank 3 features $G_j := Z_j X_{j+1} Z_{j+2}$. For example, the $r = 5$ feature $\langle Z_1 Y_2 X_3 Y_4 Z_5 \rangle$ equals the overlapping product $\langle G_1 G_2 G_3 \rangle$. The overlap can either span over two sites as in the previous example or a single site as in

(a) $g > 0$ rank 1

(b) $g < 0$ rank 2

(c) $g < 0$ rank 3
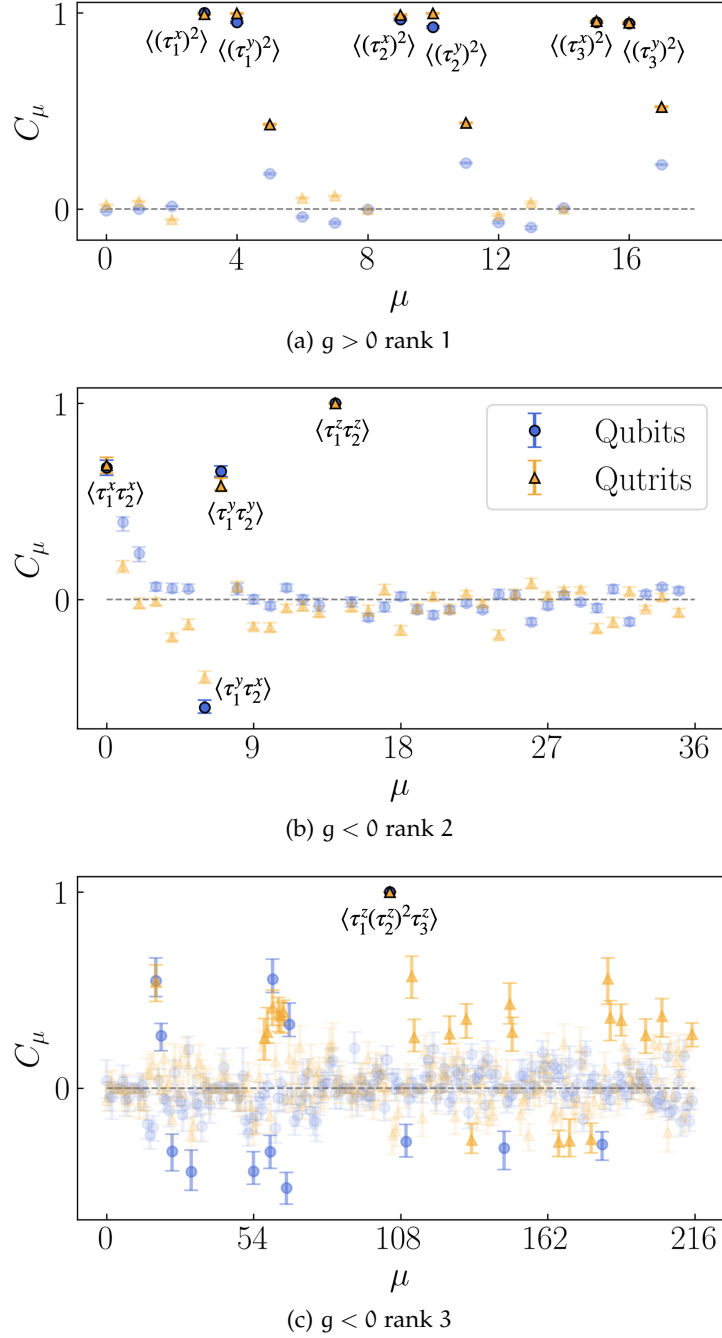
Figure 5.7: Coefficient vector for the spin-1 family of states. For (a) and (c) a cluster size $n = 3$ is used while for (b) the cluster size is $n = 2$. In all cases the qubit and qutrit experiment agree on the most dominant features. Only at rank 3 there is a discrepancy regarding the sub-leading features. An important difference to the spin-1/2 case is that the operator basis for spin-1 is able to produce the identity as $(\tau^x)^2 + (\tau^y)^2 + (\tau^z)^2 = 2I$. The rank 2 features $\langle \tau_1^z \tau_2^z \rangle$ would therefore appear in the rank 3 result disguised as a sum $\sum_{a=x,y,z} \langle \tau_1^z \tau_2^z (\tau_3^a)^2 \rangle = 2 \langle \tau_1^z \tau_2^z I_3 \rangle$. These redundant features are masked automatically to simplify interpretation.

$\langle G_1 G_3 \rangle \propto \langle Z_1 X_2 X_4 Z_5 \rangle$. Extrapolating these observations to higher ranks, the form of the underlying order can be inferred as

$$O^2_{g<0} = \prod_{i=j}^{L-2} G_j \propto Z_1 Y_2 \left[\prod_{j=3}^{L-2} X_j \right] Y_{L-1} Z_L, \tag{5.16}$$

for uniform two-site overlap, reproducing Equation 5.3, and as

$$O^1_{g<0} = \prod_{j=1}^{(L-1)/2} G_{2j-1} \propto Z_1 \left[\prod_{j=2}^{(L-1)/2} X_{2j} \right] Z_L, \tag{5.17}$$

for uniform single-site overlap, assuming L is odd. Besides these two cases, any mix between one-site and two-site overlaps are also learned by TKSVM.

Analyzing the spin-1 family of states at $r = 1$, again only the phase $g > 0$ (trivial phase) is distinguishable from the random samples, in which case the coefficient vector displayed in Figure 5.7a is obtained. Analogous to the spin-1/2 states, periodicity is observed with respect to all the sites of the cluster, which implies that the underlying order extends merely over a single site. Hence, the order parameter of the trivial phase can be inferred to be

$$O_{g>0} = \left[(\tau^x_j)^2 + (\tau^y_j)^2 \right]. \tag{5.18}$$

As expected for the trivial phase, the predominant features at higher ranks are always reducible to products of the rank 1 features. This indicates the proximity to a product state throughout the phase, and the product state saturating the order parameter is precisely the state $\bigotimes |\circ\rangle$ of the $g = 1$ limit.

In the phase $g < 0$, the dominant features $\langle \tau^z_1 \tau^z_2 \rangle$ are observed at $r = 2$ and $\langle \tau^z_1 (\tau^z_2)^2 \tau^z_3 \rangle$ at $r = 3$, see Figures 5.7b and 5.7c, respectively. Extrapolating these observations, the string order parameter for the SPT phase

$$O_{g<0} = \tau^z_1 \left[\prod_{j=2}^{(L-1)} (\tau^z_j)^2 \right] \tau^z_L \tag{5.19}$$

can be inferred. In contrast to the spin-1/2 case, not all the elements of the spin-1 operator basis are unitary. Therefore, it is crucial to normalize the string operator to obtain a non-vanishing string order in the limit $L \to \infty$. Specifically, the operator $(\tau^z)^2$ must be normalized as

$$\mathcal{N} \int d\psi \, \langle \psi | (\tau^z)^2 | \psi \rangle = 1 \tag{5.20}$$

where the states $|\psi\rangle$ are Haar-distributed. After rescaling with the resulting normalization factor $\mathcal{N} = 3/2$, the expectation value of the
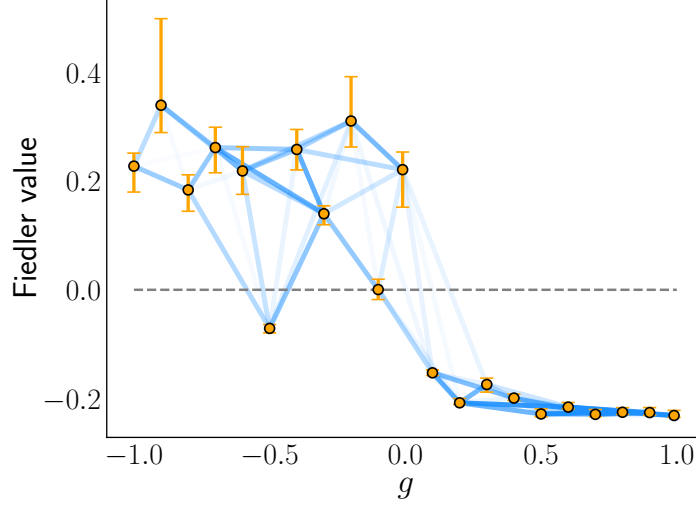
Figure 5.8: Rank 1 phase classification based on qutrit data. The sign of the Fiedler value indicates to which part of the graph each vertex belongs. Compared to the strongly connected part corresponding to the trivial para-magnetic phase, the part corresponding to the SPT phase is less connected.

string operator matches the expectation value of the well known AKLT string order parameter

$$\langle\psi(g)|\tau_1^z\left[\prod_{j=2}^{(L-1)}\frac{3}{2}(\tau_j^z)^2\right]\tau_L^z|\psi(g)\rangle =$$

$$(-1)^L\,\langle\psi(g)|\tau_1^z\left[\prod_{j=2}^{(L-1)}e^{i\pi\tau_j^z}\right]\tau_L^z|\psi(g)\rangle \quad (5.21)$$

up to a sign $(-1)^L$, cf. the left-hand side of Equation 5.11. This equation holds because $e^{i\pi\tau^z} = -\frac{3}{2}(\tau^z)^2 + \frac{1}{2}(\tau^x)^2 + \frac{1}{2}(\tau^y)^2$ and the expectation value of all observables containing $(\tau^x)^2$ or $(\tau^y)^2$ vanishes, when expanding the exponential in the right-hand side.

5.5   COMPARISON OF QUBIT AND QUTRIT IMPLEMENTATIONS

In this section, the two different experimental realizations of the spin-1 family of states are compared. Inspecting the respective phase classification results, Figure 5.5b and Figure 5.8, it becomes clear that the qubit implementation yields more consistent results. In case of the qutrit implementation, the trivial phase $g > 0$ is represented by one strongly connected part of the graph, as expected, but the part representing the SPT phase $g < 0$ is less strongly connected than for the qubit implementation. Nonetheless, only a single phase point $g = -0.5$ is misclassified and another one at $g = -0.1$ has an ambiguous Fiedler value close to zero. Hence, the phase classification can still be con-

(a) r = 1

(b) r = 2

(c) r = 3
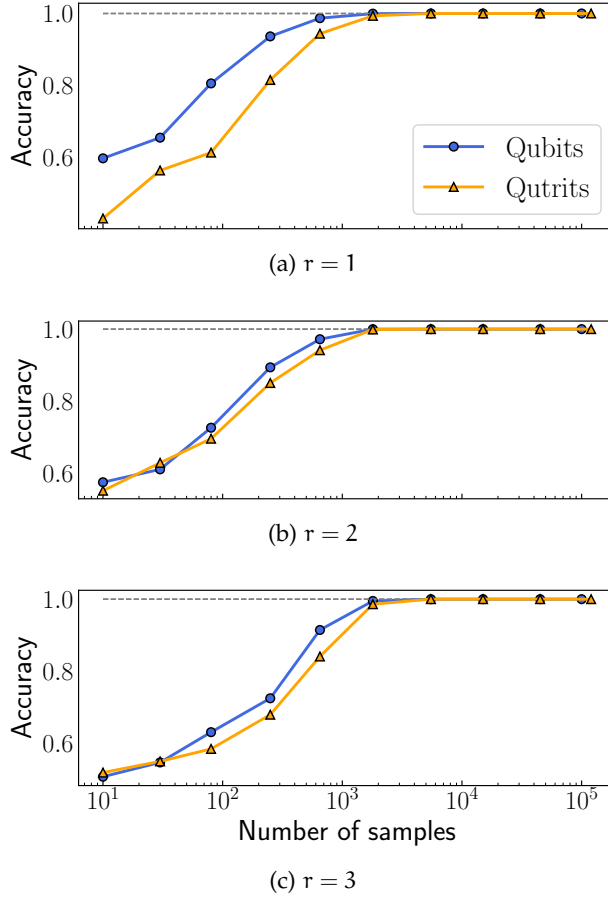
Figure 5.9: Prediction accuracy for the binary classification task between the AKLT state and a product-state. For each classification task, experimental data is used to train the learning model, but simulated data is used to test it. The qubit implementation is slightly more efficient than the qutrit implementation at rank 1.

sidered successful and one can proceed with merging the datasets to extract the characteristic features of each phase. At ranks higher than $r = 1$, however, the phase classification algorithm fails to identify two distinct parts of the graph for the qutrit data (therefore not shown), while it still works for the qubit data.

Although the qutrit implementation yields slightly less consistent results regarding the phase classification, once the datasets are merged according to their identified phases, it performs equally well in regard to identifying the correct local observables characterizing each phase, see Figure 5.7. One possible explanation besides the lower benchmark values of coherence and gate fidelities, could be that the fraction of samples containing the invalid outcome $|11\rangle$ is discarded, effectively reducing the relative amount of noise in the qubit data.

5.6   ACCURACY SCALING

In this final section, the prediction accuracy in dependence of the number of training samples is investigated. Consider the binary classification task for the two datasets that are deepest in each phase of the spin-1 family of states, namely $g = -1$ and $g = +1$. Once the training stage is completed, the TKSVM decision function is used to predict the class of a set of test samples. The fraction of correctly classified test samples is called *accuracy*. These test samples are generated in a simulation, therefore free of any noise that might be manifest in the experimental training set. This approach not only quantifies the minimum number of samples required for successful classification, but also demonstrates the ability of TKSVM to learn the correct decision boundary from noisy data, see Figure 5.9.

Finally, it is important to note that the number of required samples can be drastically reduced if the size of the system is increased. When computing the feature vector components, the average is taken over several independent samples (sample average). Assuming translational invariance, this average can also be taken over different clusters within the same sample, thereby improving the precision to which the features are estimated (cluster average). Considering the spin-1/2 family of states, the experimentally realized system consists of $L = 5$ sites and admits merely a single $n = 5$-site cluster, whereas for the simulation of a system with $L = 72$ sites, there are $L - n + 1 = 68$ overlapping clusters or $\lfloor L/n \rfloor = 14$ non-overlapping clusters.

For details on the sample and cluster averaging techniques, refer to Section 3.1.

As before, consider a binary classification task, but this time classifying the whole SPT phase against the whole trivial phase by merging datasets with $g < 0$ and $g > 0$, respectively. The test set always consists of *simulated* data, whereas for training both simulated and experiment data are used. The results are displayed in Figure 5.10. As expected, the overlapping cluster average reduces the number of required samples by roughly a factor of 68. Moreover, the learning model classifies simulated test samples with comparable accuracy when trained with simulated and experimental data.

5.7   CONCLUSION

This application demonstrates the ability of TKSVM to analyze quantum data generated using state-of-the-art trapped ion quantum computers. In particular, two families of MPS featuring a transition between a symmetry-protected topological phase and a trivial phase were implemented. The first family of states represents a well-studied spin-1/2 system that interpolates between the cluster state and a trivial product state, while the second family of states is a less-explored spin-1 system containing the AKLT state as a paradigmatic instance of SPT order. The learning model correctly distinguishes between the two phases in
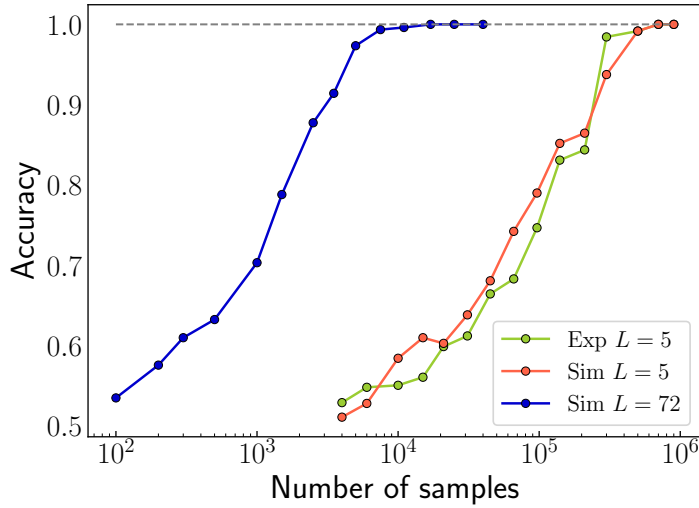
Figure 5.10: TKSVM prediction accuracy for the spin-1/2 family of states. The learning model is trained at rank 5 with cluster size $n = 5$ in all cases. For physical system size $L = 5$, the learning model is trained using simulated as well as experiment data, while for system size $L = 72$ only simulated data is available. The test-sets used to determine the prediction accuracy consists of simulated data in all cases.

both setups in an interpretable way. As a result, the string order parameters characterizing the SPT phase could be successfully extracted. The results highlight the practical utility of TKSVM for the efficient investigation of quantum phase diagrams using NISQ devices, as well as its robustness against the unavoidable experimental imperfections.

Natural extensions include analysis of quantum data generated from adiabatically prepared quantum states in analog quantum simulators. Especially in the context of Rydberg atoms, significant progress with regard to ensuring informationally completeness of measurements was recently reported [11]. In this setting, TKSVM would enlarge the toolbox of available techniques to analyze large-size quantum systems in an unbiased fashion, with potential implications for the solution of long-standing problems not amenable to classical algorithms for many-body systems.

LEARNING FROM SYNTHETIC DATA

A shared central theme of quantum computation and quantum many-body physics is the estimation of a quantum state from a large number of measurements. Standard quantum state tomography quickly becomes infeasible for this task due to the exponential scaling of the many-body Hilbert space [5, 49, 123]. Various strategies, such as matrix product state tomography [28, 70], neural network tomography [18, 127] and randomized measurements [15, 35, 129] have been proposed to improve the efficiency by restricting the target functions to particular information or specific types of quantum states. A noticeable recent advance in efficiency is the development of shadow tomography [1] and the classical shadow scheme [53], which promise to estimate a range of observables accurately from considerably less measurements. Nevertheless, the resources required to achieve a certain accuracy still strongly depend on the entanglement properties of the target state and the complexity of the observable [53, 119]. Therefore, an algorithm like TKSVM, wherein the defining features of an unknown complex quantum state can be detected and returned as an interpretable output, is highly desirable. Those features can then be used to make predictions for new measurements and serve as input for more sophisticated tomography processes. In this chapter, TKSVM is benchmarked by analyzing exclusively synthetic data, thereby lifting limitations imposed by experimental imperfections.

6.1 CLUSTER MODEL WITH TWO PARAMETERS

Consider the spin-1/2 cluster model subject to an external field and two-site Ising interaction

$$H = -\sum_{i=2}^{N-1} Z_{i-1} X_i Z_{i+1} - h_1 \sum_{i=1}^{N} X_i - h_2 \sum_{i=1}^{N-1} X_i X_{i+1} \tag{6.1}$$

where the sums run over the lattice sites of a chain with open boundary conditions, and the external field parameter is non-negative. This Hamiltonian exhibits a trivial paramagnetic, a symmetry-broken, and a SPT phase. The distinct phases of the Hamiltonian can be understood by considering the following limits. For $h_1 \to \infty$, the ground state is a trivial paramagnet with all spins in the eigenstate of X with positive eigenvalue, while for $h_2 \to -\infty$ it is an Ising anti-ferromagnet. The ground state at $h_1 = h_2 = 0$ is known as the cluster state protected by a $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry and constitutes the basis for measurement-based
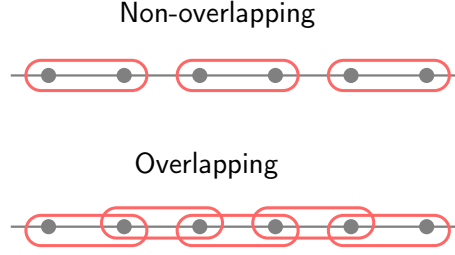
Non-overlapping

Overlapping

Figure 6.1: Cluster averaging technique with clusters of size $n = 2$ for the chain lattice. If the translational symmetry of the quantum state permits, overlapping clusters can be used to achieve higher precision of the cluster average estimate.

quantum computers also known as *one-way quantum computers* [14, 84, 85, 107, 131].

In order to determine the phase diagram of the model, a grid of 17×15 equidistant points in parameter space are considered. At each phase point, the ground state for a system of size $N = 72$ is represented as MPS, obtained using density matrix renormalization group [39, 114]. Sampling the MPS according to Section 2.2, $N_s = 5\,000$ snapshots of the tetrahedral POVM are generated per phase point.

Since the system size is large enough, the cluster averaging technique is used to accurately estimate features at low ranks. At ranks greater than 4, the number of local observables, i.e. the feature vector dimension, becomes too large such that the cluster averaging must be combined with sample averaging to ensure good estimates. For example, the rank 1 estimators for non-overlapping n-site clusters can be computed as (cf. Equation 3.10)

$$\text{tr}(O_j^a \rho) \approx \frac{1}{\lfloor N/n \rfloor} \sum_{k=1}^{\lfloor N/n \rfloor} \text{tr}\big[O_j^a Q^{-1}(M_j^{(k)})\big] \tag{6.2}$$

with O^a an element of the operator basis $B = \{X, Y, Z\}$, and where k labels the different clusters while j labels the site within one cluster. The total number of cluster within a single sample is $\lfloor N/n \rfloor$, if the clusters are non-overlapping, and $N - n + 1$ if the clusters overlap. These scenarios are compared in Figure 6.1.

Assuming no knowledge of the model, it is sensible to begin the phase classification task using the lowest values for the hyperparameters (rank and cluster), capturing simple orders such as magnetization. In general, there is no single choice of hyperparameters, such that all phases of a model can be distinguished simultaneously. Therefore, the phase classification task must be repeated several times until a converged phase diagram is obtained, in the sense that increasing the hyperparameters does not yield an even finer partition of the graph. For the cluster model, the phase diagram is converged at rank $r = 3$ and cluster size $n = 3$, see Figure 6.2. After obtaining the phase diagram, the underlying orders characterizing each phase must
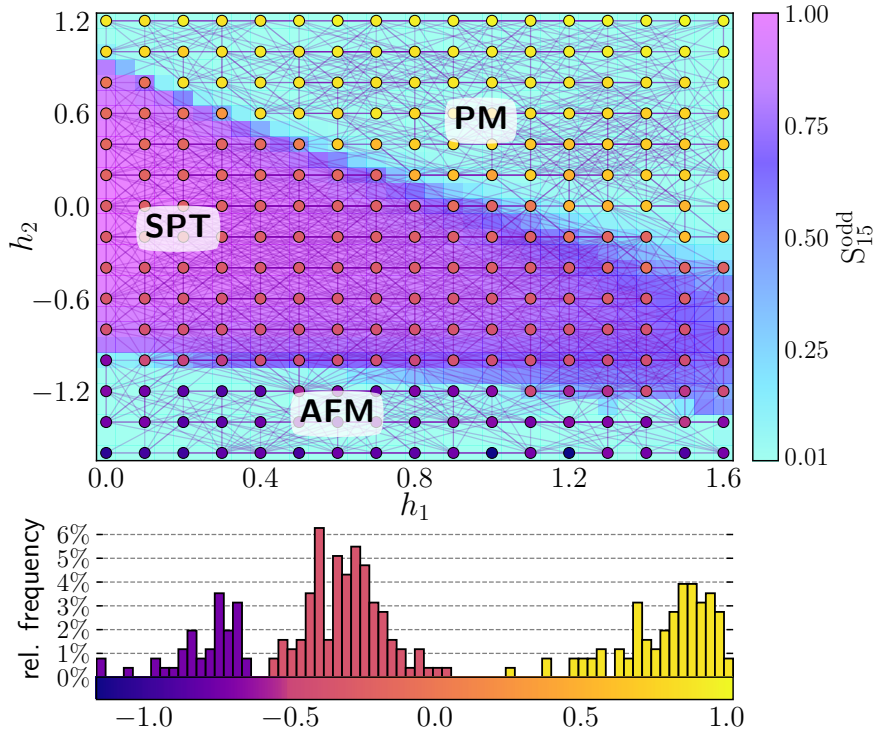
Figure 6.2: Upper panel: Partitioned graph corresponding to the cluster model phase diagram, learned at rank $r = 3$ and cluster size $n = 3$. Partitioning the graph leads to three components, which are identified as the paramagnetic (PM), SPT ordered, and anti-ferromagnetic (AFM) phases. Every phase point is colored according to the appropriate Fiedler vector entry, which indicates its phase correspondence. The background displays the expectation value of the string order parameter S_n^{odd} defined in Equation 6.6 with length $n = 15$, which is finite in the SPT phase but vanishes elsewhere. Lower panel: A histogram of the Fiedler vector entries, whose values are color-coded. The three separated regions correspond to the three phases.

r, n	$\dim(\boldsymbol{\phi})$	$N_{\phi \neq 0}$	Example
3, 3	27	1	$B_2 = Z_1 X_2 Z_3$
3, 6	540	4	$B_4 = Z_3 X_4 Z_5$
4, 5	405	3	$B_3 B_4 = Z_2 Y_3 Y_4 Z_5$
5, 5	243	1	$B_2 B_3 B_4 = -Z_1 Y_2 X_3 Y_4 Z_5$
5, 7	5103	8	$B_2 B_4 B_6 = Z_1 X_2 X_4 X_6 Z_7$
6, 9	61236	30	$B_2 B_4 B_5 B_7 = Z_1 X_2 Y_4 Y_5 X_7 Z_8$
7, 9	78732	35	$B_2 B_3 B_5 B_6 B_8 = Z_1 Y_2 Y_3 Y_5 Y_6 X_8 Z_9$
8, 9	59049	22	$B_2 B_3 B_4 B_6 B_7 B_8 = Z_1 Y_2 X_3 Y_4 Y_6 X_7 Y_8 Z_9$
9, 9	19683	6	$\prod_{j=2}^{8} B_j = -Z_1 Y_2 X_3 X_4 X_5 X_6 X_7 Y_8 Z_9$

Table 6.1: Excerpt of non-trivial features learned at different ranks r and cluster sizes n. The dimension of the feature space is $\dim(\boldsymbol{\phi}) = \binom{n}{r} \times 3^r$. $N_{\phi \neq 0}$ denotes the number of non-vanishing features at fixed r and n.

be extracted. The discussion shall be centered around the SPT phase, because the paramagnetic and anti-ferromagnetic phases only lead to simple rank 1 orders $\langle \sum_{i=1}^{N} X_i \rangle$ and $\langle \sum_{i=1}^{N/2} X_{2i-1} - X_{2i} \rangle$, captured with a minimal cluster size of $n = 1$ and $n = 2$, respectively. The pure cluster model in the limit $h_1 = h_2 = 0$ shall be analyzed in isolation first. Systematically increasing ranks and clusters up to $r, n = 9$, the first non-vanishing features emerges at $n = r = 3$ and take the form

$$B_j := Z_{j-1} X_j Z_{j+1} \tag{6.3}$$

where $j = 2, \dots, n-1$ labels sites within the cluster. Increasing the rank to $r = 4$ and cluster size to $n \geqslant 5$, two additional structures

$$
\begin{aligned}
B_j B_{j+1} &= Z_{j-1} Y_j Y_{j+1} Z_{j+2} \quad j \in \{2, \dots, n-2\} \\
B_j B_{j+2} &= Z_{j-1} X_j X_{j+2} Z_{j+3} \quad j \in \{2, \dots, n-3\}.
\end{aligned}
\tag{6.4}
$$

are detected. In Table 6.1 the features learned for ranks and cluster sizes $r, n \in \{3, \dots, 9\}$ are summarized. It appears that all features learned at any rank and cluster size can be expressed as products of the 3-site building blocks learned at rank 3. Extrapolating $n, r \to \infty$, they reproduce, among many other kinds of string orders, the commonly used *dense* and *odd* string order parameters [122, 131]

The odd string order parameter requires the number of sites to be odd.

$$S_{n \to \infty}^{\text{dense}} = \langle \prod_{j=2}^{n-1} B_j \rangle = (-1)^n \langle Z_1 Y_2 \big(\prod_{j=3}^{n-2} X_j \big) Y_{n-1} Z_n \rangle \tag{6.5}$$

$$S_{n \to \infty}^{\text{odd}} = \langle \prod_{j=1}^{\frac{n-1}{2}} B_{2j} \rangle = \langle Z_1 \big(\prod_{j=2}^{\frac{n-1}{2}} X_{2j} \big) Z_n \rangle. \tag{6.6}$$

It is remarkable, that even though string order is defined in the limit of $n \to \infty$, thus technically a non-local feature, it is still possible to infer it by extrapolation. This is achieved by combining the structure of the learned local variants of the string order, with the observation that at every rank new features appear, which can't be expressed as a trivial combination of lower rank features.

So far, only one specific point in the SPT phase has been considered. In principle some of the observed features might decay as soon as one moves away from the pure cluster state by making h_1 or h_2 finite. To find the order parameter for the whole phase, the standard procedure is to merge all data sets belonging to the SPT according to the machine-learned phase diagram. Note that at this stage, the phase is not yet characterized. The phase can only be identified as SPT ordered, once the string orders are confirmed to be the correct order parameter of the entire phase. The resulting pooled data set contains $96 \times 5\,000$ samples. Classifying the pooled data set against random samples reveals that the same operators as in Table 6.1 remain dominant throughout the phase. Secondary features caused by finite h_1, h_2 are significantly weaker, see Figure 6.3.

After the phase diagram is determined and the phases are characterized, the learning efficiency of the algorithm can be investigated. For classification tasks, the learning success is typically quantified by the *accuracy*, representing the fraction of correctly classified samples from an independent and sufficiently large test-set. In the following discussion, only one specific classification task shall be considered, namely the classification of the pure cluster state ($h_1 = h_2 = 0$) against random samples. The results transfer to other classification tasks, as long as the underlying datasets are truly separable by means of local observables. In case of non-separable datasets, the accuracy will always be close to $1/2$, independent of all parameters.

Figure 6.4a displays the accuracy scaling with respect to the number of clusters used during the training state, for different system sizes. The rank and cluster size, thus the feature vector dimension, remain fixed. Clearly, the accuracy only depends on the total number of clusters and not on the number of samples. A larger system requires less samples than a small one to reach the same total number of clusters and hence to reach the same accuracy.

An accuracy ~ 0.8 is usually enough for a human to interpret the underlying order from the coefficient matrix.

Next, one could investigate the scaling for different feature vector dimensions, i. e. the number of observables, but it is more sensible to consider the rank instead. This is reasonable, because for feature vector computation using classical shadows, the number of observables enters logarithmically, but there is an additional exponential scaling factor 3^r only depending on the rank. Therefore, the rank rather than the feature vector dimension is investigated, always choosing the minimal cluster size (matching the rank). As shown in Figure 6.4b, the required number of samples to reach good accuracy scales exponentially, confirming

(a) CLUSTER STATE $h_1 = h_2 = 0$

(b) POOLED SPT PHASE

Figure 6.3: Non-trivial features in the SPT phase learned at $r = 4, n = 5$. The figures show a prominent column $\bar{v}$ of the associated $C_{\mu\nu}$ matrix. (a) The pure cluster state in the limit $h_1 = h_2 = 0$. Characteristic features are labeled by their analytical expression and correspond to the patterns B_2B_4, B_2B_3, and B_3B_4, cf. Equation 6.4. (b) All phase points belonging to the SPT phase according to the graph partition are pooled and treated as one dataset. The features from the pure limit remain dominant, as they reflect the local correlation structure of the whole phase.
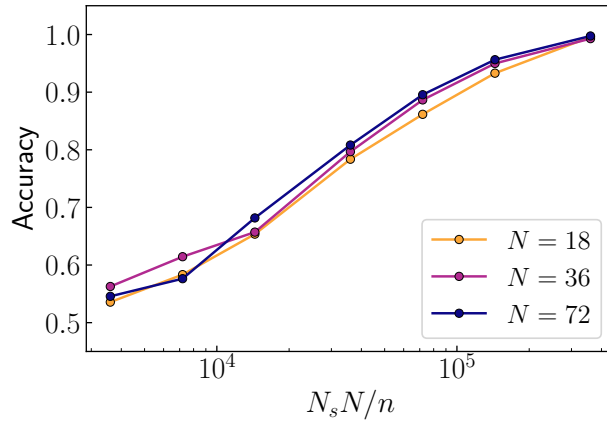
(a) SYSTEM SIZES
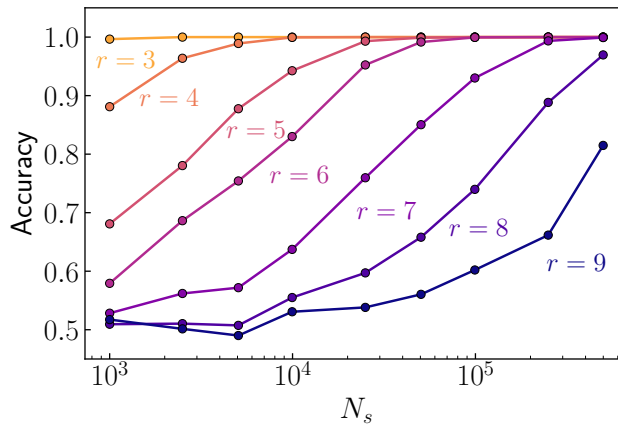
(b) RANKS

Figure 6.4: (a) Dependence of the binary prediction accuracy on the total number of clusters $N_s N/n$ for different system sizes. The hyperparameters used here are $r = n = 5$ and the clusters are overlapping. (b) Accuracy dependence on the number of training samples for different ranks. The system size is held constant at $N = 72$, using overlapping clusters with the size always matching the rank, $n = r$.

the expectation. This has a clear physical implication: in general, the efficiency of an unbiased machine learning algorithm strongly depends on the nature of the phase and the complexity of the target functions. Few samples are typically enough for learning the simplest orders, but complicated orders and entanglement patterns require more. It also reflects an intrinsic limitation for unbiased algorithms to learn arbitrary high-rank quantities and long-range entanglement structures.

6.2   TORIC CODE MODEL WITH EXTERNAL FIELDS

In this section, TKSVM is applied to a toric code model subject to a magnetic field [62] showing that it can distinguish the topologically ordered phase from a trivially disordered phase and identify the stabilizer operators even far away from the zero field limit.

Characterizing long-range entangled topological phases by identifying their anyonic statistics [62, 63] or topological classes [68, 69] in a purely data-driven manner, arguably remains beyond the capabilities of existing machine learning algorithms. Even though it is possible to discriminate topological and trivial phases, in the sense that a learning model can accurately make predictions [54], typically no physical information can be gained due to non-interpretability of the model. Nevertheless, in case of TKSVM, information about local correlation patterns can bring valuable insight into the state and guide further tomography processes specialized for topological features. The Hamiltonian of the toric code model subject to magnetic fields $h_x, h_z > 0$ is given by [62]

$$H = -\sum_{v}\prod_{i\in v} Z_i - \sum_{p}\prod_{i\in p} X_i - h_x \sum_i X_i - h_z \sum_i Z_i. \qquad (6.7)$$

Qubits are located on the links of a square lattice, and v and p denote vertices and plaquettes of the lattice, see Figure 6.5.

This model is topologically ordered for $h_x, h_z \lesssim 0.33$ and possesses two trivial phases at related by symmetry [136], see the phase diagram in Figure 6.6. The training data is produced by exactly diagonalizing the Hamiltonian Equation 6.7 on a system consisting of 18 qubits with periodic boundary conditions. Since the phase transitions from the topological into any of the trivial phases are not fully characterized by a local observable, the phase diagram can't be reliably determined by TKSVM. Therefore specific points are selected and investigated separately. First the pure toric code limit $h_x = h_z \to 0$ is considered. Using $N_s = 1\,000$ snapshots of the POVM based on MUB, the learning model is trained against the same number of random snapshots. A cluster of 2×2 unit cells yields a conclusive coefficient matrix at $r = 4$, and the machine captures two features to be interpreted as the toric code stabilizers $A_v = \prod_{i\in v} Z_i$ and $B_p = \prod_{i\in p} X_i$, as shown

Vertex   Plaquette   8-site (4 unit-cells)
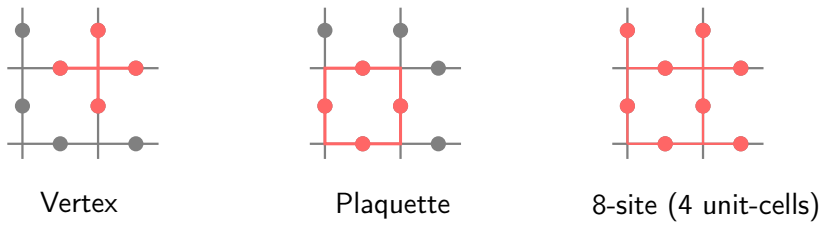
Figure 6.5: Different possible clusters of the square lattice. For an unknown quantum state, natural choices for a cluster include multiple lattice cells and the nearest neighbors of a site. Like for the chain lattice, the clusters may or may not be overlapping by choice. Vertex and plaquette clusters are sufficient to learn the stabilizer constraints $\langle A_v \rangle = 1$ and $\langle B_p \rangle = 1$, respectively. The most unbiased cluster choice when no information about the Hamiltonian is available, however, is a multiple of unit-cells of the lattice such as the 8-site cluster, which is large enough to capture both local constraints simultaneously.
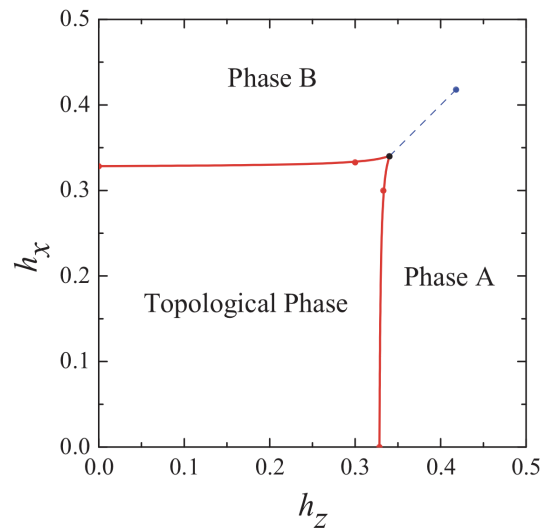
Figure 6.6: Phase diagram of the toric code model subject to magnetic fields, determined by Monte Carlo methods. Reproduced from Wu *et al.* [136].

(a) $h_x = 0, h_z = 0.02$

(b) $h_x = 0.3, h_z = 0.3$

(c) $h_x = 0, h_z = 1.5$
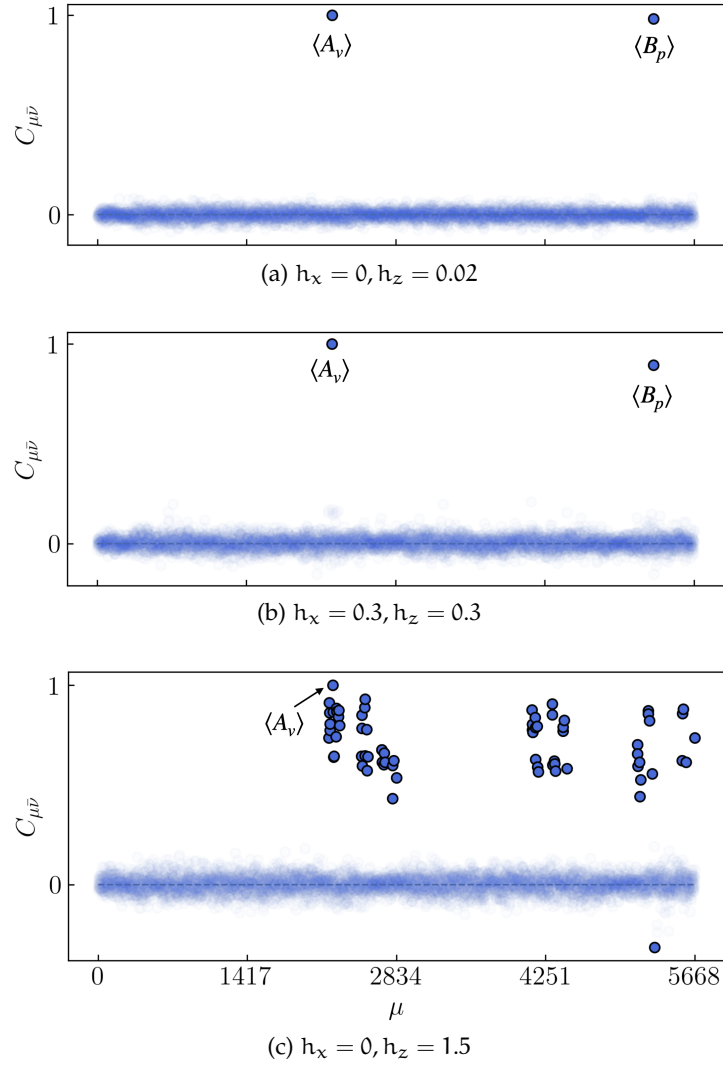
Figure 6.7: Features detected by TK-SVM for the toric code model using a 2×2 unit cell cluster at rank 4. The prominent coefficient matrix column $\bar{\nu}$ is given by $\phi_{\bar{\nu}} = A_\nu$ with ν the only vertex contained in an 8-site cluster. (a) In the zero field limit $h_x, h_z \to 0$, two features representing the vertex and plaquette operators A_ν and B_p are captured. (b) The same features remain prominent even at sizeable magnetic fields $h_x = h_z = 0.3$ close to the phase boundary. (c) In the non-topological phase at $(h_x, h_z) = (0, 1.5)$, multiple trivial features appear due to the strong polarization.

in Figure 6.7. The same procedure is carried out at sizeable fields $(h_x, h_z) = (0.3, 0.3)$, where the same primal features are identified. Thus, the learning model can successfully detect the correct stabilizers from local measurements sampled far away from the pure toric code limit. Meanwhile for a third phase point $(h_x, h_z) = (0, 1.5)$ outside the topological phase, one of the stabilizers $\langle B_p \rangle$ vanishes, while the other one remains finite. Additionally many less meaningful features arise due to the strong polarization. The stabilizers cannot be viewed as order parameters, since one of them always remains finite depending on the phase transition. Crossing into phase A, implying large h_z field, $\langle B_p \rangle$ vanishes as one would expect from a order parameter, but $\langle A_v \rangle$ remains finite. Crossing into phase B, however, the situation is entirely reversed.

Although the stabilizers do not directly characterize the topological phase, their explicit forms indicate the underlying gauge structure and can inspire specialized feature mapping and kernel designs. For instance, one may generate training sets by sampling closed loops [116, 142] or define the kernel of a gauge symmetry [110, 112], which are strategies employed in previous studies of machine learning intrinsic topological orders.

The simulated system considered here is fairly small consisting of 3×3 unit cells yielding a total of 18 sites. Hence the entire system may be used as a cluster with size $n = 18$ in conjunction with sample averaging for feature estimation. For further discussion, let γ represent non-contractible closed loops on the square lattice and let ζ represent non-contractible closed cuts on the dual lattice, see Figure 6.8. In the topological phase the loop and cut operators

$$S_\gamma^z = \prod_{j \in \gamma} Z_j \quad \text{and} \quad S_\zeta^x = \prod_{j \in \zeta} X_j \qquad (6.8)$$

are detected at rank 3, see Figures 6.9 and 6.10. At first it seems surprising, that changing the shape of the cluster reveals features at lower rank than before. But this is simply attributed to the small system size and the fact that the shortest closed loops and cuts that are non-contractible act non-trivially on less sites than the stabilizers.

At higher ranks, products of stabilizers become apparent. For example, the product $\langle A_v B_p \rangle$ is detected at rank 6 if the vertex v and the plaquette p overlap on two sites and at rank 8 if they don't overlap. Since any contractible closed path can be expressed as product of vertex or plaquette operators (depending if they are located on the lattice or its dual), they will all be detected when the rank matches the number of sites that the path acts on. Non-contractible closed paths, on the other hand, can not be expressed as products of stabilizers, and directly probe the topological numbers of the state. Thus in this instance, if the system is small enough to use a cluster containing non-contractible loops, TKSVM is able to detect topological properties. However, the rank required to capture these features must be $\geqslant L$ for
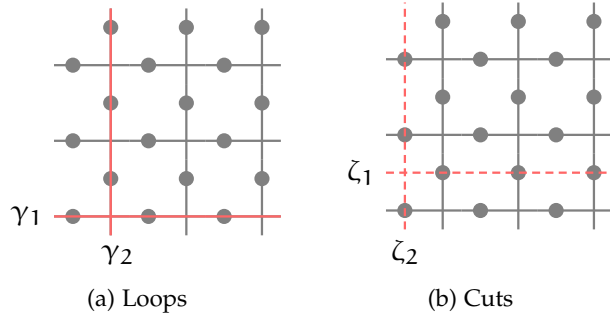
(a) Loops   (b) Cuts

Figure 6.8: Loops and cuts defined on the square lattice and the dual square lattice, respectively. The system has periodic boundary conditions along both spatial dimensions, hence a closed path on the lattice or its dual can wind around the torus either horizontally or vertically.

a system containing $L \times L$ unit cells, implying exponential scaling with respect to L.

6.3 CONCLUSION

Machine learning techniques exhibit growing abilities of analyzing complex classical and quantum data. This application demonstrated the potential of TKSVM as a first-principle method to detect entanglement structures in many-body qubit systems. The phase diagram of the cluster model was constructed from experimentally accessible POVM snapshots without supervision. Furthermore the respective order parameters of the phases were extracted systematically. In particular, the local entanglement patterns and string order parameters of its SPT phase were discussed. Finally, an intrinsic topological phase subject to magnetic fields was examined, and the ability of the learning model to detect the explicit stabilizers relying on data sampled far from the zero field limit was shown.

The results pave the way to investigate membrane-like order parameters in higher dimensional SPT phases [22] and to analyze local measurements of general topological models such as lattice gauge theories [65] and fracton models [132]. In summary, the characteristic local entanglement structures in a quantum phase were identified, without relying on particular known limits or any other information about the Hamiltonian. This marks a crucial difference compared to popular approaches in neural-network based algorithms [10, 27, 67], which make use of special known limits during the training stage.

(a) $h_x = 0.02, h_z = 0$

(b) $h_x = 0, h_z = 0.02$
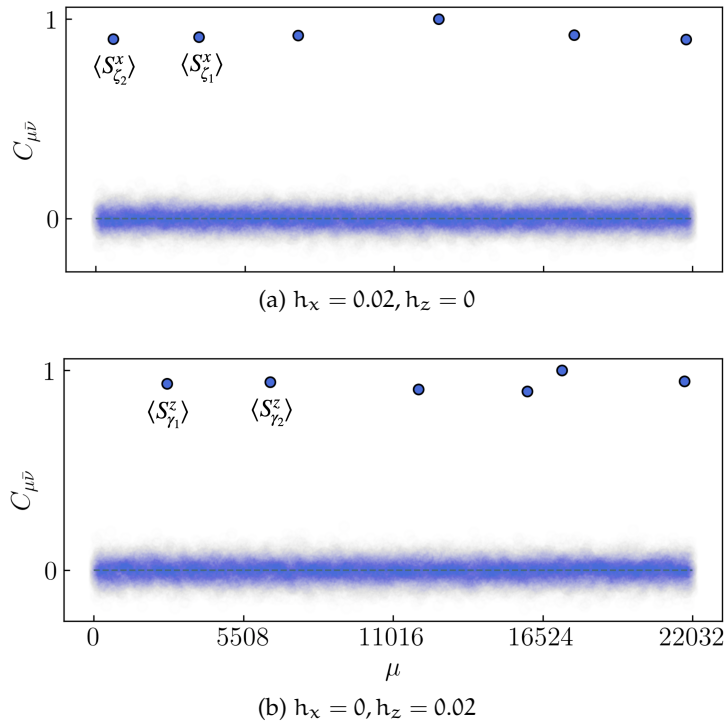
Figure 6.9: Coefficient matrix at rank 3 with a cluster consisting of 3×3 unit cells, i.e. the whole system. Away from the self dual line where $h_x = h_z$, either the operators $\langle S_\zeta^x \rangle$ or $\langle S_\gamma^z \rangle$ are detected, depending on whether $h_x > h_z$ or $h_x < h_z$.

(a) $h_x = 0.02, h_z = 0.02$

(b) $h_x = 0.3, h_z = 0.3$
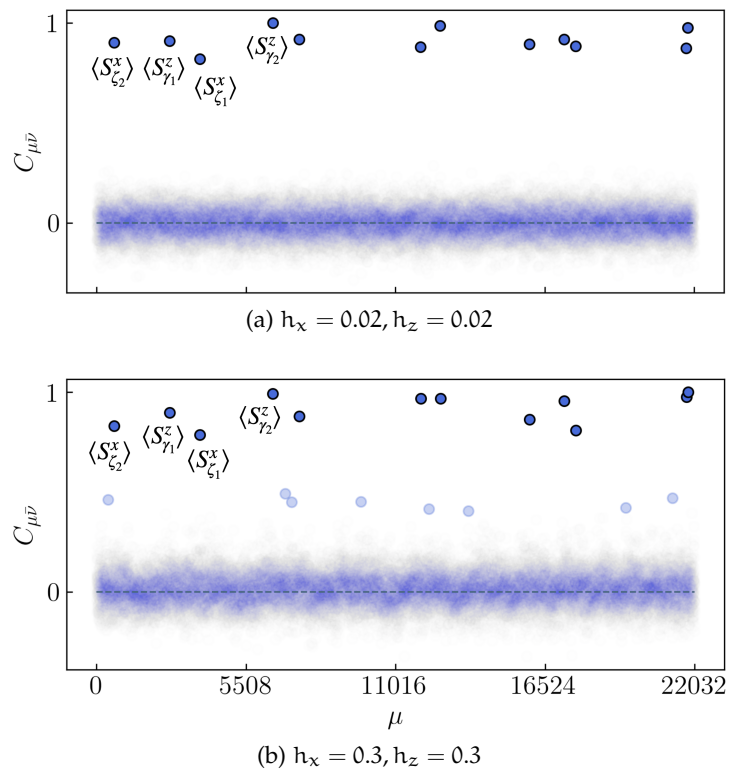
Figure 6.10: On the self dual line where $h_x = h_z$, the operators $\langle S_\gamma^z \rangle$ or $\langle S_\zeta^x \rangle$ are detected simultaneously. For strong fields sub-dominant features become apparent.

CLASSICAL SPIN MODEL ON THE BREATHING PYROCHLORE LATTICE

Classical spin models hosting exotic higher rank gauge fields [99, 101, 105] often lead to emergent topological excitations known as fractons [46, 82, 102]. Any model giving rise to emergent gauge fields or fractons necessarily features competing interactions. This type of interactions causes frustration, impeding the formation of long-range order. Numerical simulations of frustrated models are notoriously difficult, because of their strong tendency to fall out of equilibrium at low temperatures. And even when simulations converge, interpretation of the results can still be challenging. In this chapter, the learning model of TKSVM is used to characterize the low temperature phase of a frustrated spin model defined on the breathing pyrochlore lattice, thereby discovering an intricate and previously unknown form of order.

The spin model in question is the Heisenberg anti-ferromagnet on the breathing pyrochlore lattice with Dzyaloshinskii-Moriya interactions on alternating tetrahedra, which is known to support a rank 2 gauge field theory [137, 141]. Descending from this higher-rank gauge field, Monte Carlo simulations show the onset of an undefined magnetic order at very low temperature $T_c \sim 10^{-3}J$ where J is the antiferromagnetic exchange coupling, see Figure 7.1. The model is essentially a worst-case scenario for Monte Carlo simulations, where a joint heat bath, parallel tempering and over-relaxation algorithm is insufficient to satisfactorily thermalize the ordered phase. This difficulty to thermalize reveals the naturally complex magnetic texture of a higher-rank gauge field. In this situation, TKSVM allows to extract the relevant information out of noisy and incomplete numerical data and, in combination with analytical support, refine the Monte Carlo simulations.

The model is the classical Heisenberg anti-ferromagnet with additional Dzyaloshinskii-Moriya interactions, defined on the breathing pyrochlore lattice visualized in Figure 7.2. Its Hamiltonian reads

$$H = J \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + D \sum_{\langle ij \rangle \in A} \mathbf{d}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j) \tag{7.1}$$

where the sites of an A-tetrahedron are located at

$$\begin{aligned}
\mathbf{r}_0 &= \frac{a}{8}(1,1,1) & \mathbf{r}_1 &= \frac{a}{8}(1,-1,-1) \\
\mathbf{r}_2 &= \frac{a}{8}(-1,1,-1) & \mathbf{r}_3 &= \frac{a}{8}(-1,-1,1)
\end{aligned} \tag{7.2}$$
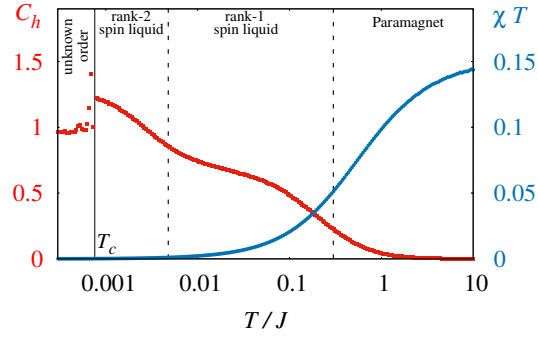
Figure 7.1: The phase diagram of the model presents three distinct regimes before ordering upon cooling [137]: A paramagnet, a rank 1 spin liquid (the Heisenberg anti-ferromagnet) and a rank 2 spin liquid, as illustrated from the evolution of the specific heat C_h and reduced susceptibility χT, computed for a system size of $N = 8\,192$ sites with a fixed the Dzyaloshinskii-Moriya coupling of $D = -0.0141 J$. The ordered phase is separated from paramagnetic fluctuations by three orders of magnitude in temperature. The vertical dashed and solid lines are respectively crossovers and a phase transition. With the aid of TKSVM, the nature of the unknown order below the transition at T_c is determined.

(a) Ground state   (b) Unit cell
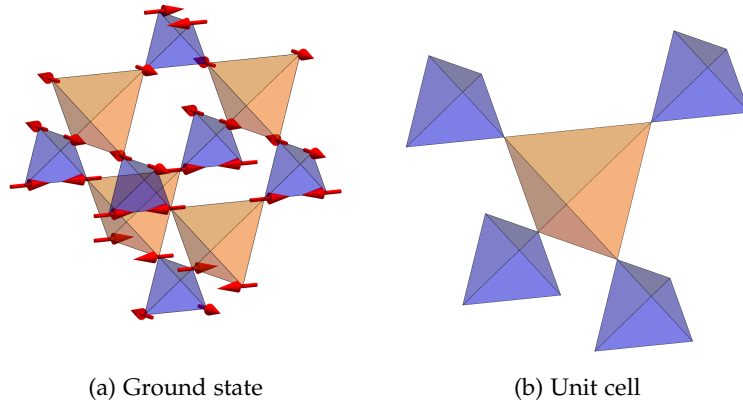
Figure 7.2: (a) Example of a ground-state spin configuration. The breathing pyrochlore lattice is made of two inequivalent types of corner-sharing tetrahedra. The ground state is coplanar and nematic in the sense that it breaks lattice-rotation symmetry, and the ordering mechanism comes from a thermal order-by-disorder selection within the emergent rank 2 U(1) gauge field. (b) The unit cell comprises 16 sites forming four A- and one B-tetrahedron. It forms the minimal unit cell of the face-centred-cubic lattice with respect to the centers of the four A—tetrahedra.

relative to its center, and where a is the length of the face-centered cubic unit cell [111]. Meanwhile, the sites of a B-tetrahedron are located at $-\mathbf{r}_0, -\mathbf{r}_1, -\mathbf{r}_2, -\mathbf{r}_3$ relative to its center. The bond dependent Dzyaloshinskii-Moriya interaction vectors are defined as [16, 66]

$$\mathbf{d}_{01} = \frac{(0,-1,1)}{\sqrt{2}} \quad \mathbf{d}_{02} = \frac{(1,0,-1)}{\sqrt{2}} \quad \mathbf{d}_{03} = \frac{(-1,1,0)}{\sqrt{2}}$$
$$\mathbf{d}_{12} = \frac{(-1,-1,0)}{\sqrt{2}} \quad \mathbf{d}_{13} = \frac{(1,0,1)}{\sqrt{2}} \quad \mathbf{d}_{23} = \frac{(0,-1,-1)}{\sqrt{2}}. \tag{7.3}$$

This model is of interest because of the possibility of realizing an exotic rank 2 $U(1)$ spin liquid for negative $D < 0$ [137]. Furthermore, materials such as $Ba_3Yb_2Zn_5O_{11}$ have parameters in this regime for the A-tetrahedra, although the antiferromagnetic interactions on the B-tetrahedra might be too small [24, 50, 106]. For $D = 0$ the model reduces to the standard Heisenberg antiferromagnet on the pyrochlore lattice, exhibiting the well understood rank 1 classical spin liquid phase which crosses over to a standard paramagnet with increasing temperature. For $D > 0$, the ground state is a simple *all-in all-out* ordered phase at low temperature [137]. However, at $D < 0$ below the rank 2 $U(1)$ spin liquid, the system orders in some manner that is not well understood because of thermalisation issues. The presence of Bragg peaks in the structure factor at finite $\mathbf{q} = W$ wave vector is the only indication of some kind of ordering, possibly co-existing with other phases.

7.1 THERMALIZATION ISSUES

The initial Monte Carlo simulations are performed following the procedure of Reference [137] for system size up to $27\,648$ classical spins normalized as $\|\mathbf{S}\| = 1/2$. Starting from a random spin configuration, the system is annealed to low temperature during 10^6 Monte Carlo steps, then thermalised for another 10^6 Monte Carlo steps, and finally data is collected for statistical averaging during 10^7 Monte Carlo steps. Each step comprises N spin-flip updates via the rejection-free heat bath algorithm and five over-relaxation updates sweeping through the entire lattice. Every 100 Monte Carlo steps, parallel tempering is employed between neighbouring temperatures, with a total of 126 distinct temperatures equally spaced between 0 and 0.0025J. Based on previous experience with classical Monte Carlo simulations, in addition to the generic and powerful over-relaxation and parallel-tempering algorithms, these simulation parameters are an order of magnitude longer in time and bigger in system size than what is usually necessary to completely characterise a typical phase transition in a frustrated magnet. Nonetheless, Figure 7.3 shows that while it is possible to spot the presence of long-range order, the magnetic order cannot be properly thermalised at very low temperatures. As

Over-relaxation is a micro-canonical update with spin rotation around the local molecular field for each spin.

a consequence it is unclear whether the order with Bragg peaks at $\mathbf{q} = W$ is really (part of) the ground state and whether it co-exists with other phases. The reasons for this issue are multiple. On a fundamental level, the problem is the ordering mechanism descending from a higher-rank gauge field. As shown in Reference [137], the rank 2 gauge field itself descends from a rank 1 gauge field with a broader phase space manifold, namely the Coulomb spin liquid of the Heisenberg anti-ferromagnet on pyrochlore lattice. It is the Dzyaloshinskii-Moriya term on the $A-$tetrahedra of the Hamiltonian that selects the rank 2 gauge field. This means that the unknown ordered phase is separated from paramagnetic fluctuations by two successive crossovers into more constrained configurational manifolds. Monte Carlo simulations are thus particularly constrained in phase space around T_c and easily become trapped in local free-energy minima. Usually this is where parallel tempering would help, by shuffling spin configurations across temperatures. But here the issue is not only that the transition temperature is far from paramagnetic fluctuations. The visible jumps in energy and order parameter in Figure 7.3 suggest a first-order transition. This strongly hinders the efficiency of parallel tempering, because the discontinuity in energy essentially prevents spin configurations from crossing the transition temperature. Hence one cannot rely on parallel tempering to help thermalise the ordered phase. In addition, the energy jump in Figure 7.3a is of the order of 10^{-4} J. Such a small energy selection is consistent with the double crossover mentioned above but, keeping in mind the proximity of a highly degenerate spin liquid, it also suggests a competition between multiple phases that are quasi-degenerate in free energy.

Even if not necessarily systematic, thermalization issues arise as relatively natural consequences of higher-rank gauge fields. Being governed by tensorial constraints and a multiplicity of conserved quantities, these exotic phases are inherently complex, and their ordering mechanisms are expected to be unconventional. With these caveats in mind, the goal is to demonstrate how TKSVM facilitates the analysis, taking advantage of its ability to extract useful information out of noisy and incomplete data.

7.2   TETRAHEDRAL ORDER

Without prior knowledge of the phase, it is natural to start at rank 1 to probe potential magnetic orders. Since the dimension of the feature vector grows only linearly with respect to the cluster size n at rank 1 [45], very large clusters consisting of multiple lattice unit cells can be used. Provided a phase is purely magnetic and has a perfect translational symmetry, the rank 1 coefficient matrices learned with different cluster sizes should converge to a regular structure. The magnetic order parameter could then be inferred and justified a
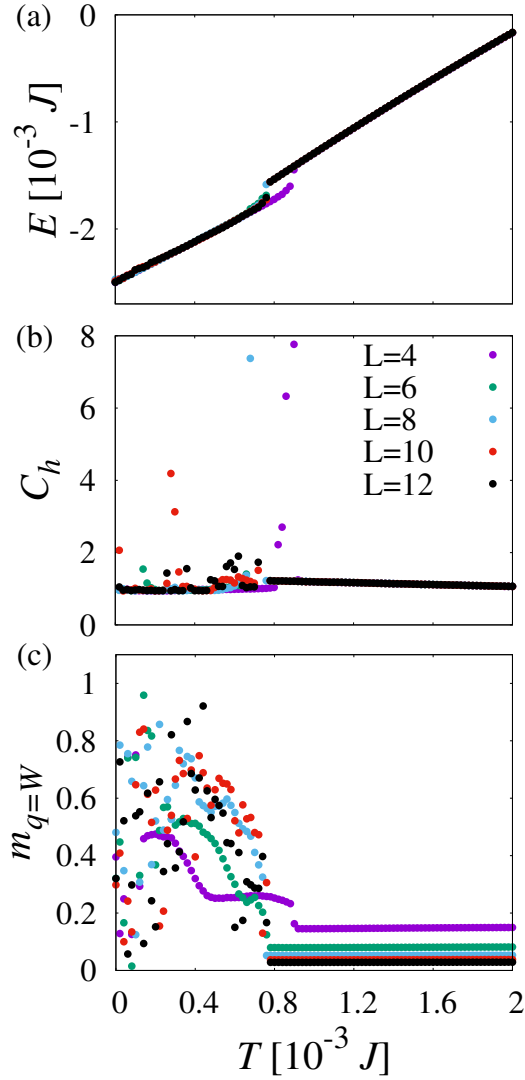
Figure 7.3: Monte Carlo simulations for different system sizes $N = 16L^3$ *before* the machine learning stage show a phase transition at $T_c \approx 8 . 10^{-4} J$ in (a) the energy, (b) the specific heat, and (c) the order parameter for wave vector $\mathbf{q} = W$ order (corners of the Brillouin zone). There are noticeable finite-size effects and thermalisation issues below $T \sim 10^{-3} J$. The energy in panel (a) has been shifted by $J \|\mathbf{S}\|^2 = 1/4$ for convenience.
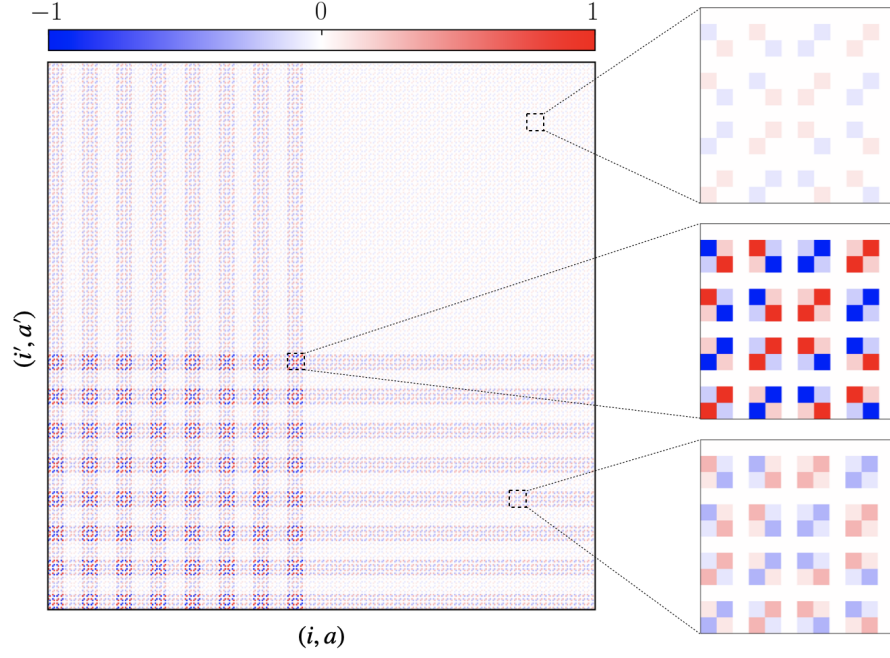
Figure 7.4: Rank 1 coefficient matrix. Each pixel represents the weight of a contraction $\langle S_i^a \rangle \langle S_{i'}^{a'} \rangle$ of features learned by the machine. The indices i and i' range from 0 to 127 labelling the spins within the 8-unit-cell cluster and a, a' label the spin components x, y, z. Magnification of three different spin-contraction blocks reveals their similar structure and different overall weight.

posteriori by measuring it in new Monte Carlo simulations. However, in the unknown phase below T_c, there is no evidence of a stable rank 1 pattern even when using very large clusters of up to 128 sites in Figure 7.4. Instead, the learned patterns display sample-dependent irregular weights that are inconsistent with long-range dipolar order. This suggests that magnetic orders do not reflect the correlations in the system completely, and the data must be analyzed further at rank 2.

The choice of the cluster at rank 2 is also guided by the lattice structure. Natural choices include the A-tetrahedron, the B-tetrahedron, and the face-centered cubic unit cell of the breathing pyrochlore lattice consisting of 16 spins, see Figure 7.2b. As it turns out, analyzing the A-tetrahedron and B-tetrahedron separately reveals all the information about the phase. After obtaining results for these clusters separately, they need to be combined to understand the structure of the ground states.

Using the A-tetrahedron as cluster, the decision function becomes

$$d_A^s \sim (c_1^s)^2 + (c_2^s)^2 \tag{7.4}$$

where $s \in \{xy, yz, zx\}$ labels a global spin plane which spontaneously breaks the spin permutation symmetry of the Hamiltonian. The two
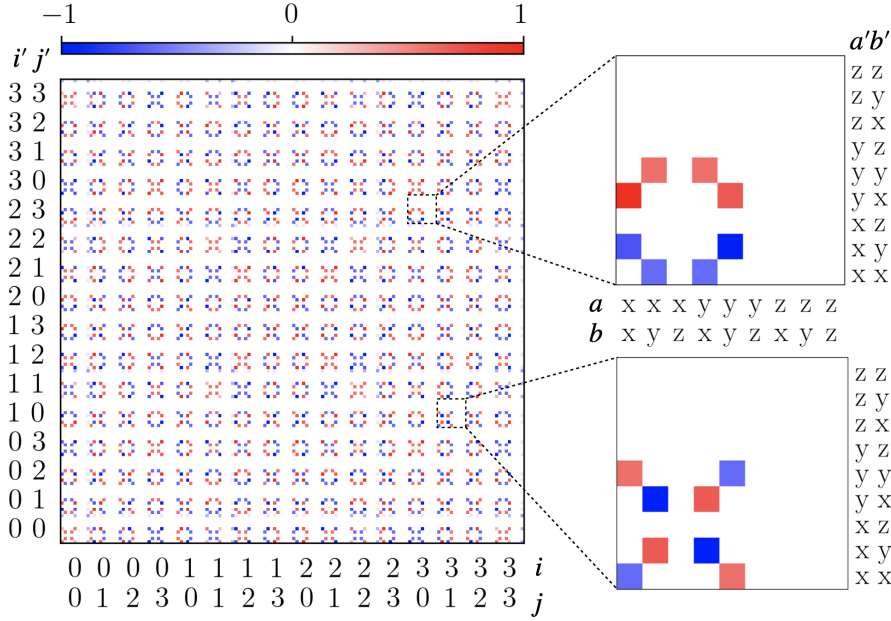
Figure 7.5: Rank 2 coefficient matrix corresponding to the decision function d_A^{xy}. Each pixel represents the weight of a contraction $\langle S_i^a S_j^b \rangle \langle S_{i'}^{a'} S_{j'}^{b'} \rangle$ learned by the machine.

order parameters c_1 and c_2 are quadratic functions of spin components

$$c_1^{yz} = \frac{1}{16}\left[(S_0^y - S_1^y + S_2^z - S_3^z)^2 + (S_0^z - S_1^z - S_2^y + S_3^y)^2\right]$$

$$c_1^{xz} = \frac{1}{16}\left[(S_0^x + S_1^z - S_2^x - S_3^z)^2 + (S_0^z - S_1^x - S_2^z + S_3^x)^2\right] \qquad (7.5)$$

$$c_1^{xy} = \frac{1}{16}\left[(S_0^x + S_1^y - S_2^y - S_3^x)^2 + (S_0^y - S_1^x + S_2^x - S_3^y)^2\right]$$

$$c_2^{yz} = \frac{2}{16}(S_0^y - S_1^y + S_2^z - S_3^z)(S_0^z - S_1^z - S_2^y + S_3^y)$$

$$c_2^{xz} = \frac{2}{16}(S_0^x + S_1^z - S_2^x - S_3^z)(S_0^z - S_1^x - S_2^z + S_3^x) \qquad (7.6)$$

$$c_2^{xy} = \frac{2}{16}(S_0^x + S_1^y - S_2^y - S_3^x)(S_0^y - S_1^x + S_2^x - S_3^y).$$

Since TKSVM is conceived to learn the optimal order parameters, the maximally ordered spin configurations can be inferred by maximizing d_A^s. The resulting fully ordered states represent potential ground states. With no loss of generality, consider a state where the ordering develops in the spin xy plane, whose corresponding TKSVM coefficient matrix is displayed in Figure 7.5.

The interpretation step of TKSVM consists of constructing the analytical expression of the underlying order from the internal parameters of the learning model. This is achieved by reading off and interpreting the graphical representation of the coefficient matrix. Since the underlying order in this case is fairly complex, the procedure is discussed

for a subset of the full pattern of the A-tetrahedra only. Specifically, the block with spin indices $(23, 30)$ is considered, see the upper zoomed-in panel of Figure 7.5. Reading off the terms from the block pattern with coefficients approximated as ± 1 yields the expression

$$
\begin{aligned}
[d_A]_{(23,30)} \sim{} & \\
& + (S_2^y S_3^x)(S_3^x S_0^x) + (S_2^y S_3^x)(S_3^y S_0^y) - (S_2^x S_3^y)(S_3^x S_0^x) \\
& - (S_2^x S_3^y)(S_3^y S_0^y) + (S_2^y S_3^y)(S_3^x S_0^y) + (S_2^y S_3^y)(S_3^y S_0^x) \\
& - (S_2^x S_3^x)(S_3^x S_0^y) - (S_2^x S_3^x)(S_3^y S_0^x).
\end{aligned}
\tag{7.7}
$$

In order to reshape the expression into a sum over square magnitudes of rank 2 order parameters, one must factorize the feature components and assign the coefficients (signs) in consistency with other block patterns $(23, 23)$, $(30, 30)$ and $(30, 23)$

$$
\begin{aligned}
[d_A]_{(23,30)} + [d_A]_{(30,23)} + [d_A]_{(23,23)} + [d_A]_{(30,30)} \sim{} & \\
(-S_2^y S_3^x + S_2^x S_3^y - S_3^x S_0^x - S_3^y S_0^y)^2 & \\
+ (+S_2^x S_3^x - S_2^y S_3^y - S_3^x S_0^y - S_3^y S_0^x)^2. &
\end{aligned}
\tag{7.8}
$$

Factorizing even further requires taking some more block patterns into account

$$
\begin{aligned}
\sum_{ij,i'j' \in \{0,2,3\}} [d_A]_{(ij,i'j')} \sim{} & \\
- \left((S_0^x - S_2^y - S_3^x)^2 + (S_0^y + S_2^x - S_3^y)^2 \right)^2 & \\
- \left(2(S_0^x - S_2^y - S_3^x) \cdot (S_0^y + S_2^x - S_3^y) \right)^2. &
\end{aligned}
\tag{7.9}
$$

This expression already contains a substantial part of the full decision function. Comparing to the definition of c_1^{xy} and c_2^{xy} in Equations 7.5 and 7.6, respectively, reveals that only the terms including $\mathbf{S}_1$ are missing.

Extending the interpretation to the full pattern yields the expression

$$
d_A = \sum_{\substack{ij,i'j' \\ \in \{0,1,2,3\}}} [d_A]_{(ij,i'j')} \propto -(c_1^{xy})^2 - (c_2^{xy})^2.
\tag{7.10}
$$

Note that the overall minus sign is of technical origin and is arbitrary in each TKSVM run, hence it can be dropped. The sign convention in Equation 7.4 is chosen to match with the signs in the existing definition of $\mathbf{m}_{T_{1-}}$ (used later), see its third component in Table 7.2 for comparison. Furthermore, a factor of $1/16$ is introduced to normalize the order parameters to 1 when saturated (deep in phase).

The form of Equation 7.4 suggests c_1^s and c_2^s are independent order parameters. Therefore, they can be maximized separately and the respective result can be checked for consistency. Maximizing $(c_1^{xy})^2$
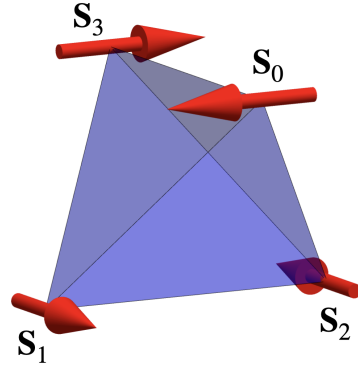
(a) A-tetrahedron
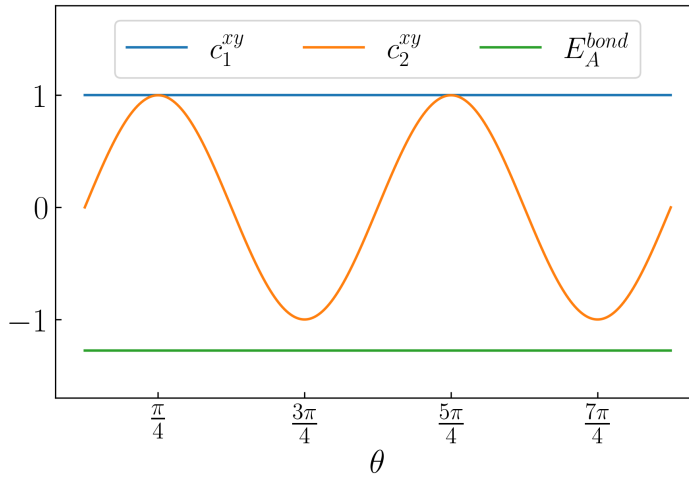
(b) Parametrization

Figure 7.6: Order-by-disorder on A-tetrahedra. (a) Example of a state maximising both $(c_1^{xy})^2$ and $(c_2^{xy})^2$ with $\theta = 7\pi/4$. The bonds lying within the spin-plane $(0-3$ and $1-2)$ have anti-parallel spins. (b) The spin-plane specific quantities c_1^{xy} and c_2^{xy} as well as the energy per bond Equation 7.13 with $J = 1$ and $D = -2$ as functions of the parametrization angle θ. Note that in the decision function Equation 7.4 the term c_2^{xy} appears squared, meaning that configurations with $c_2^{xy} = -1$ are also maxima of the decision function.
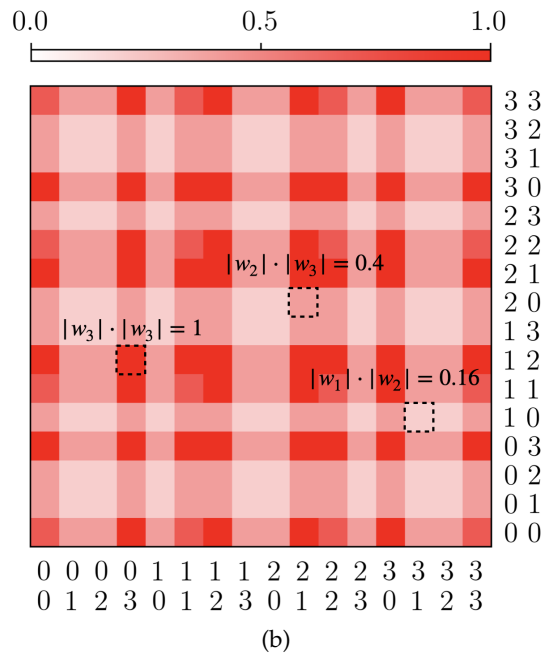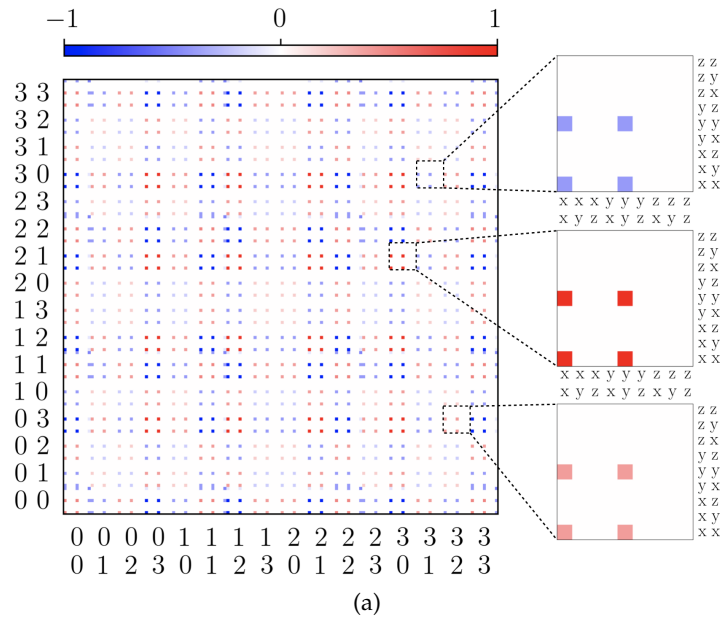
Figure 7.7: Rank 2 coefficient matrix extracted from the decision function for the B-tetrahedron. (a) All blocks have the same structure but different weights. In this example, the z-component vanishes because of the global spin-plane selection. (b) Absolute block-weight products of the left pattern.

Spatial layer	yz	xz	xy
Maximal weight	$w_1 = 1$	$w_2 = 1$	$w_3 = 1$
$(\mathbf{S}_0 \cdot \mathbf{S}_1)$	-1	± 1	± 1
$(\mathbf{S}_0 \cdot \mathbf{S}_2)$	± 1	-1	∓ 1
$(\mathbf{S}_0 \cdot \mathbf{S}_3)$	∓ 1	∓ 1	-1
$(\mathbf{S}_1 \cdot \mathbf{S}_2)$	∓ 1	∓ 1	-1
$(\mathbf{S}_1 \cdot \mathbf{S}_3)$	± 1	-1	∓ 1
$(\mathbf{S}_2 \cdot \mathbf{S}_3)$	-1	± 1	± 1
Missing conf.	$P(\Lambda_E^{\star 1}) = 0$	$P(\Lambda_E^{\star 2}) = 0$	$P(\Lambda_E^{\star 3}) = 0$

Table 7.1: Collinear ground states configurations on B-tetrahedra. In each case there are two possible solutions, reflecting the $\mathbb{Z}_2$ symmetry. The uppermost row describes which spatial layer possesses the $\mathbb{Z}_2$ symmetry, while the lowertmost row indicates which of the three possible configurations of the B-tetrahedra, defined in Equation 7.24, is not allowed.
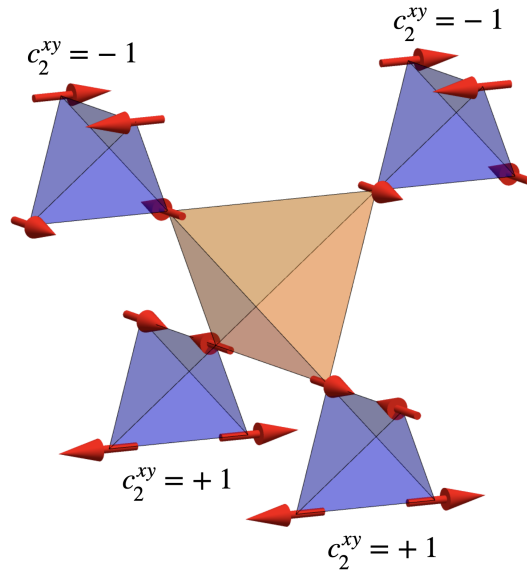
Figure 7.8: Ground-state configuration with ordering in the xy plane. The value of c_2^{xy} is uniform within an A-tetrahedron layer, but alternates over different layers. Compare the upper two blue tetrahedra to the lower two. In this example, the spin ordering plane coincides with the spatial plane.

leads to a manifold of spin configurations parametrized by an angle $\theta \in [0, 2\pi]$

$$\mathbf{S}_0 = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \quad \mathbf{S}_1 = \begin{pmatrix} \cos\theta + \frac{\pi}{2} \\ \sin\theta + \frac{\pi}{2} \\ 0 \end{pmatrix}$$

$$\mathbf{S}_2 = \begin{pmatrix} \cos\theta - \frac{\pi}{2} \\ \sin\theta - \frac{\pi}{2} \\ 0 \end{pmatrix} \quad \mathbf{S}_3 = \begin{pmatrix} \cos\theta + \pi \\ \sin\theta + \pi \\ 0 \end{pmatrix} \tag{7.11}$$

as illustrated in Figure 7.6a. Maximizing $(c_2^{xy})^2$ yields the same structure as Equation 7.11. Its evolution as a function of θ is plotted in Figure 7.6b and shows that $(c_2^{xy})^2$ is maximised for four discrete values

$$\theta \in \left\{ \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4} \right\} \tag{7.12}$$

only. The $U(1)$ manifold of Equation 7.11 belongs to the ground state of the traditional pyrochlore antiferromagnet with negative Dzyaloshinskii-Moriya interactions on *all* tetrahedra [16, 23, 89] without "breathing" anisotropy. Indeed, applying the spin configuration of Equation 7.11 to the Hamiltonian Equation 7.1, its ground state energy per bond is recovered

$$E_A^{bond} = \frac{1}{6}(-2J + 2\sqrt{2}D). \tag{7.13}$$

Since the energy per bond is independent of θ, any selection of specific θ values is necessarily due to thermal order-by-disorder, meaning the selection is entropic rather than energetic. In the pyrochlore model without breathing anisotropy, the order-by-disorder mechanism selects the $\mathbf{q} = 0$ long-range order, which corresponds to either $\theta = \frac{3\pi}{4}$ or $\theta = \frac{7\pi}{4}$ for all tetrahedra, as illustrated in Figure 7.6a. However, in the present model where the Dzyaloshinskii-Moriya terms vanish on B-tetrahedra, the ground state manifold is enlarged and the two additional solutions $\theta \in \left\{ \frac{\pi}{4}, \frac{5\pi}{4} \right\}$ on the A$-$tetrahedra exist. The ordering mechanism is thus a two-step process: First, the selection of the $U(1)$ manifold Equation 7.11, and second, the coalescence on special points of the manifold via thermal order-by-disorder.

Using the B-tetrahedron as cluster, the decision function is identified as

$$d_B \sim \left[w_1 \left(\mathbf{S}_0 \cdot \mathbf{S}_1 + \mathbf{S}_2 \cdot \mathbf{S}_3 \right) + w_2 \left(\mathbf{S}_0 \cdot \mathbf{S}_2 + \mathbf{S}_1 \cdot \mathbf{S}_3 \right) \right.$$
$$\left. + w_3 \left(\mathbf{S}_0 \cdot \mathbf{S}_3 + \mathbf{S}_1 \cdot \mathbf{S}_2 \right) \right]^2. \tag{7.14}$$

The values of the weights w_1, w_2, w_3 can be read off from the TKSVM patterns, satisfying

$$w_1 + w_2 + w_3 = 1, \quad \max w_i = 1 \tag{7.15}$$

where the maximal w_i determines which spatial layer possesses a planar spin-flip symmetry. Under this constraint, d_B can be intuitively maximized if the four spins in a B-tetrahedron are collinear. The solutions are listed in Table 7.1, and Figure 7.7 shows an example of the coefficient matrix when the ordering is in the spin xy plane and $w_3 = 1$.

7.3 QUENCHED SIMULATIONS

Instead of slowly annealing from high temperature, the system is now quenched into the configuration of Figure 7.2a, followed by 10^6 Monte Carlo steps of thermalisation, and 10^7 Monte Carlo steps for measurements. Paving the lattice with the ground state found by the machine requires alternating the sign of the spin configurations of xy-layers along the z direction, see Figure 7.8. This naturally forms a 32-site magnetic unit cell. The simulations are run again with 126 temperatures equally spaced between 0 and 0.0025J. The results are shown in Figure 7.9, computed with the same physical parameters as for the Monte Carlo simulations without quenching.

These new Monte Carlo simulations converge nicely and confirm the stability of the ground state found by the machine. The transition is now violently first order, and its hysteresis explains the shift of the transition temperature T_c from $0.8 \, 10^{-3}$J in Figure 7.1 to $1.5 \, 10^{-3}$J in Figure 7.9. Quenched simulations provide an upper bound of T_c, while slow annealing has more difficulty in finding the ordered phase and provides a lower bound to T_c. The c_1 and c_2 order parameters correctly describe the ground state, with a noticeably stronger finite-size dependence for the latter, which is a common consequence of order-by-disorder [138]. Finally, the order parameter $\mathbf{m}_{q=W}$ saturates at $T = 0$, which means there is a priori no co-existence with other phases.

7.4 EMERGENT PLANAR SYMMETRY

Investigating the origin of the irregular weights in the decision function of the B-tetrahedron, cf. Figure 7.7, reveals an emergent planar-flip symmetry. For simplicity, continue to consider the ground state as Figure 7.2a as an example, where the order is developed in the xy plane and the planar symmetry acts on the spatial xy planes. In general the plane in which the spins are confined and the spatial plane possessing the planar-flip symmetry do not need to coincide, but this does not affect the discussion. The spin-order plane is manifest from the c_1 and
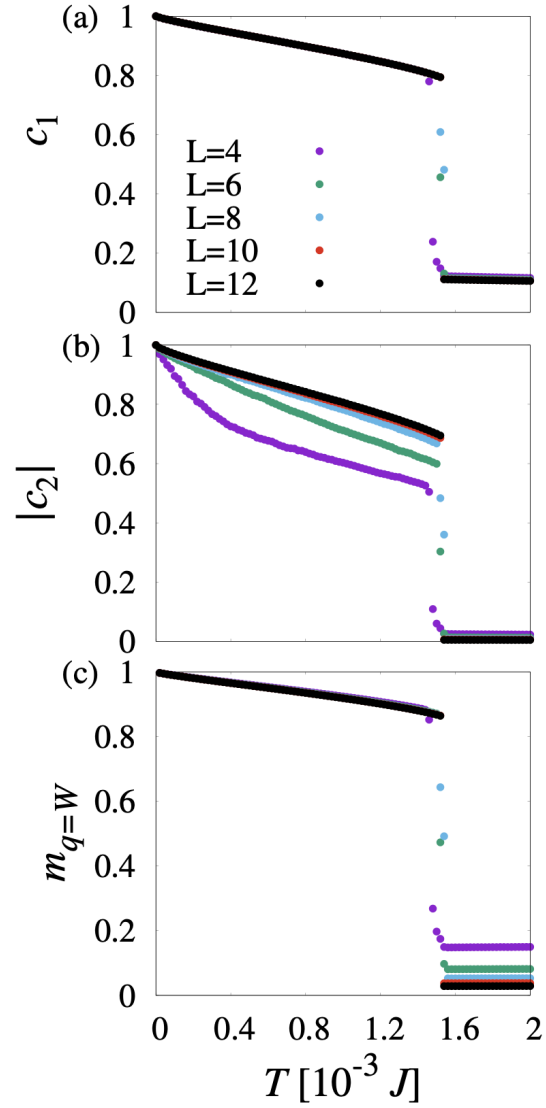
Figure 7.9: The quenched Monte Carlo simulations exhibit a phase transition at $T_c \approx 1.5 \, 10^{-3} J$ for all three order parameters: (a) c_1 in the xy plane, (b) $|c_2|$ in the xy plane, and (c) the order parameter for $\mathbf{q} = W$ order. In case of c_2 the absolute is computed to avoid cancellation between A-tetrahedron layers, cf. Figure 7.8.
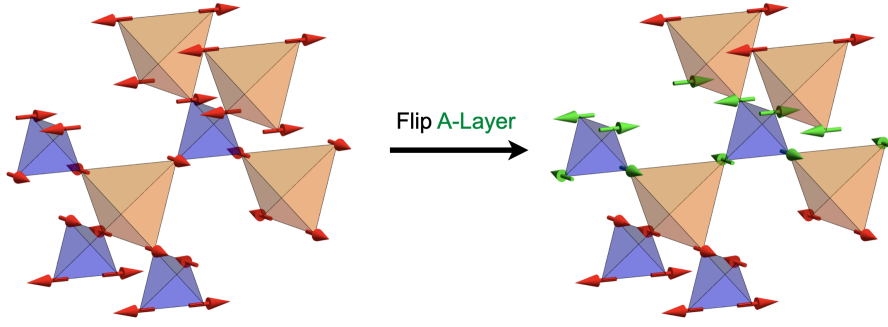
Figure 7.10: Planar $\mathbb{Z}_2$ symmetry. All A-tetrahedra within one layer are flipped (indicated in green), thereby changing the value of Λ_E in the two neighbouring B-layers. On B-tetrahedra, only intra-layer bonds connecting two A-tetrahedra from the same layer are affected by the transformation.

c_2 parameters, while the direction of the spatial planar-flip symmetry can be inferred from the largest weight in d_B. Given the collinearity on the B-tetrahedron, the decision function Equation 7.14 reduces to

$$d_B \sim (-w_1 - w_2 + w_3)^2 = (w_3)^2 \tag{7.16}$$

where the values from the corresponding configuration in Table 7.1 and the weight relation in Equation 7.15 were used. As in this example $\max w_i = w_3 = 1$, the relation Equation 7.15 reduces to $w_1 + w_2 = 0$. Equation 7.16 manifests a property of d_B that it is invariant under flipping a specific pair of spins, which can be either (S_0, S_3) or (S_1, S_2) in the current example. Nevertheless, as spins in a B-tetrahedron belong to different A-tetrahedra, in order to preserve the value of the order parameters c_1 and c_2, one has to flip all spins in the two neighbouring A-tetrahedra. This procedure is then repeated to further A-tetrahedra neighbours, and closes only after flipping all A-tetrahedra in an entire layer, as illustrated in Figure 7.10.

As a consequence, the system hybridizes rank 1 and rank 2 orders. This is a rather unconventional emergent property for an ordered phase. Such sub-extensive zero modes usually require to be artificially enforced via either a global or local symmetry of the system [46, 82, 102]. Since this planar spin flip is sub-extensive (it scales with L^2), it is expected to be dynamically robust. Assume that while cooling the system, a planar spin flip takes place. This is possible since it costs zero energy and the transition is violently first order. Once the cubic symmetry is broken in favour of a given plane, say the xy plane, two layers of A−tetrahedra far from each other may likely order independently. As a result, the long-range dipolar order with Bragg peaks at $\mathbf{q} = W$ found in Figure 7.9 will be perturbed by multiple planar spin flips at random positions in the system. And since there is a vanishingly small probability to move L^2 spins coherently in the thermodynamic limit, such energetically degenerate spin configuration should remain stable

over very long time scales. This mechanism explains the origin of the irregular weights in Figure 7.4, and likely plays an important role in the difficulty to interpret previous Monte Carlo simulations, since it conceals the magnetic dipolar order. As a result, the magnetic order will be long-range only in the spatial plane, but not in the orthogonal direction. Such states break lattice rotation symmetry and may be dubbed *hybridized nematic order*, with the direction of the nematicity orthogonal to the spatial layer possessing the spin-flip symmetry.

It should be mentioned that among all of the energetically degenerate states connected by the $\mathbb{Z}_2$ symmetry, three of them have the 16-site cubic unit cell of Figure 7.8 paving the entire lattice. They possess two planar symmetries rather than only one, e.g. planar xy- and yz-symmetries with the corresponding weights $w_1 = w_3 = 1, w_2 = -1$, cf. Table 7.1. But any planar spin flip as in Figure 7.10 would immediately break the fragile cubic symmetry.

7.5 RELATION TO T_d IRREPS

To gain more insight, it is sensible to re-express the machine-learned quantities in terms of order parameters transforming according to the irreducible representations of the point group T_d, defined in Table 7.2. In terms of these expressions, the two order parameters found in the decision function in d_A^s can be rewritten as

$$c_1^s = \frac{1}{4}(\|\mathbf{m}_{T_{1-}}\|^2 + \|\mathbf{m}_E\|^2) \tag{7.17}$$

$$c_2^s = \frac{1}{4}(\|\mathbf{m}_{T_{1-}}\|^2 - \|\mathbf{m}_E\|^2). \tag{7.18}$$

The meaning of the ground-state condition $c_2^s = \pm 1$ becomes more intuitive in the irrep basis. Order-by-disorder ensures that an A-tetrahedron is either fully on $\mathbf{m}_{T_{1-}}$ or on $\mathbf{m}_E$. The alteration of $c_2^s = \pm 1$ further means that the system can be viewed as staggered layers of $\mathbf{m}_{T_{1-}}$ and $\mathbf{m}_E$ A-tetrahedra.

For the B-tetrahedron, it is more convenient to use bond order parameters

$$\begin{pmatrix} \Lambda_{A_1} \\ \Lambda_{E,1} \\ \Lambda_{E,2} \\ \Lambda_{T_2,1} \\ \Lambda_{T_2,2} \\ \Lambda_{T_2,3} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \mathbf{S}_0 \cdot \mathbf{S}_1 \\ \mathbf{S}_0 \cdot \mathbf{S}_2 \\ \mathbf{S}_0 \cdot \mathbf{S}_3 \\ \mathbf{S}_1 \cdot \mathbf{S}_2 \\ \mathbf{S}_1 \cdot \mathbf{S}_3 \\ \mathbf{S}_2 \cdot \mathbf{S}_3 \end{pmatrix} \tag{7.19}$$

Op.	Definition in terms of spin components
m_{A_2}	$\frac{1}{2\sqrt{3}}(S_0^x + S_0^y + S_0^z + S_1^x - S_1^y - S_1^z - S_2^x + S_2^y - S_2^z - S_3^x - S_3^y + S_3^z)$
$\mathbf{m}_E$	$\begin{pmatrix} \frac{1}{2\sqrt{6}}(-2S_0^x + S_0^y + S_0^z - 2S_1^x - S_1^y - S_1^z + 2S_2^x + S_2^y - S_2^z + 2S_3^x - S_3^y + S_3^z) \\ \frac{1}{2\sqrt{2}}(-S_0^y + S_0^z + S_1^y - S_1^z - S_2^y - S_2^z + S_3^y + S_3^z) \end{pmatrix}$
$\mathbf{m}_{T_{1+}}$	$\begin{pmatrix} \frac{1}{2}(S_0^x + S_1^x + S_2^x + S_3^x) \\ \frac{1}{2}(S_0^y + S_1^y + S_2^y + S_3^y) \\ \frac{1}{2}(S_0^z + S_1^z + S_2^z + S_3^z) \end{pmatrix}$
$\mathbf{m}_{T_{1-}}$	$\begin{pmatrix} \frac{-1}{2\sqrt{2}}(S_0^y + S_0^z - S_1^y - S_1^z - S_2^y + S_2^z + S_3^y - S_3^z) \\ \frac{-1}{2\sqrt{2}}(S_0^x + S_0^z - S_1^x + S_1^z - S_2^x - S_2^z + S_3^x - S_3^z) \\ \frac{-1}{2\sqrt{2}}(S_0^x + S_0^y - S_1^x + S_1^y + S_2^x - S_2^y - S_3^x - S_3^y) \end{pmatrix}$
$\mathbf{m}_{T_2}$	$\begin{pmatrix} \frac{1}{2\sqrt{2}}(-S_0^y + S_0^z + S_1^y - S_1^z + S_2^y + S_2^z - S_3^y - S_3^z) \\ \frac{1}{2\sqrt{2}}(S_0^x - S_0^z - S_1^x - S_1^z - S_2^x + S_2^z + S_3^x + S_3^z) \\ \frac{1}{2\sqrt{2}}(-S_0^x + S_0^y + S_1^x + S_1^y - S_2^x - S_2^y + S_3^x - S_3^y) \end{pmatrix}$

Table 7.2: Order parameters for breaking the point-group symmetry of a single tetrahedron, used to describe learned TKSVM order parameters on an A-tetrahedron. The fields $\mathbf{m}_X$ transform according to irreducible representation X of the tetrahedral point group T_d. Reproduced from [137, 138].

transforming according to the irreps A_1, E and T_2 of the tetrahedral point-group [117]. The decision function d_B expressed in terms of bond order parameters reads

$$d_B \sim \left[a_{A_1} \Lambda_{A_1} + a_{E,1} \Lambda_{E,1} + a_{E,2} \Lambda_{E,2} \right]^2 \tag{7.20}$$

with

$$a_{A_1} = \sqrt{\frac{2}{3}} \, (w_1 + w_2 + w_3) \tag{7.21}$$

$$a_{E,1} = \sqrt{\frac{1}{3}} \, (2w_1 - w_2 - w_3) \tag{7.22}$$

$$a_{E,2} = w_2 - w_3. \tag{7.23}$$

Maximizing d_B then requires $\Lambda_{A_1} \equiv -\sqrt{\frac{2}{3}}$ and that $\boldsymbol{\Lambda}_E$ can only be in one of the following three configurations

$$\boldsymbol{\Lambda}_E^{\star 1} = \begin{pmatrix} \frac{4}{\sqrt{3}} \\ 0 \end{pmatrix} \quad \boldsymbol{\Lambda}_E^{\star 2} = \begin{pmatrix} \frac{-2}{\sqrt{3}} \\ 2 \end{pmatrix} \quad \boldsymbol{\Lambda}_E^{\star 3} = \begin{pmatrix} \frac{-2}{\sqrt{3}} \\ -2 \end{pmatrix} \tag{7.24}$$

which are the three maxima of $\|\boldsymbol{\Lambda}_E\|^2$ under the condition of minimal Λ_{A_1}. These three configurations are transformed by the $\mathbb{Z}_2$ planar symmetry as depicted in Figure 7.11. The distribution $\{P(\boldsymbol{\Lambda}_E^{\star 1}), P(\boldsymbol{\Lambda}_E^{\star 2}), P(\boldsymbol{\Lambda}_E^{\star 3})\}$ of $\boldsymbol{\Lambda}_E$ over all the B-tetrahedra may be inferred from the weights of d_B in Equation 7.14. In the extreme cases where all B-tetrahedra are in the same $\boldsymbol{\Lambda}_E$ configuration, the coefficients take the special values (denoted by a star)

$$\mathbf{a}_E^{\star 1} = \begin{pmatrix} \frac{-4}{\sqrt{3}} \\ 0 \end{pmatrix}, \quad \mathbf{a}_E^{\star 2} = \begin{pmatrix} \frac{2}{\sqrt{3}} \\ -2 \end{pmatrix}, \quad \text{or} \quad \mathbf{a}_E^{\star 3} = \begin{pmatrix} \frac{2}{\sqrt{3}} \\ 2 \end{pmatrix}. \tag{7.25}$$

In addition, $a_{A_1} = \sqrt{2/3}$ reflects the ground state condition $\Lambda_{A_1} = -\sqrt{2/3}$ which is independent of $\boldsymbol{\Lambda}_E$. The coefficients $a_{A_1}, a_{E,1}, a_{E,2}$ for the general case are then obtained by solving a linear equation system

$$\begin{pmatrix} a_{A_1} & a_{A_1} & a_{A_1} \\ a_{E,1}^{\star 1} & a_{E,1}^{\star 2} & a_{E,1}^{\star 3} \\ a_{E,2}^{\star 1} & a_{E,2}^{\star 2} & a_{E,2}^{\star 3} \end{pmatrix} \begin{pmatrix} P(\boldsymbol{\Lambda}_E^{\star 1}) \\ P(\boldsymbol{\Lambda}_E^{\star 2}) \\ P(\boldsymbol{\Lambda}_E^{\star 3}) \end{pmatrix} = \begin{pmatrix} a_{A_1} \\ a_{E,1} \\ a_{E,2} \end{pmatrix}. \tag{7.26}$$

Here the first equation reduces to the normalisation of the distribution $\sum_i P(\boldsymbol{\Lambda}_E^{\star i}) = 1$, equivalent to $\sum_i w_i = 1$. In the example of Figure 7.7, the weights are given by $w_1 = -w_2 = 0.4$ and $w_3 = 1$, which translates to $a_{A_1} = \sqrt{2/3}$, $a_{E,1} = \sqrt{1/3} \cdot 0.2$ and $a_{E,2} = -1.4$. Solving the linear system for these values yields

$$P(\boldsymbol{\Lambda}_E^{\star 1}) = 0.3, \ P(\boldsymbol{\Lambda}_E^{\star 2}) = 0.7, \ P(\boldsymbol{\Lambda}_E^{\star 3}) = 0. \tag{7.27}$$
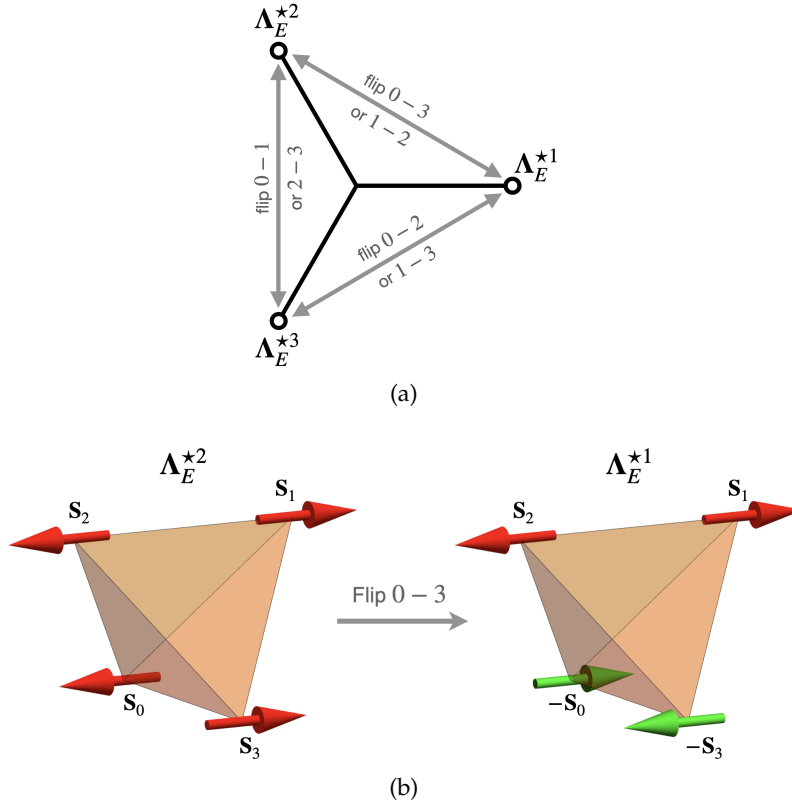
Figure 7.11: (a) Triangle formed by the maxima of $\|\Lambda_E\|^2$ together with the $\mathbb{Z}_2$ bond flip relations. (b) $\mathbb{Z}_2$ symmetry operation acting on a bond, thereby mapping one maximum to another.

In general there is always one vanishing $P(\Lambda_E^{\star i})$, which is equivalent to the statement that the weight $\max w_i = 1$ is maximal, and can be associated with the spatial orientation of the planar $\mathbb{Z}_2$ symmetry, as listed in Table 7.1.

7.6 CONCLUSION

In conclusion, the nature of the long-range ordered phase was successfully determined, and the learning model identified two distinct order parameters that bring a two-step ordering mechanism to light. The first step is a traditional energetic selection into a continuously degenerate ground state manifold, followed by an entropic selection via thermal order-by-disorder [52, 118] as the second step.

Quenching the Monte Carlo simulations into the learned spin configurations led to well converged data. Subsequent analytical considerations finally unveiled zero-energy sub-extensive dynamics, which represent neither a global nor a local symmetry of the system, but rather an emergent $\mathbb{Z}_2$ subsystem symmetry, which leads to hybridization of linear (dipolar) and quadratic (quadrupolar) magnetic orders. The emergence of a sub-extensive symmetry is markedly unusual for a

realistic two-body Hamiltonian, since it typically requires an artificial and complex set of interactions, like in models of fractonic matter [46, 82, 102] for instance.

Poorly thermalized Monte Carlo data served as input to the learning model, which discerned certain patterns from which the unconventional and complex magnetic structure of the phase could be laid bare. The study illustrates how numerical simulations can be enhanced from repeated combined iterations of machine learning and analysis by a human. The strong interpretability of TKSVM played a crucial role, but also analytical and group-theoretical arguments were indispensable. In summary, this collaborative approach offers one possible paradigm for research driven by machine learning in an age in which humans and machine learning models have complementary strengths.

BIBLIOGRAPHY

[1] Scott Aaronson. "Shadow Tomography of Quantum States." In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. New York, NY, USA: Association for Computing Machinery, 2018, 325–338. ISBN: 9781450355599. DOI: 10.1145/3188745.3188802. URL: https://doi.org/10.1145/3188745.3188802.

[2] Rajeev Acharya et al. "Suppressing quantum errors by scaling a surface code logical qubit." In: *Nature* 614.7949 (2023), pp. 676–681. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05434-1. URL: https://doi.org/10.1038/s41586-022-05434-1.

[3] Ian Affleck, Tom Kennedy, Elliott H. Lieb, and Hal Tasaki. "Rigorous results on valence-bond ground states in antiferromagnets." In: *Phys. Rev. Lett.* 59 (7 1987), pp. 799–802. DOI: 10.1103/PhysRevLett.59.799. URL: https://link.aps.org/doi/10.1103/PhysRevLett.59.799.

[4] Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. "Quantum State Tomography with Conditional Generative Adversarial Networks." In: *Phys. Rev. Lett.* 127 (14 2021), p. 140502. DOI: 10.1103/PhysRevLett.127.140502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.140502.

[5] K Banaszek, M Cramer, and D Gross. "Focus on quantum tomography." In: *New Journal of Physics* 15.12 (2013), p. 125020. DOI: 10.1088/1367-2630/15/12/125020. URL: https://doi.org/10.1088/1367-2630/15/12/125020.

[6] F. Barratt, James Dborin, Matthias Bal, Vid Stojevic, Frank Pollmann, and A. G. Green. "Parallel quantum simulation of large systems on small NISQ computers." In: *npj Quantum Information* 7.1 (2021), p. 79. ISSN: 2056-6387. DOI: 10.1038/s41534-021-00420-3. URL: https://doi.org/10.1038/s41534-021-00420-3.

[7] Paul Benioff. "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines." In: *Journal of Statistical Physics* 22.5 (1980), pp. 563–591. ISSN: 1572-9613. DOI: 10.1007/BF01011339. URL: https://doi.org/10.1007/BF01011339.

[8] Robin Blume-Kohout and Kevin C. Young. "A volumetric framework for quantum computer benchmarks." In: *Quantum* 4 (Nov. 2020), p. 362. ISSN: 2521-327X. DOI: 10.22331/q-2020-11-15-362. URL: https://doi.org/10.22331/q-2020-11-15-362.

[9] Dolev Bluvstein et al. "Logical quantum processor based on reconfigurable atom arrays." In: *Nature* 626.7997 (2024), pp. 58–65. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06927-3. URL: https://doi.org/10.1038/s41586-023-06927-3.

[10] Annabelle Bohrdt, Christie S. Chiu, Geoffrey Ji, Muqing Xu, Daniel Greif, Markus Greiner, Eugene Demler, Fabian Grusdt, and Michael Knap. "Classifying snapshots of the doped Hubbard model with machine learning." In: *Nature Physics* 15.9 (2019), pp. 921–924. DOI: 10.1038/s41567-019-0565-x. URL: https://doi.org/10.1038/s41567-019-0565-x.

[11] Guillaume Bornet et al. "Enhancing a Many-Body Dipolar Rydberg Tweezer Array with Arbitrary Local Controls." In: *Physical Review Letters* 132.26 (June 2024). ISSN: 1079-7114. DOI: 10.1103/physrevlett.132.263601. URL: http://dx.doi.org/10.1103/PhysRevLett.132.263601.

[12] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. "A Training Algorithm for Optimal Margin Classifiers." In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. COLT '92. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 1992, 144–152. ISBN: 089791497X. DOI: 10.1145/130385.130401. URL: https://doi.org/10.1145/130385.130401.

[13] Sergey Bravyi, Andrew W. Cross, Jay M. Gambetta, Dmitri Maslov, Patrick Rall, and Theodore J. Yoder. "High-threshold and low-overhead fault-tolerant quantum memory." In: *Nature* 627.8005 (2024), pp. 778–782. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07107-7. URL: https://doi.org/10.1038/s41586-024-07107-7.

[14] Hans J. Briegel and Robert Raussendorf. "Persistent Entanglement in Arrays of Interacting Particles." In: *Phys. Rev. Lett.* 86 (5 2001), pp. 910–913. DOI: 10.1103/PhysRevLett.86.910. URL: https://link.aps.org/doi/10.1103/PhysRevLett.86.910.

[15] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos. "Probing Renyi entanglement entropy via randomized measurements." In: *Science* 364.6437 (2019), pp. 260–263. DOI: 10.1126/science.aau4963. eprint: https://www.science.org/doi/pdf/10.1126/science.aau4963. URL: https://www.science.org/doi/abs/10.1126/science.aau4963.

[16] B. Canals, M. Elhajal, and C. Lacroix. "Ising-like order by disorder in the pyrochlore antiferromagnet with Dzyaloshinskii-Moriya interactions." In: *Phys. Rev. B* 78 (2008), p. 214431. DOI: 10.1103/PhysRevB.78.214431.

[17]   Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. "Machine learning and the physical sciences." In: *Rev. Mod. Phys.* 91 (4 2019), p. 045002. DOI: `10.1103/RevModPhys.91.045002`. URL: `https://link.aps.org/doi/10.1103/RevModPhys.91.045002`.

[18]   Juan Carrasquilla, Giacomo Torlai, Roger G. Melko, and Leandro Aolita. "Reconstructing quantum states with generative models." In: *Nature Machine Intelligence* 1.3 (2019), pp. 155–161. ISSN: 2522-5839. DOI: `10.1038/s42256-019-0028-1`. URL: `https://doi.org/10.1038/s42256-019-0028-1`.

[19]   Chih-Chung Chang and Chih-Jen Lin. "Training v-support vector classifiers: theory and algorithms." In: *Neural Comput.* 13.9 (2001), pp. 2119–2147. DOI: `10.1162/089976601750399335`.

[20]   Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: A Library for Support Vector Machines." In: *ACM Trans. Intell. Syst. Technol.* 2.3 (May 2011), 27:1–27:27. ISSN: 2157-6904. DOI: `10.1145/1961189.1961199`. URL: `http://doi.acm.org/10.1145/1961189.1961199`.

[21]   Cheng Chen et al. "Continuous symmetry breaking in a two-dimensional Rydberg array." In: *Nature* 616.7958 (2023), pp. 691–695. ISSN: 1476-4687. DOI: `10.1038/s41586-023-05859-2`. URL: `https://doi.org/10.1038/s41586-023-05859-2`.

[22]   Xie Chen, Zheng-Cheng Gu, Zheng-Xin Liu, and Xiao-Gang Wen. "Symmetry protected topological orders and the group cohomology of their symmetry group." In: *Phys. Rev. B* 87 (15 2013), p. 155114. DOI: `10.1103/PhysRevB.87.155114`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.87.155114`.

[23]   G.-W. Chern. "Pyrochlore antiferromagnet with antisymmetric exchange interactions: critical behavior and order from disorder." In: *arXiv* (Aug. 2010). arXiv: `1008.3038 [cond-mat.str-el]`.

[24]   Li Ern Chern, Yong Baek Kim, and Claudio Castelnovo. "Competing quantum spin liquids, gauge fluctuations, and anisotropic interactions in a breathing pyrochlore lattice." In: *Phys. Rev. B* 106 (13 2022), p. 134402. DOI: `10.1103/PhysRevB.106.134402`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.106.134402`.

[25]   A. Chi-Chih Yao. "Quantum circuit complexity." In: *Proceedings of the 1993 IEEE 34th Annual Foundations of Computer Science.* SFCS '93. USA: IEEE Computer Society, 1993, 352–361. ISBN: 0818643706. DOI: `10.1109/SFCS.1993.366852`. URL: `https://doi.org/10.1109/SFCS.1993.366852`.

[26] Craig R. Clark et al. "High-Fidelity Bell-State Preparation with $^{40}\text{Ca}^+$ Optical Qubits." In: *Phys. Rev. Lett.* 127 (13 2021), p. 130505. DOI: 10.1103/PhysRevLett.127.130505. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.130505.

[27] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. "Quantum convolutional neural networks." In: *Nature Physics* 15.12 (2019), pp. 1273–1278. DOI: 10.1038/s41567-019-0648-8. URL: https://doi.org/10.1038/s41567-019-0648-8.

[28] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. "Efficient quantum state tomography." In: *Nature Communications* 1.1 (2010), p. 149. DOI: 10.1038/ncomms1147. URL: https://doi.org/10.1038/ncomms1147.

[29] Thomas Decker, Dominik Janzing, and Thomas Beth. "Quantum circuits for single-qubit measurements corresponding to platonic solids." In: *International Journal of Quantum Information* 02.03 (2004), pp. 353–377. DOI: 10.1142/S0219749904000298. eprint: https://doi.org/10.1142/S0219749904000298. URL: https://doi.org/10.1142/S0219749904000298.

[30] David Elieser Deutsch and Roger Penrose. "Quantum computational networks." In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 425.1868 (1989), pp. 73–90. DOI: 10.1098/rspa.1989.0099. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1989.0099. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1989.0099.

[31] David Deutsch and Roger Penrose. "Quantum theory, the Church–Turing principle and the universal quantum computer." In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (1985), pp. 97–117. DOI: 10.1098/rspa.1985.0070. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1985.0070. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1985.0070.

[32] David P. DiVincenzo. "The Physical Implementation of Quantum Computation." In: *Fortschritte der Physik* 48.9-11 (2000), pp. 771–783. DOI: https://doi.org/10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1521-3978%28200009%2948%3A9/11%3C771%3A%3AAID-PROP771%3E3.0.CO%3B2-E. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/1521-3978%28200009%2948%3A9/11%3C771%3A%3AAID-PROP771%3E3.0.CO%3B2-E.

[33] Yulong Dong, Lin Lin, and Yu Tong. "Ground-State Preparation and Energy Estimation on Early Fault-Tolerant Quantum Computers via Quantum Eigenvalue Transformation of Unitary Matrices." In: *PRX Quantum* 3 (4 2022), p. 040305. DOI: 10.1103/PRXQuantum.3.040305. URL: https://link.aps.org/doi/10.1103/PRXQuantum.3.040305.

[34] B. Efron and C. Stein. "The Jackknife Estimate of Variance." In: *The Annals of Statistics* 9.3 (1981), pp. 586 –596. DOI: 10.1214/aos/1176345462. URL: https://doi.org/10.1214/aos/1176345462.

[35] Andreas Elben, Jinlong Yu, Guanyu Zhu, Mohammad Hafezi, Frank Pollmann, Peter Zoller, and Benoît Vermersch. "Many-body topological invariants from randomized measurements in synthetic quantum matter." In: *Science Advances* 6.15 (2020), eaaz3666. DOI: 10.1126/sciadv.aaz3666. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aaz3666. URL: https://www.science.org/doi/abs/10.1126/sciadv.aaz3666.

[36] Richard P. Feynman. "Simulating physics with computers." In: *International Journal of Theoretical Physics* 21.6 (1982), pp. 467–488. ISSN: 1572-9575. DOI: 10.1007/BF02650179. URL: https://doi.org/10.1007/BF02650179.

[37] Miroslav Fiedler. "Algebraic connectivity of graphs." In: *Czechoslovak Mathematical Journal* 23.2 (1973), pp. 298–305. URL: http://eudml.org/doc/12723.

[38] Miroslav Fiedler. "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory." eng. In: *Czechoslovak Mathematical Journal* 25.4 (1975), pp. 619–633.

[39] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. "The ITensor Software Library for Tensor Network Calculations." In: *SciPost Phys. Codebases* (2022), p. 4. DOI: 10.21468/SciPostPhysCodeb.4. URL: https://scipost.org/10.21468/SciPostPhysCodeb.4.

[40] Christopher A. Fuchs, Michael C. Hoang, and Blake C. Stacey. "The SIC Question: History and State of Play." In: *Axioms* 6.3 (2017). ISSN: 2075-1680. DOI: 10.3390/axioms6030021. URL: https://www.mdpi.com/2075-1680/6/3/21.

[41] Katerina Gratsea, Jakob S. Kottmann, Peter D. Johnson, and Alexander A. Kunitsa. *Comparing Classical and Quantum Ground State Preparation Heuristics*. 2024. arXiv: 2401.05306 [quant-ph]. URL: https://arxiv.org/abs/2401.05306.

[42] Jonas Greitemann. "Investigation of hidden multipolar spin order in frustrated magnets using interpretable machine learning techniques." PhD thesis. Ludwig-Maximilians-Universität München, 2019.

[43]   Jonas Greitemann, Ke Liu, Ludovic D. C. Jaubert, Han Yan, Nic Shannon, and Lode Pollet. "Identification of emergent constraints and hidden order in frustrated magnets using tensorial kernel methods of machine learning." In: *Physical Review B* 100.17 (2019). ISSN: 2469-9969. DOI: 10.1103/physrevb.100.174408. URL: http://dx.doi.org/10.1103/PhysRevB.100.174408.

[44]   Jonas Greitemann, Ke Liu, and Lode Pollet. tensorial-kernel SVM library, https://gitlab.physik.uni-muenchen.de/tk-svm/tksvm-op.

[45]   Jonas Greitemann, Ke Liu, and Lode Pollet. "Probing hidden spin order with interpretable machine learning." In: *Physical Review B* 99.6 (2019). ISSN: 2469-9969. DOI: 10.1103/physrevb.99.060404. URL: http://dx.doi.org/10.1103/PhysRevB.99.060404.

[46]   Andrey Gromov and Leo Radzihovsky. "Colloquium: Fracton matter." In: *Rev. Mod. Phys.* 96 (1 2024), p. 011001. DOI: 10.1103/RevModPhys.96.011001. URL: https://link.aps.org/doi/10.1103/RevModPhys.96.011001.

[47]   D. Gross, F. Krahmer, and R. Kueng. "A Partial Derandomization of PhaseLift Using Spherical Designs." In: *Journal of Fourier Analysis and Applications* 21.2 (2015), pp. 229–266. ISSN: 1531-5851. DOI: 10.1007/s00041-014-9361-2. URL: https://doi.org/10.1007/s00041-014-9361-2.

[48]   Lov K. Grover. "A fast quantum mechanical algorithm for database search." In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '96. Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, 212–219. ISBN: 0897917855. DOI: 10.1145/237814.237866. URL: https://doi.org/10.1145/237814.237866.

[49]   Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. "Sample-Optimal Tomography of Quantum States." In: STOC '16. https://doi.org/10.1145/2897518.2897585. Cambridge, MA, USA: Association for Computing Machinery, 2016, 913–925. ISBN: 9781450341325. DOI: 10.1145/2897518.2897585. URL: https://doi.org/10.1145/2897518.2897585.

[50]   T. Haku et al. "Low-energy excitations and ground-state selection in the quantum breathing pyrochlore antiferromagnet $Ba_3Yb_2Zn_5O_{11}$." In: *Phys. Rev. B* 93 (22 2016), p. 220407. DOI: 10.1103/PhysRevB.93.220407. URL: http://link.aps.org/doi/10.1103/PhysRevB.93.220407.

[51] T. P. Harty, D. T. C. Allcock, C. J. Ballance, L. Guidoni, H. A. Janacek, N. M. Linke, D. N. Stacey, and D. M. Lucas. "High-Fidelity Preparation, Gates, Memory, and Readout of a Trapped-Ion Quantum Bit." In: *Phys. Rev. Lett.* 113 (22 2014), p. 220501. DOI: `10.1103/PhysRevLett.113.220501`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.113.220501`.

[52] Christopher L. Henley. "Ordering due to disorder in a frustrated vector antiferromagnet." In: *Phys. Rev. Lett.* 62 (17 1989), pp. 2056–2059. DOI: `10.1103/PhysRevLett.62.2056`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.62.2056`.

[53] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." In: *Nature Physics* 16.10 (2020), pp. 1050–1057. ISSN: 1745-2481. DOI: `10.1038/s41567-020-0932-7`. URL: `https://doi.org/10.1038/s41567-020-0932-7`.

[54] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V. Albert, and John Preskill. "Provably efficient machine learning for quantum many-body problems." In: *Science* 377.6613 (2022), eabk3333. DOI: `10.1126/science.abk3333`. eprint: `https://www.science.org/doi/pdf/10.1126/science.abk3333`. URL: `https://www.science.org/doi/abs/10.1126/science.abk3333`.

[55] Mohsin Iqbal et al. "Non-Abelian topological order and anyons on a trapped-ion processor." In: *Nature* 626.7999 (2024), pp. 505–511. ISSN: 1476-4687. DOI: `10.1038/s41586-023-06934-4`. URL: `https://doi.org/10.1038/s41586-023-06934-4`.

[56] Raban Iten, Roger Colbeck, Ivan Kukuljan, Jonathan Home, and Matthias Christandl. "Quantum circuits for isometries." In: *Phys. Rev. A* 93 (3 2016), p. 032318. DOI: `10.1103/PhysRevA.93.032318`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.93.032318`.

[57] Raban Iten, Oliver Reardon-Smith, Emanuel Malvetti, Luca Mondada, Gabrielle Pauvert, Ethan Redmond, Ravjot Singh Kohli, and Roger Colbeck. *Introduction to UniversalQCompiler*. 2021. arXiv: `1904.01072 [quant-ph]`. URL: `https://arxiv.org/abs/1904.01072`.

[58] Ali Javadi-Abhari et al. *Quantum computing with Qiskit*. 2024. arXiv: `2405.08810 [quant-ph]`. URL: `https://arxiv.org/abs/2405.08810`.

[59] Been Kim, Rajiv Khanna, and Oluwasanmi Koyejo. "Examples are not enough, learn to criticize! Criticism for interpretability." English (US). In: *Advances in Neural Information Processing Systems* (2016). 30th Annual Conference on Neural Information

Processing Systems, NIPS 2016 ; Conference date: 05-12-2016 Through 10-12-2016, pp. 2288–2296. ISSN: 1049-5258.

[60]   Youngseok Kim et al. "Evidence for the utility of quantum computing before fault tolerance." In: *Nature* 618.7965 (2023), pp. 500–505. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06096-3. URL: https://doi.org/10.1038/s41586-023-06096-3.

[61]   Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: 1312.6114 [stat.ML]. URL: https://arxiv.org/abs/1312.6114.

[62]   A.Yu. Kitaev. "Fault-tolerant quantum computation by anyons." In: *Annals of Physics* 303.1 (2003), pp. 2–30. ISSN: 0003-4916. DOI: https://doi.org/10.1016/S0003-4916(02)00018-0. URL: https://www.sciencedirect.com/science/article/pii/S0003491602000180.

[63]   Alexei Kitaev. "Anyons in an exactly solved model and beyond." In: *Ann. Phys. (N. Y.)* 321.1 (2006). January Special Issue, pp. 2–111. ISSN: 0003-4916. DOI: 10.1016/j.aop.2005.10.005. URL: http://www.sciencedirect.com/science/article/pii/S0003491605002381.

[64]   Andreas Klappenecker and Martin Roetteler. "Mutually unbiased bases, spherical designs, and frames." In: *Wavelets XI*. Ed. by Manos Papadakis, Andrew F. Laine, and Michael A. Unser. Vol. 5914. International Society for Optics and Photonics. SPIE, 2005, 59140P. DOI: 10.1117/12.615759. URL: https://doi.org/10.1117/12.615759.

[65]   John B. Kogut. "An introduction to lattice gauge theory and spin systems." In: *Rev. Mod. Phys.* 51 (4 1979), pp. 659–713. DOI: 10.1103/RevModPhys.51.659. URL: https://link.aps.org/doi/10.1103/RevModPhys.51.659.

[66]   Valeri N. Kotov, Maged Elhajal, Michael E. Zhitomirsky, and Frédéric Mila. "Dzyaloshinsky-Moriya-induced order in the spin-liquid phase of the $S = 1/2$ pyrochlore antiferromagnet." In: *Phys. Rev. B* 72 (1 2005), p. 014421. DOI: 10.1103/PhysRevB.72.014421. URL: https://link.aps.org/doi/10.1103/PhysRevB.72.014421.

[67]   Korbinian Kottmann, Patrick Huembeli, Maciej Lewenstein, and Antonio Acín. "Unsupervised Phase Discovery with Deep Anomaly Detection." In: *Phys. Rev. Lett.* 125 (17 2020), p. 170603. DOI: 10.1103/PhysRevLett.125.170603. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.170603.

[68]   Tian Lan, Liang Kong, and Xiao-Gang Wen. "Classification of $(3+1)$D Bosonic Topological Orders: The Case When Point-like Excitations Are All Bosons." In: *Phys. Rev. X* 8 (2 2018),

p. 021074. DOI: `10.1103/PhysRevX.8.021074`. URL: `https://link.aps.org/doi/10.1103/PhysRevX.8.021074`.

[69] Tian Lan and Xiao-Gang Wen. "Classification of $3+1$D Bosonic Topological Orders (II): The Case When Some Pointlike Excitations Are Fermions." In: *Phys. Rev. X* 9 (2 2019), p. 021005. DOI: `10.1103/PhysRevX.9.021005`. URL: `https://link.aps.org/doi/10.1103/PhysRevX.9.021005`.

[70] B. P. Lanyon et al. "Efficient tomography of a quantum many-body system." In: *Nature Physics* 13.12 (2017), pp. 1158–1162. DOI: `10.1038/nphys4244`. URL: `https://doi.org/10.1038/nphys4244`.

[71] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." In: *Nature* 521.7553 (2015), pp. 436–444. ISSN: 1476-4687. DOI: `10.1038/nature14539`. URL: `https://doi.org/10.1038/nature14539`.

[72] Seunghoon Lee et al. "Evaluating the evidence for exponential quantum advantage in ground-state quantum chemistry." In: *Nature Communications* 14.1 (2023), p. 1952. ISSN: 2041-1723. DOI: `10.1038/s41467-023-37587-6`. URL: `https://doi.org/10.1038/s41467-023-37587-6`.

[73] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. *Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design*. 2017. arXiv: `1704.01552 [cs.LG]`.

[74] Lin Lin and Yu Tong. "Near-optimal ground state preparation." In: *Quantum* 4 (Dec. 2020), p. 372. ISSN: 2521-327X. DOI: `10.22331/q-2020-12-14-372`. URL: `https://doi.org/10.22331/q-2020-12-14-372`.

[75] Ke Liu, Jonas Greitemann, and Lode Pollet. "Learning multiple order parameters with interpretable machines." In: *Physical Review B* 99.10 (2019). ISSN: 2469-9969. DOI: `10.1103/physrevb.99.104410`. URL: `http://dx.doi.org/10.1103/PhysRevB.99.104410`.

[76] Ke Liu, Nicolas Sadoune, Nihal Rao, Jonas Greitemann, and Lode Pollet. "Revealing the phase diagram of Kitaev materials by machine learning: Cooperation and competition between spin liquids." In: *Phys. Rev. Res.* 3 (2 2021), p. 023016. DOI: `10.1103/PhysRevResearch.3.023016`. URL: `https://link.aps.org/doi/10.1103/PhysRevResearch.3.023016`.

[77] Dmitri Maslov. "Basic circuit compilation techniques for an ion-trap quantum machine." In: *New Journal of Physics* 19.2 (2017), p. 023035. DOI: `10.1088/1367-2630/aa5e47`. URL: `https://dx.doi.org/10.1088/1367-2630/aa5e47`.

[78]   Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. "A high-bias, low-variance introduction to Machine Learning for physicists." In: *Physics Reports* 810 (2019). A high-bias, low-variance introduction to Machine Learning for physicists, pp. 1–124. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2019.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157319300766.

[79]   Roger G. Melko, Giuseppe Carleo, Juan Carrasquilla, and J. Ignacio Cirac. "Restricted Boltzmann machines in quantum physics." In: *Nature Physics* 15.9 (2019), pp. 887–892. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0545-1. URL: https://doi.org/10.1038/s41567-019-0545-1.

[80]   Francesco Mezzadri. "How to Generate Random Matrices from the Classical Compact Groups." In: *NOTICES of the AMS* 54.5 (2007), pp. 592–604. DOI: 10.48550/arXiv.math-ph/0609050. URL: https://www.ams.org/notices/200705/fea-mezzadri-web.pdf.

[81]   Abel Molina and John Watrous. "Revisiting the simulation of quantum Turing machines by quantum circuits." In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475.2226 (2019), p. 20180767. DOI: 10.1098/rspa.2018.0767. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2018.0767. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0767.

[82]   Rahul M. Nandkishore and Michael Hermele. "Fractons." In: *Annual Review of Condensed Matter Physics* 10.1 (2019), pp. 295–313. DOI: 10.1146/annurev-conmatphys-031218-013604. URL: https://doi.org/10.1146/annurev-conmatphys-031218-013604.

[83]   Samuel Neyens et al. "Probing single electrons across 300-mm spin qubit wafers." In: *Nature* 629.8010 (2024), pp. 80–85. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07275-6. URL: https://doi.org/10.1038/s41586-024-07275-6.

[84]   Michael A. Nielsen. "Optical Quantum Computation Using Cluster States." In: *Phys. Rev. Lett.* 93 (4 2004), p. 040503. DOI: 10.1103/PhysRevLett.93.040503. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.040503.

[85]   Michael A. Nielsen. "Cluster-state quantum computation." In: *Reports on Mathematical Physics* 57.1 (2006), pp. 147–161. ISSN: 0034-4877. DOI: https://doi.org/10.1016/S0034-4877(06)80014-5. URL: https://www.sciencedirect.com/science/article/pii/S0034487706800145.

[86]   Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: `10.1017/CB09780511976667`.

[87]   Jaakko Nissinen, Ke Liu, Robert-Jan Slager, Kai Wu, and Jan Zaanen. "Classification of point-group-symmetric orientational ordering tensors." In: *Phys. Rev. E* 94 (2 2016), p. 022701. DOI: `10.1103/PhysRevE.94.022701`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.94.022701`.

[88]   J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. ISBN: 9780387400655. URL: `https://books.google.de/books?id=VbHYoSyelFcC`.

[89]   Vincent Noculak, Daniel Lozano-Gomez, Jaan Oitmaa, Rajiv R. P. Singh, Yasir Iqbal, Michel J. P. Gingras, and Johannes Reuther. "Classical and quantum phases of the pyrochlore S=1/2 magnet with Heisenberg and Dzyaloshinskii-Moriya interactions." In: *Physical Review B* 107.21 (June 2023), p. 214414. DOI: `10.1103/PhysRevB.107.214414`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.107.214414`.

[90]   Ryan O'Donnell and John Wright. "Efficient Quantum Tomography." In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '16. Cambridge, MA, USA: Association for Computing Machinery, 2016, 899–912. ISBN: 9781450341325. DOI: `10.1145/2897518.2897544`. URL: `https://doi.org/10.1145/2897518.2897544`.

[91]   Vern Paulsen. "Dilation Theorems." In: *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2003, 43–57. DOI: `10.1017/CB09780511546631.005`.

[92]   "Information and Thermodynamics." In: *Quantum Theory: Concepts and Methods*. Ed. by Asher Peres. Dordrecht: Springer Netherlands, 2002, pp. 285–288. ISBN: 978-0-306-47120-9. DOI: `10.1007/0-306-47120-5_9`. URL: `https://doi.org/10.1007/0-306-47120-5_9`.

[93]   D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. "Matrix Product State Representations." In: *Quantum Info. Comput.* 7.5 (2007), 401–430. ISSN: 1533-7146.

[94]   D. Pérez-García, M. M. Wolf, M. Sanz, F. Verstraete, and J. I. Cirac. "String Order and Symmetries in Quantum Spin Lattices." In: *Phys. Rev. Lett.* 100 (16 2008), p. 167202. DOI: `10.1103/PhysRevLett.100.167202`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.100.167202`.

[95] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor." In: *Nature Communications* 5.1 (2014), p. 4213. ISSN: 2041-1723. DOI: 10.1038/ncomms5213. URL: https://doi.org/10.1038/ncomms5213.

[96] I. Pogorelov et al. "Compact Ion-Trap Quantum Computing Demonstrator." In: *PRX Quantum* 2 (2 2021), p. 020343. DOI: 10.1103/PRXQuantum.2.020343. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.020343.

[97] Pedro Ponte and Roger G. Melko. "Kernel methods for interpretable machine learning of order parameters." In: *Phys. Rev. B* 96 (20 2017), p. 205146. DOI: 10.1103/PhysRevB.96.205146. URL: https://link.aps.org/doi/10.1103/PhysRevB.96.205146.

[98] Lukas Postler et al. "Demonstration of fault-tolerant universal quantum gate operations." In: *Nature* 605.7911 (2022), pp. 675–680. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04721-1. URL: https://doi.org/10.1038/s41586-022-04721-1.

[99] Abhinav Prem, Sagar Vijay, Yang-Zhi Chou, Michael Pretko, and Rahul M. Nandkishore. "Pinch point singularities of tensor spin liquids." In: *Phys. Rev. B* 98 (16 2018), p. 165140. DOI: 10.1103/PhysRevB.98.165140. URL: https://link.aps.org/doi/10.1103/PhysRevB.98.165140.

[100] John Preskill. "Quantum Computing in the NISQ era and beyond." In: *Quantum* 2 (Aug. 2018), p. 79. ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79. URL: https://doi.org/10.22331/q-2018-08-06-79.

[101] Michael Pretko. "Generalized electromagnetism of subdimensional particles: A spin liquid story." In: *Physical Review B* 96.3 (July 2017), p. 035119. DOI: 10.1103/PhysRevB.96.035119. URL: https://link.aps.org/doi/10.1103/PhysRevB.96.035119 (visited on 07/18/2017).

[102] Michael Pretko, Xie Chen, and Yizhi You. "Fracton phases of matter." en. In: *International Journal of Modern Physics A* 35.06 (Feb. 2020), p. 2030003. ISSN: 0217-751X, 1793-656X. DOI: 10.1142/S0217751X20300033. URL: https://www.worldscientific.com/doi/abs/10.1142/S0217751X20300033.

[103] Nihal Rao, Ke Liu, Marc Machaczek, and Lode Pollet. "Machine-learned phase diagrams of generalized Kitaev honeycomb magnets." In: *Phys. Rev. Res.* 3 (3 2021), p. 033223. DOI: 10.1103/PhysRevResearch.3.033223. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.3.033223.

[104] Nihal Rao, Ke Liu, and Lode Pollet. "Inferring hidden symmetries of exotic magnets from detecting explicit order parameters." In: *Phys. Rev. E* 104 (1 2021), p. 015311. DOI: `10.1103/PhysRevE.104.015311`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.104.015311`.

[105] Alex Rasmussen, Yi-Zhuang You, and Cenke Xu. *Stable Gapless Bose Liquid Phases without any Symmetry*. 2016. arXiv: `1601.08235 [cond-mat.str-el]`. URL: `https://arxiv.org/abs/1601.08235`.

[106] J. G. Rau et al. "Anisotropic Exchange within Decoupled Tetrahedra in the Quantum Breathing Pyrochlore $Ba_3Yb_2Zn_5O_{11}$." In: *Phys. Rev. Lett.* 116 (25 2016), p. 257204. DOI: `10.1103/PhysRevLett.116.257204`. URL: `http://link.aps.org/doi/10.1103/PhysRevLett.116.257204`.

[107] Robert Raussendorf and Hans J. Briegel. "A One-Way Quantum Computer." In: *Phys. Rev. Lett.* 86 (22 2001), pp. 5188–5191. DOI: `10.1103/PhysRevLett.86.5188`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.86.5188`.

[108] Joseph M. Renes, Robin Blume-Kohout, A. J. Scott, and Carlton M. Caves. "Symmetric informationally complete quantum measurements." In: *Journal of Mathematical Physics* 45.6 (May 2004), pp. 2171–2180.

[109] Martin Ringbauer, Michael Meth, Lukas Postler, Roman Stricker, Rainer Blatt, Philipp Schindler, and Thomas Monz. "A universal qudit quantum processor with trapped ions." In: *Nature Physics* 18.9 (2022), pp. 1053–1057. ISSN: 1745-2481. DOI: `10.1038/s41567-022-01658-0`. URL: `https://doi.org/10.1038/s41567-022-01658-0`.

[110] Joaquin F. Rodriguez-Nieva and Mathias S. Scheurer. "Identifying topological order through unsupervised machine learning." In: *Nature Physics* 15.8 (2019), pp. 790–795. DOI: `10.1038/s41567-019-0512-x`. URL: `https://doi.org/10.1038/s41567-019-0512-x`.

[111] K. A. Ross, L. Savary, B. D. Gaulin, and L. Balents. "Quantum Excitations in Quantum Spin Ice." In: *Physical Review X* 1 (2011), p. 021002.

[112] Mathias S. Scheurer and Robert-Jan Slager. "Unsupervised Machine Learning and Band Topology." In: *Phys. Rev. Lett.* 124 (22 2020), p. 226401. DOI: `10.1103/PhysRevLett.124.226401`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.124.226401`.

[113] Bernhard Schölkopf, Alex J Smola, Robert C Williamson, and Peter L Bartlett. "New support vector algorithms." In: *Neural Comput.* 12.5 (2000), pp. 1207–1245. DOI: `10.1162/089976600300015565`.

[114] Ulrich Schollwöck. "The density-matrix renormalization group in the age of matrix product states." In: *Annals of Physics* 326.1 (2011). January 2011 Special Issue, pp. 96–192. ISSN: 0003-4916. DOI: https://doi.org/10.1016/j.aop.2010.09.012. URL: https://www.sciencedirect.com/science/article/pii/S0003491610001752.

[115] A J Scott. "Tight informationally complete quantum measurements." In: *Journal of Physics A: Mathematical and General* 39.43 (2006), p. 13507. DOI: 10.1088/0305-4470/39/43/009. URL: https://dx.doi.org/10.1088/0305-4470/39/43/009.

[116] Dan Sehayek and Roger G. Melko. "Persistent homology of $\mathbb{Z}_2$ gauge theories." In: *Phys. Rev. B* 106 (8 2022), p. 085111. DOI: 10.1103/PhysRevB.106.085111. URL: https://link.aps.org/doi/10.1103/PhysRevB.106.085111.

[117] Nic Shannon, Karlo Penc, and Yukitoshi Motome. "Nematic, vector-multipole, and plateau-liquid states in the classical O(3) pyrochlore antiferromagnet with biquadratic interactions in applied magnetic field." In: *Phys. Rev. B* 81.18 (May 2010), p. 184409. DOI: DOI10.1103/PhysRevB.81.184409. URL: http://dx.doi.org/10.1103/PhysRevB.81.184409.

[118] E.F. Shender. "Antiferromagnetic garnets with fluctuationally interacting sublattices." In: *Sov. Phys. JETP* 56 (1982), p. 178.

[119] Saumya Shivam, Curt W. von Keyserlingk, and Shivaji L. Sondhi. "On classical and hybrid shadows of quantum states." In: *SciPost Phys.* 14 (2023), p. 094. DOI: 10.21468/SciPostPhys.14.5.094. URL: https://scipost.org/10.21468/SciPostPhys.14.5.094.

[120] Peter W. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer." In: *SIAM Journal on Computing* 26.5 (1997), pp. 1484–1509. DOI: 10.1137/S0097539795293172. eprint: https://doi.org/10.1137/S0097539795293172. URL: https://doi.org/10.1137/S0097539795293172.

[121] M. P. da Silva et al. *Demonstration of logical qubits and repeated error correction with better-than-physical error rates.* 2024. arXiv: 2404.02280 [quant-ph]. URL: https://arxiv.org/abs/2404.02280.

[122] Adam Smith, Bernhard Jobst, Andrew G. Green, and Frank Pollmann. "Crossing a topological phase transition with a quantum computer." In: *Phys. Rev. Research* 4 (2 2022), p. L022020. DOI: 10.1103/PhysRevResearch.4.L022020. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.L022020.

[123]   Chao Song et al. "10-Qubit Entanglement and Parallel Logic Operations with a Superconducting Circuit." In: *Phys. Rev. Lett.* 119 (18 2017), p. 180511. DOI: 10.1103/PhysRevLett.119.180511. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.180511.

[124]   Anders Sørensen and Klaus Mølmer. "Quantum Computation with Ions in Thermal Motion." In: *Phys. Rev. Lett.* 82 (9 1999), pp. 1971–1974. DOI: 10.1103/PhysRevLett.82.1971. URL: https://link.aps.org/doi/10.1103/PhysRevLett.82.1971.

[125]   R. Srinivas et al. "High-fidelity laser-free universal control of trapped ion qubits." In: *Nature* 597.7875 (2021), pp. 209–213. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03809-4. URL: https://doi.org/10.1038/s41586-021-03809-4.

[126]   Aaron Szasz, Ed Younis, and Wibe De Jong. "Numerical Circuit Synthesis and Compilation for Multi-State Preparation." In: *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 01. 2023, pp. 768–778. DOI: 10.1109/QCE57702.2023.00092.

[127]   Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. "Neural-network quantum state tomography." In: *Nature Physics* 14.5 (2018), pp. 447–450. DOI: 10.1038/s41567-018-0048-5. URL: https://doi.org/10.1038/s41567-018-0048-5.

[128]   Vladimir N. Vapnik. *Statistical Learning Theory.* Wiley, 1998. ISBN: 9780471030034.

[129]   B. Vermersch, A. Elben, L. M. Sieberer, N. Y. Yao, and P. Zoller. "Probing Scrambling Using Statistical Correlations between Randomized Measurements." In: *Phys. Rev. X* 9 (2 2019), p. 021061. DOI: 10.1103/PhysRevX.9.021061. URL: https://link.aps.org/doi/10.1103/PhysRevX.9.021061.

[130]   Ruben Verresen, Roderich Moessner, and Frank Pollmann. "One-dimensional symmetry protected topological phases and their transitions." In: *Phys. Rev. B* 96 (16 2017), p. 165124. DOI: 10.1103/PhysRevB.96.165124. URL: https://link.aps.org/doi/10.1103/PhysRevB.96.165124.

[131]   F. Verstraete and J. I. Cirac. "Valence-bond states for quantum computation." In: *Phys. Rev. A* 70 (6 2004), p. 060302. DOI: 10.1103/PhysRevA.70.060302. URL: https://link.aps.org/doi/10.1103/PhysRevA.70.060302.

[132]   Sagar Vijay, Jeongwan Haah, and Liang Fu. "Fracton topological order, generalized lattice gauge theory, and duality." In: *Phys. Rev. B* 94 (23 2016), p. 235157. DOI: 10.1103/PhysRevB.94.235157. URL: https://link.aps.org/doi/10.1103/PhysRevB.94.235157.

[133] John Watrous. *Quantum Computational Complexity*. 2008. arXiv: 0804.3401 [quant-ph]. URL: https://arxiv.org/abs/0804.3401.

[134] Michael M. Wolf, Gerardo Ortiz, Frank Verstraete, and J. Ignacio Cirac. "Quantum Phase Transitions in Matrix Product Systems." In: *Phys. Rev. Lett.* 97 (11 2006), p. 110403. DOI: 10.1103/PhysRevLett.97.110403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.97.110403.

[135] William K Wootters and Brian D Fields. "Optimal state-determination by mutually unbiased measurements." In: *Annals of Physics* 191.2 (1989), pp. 363–381. ISSN: 0003-4916. DOI: https://doi.org/10.1016/0003-4916(89)90322-9. URL: https://www.sciencedirect.com/science/article/pii/0003491689903229.

[136] Fengcheng Wu, Youjin Deng, and Nikolay Prokof'ev. "Phase diagram of the toric code model in a parallel magnetic field." In: *Phys. Rev. B* 85 (19 2012), p. 195104. DOI: 10.1103/PhysRevB.85.195104. URL: https://link.aps.org/doi/10.1103/PhysRevB.85.195104.

[137] Han Yan, Owen Benton, L. D. C. Jaubert, and Nic Shannon. "Rank–2 $U(1)$ Spin Liquid on the Breathing Pyrochlore Lattice." In: *Phys. Rev. Lett.* 124 (12 2020), p. 127203. DOI: 10.1103/PhysRevLett.124.127203. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.127203.

[138] Han Yan, Owen Benton, Ludovic Jaubert, and Nic Shannon. "Theory of multiple-phase competition in pyrochlore magnets with anisotropic exchange with application to $Yb_2Ti_2O_7$, $Er_2Ti_2O_7$, and $Er_2Sn_2O_7$." In: *Phys. Rev. B* 95 (9 2017), p. 094422. DOI: 10.1103/PhysRevB.95.094422. URL: https://link.aps.org/doi/10.1103/PhysRevB.95.094422.

[139] Ed Younis, Costin C Iancu, Wim Lavrijsen, Marc Davis, Ethan Smith, and USDOE. *Berkeley Quantum Synthesis Toolkit (BQSKit) v1*. Apr. 2021. DOI: 10.11578/dc.20210603.2. URL: https://www.osti.gov//servlets/purl/1785933.

[140] "Yuri I. Manin. Mathematics as Metaphor: Selected Essays of Yuri I. Manin." In: *Philosophia Mathematica* 17.1 (Apr. 2008), pp. 122–123. ISSN: 0031-8019. DOI: 10.1093/philmat/nkn013. eprint: https://academic.oup.com/philmat/article-pdf/17/1/122/4237389/nkn013.pdf. URL: https://doi.org/10.1093/philmat/nkn013.

[141] Emily Z. Zhang, Finn Lasse Buessen, and Yong Baek Kim. "Dynamical signatures of rank-2 $U(1)$ spin liquids." In: *Phys. Rev. B* 105 (6 2022), p. L060408. DOI: 10.1103/PhysRevB.105.L060408. URL: https://link.aps.org/doi/10.1103/PhysRevB.105.L060408.

[142] Yi Zhang, Roger G. Melko, and Eun-Ah Kim. "Machine learn-
       ing $\mathbb{Z}_2$ quantum spin liquids with quasiparticle statistics." In:
       Phys. Rev. B 96 (24 2017), p. 245119. DOI: 10.1103/PhysRevB.96.
       245119. URL: https://link.aps.org/doi/10.1103/PhysRevB.
       96.245119.