SISSA

Scuola
Internazionale
Superiore di
Studi Avanzati

Physics Area - PhD course in
Theory and Numerical Simulation of Condensed Matter

Quantum Approximate Optimization Algorithm and Variational Quantum Computing: from binary neural networks to ground state preparation

Thesis submitted for the degree of
Doctor Philosophiae

Candidate:

Pietro Torta

Advisor:

Giuseppe E. Santoro

Academic Year 2022-2023

Abstract

In this Thesis, I explore the domain of hybrid quantum-classical computation, the foremost approach for utilizing Noisy Intermediate-Scale Quantum (NISQ) devices. The opening Chapter presents an overview of Variational Quantum Algorithms (VQAs), highlighting the primary algorithmic challenges. It offers an in-depth review of the Quantum Approximate Optimization Algorithm (QAOA), including a discussion of its variants for ground state preparation. My presentation focuses on the connection of QAOA with Adiabatic Quantum Computation, the patterns in optimal QAOA parameters, and the ability to transfer optimal schedules across typical instances or from smaller to larger problem sizes. In the second Chapter, similar techniques are proven effective in training QAOA for the supervised learning of a simple Binary Neural Network. This model represents an idealized yet prototypical example of classical combinatorial optimization problems involving multi-spin interactions. I remark on potential extensions of our work that could be amenable to direct implementation on NISQ devices. In the third Chapter, the discussion shifts toward quantum many-body ground state preparation, focusing on the one-dimensional Heisenberg XYZ model and Longitudinal-Transverse-Field Ising Model (LTFIM). We have developed a novel technique that, at any point in the phase diagram, leverages the transferability of a specific class of optimal schedules from systems with small to those with larger numbers of qubits. Remarkably, a classically computed solution for a small system may serve as a warm start for large-scale ground state preparation on actual quantum devices. A warm-start optimization is shown to be free from trainability issues, specifically vanishing gradients (Barren Plateaus). In the fourth Chapter, we tailor a QAOA scheme to characterize a topological quantum phase transition within a $\mathbb{Z}_2$ lattice gauge theory model. This investigation is particularly significant due to its implications for high-energy physics and relevance to quantum error correction and surface codes. Our methods may serve as a solid ground to develop techniques aimed at investigating the dynamical properties and excitations of analogous models. Appendices include supplementary results on our case studies. In the concluding Chapter, I propose new stimulating research directions and help to identify core challenges and unresolved questions in variational quantum computing that transcend any particular application domain.

Acknowledgements

For the non-Italian friends: copy and paste the Italian text on GPT. Indeed, borrowing its own words: "It is capable of delivering a more accurate translation than the thesis Author."

Non ho mai scritto una sezione di ringraziamenti prima d'ora, né per la tesi triennale né per quella magistrale. Non perché non avessi persone da ringraziare, figuriamoci, ma perché mi è sempre sembrato un po' autocelebrativo e fine a se stesso. Ma crescere significa anche mettere in discussione le proprie convinzioni, quindi eccoci qui.

La mia avventura di vita a Trieste è iniziata poco prima del 2020, un periodo non facile per cambiare città. Ma poi la vita si è fatta più semplice, ho scoperto una città meravigliosa, che qualcuno definisce un idillio borghese. San Giacomo ha poco di borghese ma molto di idilliaco: grazie Trieste.

E grazie anche alla SISSA. Senza retorica, il livello scientifico delle persone incontrate è *remarkable*, come a molti scienziati piace scrivere frequentemente nei propri articoli. In pochi hanno la fortuna di vivere e lavorare immersi in un parco naturale con vista dall'alto sul golfo e sul mare. Tutto ciò mi manca già.

Ci tengo a ringraziare anche le persone, non solo i luoghi. Ringrazio il mio supervisor Giuseppe Santoro, per aver creato un ambiente lavorativo privo di stress e formalità, e per la sua fiducia nell'avermi lasciato grande libertà accademica e di esplorazione in questi anni. Ringrazio anche Mario Collura, che è stato (non formalmente, ma nella sostanza) un ottimo co-supervisor.

Naturalmente ringrazio i miei amici e collaboratori che hanno fatto parte della mia vita a Trieste: Gabriele, Guglielmo, Francesco, Nishan, Mina, Alì, Matteo, Giovanni, Ruiyi e tanti altri, in forme e tempi diversi.

Ringrazio infine i miei genitori Anna e Freddy per il supporto e il sostegno in questi anni, soprattutto nei momenti più complessi. Grazie anche ai miei amici e alle persone che mi vogliono bene, soprattutto a quelle con cui i rapporti sono cambiati, ma ci sono molte cose che restano.

Publications

This thesis reports — in Chapters 2, 3 and 4 — results from the following publications:

[1] P. Torta, G. B. Mbeng, C. Baldassi, R. Zecchina, and G. E. Santoro, *Quantum approximate optimization algorithm applied to the binary perceptron*, Phys. Rev. B **107**, 094202 (2023)

[2] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, *Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz*, Phys. Rev. A **106**, L060401 (2022)

[3] L. Lumia, P. Torta, G. B. Mbeng, G. E. Santoro, E. Ercolessi, M. Burrello, and M. M. Wauters, *Two-Dimensional $\mathbb{Z}_2$ Lattice Gauge Theory on a Near-Term Quantum Simulator: Variational Quantum Optimization, Confinement, and Topological Order*, PRX Quantum **3**, 020320 (2022)

During the course of the PhD, I have further authored and contributed to the following publication, which will not be discussed in this thesis.

[4] G. Lami, P. Torta, G. E. Santoro, and M. Collura, *Quantum annealing for neural network optimization problems: A new approach via tensor network simulations*, SciPost Phys. **14**, 117 (2023)

New works are currently being prepared with two main focuses. The first extends the methods of [1] to general binary neural network architectures. The second explores the interplay between the techniques employed in [2, 3] and quantum optimal control.

Contents

Contents

Acronym	Full Form
NISQ	Noisy Intermediate-Scale Quantum
VQA	Variational Quantum Algorithms
QML	Quantum Machine Learning
VQE	Variational Quantum Eigensolver
QAOA	Quantum Approximate Optimization Algorithm
PQC	Parameterized Quantum Circuit
AQC	Adiabatic Quantum Computation
QA	Quantum Annealing
dQA	digitized Quantum Annealing
IID	Independent and Identically Distributed
ANN	Artificial Neural Network
BP	Barren Plateaus
HVA	Hamiltonian Variational Ansatz
QUBO	Quadratic Unconstrained Binary Optimization
BiNNs	Binary Neural Networks
TFIM	Transverse-Field Ising Model
LTFIM	Longitudinal-Transverse-Field Ising Model

1. Introduction

> *Can you do it with a new kind of computer – a quantum computer?*
> *[. . .] It's not a Turing machine, but a machine of a different kind.*
>
> —Richard P. Feynman, May 1981

Computation and algorithms naturally arise in any logical and quantitative endeavor and have been the subject of speculation and practical investigation for centuries (or millennia), even before the "recent" advent of the modern scientific method. The undisputed leap forward in the computational capabilities of humankind was the dawn of the Information Age in the middle of the twentieth century: a symbolic starting date was 1947, the year of the invention of the transistor.

In the last 70 years, digital computers have constantly been at the forefront of technological advancement, reshaping industries and our daily lives, and leading to an unprecedented pace of scientific progress and technological discoveries, including the current ongoing revolution of Artificial Intelligence (AI).

From a historical perspective, it might be surprising that already in the 1980s, well before the advent of the World Wide Web, physicists started to conjecture about *quantum computers*. Quantum computers can be broadly intended as computational models leveraging inherent properties of quantum mechanics, namely *quantum superposition* and *entanglement* [5], to perform useful computation.

Perhaps ironically, the basic building block of every day "classical" computers (your laptop), namely the aforementioned transistor, is ultimately a quantum device, and its progressive technological improvement and miniaturization did require an operative and theoretical understanding of quantum mechanics. Nonetheless, even the most recent nanometer-scale transistors – featuring less than 10^{11} atoms – are operated classically, as their usage is rooted in classical physics and Boolean algebra. In simple terms, a transistor can only be in either of two distinct logical states: it can conduct current if a bias voltage is applied, and this is a bit 1, or not, and this is a bit 0. A bit is the basic unit of classical information.

In modern terms, a quantum computer is, borrowing Feynman's words "*a machine of a different kind*", which relies on a new unit of *quantum* information, the quantum bit or *qubit*. Intuitively, a qubit can be not only in the states 0 and 1, but also in a quantum superposition of the two: elementary quantum mechanics proves that

the possible states of a qubit are in one-to-one correspondence with the points of a (Bloch) sphere, specified by two angles (very much alike the latitude and the longitude of a globe) [5].

In what follows, we briefly summarize the state of the art of quantum computing, focusing on the difference between fault-tolerant quantum computation and the newer and faster-developing field of Noisy Intermediate-Scale Quantum (NISQ) devices [6].

1.1. Two perspectives on quantum computing

The early scientific speculation on quantum computing [7] focused on a simple yet far-reaching suggestion: if Nature behaves quantum-mechanically, it might be more convenient to simulate it with a quantum computer, rather than a classical one. This is what we call quantum simulation: it is perhaps the most natural and straightforward application of quantum computing, rooted in the seminal ideas of Feynman in 1980s [8] and later proved possible by S. Lloyd [9].

An even more ambitious question was posed in the same early years, notably by D. Deutsch in 1985 [10]: could quantum computers offer any advantage over classical computation at solving problems that have *nothing to do* with quantum physics? In other terms, could a hypothetical perfectly working quantum computer impact technological and scientific endeavors beyond the mere simulation of quantum physics?

The first positive answer dates back to 1994, when P. Shor invented an *efficient* quantum algorithm for determining prime factors of composite integers [11], a classically hard problem that lies at the heart of some of the most used cryptographic protocols. The interest in quantum computing skyrocketed. The exponential computational advantage offered by a hypothetical quantum computer over the best-known classical algorithm would have entailed the crack of many secure IT systems [1].

However, even back then, it was immediately clear that these neat theoretical and algorithmic advances were at least decades ahead of the experimental platforms required for their functioning: a quantum computer simply did not exist, and there were serious concerns and a strong skepticism that it could ever be realized. Indeed, transistors (the classical bits of our computers) are quite resilient to thermal and electrical noise, and the classical errors (essentially, unwanted bit flips from 0 to 1 or vice versa) are quite easily corrected.

The opposite is true for qubits: these microscopic bits of quantum information

---

[1]By now, different post-quantum cryptographic algorithms have been devised, which would not be vulnerable to quantum attacks. It is suggestive that the first commercial application of quantum technologies is usually associated with quantum cryptography, specifically Quantum Key Distribution (QKD).

are extremely sensitive to noise and very prone to lose their quantum information, a phenomenon known as *decoherence*. Additionally, even if the quantum information is preserved for a certain period of *coherence time*, the qubit state can be scrambled in many more ways (the latitude and the longitude of the point on the Bloch sphere can be randomly altered). If more qubits are present, the errors might even correlate, potentially leading to a nightmare.

It was only two years later, in 1996, when the first quantum error-correcting codes were devised [12, 13], leading to *fault-tolerant quantum computation*. This is a true milestone in quantum computing: it was proven, at least in principle, that a quantum computer built with noisy hardware can be scaled up to large devices and be employed to solve hard problems, as long as the frequency and the correlations of errors are kept under certain thresholds [14].

Let us fast-forward to today. Celebrated quantum algorithms with exponential speed-up over classical methods include Shor's algorithm for factoring integers [11], the quantum simulation of quantum systems [9], and the HHL algorithm for the solution of some kinds of linear systems of equations [15]. Quantum simulation, in particular quantum chemistry [16], may provide the first real-world impact in the not-so-distant future: it has the true potential of innovation for strategic areas, ranging from improvements in human health (pharmaceuticals), agriculture, and pollution (nitrogen fixation), and the sustainability of ecosystems (energy storage [17] and production). Another renowned example is Grover's search algorithm [18], which exhibits a *provable* quadratic speedup over the *best possible* classical algorithm for search in an unstructured database [2].

However, to run these algorithms with quantum error correction, *millions* of physical qubits would be required. Remarkably, and against the bets of the most skeptical scientists, it took only a few decades to implement the first "quantum computers" based on a variety of different technological platforms. Regrettably, at the moment, these devices feature at most a few hundred (or thousands, in the near-term future) qubits, with significant shortcomings in terms of noise levels and usability [3]. See Ref. [19] for a review of the state of the art of experimental quantum technologies.

In summary, a universal fault-tolerant quantum computer is still far from experimental reach. For this reason, the current "quantum computers" are better known as Noisy Intermediate-Scale Quantum(NISQ) devices [6]. In the words of distinguished scientist J. Preskill, inventor of the acronym: *"Intermediate-scale" conveys that to-*

---

[2]Despite being one of the conceptual milestones of quantum computing, this theoretical quadratic speedup relies on a highly idealized assumption: the classical and quantum computers should have the same clock speed (number of elementary operations per second). This and other practical issues led to some skepticism on the actual realization of this speedup.

[3]In particular, depending on the experimental platform, these shortcomings include but are not limited to: the qubit-connectivity map to perform entangling 2-qubit gates, short coherence time, error rates for native single- and two-qubit gates, and readout (measurement) errors.

day's quantum devices with more than 50 well-controlled qubits cannot be simulated by brute force using the most powerful currently existing classical supercomputers; "noisy" reminds us that these devices are not error corrected and that the noise limits their computational power [7].

A novel perspective on quantum computing naturally emerges from the first experimental implementations of quantum devices: can we employ this NISQ technology to do anything useful? Are NISQ devices capable of providing some kind of quantum speed-up over classical computation [20]? These questions, at the moment, remain open [4]. Nowadays, most NISQ algorithms [21] are *hybrid* in nature: they rely both on a *classical* and a *quantum* processor, delegating the quantum device to solve only the classically-hard tasks of a larger computation. These hybrid quantum-classical algorithms, known as Variational Quantum Algorithms (VQAs) [22], are covered in the next section. Very recently, there has been substantial progress in the application of quantum error-correcting strategies to near-term devices [23], paving the way to the first implementations of error-corrected algorithms.

Here ends the non-technical introduction and motivation for this Thesis. A summary of quantum mechanics and quantum computation is well beyond the present scope, and very good references exist [5, 24, 25]. From now on, we assume knowledge of quantum circuits and quantum gates, the difference between digital gate-based and analog quantum devices (such as quantum annealers), and the main tools and concepts in quantum many-body physics.

1.2. Variational Quantum Algorithms

> *All you need in this life is ignorance and confidence; then success is sure.*
>
> —Mark Twain

In broad terms, Variational Quantum Algorithms (VQAs) are a class of hybrid quantum-classical computational schemes that do not require quantum error correction, certainly among the main candidates for near-term practical applications of NISQ devices. These algorithms have recently raised an unprecedented interest in the scientific and industrial community, leading to an ever-growing body of litera-

---

[4]Recent years have witnessed major investments in quantum computing by big tech companies, intense funding for public and private institutions, and the rise of many quantum start-ups. Although some of these processes might be reminiscent of speculative (or hype) bubbles, I find the view of J. Preskill again inspiring: *Though I'm confident that quantum computing will have a transformative impact on society eventually, that impact might still be a few decades away. No one knows for sure how long it will take to get there. Keep in mind, though, that the technology is still at an early stage, with many competing approaches, so an unanticipated breakthrough could change the outlook suddenly* [7].
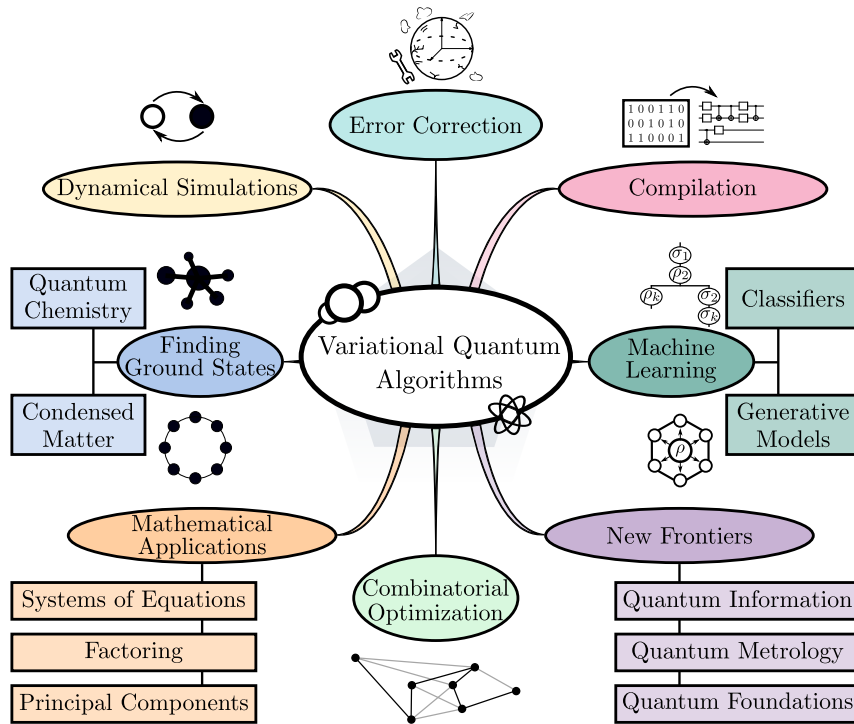
Figure 1.1.: An overview of the main fields of application for Variational Quantum Algorithms. In this Thesis, our primary focus is on finding ground states for condensed matter systems and solving combinatorial optimization problems. Although these problems appear quite distinct, they can be unified under a simplified variational framework, by defining an appropriate quantum or classical *target* Hamiltonian, as described in this Introduction. This figure is adapted from Ref. [22].

ture on many diverse applications. The main active research areas are summarized in Fig. 1.1, which is borrowed from Ref. [22]. A few examples include finding quantum ground states [3, 26, 27], classical combinatorial optimization [28], the solution of linear systems of equations [29], and applications to quantum information [30]. Other promising VQA applications are represented by finding excited states and approximately simulating quantum dynamics in real or imaginary time. A particularly active field is Quantum Machine Learning (QML), often based on hybrid quantum-classical schemes. Recently, dequantization techniques [31] ruled out an exponential advantage for promising QML applications on classical data. Despite the possibility of a polynomial advantage of QML algorithms over classical computation, QML on *quantum data* (either from a physical quantum system or a NISQ device) looks a more promising avenue for a relevant quantum speedup. Any attempt to provide a comprehensive list of publications on even a single area of application for VQAs would be a daunting task. We refer the interested reader to two accurate reviews on the state of the art [21, 22].

The two earliest proposals of VQAs have been the Variational Quantum Eigensolver (VQE) [32] and the Quantum Approximate Optimization Algorithm (QAOA) [28]. Historically, both were devised in 2014: the former was designed for quantum

ground state preparation, the latter for classical combinatorial optimization. Nevertheless, on a practical level, they both rely on the same VQA framework: the classically difficult part of the algorithm — i.e. preparing and measuring non-trivial quantum states — is delegated to a quantum device, whereas the remaining computations are performed on a standard classical computer. Depending on the experimental technology, it is often assumed that a gate-based quantum device with a universal set of quantum gates is used. The quantum state is represented as a Parameterized Quantum Circuit (PQC), iteratively updated by classical machinery, with the goal of minimizing a cost function.

In what follows, we will dive into the general formulation of VQAs. We will then introduce the concepts of Adiabatic Quantum Computation (AQC) and Quantum Annealing (QA), focusing on a digitized version of QA and its connection with QAOA. The latter is employed in the following chapters, both in its original [28] and modified formulations [33].

1.2.1. VQA general framework

VQAs are designed to mitigate the principal limitations of contemporary quantum hardware, including a limited number of qubits, limited qubit connectivity, and noise processes (coherent and incoherent errors) that limit the circuit depth [5]. Indeed, a shallow-depth PQC is employed, and the optimization of its parameters is performed using a classical algorithm.

Proposals for VQAs span a broad spectrum of scientific and technological disciplines, wherever quantum computing is expected to offer advantages over classical algorithms, in the quest for a quantum speed up [20]. This class of algorithms relies on a hybrid quantum-classical optimization scheme that is largely independent, in its abstract formulation, of the specific domain of application or the experimental support for the quantum circuit implementation.

Let us now describe the building blocks of a VQA, as schematically depicted in Fig 1.2. To begin with, the problem should be formulated as the optimization of a cost function C, which depends on a set of continuous or discrete real parameters $\boldsymbol{\theta}$ embedded in the PQC. The latter plays the role of an ansatz, which is iteratively updated to solve the optimization task

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} C(\boldsymbol{\theta}).\tag{1.1}$$

---

[5]In a quantum circuit, each qubit is sequentially involved in single or two-qubit operations, e.g. Pauli rotations and CNOTs. The circuit depth refers to the maximum number of quantum gates that need to be applied in sequence since they can not be parallelized. It is related to the total time required to perform the computation, which can be lowered by optimizing the quantum compilation of a high-level quantum algorithm into native gates of a given quantum hardware.
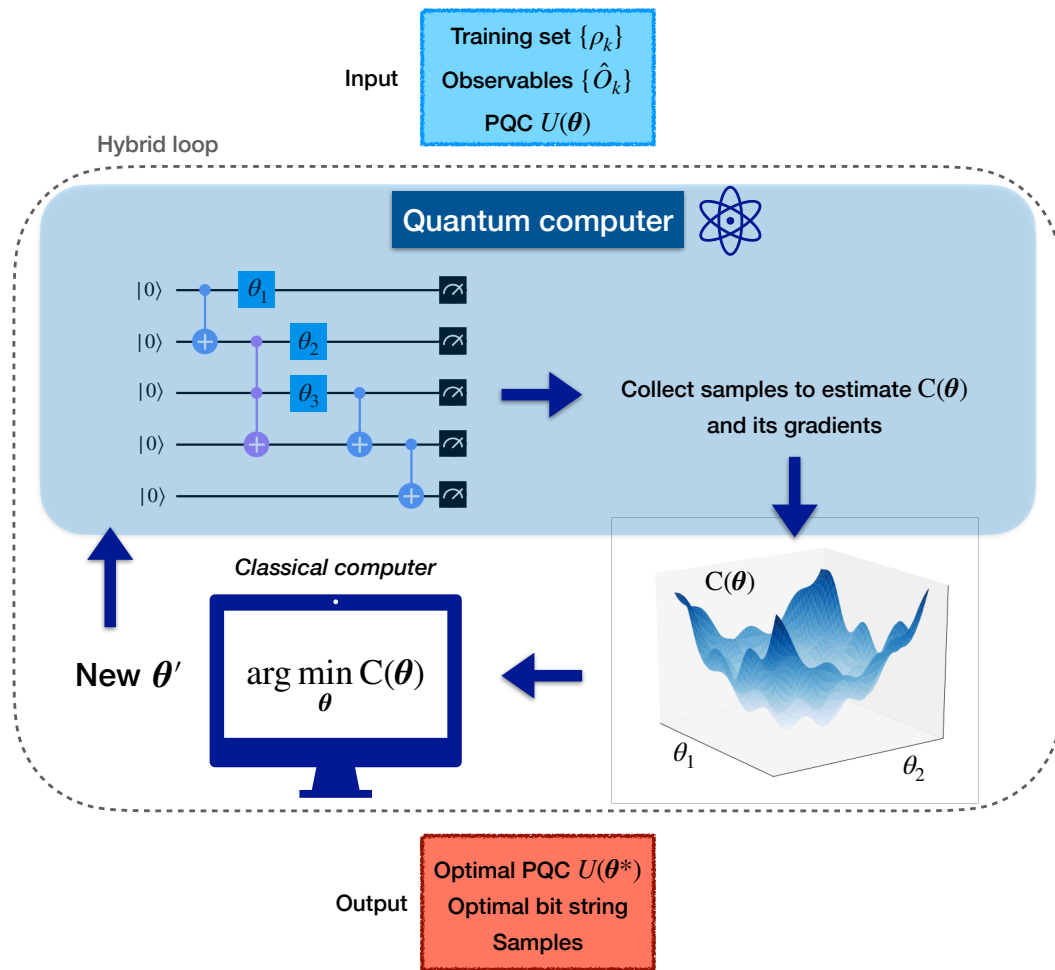
Figure 1.2.: Hybrid quantum-classical loop of a VQA, alongside its typical input and output. The quantum device is utilized only for classically hard tasks, as depicted in the shaded region. At each iterative step, the variational angles $\boldsymbol{\theta}$ — encoded in the PQC — are fixed: The quantum device is used for *repeated* state preparations and measurements. Here, we show the simpler case of a single input state $\rho_k = |\psi_0\rangle\langle\psi_0|$ (with an initial blank register), see Eq. 1.4. Once the state is prepared, qubits are measured in some bases, and the outcomes are collected to estimate the cost function $C(\boldsymbol{\theta})$ and its gradients. A pictorial two-dimensional cut of the cost function $C(\boldsymbol{\theta})$ is shown: the cost function is non-convex, implying a challenging optimization. The estimated value of $C(\boldsymbol{\theta})$ (and its gradients) are passed to a classical computer, where an optimization routine performs the step $\boldsymbol{\theta} \to \boldsymbol{\theta}'$, proposing a new improved set of variational parameters. A new PQC is then repeatedly prepared and measured in the next iteration.

The crucial feature of a VQA is the synergy between quantum and classical computational resources. At each iteration, the quantum hardware is used to prepare the ansatz and perform measurements on it, yielding estimates for $C(\boldsymbol{\theta})$ and its gradients, if needed. These values are then passed to a classical computer, which performs an optimization step by updating the set of variational parameters $\boldsymbol{\theta}$. This procedure is iterated up to convergence, or an exit condition is met. The outcome of

the algorithm is usually the optimal PQC, or related quantities, such as a collection of bit strings (or a single optimal bit string, like in QAOA).

Following Ref. [22], in very general terms, the cost function may be written as

$$C(\boldsymbol{\theta}) = f\big(\{\rho_k\}, \{\widehat{O}_k\}, U(\boldsymbol{\theta})\big)\,, \tag{1.2}$$

where f is some function, $U(\boldsymbol{\theta})$ is a parameterized unitary, $\{\rho_k\}$ is a set of input (or initial) states that can be interpreted as a training set, and $\{\widehat{O}_k\}$ is a set of observables. In practice, most implementations of a variational scheme rely on a cost function in the form

$$C(\boldsymbol{\theta}) = \sum_k f_k\big(\mathrm{Tr}[\widehat{O}_k U(\boldsymbol{\theta}) \rho_k U^\dagger(\boldsymbol{\theta})]\big)\,. \tag{1.3}$$

More explicitly, each initial state ρ_k is time-evolved by the parameterized unitary, and then used to compute an expectation value for each observable $\widehat{O}_k$: these expectation values are post-processed by applying functions f_k and then summed together, to yield the total cost. The most common case, e.g. for quantum ground state preparation and classical optimization, reduces to the well-known variational principle, adopted for instance in Variational Monte Carlo techniques. This is obtained by simply dropping the sum over k, and choosing $f_k(x) = x$, $\rho_k = |\psi_0\rangle\langle\psi_0|$, leading to

$$C(\boldsymbol{\theta}) = \langle\psi_0|U^\dagger(\boldsymbol{\theta})\widehat{O}U(\boldsymbol{\theta})|\psi_0\rangle\,, \tag{1.4}$$

where the state ψ_0 is an easily prepared initial state (often a product state), and the observable $\widehat{O}$ coincides with the quantum Hamiltonian, whose ground state we want to find, or the classical optimization cost function embedded in a quantum setup via standard base encoding [5] (see Eq. 1.13).

This abstract formulation may hide several complex tasks and open challenges that are involved, in practice, to devise a meaningful VQA. Let us now focus on the specific case of Eq. 1.4 for simplicity of notation, even though the discussion extends straightforwardly to the general case. The definition of the cost function C is crucial for many reasons. Obviously, it must be "faithful", meaning that its minimum should coincide with the solution of the problem. However, since the global minimum is usually found only in simple idealized scenarios, it should also be "operationally meaningful", i.e. lower values should indicate progressively better solutions. The choice of the observable $\widehat{O}$ is usually quite obvious, as in the aforementioned cases of ground state preparation or classical optimization tasks. Incidentally, one should usually avoid *global* cost functions, such as the fidelity with respect to a known quantum state, as this would lead to a poorly trainable regime and vanishing gradients of C [34].

The real challenge resides in the definition of a PQC $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$, i.e. of the variational quantum state obtained by applying the parameterized unitary $U(\boldsymbol{\theta})$ to the simple initial state ψ_0. First, the physical hardware limitations should be

accounted for, both concerning qubit number and connectivity and in terms of circuit depth. Secondly, the PQC should be *efficiently* implementable on the hardware in terms of a set of native gates, and it should be scalable up to a large number of qubits with limited resources (e.g. gate counts). Moreover, at each iteration, the cost function $C(\boldsymbol{\theta})$ is estimated by repeated measurements on the *same* quantum state $|\psi(\boldsymbol{\theta})\rangle$, and this set of measurements (together with classical post-processing) should also be efficient. Despite small-scale classical simulations are often run by exactly computing the PQC, by employing dedicated software likewise *Qiskit* [35] or *Pennylane* [36], in a real-world scenario it is not possible to access the exact values of the cost function, which should be estimated as a sample mean (empirical mean estimator) over a certain number of "shots" K:

$$C^{(K)}(\boldsymbol{\theta}) := \frac{1}{K} \sum_{j=1}^{K} C_j(\boldsymbol{\theta}) . \tag{1.5}$$

We remark that the right-hand side of the previous equation is the average over different measurement outcomes on $|\psi(\boldsymbol{\theta})\rangle$, which might have to be performed on different bases, typically depending on the decomposition of the operator $\widehat{O}$ into a weighted sum of Pauli strings. We dub "shot noise" the statistical noise that is inherently present due to a finite value of K, even in the idealized scenario of a perfectly working quantum device without any physical noise or gate errors: by elementary statistics, we expect the statistical error to scale as $K^{-1/2}$.

The computation of gradients $\nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta})$ (if needed for the classical optimization routine) is also a critical task, which can be tackled via finite differences approaches (very sensitive to shot and physical noise), analytical tools such as the parameter shift rule [37] and its generalized versions [38], or other techniques based on natural gradients and their approximation [39, 40].

Another key point is the question of classical simulability: if any practical quantum speedup is sought, certainly the quantum state $\psi(\boldsymbol{\theta})$ should *not* be classically simulable. What might seem an obvious statement, is actually an open research question of paramount importance in variational quantum computing and, more generally, in quantum many-body physics. Indeed, it is still an open challenge to characterize what makes a quantum state hard to be classically simulated, either in the *weak* sense (efficient sampling of the square modulus of the wavefunction) or in the *hard* sense (computing outcome probabilities efficiently) [41]. Note that, to simulate a VQA classically, it would be enough for the PQC to be simulable in the *weak* sense (i.e. sampled efficiently, to compute the cost function C and its gradients), a concept that has not been fully explored, yet [41]. For example, low-entangled quantum states can be simulated with tensor network techniques [42], whereas stabilizer circuits (composed by Clifford gates only) are classically simulable due to the Gottesman-Knill theorem: quantum entanglement and the so-called quantum magic are two necessary resources (but not sufficient, on their own) to make a quantum

state classically hard. Another winning strategy to simulate classes of quantum wavefunctions is neural network quantum states [43].

In summary, a good PQC (or ansatz) should be *expressive* enough to encode non-trivial (and non-classically simulable) quantum states, including states that yield a solution to the problem at hand. Nevertheless, there is a trade-off between high "expressivity" — i.e. the portion of Hilbert space that is covered by the trial wavefunction — and low *trainability* — i.e. the ease in finding good parameters, as a result of the classical optimization of the cost function.

Ansatz design is an intricate subject, that is highly problem-dependent: an exhaustive review of different strategies is well beyond the current scope, see Refs. [21, 22]. A simple classification can be done in terms of "problem-inspired" and "problem-agnostic" ansatzes, where the first class includes PQCs that draw inspiration from the physics of the problem. An example of this class is QAOA, which is detailed in Sec. 1.3. On the other hand, some implementations of VQE are problem-agnostic and rather focus on clever use of hardware resources, namely the "hardware-efficient ansatz", which is crafted to minimize the circuit depth on a specific experimental platform.

1.2.2. Algorithmic and theoretical challenges

Recent years have witnessed tremendous advances in the theory of hybrid quantum-classical computation, with several algorithmic proposals and steady experimental progress. However, conclusive evidence of practical quantum speedup over the best classical algorithms is still lacking or highly debated in all application areas depicted in Fig. 1.1. Unfortunately, any provable exponential speedup seems hard to obtain. Short of sharp theoretical arguments or mathematical proofs, the best way to prove practical quantum advantage (or quantum utility) in real-world use cases for science or technology is to implement VQAs in practice. Nowadays, NISQ devices are nearly beyond the threshold of classical simulations [44], even though convincing evidence remains elusive [45]. Nevertheless, the current technology appears mature enough for a direct investigation of the quantum utility of VQAs: To do so, many challenges should be addressed.

Here, we summarize the main algorithmic and theoretical challenges that arise in a realistic VQA implementation. We do not discuss coherent noise, gate errors, or incoherent noise such as state-preparation-and-measurement (SPAM) errors, which represent additional hurdles to quantum utility. These issues can be tackled, at least in principle, by utilizing quantum error correction or error-mitigation techniques. Since these tools are not used in the following Chapters, we refer the reader to existing literature.

Local minima traps

First and foremost, the main challenge for a VQA is *trainability*, that is the effectiveness of a classical optimization routine in the minimization of the cost function. Even in the simplified setting of Eq. 1.4, e.g. QAOA (see the next section), the optimization landscape is usually highly non-convex, and *local minima traps* represent a serious issue. In practice, for a realistic application, a heuristic variational approach is not expected to yield the global minimum. Hence, we remark the distinction between "good" local minima solutions $\boldsymbol{\theta}^*$ (with low-enough values of the cost function $C(\boldsymbol{\theta}^*)$) and local minima "traps", where the classical optimizer may get stuck, without providing any reasonable solution. The presence of local minima traps in the variational optimization of PQCs is a ubiquitous phenomenon experienced by all practitioners. This empirical evidence led to incremental advances in devising more effective optimizers [27, 46]. A systematic characterization of the presence of local minima traps in the cost function landscape for certain PQCs was derived in Ref. [47], and VQA training was formally proven to be NP-hard in Ref. [48]. On the other hand, classical neural networks can be trained effectively using techniques like the ADAM optimizer, especially when they have more parameters than are strictly necessary — a condition known as benign overparameterization. Despite seeming similarities, PQCs show substantial differences: To date, researchers have not identified an effective overparameterized regime and a reliable classical optimization routine for VQAs. Therefore, current VQAs rely on a careful design of the ansatz — usually problem-inspired, or encoding model symmetries — and on a careful training procedure. In addition to local minima traps, the issue of vanishing gradients also affects the cost function landscape of VQAs: this is inherently due to their quantum nature.

Barren plateaus

Another trainability issue has become a whole new research field, namely *Barren Plateaus* (BPs): in a nutshell, the magnitude of the gradients of the cost function $C(\boldsymbol{\theta})$ tend to vanish exponentially almost everywhere by increasing the number of qubits N. This is a serious drawback for a learning model, such as VQAs, where the cost function $C(\boldsymbol{\theta})$ and its gradients can be estimated only via sampling, namely by repeated preparation and measurement of the parameterized quantum state $\psi(\boldsymbol{\theta})$.

To understand the essence of the BP phenomenon more formally, let us consider a cost function $C(\boldsymbol{\theta})$ and its partial derivatives (gradient components) $\partial_{\theta_j} C(\boldsymbol{\theta})$. Under mild assumptions, by averaging any partial derivative over the whole variational landscape, one would expect to obtain nearly zero due to sign cancellations. If we denote as $\langle \cdot \rangle_{\boldsymbol{\theta}}$ the average over the variational space with a flat (uniform) distribution,

this condition reads

$$\langle \partial_{\theta_j} \mathrm{C}(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}} = 0 \,. \tag{1.6}$$

This is formally true for 1-design PQCs, i.e. for sufficiently expressive and deep ansatzes, as proven in Ref. [49]. By definition, BPs occur when also the gradient magnitudes, quantified by the variance of a generic partial derivative, vanish *exponentially* with the qubit number N:

$$\mathrm{Var}\left[\partial_{\theta_j} \mathrm{C}(\boldsymbol{\theta})\right] = \left\langle \left(\partial_{\theta_j} \mathrm{C}(\boldsymbol{\theta})\right)^2 \right\rangle_{\boldsymbol{\theta}} \sim \frac{1}{b^N} \,, \tag{1.7}$$

for some $b > 1$. As proven analytically in the first seminal paper on BPs, Ref. [49], highly expressive PQCs suffer from vanishing gradients when randomly initialized. These results are valid in the framework of 2-design random quantum circuits [50–52], namely deep and problem-agnostic ansatzes that are, in informal words, just a bunch of parameterized rotations and entangling gates repeated for a large number of layers.

In this case, as formally stated by the Chebyshev inequality, the probability of having sizeable gradients in the variational energy landscape is also exponentially small for increasing N:

$$P\left(|\partial_{\theta_j} \mathrm{C}(\boldsymbol{\theta})| > \delta\right) \leq \frac{\mathrm{Var}\left[\partial_{\theta_j} \mathrm{C}(\boldsymbol{\theta})\right]}{\delta^2} \,, \tag{1.8}$$

for any $\delta > 0$. In classical learning models, such as artificial neural networks, the cost function and its gradient components can be computed exactly (up to numerical precision): an estimation with accuracy ε usually scales as $\mathcal{O}(\log(1/\varepsilon))$. In contrast, in the quantum setting, a gradient estimation relies on sampling, very much like the cost function estimation in Eq. 1.5. This can be seen e.g. by applying the parameter shift rule or its generalized versions.

Hence, let us focus on the empirical mean estimator $\mathrm{C}^{(\mathrm{K})}(\boldsymbol{\theta})$ of the true value of the cost function $\mathrm{C}(\boldsymbol{\theta})$, which is constructed from K measurement outcomes $\mathrm{C}_j(\boldsymbol{\theta})$. To obtain an ε-accurate estimate of the true value, with a probability at least $(1-\delta)$, i.e.

$$|\mathrm{C}^{(\mathrm{K})}(\boldsymbol{\theta}) - \mathrm{C}(\boldsymbol{\theta})| \leq \varepsilon \,, \tag{1.9}$$

one needs a number of shots [6] scaling as [53]

$$\mathrm{K} = \mathcal{O}\left(\frac{1}{\epsilon^2}\right) \log\left(\frac{2}{\delta}\right) \,. \tag{1.10}$$

---

[6]Here, we are examining an idealized scenario, in which we perform measurements on the basis of eigenstates of the operator $\widehat{\mathrm{O}}$. In practice, this is rarely the case, unless $\widehat{\mathrm{O}}$ is classical, i.e. diagonal in the computational basis. In most applications, it is necessary to decompose $\widehat{\mathrm{O}}$ into a weighted sum of Pauli strings. Then, one performs repeated state preparations of $\psi(\boldsymbol{\theta})$, measuring each time in the diagonal basis of a single Pauli string: all of this is required to obtain a *single* experimental value for $\mathrm{C}_j(\boldsymbol{\theta})$. This may often represent a significant computational overhead.

In the presence of BPs, it is mandatory to estimate the cost function with accuracy at least $\varepsilon \sim b^{-N}$, leading to an unfeasible exponentially-large number of shots K = $\mathcal{O}\left(b^{2N}\right)$.

In summary, exponentially vanishing gradients imply an exponentially large number of state preparations and measurement shots at each VQA iteration, to accurately estimate the gradients of the cost function, find a loss-minimizing direction, and navigate the loss landscape. Hence, in the presence of BPs, an exponential quantum speed-up is theoretically ruled out and, on a practical level, a randomly-initialized VQA is certified to fail by scaling up the problem beyond a few tens qubits, i.e. precisely in the regime beyond the reach of classical simulations.

Let us remark that the BP phenomenon is inherently quantum and related to the sampling procedure. Indeed, in classical machine learning, one has access to float-precision values of $C(\boldsymbol{\theta})$ and its gradients: it is possible to track gradient magnitudes and rescale the whole cost function by an appropriate factor to cure vanishing or exploding values of partial derivatives.

Following the first results on 2-design PQCs, many papers investigated the occurrence or absence of BPs in different frameworks, both numerically and analytically. A remarkable link has been demonstrated between the locality of the observable $\widehat{O}$ in the cost function C and the severity of the BP phenomenon upon increasing the qubit count N: A global cost function is very problematic, as the onset of BPs is already observed for a PQC with a constant depth $\mathcal{O}(1)$. On the opposite, a local observable is BP-free up to, at least, depth $\mathcal{O}(\log N)$. Although these results are also based on the 2-design assumption, they suggest the reasonable practice of avoiding global operators $\widehat{O}$, such as projectors on a subspace.

In recent years, a lot of progress has been made in the formal understanding of this topic. In particular, BPs have been linked to highly-expressive PQCs [54], beyond the idealized framework of a 2-design. Unfortunately, they also arise in the context of less-expressive symmetry-preserving ansatzes that are problem-inspired [55, 56], or in equivariant [57] ansatzes. Other sources of exponentially vanishing gradients are entanglement in the initial state [58] and, for a large class of PQCs, the presence of noise: these noise-induced BPs [59] are conceptually distinct from other results and may prove particularly difficult to overcome. On a theoretical side, the presence of BPs has been equivalently characterized in terms of cost function concentration and the presence of narrow gorges in the training landscape [60]. Despite being formulated for first-order gradients (partial derivatives of $C(\boldsymbol{\theta})$), in practice, the BP issue has been shown to affect both higher-order gradient-based algorithms [61] and common gradient-free methods [62].

As highlighted by this partial summary, there has been a flourishing of papers investigating variants of the BP phenomenon for different PQC architectures. Very recently, however, a unified theory of all possible sources of barren plateaus for

generic deep PQCs has been proposed independently by two groups, in Refs. [63] and [64].

Despite the intrinsic scientific interest of this research branch, the outlined scenario might appear rather delusional: most currently-used PQC architectures have severe trainability issues, due to unfavorable non-convex landscapes with many local minima traps and vanishing gradients that require exponential resources to be estimated. As mentioned above, this appears to be in stark contrast with the *unreasonable* effectiveness of training classical learning systems, e.g. artificial neural networks [65]. The quest for a provably trainable VQA architecture, i.e. one that is free of the BP phenomenon, has not led to promising results: on the contrary, it appears that most (if not all) provably trainable PQCs end up being *classically simulable*, ruling out any super-polynomial quantum advantage. This conundrum has ultimately led to a proposal to somehow "pause and rethink variational quantum computing" [66].

Several proposals have been put forth by the community to mitigate or avoid BPs. A non-comprehensive list includes employing classical pre-training techniques [67, 68], layerwise learning for classification tasks [69], identity-block initialization [70], or classical shadows [71]. Moreover, some specific ansatz choices were proven to be free from the vanishing gradient problem, e.g. Quantum Convolutional Neural Networks [72] or some quantum tensor networks [73]. A nearly complete list of approaches can be found in [66]: Unfortunately, many of them turn out to lead to classically simulable algorithms that effectively live in polynomially-sized subspaces of the full Hilbert space.

Luckily, these "negative" results do not rule out polynomial quantum advantage, or even super-polynomial speed ups, if *warm starts* — i.e. clever initialization strategies for the parameters $\boldsymbol{\theta}$ — are adopted. More ambitiously, different approaches to variational quantum computing might be explored, e.g. motivated by fault-tolerant algorithms or by embracing the classical simulability and leveraging the quantum device only for an initial data acquisition procedure, which might also be iteratively optimized.

Despite these results being very recent, the importance of warm starts and smart parameter initialization, often based on iterative procedures, was already renowned by the community in the context of structured ansatzes inspired by Adiabatic Quantum Computation (AQC), such as QAOA and its generalizations. In Refs. [74, 75], the authors leverage an iterative scheme by optimizing only a subset of gate parameters at each iteration and using this result as a warm-start for the next iterative step.

In this Thesis, we mainly focus on this class of problem-inspired ansatzes related to AQC in the context of classical optimization and ground-state preparation, avoiding highly expressive and unstructured PQCs. We also employ and develop strategies

to avoid random initialization, boosting the trainability of hybrid quantum-classical schemes.

In the next section, we review the basics of AQC and its connection to digitized schemes that can be implemented on a gate-based model and trained in a VQA framework.

1.3. Adiabatic Quantum Computation, Quantum Annealing, and QAOA

Before the early proposals of VQAs in 2014, the concept of quantum optimization had already been explored in the framework of Quantum Annealing (QA) [76–80] and Adiabatic Quantum Computation (AQC) [81] [7], including the first experimental implementations [82].

In a nutshell, AQC relies on the following idea: a quantum system is initialized in the easy-to-prepare ground state of an initial Hamiltonian $\widehat{H}_0$, and then driven out of equilibrium by an external coupling, finally reaching the ground state of a target Hamiltonian $\widehat{H}_{\text{targ}}$, which encodes by construction the solution of a computational problem. In the following, we summarize the main concepts and challenges of AQC, referring to [81] for a historical introduction and a complete scientific review.

This framework was originally conceived for classical combinatorial optimization, focusing on an Ising spin-glass cost function (or Hamiltonian)

$$E_{\text{Ising}}(\sigma_1, \ldots, \sigma_N) = \sum_{i=1}^{N} h_i \sigma_i - \sum_{i<j} J_{ij} \sigma_i \sigma_j \,, \tag{1.11}$$

expressed in terms of N binary variables (or classical spins) $\sigma_i = \pm 1$. Since finding the ground state of a spin glass is an NP-complete problem, there exist mappings from every other combinatorial optimization problem in the NP complexity class to a spin glass. Several explicit examples are formulated in Ref. [83]. This class of problems is mathematically equivalent to QUBO (Quadratic Unconstrained Binary Optimization) [84], and has its distinctive feature in the presence of 2-bodies interactions only: this is a 2-local Hamiltonian, by definition. Notice that no hypotheses are made on the interaction couplings J_{ij}, which are usually disordered and long- or infinite-range.

However, other interesting classical optimization problems that can still be recast into classical spin models, do *not* admit a 2-local (or k-local, for any $k \ll N$)

---

[7]Over the last two decades, there has been a flourishing of names referring to different variants of QA (initially intended as a classical algorithm simulating quantum fluctuations) and AQC. Here, following Ref. [81], we shall dub the most general case as AQC, encompassing all alternative formulations.

Hamiltonian. In this case, higher-order interactions are present:

$$E(\sigma_1, \ldots, \sigma_N) = \sum_{i=1}^{N} h_i \sigma_i - \sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_{i<j<k} K_{ijk}\sigma_i\sigma_j\sigma_k - \ldots \ . \qquad (1.12)$$

This is, for instance, the case of supervised learning of Binary Neural Networks (see Chapter 2).

Regardless of k-locality, the standard strategy to obtain a quantum Hamiltonian is to map classical spins to quantum spin-1/2 Pauli operators $\hat{\sigma}_j^z$, in the so-called base encoding. Hence, the initial cost function is mapped to a quantum Hamiltonian that is *diagonal*, by construction, in the standard computational basis of quantum computation [5]:

$$E(\sigma_1, \ldots, \sigma_N) \rightarrow \widehat{H}_z(\hat{\sigma}_1^z, \ldots, \hat{\sigma}_N^z) \ . \qquad (1.13)$$

Coming back to AQC, let us phrase it in its standard formulation: the target Hamiltonian — whose ground state we wish to find — is $\widehat{H}_{\mathrm{targ}} = \widehat{H}_z$, while the initial Hamiltonian is often $\widehat{H}_0 = \widehat{H}_x = -\Gamma_0 \sum_j \hat{\sigma}_j^x$, i.e. a transverse field term allowing for quantum fluctuations (we set $\Gamma_0 > 0$). The idea is to construct an interpolating Hamiltonian

$$\widehat{H}(s) = s\widehat{H}_{\mathrm{targ}} + (1-s)\widehat{H}_0 \,, \qquad (1.14)$$

with $s = s(t)$ playing the role of an external driving. Ideally, the goal would be to pursue adiabatic dynamics, following the instantaneous ground state of $\widehat{H}(s)$, by slowly increasing $s(t)$ from $s(0) = 0$ to $s(\tau) = 1$ in a large total annealing time τ. Clearly, one should start from the easily prepared ground state of $\widehat{H}_x$:

$$|\psi_0\rangle = |+\rangle^{\otimes N} = \left(\frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}} \right)^{\otimes N} , \qquad (1.15)$$

where $|\uparrow\rangle$ and $|\downarrow\rangle$ denote the spin up/down eigenstates of $\hat{\sigma}^z$. Let us remark that this formulation of AQC is also valid for the ground state preparation of a many-body quantum Hamiltonian $\widehat{H}_{\mathrm{targ}}$, which is usually expressed in terms of the whole set of non-commuting Pauli operators $\{\hat{\sigma}_j^\alpha\}$ for $j = 1\cdots N$ and $\alpha = x, y, z$.

The formal solution of the Schrödinger equation

$$i\hbar\frac{d}{dt}|\psi(t)\rangle = \widehat{H}(s(t))|\psi(t)\rangle \qquad (1.16)$$

given the initial condition $|\psi(t=0)\rangle = |\psi_0\rangle$ is written as $|\psi(t)\rangle = \widehat{U}_{\mathrm{AQC}}(t,0)|\psi_0\rangle$, with a time-ordered evolution operator

$$\widehat{U}_{\mathrm{AQC}}(t,0) = \mathcal{T}\exp\left(-\frac{i}{\hbar}\int_0^t \widehat{H}(s(t'))dt' \right) . \qquad (1.17)$$

In an ideal adiabatic evolution, the time-evolved state $|\psi(t)\rangle$ would coincide with the instantaneous ground state of the time-dependent Hamiltonian at any time t, leading

to the correct ground state at $t = \tau$. However, adiabaticity can be guaranteed only under precise assumptions. In fact, the main hurdle in AQC is related to the large values of the annealing time τ required for adiabaticity, possibly diverging in the thermodynamic limit, if the system crosses a critical point or, even worse, a first-order phase transition [85]. A quantitative analysis of the relationship between the minimum spectral gap Δ encountered during the annealing and the large values of τ required for adiabaticity relies on the adiabatic theorem(s). This is a very subtle theoretical subject, which is reviewed in [81]: here, in summary, and with a great deal of simplification, we may state that (under reasonable hypotheses) the minimum annealing time τ required for adiabaticity should scale as Δ^{-2}. Whereas, in principle, the annealing schedule $s(t)$ could be chosen with some freedom, by "slowing down" close to points where the spectral gap of $\widehat{H}(s)$ is minimum, in practice this requires knowledge of spectral information, a notoriously difficult problem [86]. Hence, very often, a *linear* schedule $s(t) = t/\tau$ is assumed.

Remarkably, upon relaxing the previous hypotheses on the initial and target Hamiltonians, it was proven that AQC is universal [87], as it is polynomially equivalent to gate-based quantum computation. This is possible by considering generic non-stoquastic Hamiltonians. A stoquastic Hamiltonian is characterized by having only non-positive off-diagonal matrix elements, as in the case of $\widehat{H}_x$ and, obviously, a diagonal $\widehat{H}_{\mathrm{targ}} = \widehat{H}_z$.

Let us notice that these methods, unlike VQAs, do not require a hybrid quantum-classical optimization loop or a PQC, but rather a single coherent quantum evolution for a long time. Therefore, AQC/QA is better suited to analog quantum simulators: current devices (notably D-Wave quantum annealers) are limited to stoquastic Ising-like Hamiltonians, and therefore not universal. Sources of non-ideality include decoherence and thermal fluctuations. Nevertheless, commercial use cases of AQC/QA have been put forward, featuring many potential industrial applications [88].

We highlight, however, that it is possible to simulate an adiabatic evolution in a gate-based model of quantum computation: this can be accomplished by *digitized* Quantum Annealing (dQA), which may be regarded as a first conceptual step towards QAOA.

1.3.1. From digitized Quantum Annealing to QAOA

In this section, we provide a summary of dQA [89, 90] and QAOA, within the context of Variational Quantum Algorithms. The digitalization of the continuous-time AQC/QA dynamics (Eq. (1.17), with $t = \tau$) is a natural procedure, which requires two successive steps: time-discretization and the Trotter split-up of the exponential of a linear combination of non-commuting operators.

In the first step, we simply choose a discretization of the time interval $[0, \tau]$ into

P small time steps Δt_m, such that $\sum_m \Delta t_m = \tau$. Correspondingly, the continuous schedule $s(t)$ is discretized into a sequence of short-time evolutions utilizing a step function, attaining P values $s_1, \ldots, s_P$. In the limit of an infinite number of infinitesimal time steps, the time-ordered exponential in the evolution operator would be approximated by a time-ordered sequence of infinitesimal-time evolutions with a fixed Hamiltonian $\widehat{H}(s_m)$. By taking equal time intervals $\Delta t_m = \Delta t$, the exact dynamics is recovered in the limit

$$\widehat{U}_{\text{step}} = \prod_{m=1}^{\leftarrow P} \mathrm{e}^{-i \frac{\Delta t}{\hbar} \widehat{H}(s_m)} \xrightarrow{\Delta t \to 0} \widehat{U}_{\text{AQC}}(\tau, 0) , \tag{1.18}$$

with the arrow $\leftarrow$ denoting a time-ordered product [8]. This time-discretization works remarkably well not only for small values of Δt, but, surprisingly, even for $\Delta t \sim \mathcal{O}(1)$, as long as P is large enough: such "robustness to time discretization" has been confirmed by a few theoretical studies, see e.g. Ref. [91].

The second step towards dQA, i.e. for the simulation of AQC/QA on a gate-based device, relies on a Trotter split-up of the non-commuting terms $\widehat{H}_{\text{targ}}$ and $\widehat{H}_0$ appearing in Eq. (1.14), for each unitary operator $\mathrm{e}^{-i \frac{\Delta t}{\hbar} \widehat{H}(s_m)}$ in Eq. (1.18). In the case of classical optimization, with $\widehat{H}_{\text{targ}} = \widehat{H}_z$ (not necessarily 2-local), we would simply write:

$$\mathrm{e}^{-i \frac{\Delta t}{\hbar} \widehat{H}(s)} = \mathrm{e}^{-i\beta \widehat{H}_x} \mathrm{e}^{-i\gamma \widehat{H}_z} + \mathcal{O}\left((\Delta t)^2\right) , \tag{1.19}$$

with $\beta = (1-s)\Delta t/\hbar$ and $\gamma = s\Delta t/\hbar$, to the lowest order in the Trotter splitting. Similarly, if $\widehat{H}_{\text{targ}}$ is a combination of a $\hat{\sigma}^z$-part $\widehat{H}_z$ and a $\hat{\sigma}^x$-part $\widehat{H}_x$ — as for the ground state preparation of an Ising model (or Ising spin glass) in a transverse field — the same expression still holds, with suitable values of β and γ. Along the same lines, one can extend this procedure to other target Hamiltonians or include higher-order terms in Δt [9].

Within the standard assumption of a linear annealing schedule $s(t) = t/\tau$, the step function approximation is simply $s_m = m/P$ for $m = 1 \cdots P$, where P is called the number of *Trotter steps* (or slices). This amounts to setting, in the case of a classical $\widehat{H}_{\text{targ}} = \widehat{H}_z$,

$$\begin{cases} \beta_m = (1 - s_m)\dfrac{\Delta t}{\hbar} , \\[2mm] \gamma_m = s_m \dfrac{\Delta t}{\hbar} . \end{cases} \tag{1.20}$$

In conclusion, the digitized-QA unitary evolution is given by:

$$|\psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \widehat{U}(\beta_P, \gamma_P) \cdots \widehat{U}(\beta_1, \gamma_1)|+\rangle^{\otimes N} , \tag{1.21}$$

---

[8] Notice that similar approximations must be carried out also to *classically simulate* the AQC dynamics, e.g. via higher-order numerical integration methods for the Schrödinger Eq. (1.16).

[9] A second-order approximation that exhibits the same computational cost of the first-order formula as a function of P is easily constructed from a symmetric second-order Trotter splitting.

where $\boldsymbol{\beta} = (\beta_1, \cdots \beta_P)$, $\boldsymbol{\gamma} = (\gamma_1, \cdots \gamma_P)$ and the m-th step evolution operator reads:

$$\widehat{U}(\beta_m, \gamma_m) = \mathrm{e}^{-i\beta_m \widehat{H}_x} \mathrm{e}^{-i\gamma_m \widehat{H}_z} \ . \tag{1.22}$$

In principle, each unitary term of Eq. (1.22) can be decomposed into elementary gates with standard techniques [10]. Upon selecting sufficiently large values of P and small values of Δt, with fixed product $\tau = P\Delta t$, continuous-time AQC/QA can be simulated with arbitrary accuracy on a gate-based model.

In a complementary view, if one fixes the number of Trotter slices P (i.e. the circuit depth), this digitized-QA scheme can be regarded as a basic example of VQA with a single variational parameter, Δt in Eq. (1.20), which we can optimize to achieve the lowest possible variational energy:

$$E_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \langle \psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) | \widehat{H}_{\mathrm{targ}} | \psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) \rangle \ , \tag{1.23}$$

with $\widehat{H}_{\mathrm{targ}} = \widehat{H}_z$ for a classical optimization [11]. Indeed, as discussed in Ref. [90], a too-small value of Δt corresponds to small Trotter errors but also to a short annealing time, while a too-large Δt is associated with large Trotter errors that make the final state rather inaccurate. Consequently, there is an optimal value of Δt for performing such digitized-QA dynamics [90].

Before concluding our discussion of dQA, a comment is in order, concerning the actual decomposition into elementary gates. Whereas the implementation of $\mathrm{e}^{-i\beta_m \widehat{H}_x}$ requires a single layer of one-qubit Pauli rotations along the x-axis, the gate decomposition (and thus the depth of the resulting quantum circuit) for the unitary $\mathrm{e}^{-i\gamma_m \widehat{H}_z}$ is strongly problem-dependent. In the case of Ising (i.e. of QUBO) Hamiltonians in Eq. (1.11) the gate decomposition is very simple. However, if multi-body interactions are present, as in Eq. (1.12), this decomposition might lead to a highly impractical scaling, requiring up to an exponentially large number of elementary gates as a function of N. This may represent a bottleneck for the actual implementation of this computational paradigm.

With the introduction of dQA, we set the stage for a straightforward analysis of the Quantum Approximate Optimization Algorithm (QAOA) by Farhi *et al.* [28]. Indeed, the QAOA trial wavefunction has the same form as in Eqs. (1.21)-(1.22), where now $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are promoted to 2P *independent* variational parameters for the quantum state, rather than fixed by a Trotter split-up. Hence, QAOA can be regarded as a specific case of VQA, whose PQC draws inspiration from AQC/QA, by generalizing its digitized version. As described in Sec. 1.2, the variational parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are iteratively updated by a classical optimization routine, with the goal of minimization of Eq. (1.23).

---

[10]However, the required number of gates might scale exponentially with the qubit number N, as commented below.

[11]In the present case of dQA, $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are intended to be solely a function of Δt, through Eq. (1.20).

Clearly, the optimal energy at the global minimum $E_{\mathrm{P}}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$ is a monotonically decreasing function of P, which systematically improves on any digitized-QA approach with an equal value of P. However, the same technical shortcomings illustrated for general VQAs may occur for QAOA: determining the global minimum $(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$ is a non-trivial task, since local optimization routines tend to get trapped into one of the many local minima of the 2P-dimensional search space, and the phenomenon of barren plateaus can make the gradients of the variational energy in Eq. (1.23) exponentially small in the qubit number N.

1.3.2. Theory and scope of QAOA

Since its first proposal [28], QAOA has rapidly gained popularity as a leading candidate among NISQ algorithms. Remarkably, it has been proven computationally universal [92, 93], and even for the lowest P = 1 (under some basic complexity-theoretic assumptions) the QAOA state in Eqs. (1.21)-(1.22) cannot be efficiently sampled by classical methods [94]. Remarkably, QAOA has been successfully implemented on a trapped-ion quantum simulator [95] and superconducting quantum circuits [96, 97].

The vast majority of QAOA applications tackle classical combinatorial optimization tasks yielding 2-local models, as in Eq. (1.11). In particular, an ever-growing body of literature focuses on MaxCut for random graphs, often 3-regular graphs (with, or without random weights). See e.g. Refs [28, 74, 98] and other more recent publications. However, very little is known on average or worst-case performance guarantees for P > 1 [12], and no consistent proof of quantum advantage compared to the best classical algorithms has yet been obtained, as it may require a large number of qubits in an experimental setup [99]. QAOA for MaxCut has been intensively investigated for years as a *potential* candidate for *practical* quantum advantage: this is also due to its efficient gate decomposition. Nonetheless, a few studies focus on relevant optimization tasks that require the more general, multi-spin, formulation of Eq. (1.12). Relevant examples include Refs. [100–103]: We will outline some of these results in the next sections. Moreover, in Chapter 2, we apply QAOA to the supervised learning of simple Binary Neural Networks. This is a paradigmatic task in machine learning, which does not admit an Ising-like cost function.

We highlight that the original QAOA formulation has been modified in many ways: Ref. [104] provides an up-to-date review of the implementation of several variants of QAOA. A recently rebranded version [33], namely the Quantum Al-

---

[12]The results for P = 1 are reported in the original QAOA paper [28]. In the opposite limit P → ∞, QAOA is guaranteed to find the solution, since it can provide an arbitrary-precision digital simulation of AQC/QA, with arbitrarily large annealing time τ. Indeed, by handling such limits with care, one could choose a vanishing Δt and a large $\tau = \mathrm{P}\Delta t$. In practice, the corresponding quantum circuit would be exceedingly deep.

ternating Operator Ansatz, proposes a simple but far-reaching generalization: in Eq. (1.22), one could replace either or both the diagonal Hamiltonian $\widehat{H}_z$ and the *mixing* Hamiltonian $\widehat{H}_x$ with some other operators. For instance, $\widehat{H}_z$ could be an approximated version of the classical Hamiltonian (or cost-function) that admits a simpler gate decomposition, whereas $\widehat{H}_x$ could be modified to introduce tailored quantum fluctuations among a relevant subset of the classical states.

Although QAOA was devised to find approximate solutions to classical combinatorial optimization problems, it can be readily generalized to construct the ground state of many-body quantum Hamiltonians, as previously mentioned also for AQC. This approach will be utilized in Chapter 3 and Chapter 4. In this context, a generalized formulation of QAOA is commonly named Hamiltonian Variational Ansatz (HVA) [2, 27, 56, 105–110], with a variational state of the form:

$$|\psi_{\mathrm{P}}(\boldsymbol{\theta})\rangle = \prod_{m=1}^{\mathrm{P}} \mathrm{e}^{-i\theta_{m,M}\widehat{H}_M} \cdots \mathrm{e}^{-i\theta_{m,1}\widehat{H}_1}|\psi_0\rangle . \tag{1.24}$$

Here, $|\psi_0\rangle$ is a simple initial state, while $m = 1\cdots\mathrm{P}$ labels successive circuit layers, each in turn composed by $j = 1\cdots M$ alternating unitaries generated by Hamiltonian operators $\widehat{H}_j$. The target Hamiltonian $\widehat{H}_{\mathrm{targ}}$ can be linearly decomposed in terms of the generators and is now a quantum Hamiltonian, usually expressed in terms of Pauli operators. The trial PQC in Eq. (1.24) is iteratively optimized to approximate the ground state by leveraging the variational principle with a cost function in the same form of Eq. (1.23).

Theoretical insight on the relation between analog AQC/QA and QAOA was gained in Refs. [111, 112], leveraging optimal control techniques [113]. Although QAOA is usually implemented on a gate-based model, it can be theoretically described as an analog algorithm, in the same form as a continuous-time driving protocol in Eq. 1.14. This is done by choosing for $s = s(t)$ a bang-bang protocol (square pulse), namely a discontinuous function jumping between 0 and 1 for tunable time intervals, i.e. β_m and γ_m respectively [75]. In this case, the total evolution time amounts to (we set $\hbar = 1$)

$$\tau = \sum_{m=1}^{\mathrm{P}} (\gamma_m + \beta_m) . \tag{1.25}$$

Upon fixing a *finite value* of τ, a natural question concerns the *optimal* protocol $s(t)$: is it a continuous annealing-like schedule or a discontinuous bang-bang protocol? In a previous seminal work on the interplay between optimal control theory and the VQA hybrid scheme, the authors applied the Pontryagin principle to show that the optimal schedule is a bang-bang protocol [114]. However, it turned out that some assumptions were not as general as expected: in Ref. [111] the optimal protocol for a fixed and finite value of τ is proven to be hybrid, consisting of both bang-bang and annealing segments. Specifically, analytical and numerical evidence suggests a "bang-anneal-bang" qualitative behavior: the optimal protocol begins and ends

with a bang, and in between it usually has an extended annealing region (possibly including additional bangs). In the follow-up paper [112], the qualitative features of the optimal protocol are investigated in further detail. Initial and final bangs are found to be related to diabatic transitions. The interior region is characterized by a smooth annealing behavior, with a superposed oscillatory pattern. In the large-τ limit, the optimal curve is observed to approach a monotonic *optimized* adiabatic schedule, similar to the one derived analytically for the Grover search in Ref. [115]. Indeed, numerical simulations show that the initial and final bangs become shorter and eventually disappear, whereas the oscillations are damped and finally vanish. Additionally, some quantitative connections between these optimal features and optimal QAOA parameters are drawn: these are analyzed below in Sec. 1.3.4, alongside some limitations of this approach and the generality of its conclusions.

1.3.3. Concentration and transferability in QAOA

In Sec. 1.2 we reviewed the main algorithmic challenges for the cost function optimization in a VQA. To date, we lack a provably trainable and classically hard regime that holds across different application domains. Hence, problem-inspired ansatzes like QAOA and careful training are constitutional in the early success of VQAs.

A *leitmotif* in the QAOA literature are *concentration* effects: for fixed P, typical instances drawn from the same problem distribution may yield a similar variational energy landscape $E_P(\boldsymbol{\beta}, \boldsymbol{\gamma})$. In this case, optimal parameters computed for the first instance often serve as an excellent warm-start for local optimization for other instances, significantly reducing the computational cost.

The topics of concentration and transferability of optimal parameters in QAOA are subtle, since they may be due to different reasons, leading to possible misconceptions about the underlying mechanism. To date, I have not found a comprehensive review of these subjects, and even a recent review paper on QAOA [104] seems incomplete concerning this matter. Since warm-start QAOA and transferability of optimal parameters are a common thread of the next Chapters, I tried to provide a concise yet thorough summary of key findings.

The first seminal results were presented in 2018 by Brandao et al. in Ref. [98]. The authors studied QAOA for unweighted 3-regular MaxCut, proving rigorously that *the whole variational landscape* concentrates for typical instances in the regime of small P and large N. This result can be intuitively understood and formally derived by a simple argument: in essence, due to the locality of the model, each term in the variational energy can be computed via a "reverse causal cone", involving

only a subgraph whose size grows (geometrically) with P [13]. Then, the occurrence of concentration is explained by the statistical dominance of tree-like subgraphs. We remark on two key aspects. Firstly, the whole parameter landscape, not only local or global minima, concentrates for small P and large-N instances. Secondly, concentration holds for typical instances drawn from the same distribution (in this case, random 3-regular graphs).

A whole series of new results stemmed from Ref. [98]. For instance, in Ref. [116], the authors propose and implement a *classical* technique to compute optimal QAOA parameters in the low-P regime, in the infinite-size limit $N \to \infty$. The focus is again on unweighted 3-regular MaxCut, but the method is benchmarked also against a simple spin glass in 2 dimensions. This is accomplished by leveraging the statistical dominance of tree-like subgraphs and devising a classical tensor-network based technique to compute the QAOA expectation value. The optimal parameters do not strongly depend on the specific problem instance but rather on the general "topological" features of the problem under investigation.

The general case of weighted MaxCut has a higher relevance for real-world combinatorial optimization, yet it features a challenging QAOA variational energy landscape, with a proliferation of local minimum traps. Very recently, in Refs. [117, 118], the authors proposed a rescaling scheme that allows parameter transferability also to weighted MaxCut. Results on the transferability of QAOA optimal parameters among instances of the same problem are also found for graph clustering (i.e. network community detection) in Ref. [119].

In the regime of shallow circuits (small P) for large instances (large N), concentration and transferability of optimal parameters are due to an intuitive argument: locality and a careful classification of subgraphs, which are typically trees. In practice, however, QAOA is not expected to provide high-quality solutions in this regime, since it does not "see" the whole graph. Far more surprising results are obtained for large P, or even in the case of non-local fully connected models, i.e. when a locality argument based on a "reverse causal cone" does not hold. Already in the seminal paper in Ref. [98], empirical observation of the validity of concentration results beyond the small P regime was presented. Here, the authors explain it in terms of the law of large numbers applied to the sum of terms that defines the cost function (which is, essentially, a sum over the edges of the graph). Albeit this argument does not constitute any formal proof, it elucidates a possible mechanism behind concentration in this regime and provides a way to test it empirically: concentration certainly arises whenever the terms in the cost function sum are low-correlated.

The most impressive research line on QAOA is the one that originated from Ref. [100], where the authors devised analytical approaches to find optimal QAOA

---

[13]An example is given in Chapter 3, precisely in Sec. 3.2, in the context of local quantum spin models with nearest-neighbor interactions.

angles for the fully-connected Sherrington-Kirkpatrick (SK) model. This is an Ising spin glass in Eq. 1.11 with zero external fields $h_i = 0$, where all-to-all random couplings J_{ij} are drawn from a symmetrical distribution with unit variance and normalized by a scaling factor $N^{-1/2}$. For *any* finite depth P, it is proven that the whole QAOA landscape becomes instance-independent for $N \to \infty$. In this limit, the cost function $E_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is calculated analytically with an iterative classical algorithm with complexity $\mathcal{O}(2^{4\mathrm{P}})$. This led to a computation of the global minimum, obtained up to P = 8, and to a lower bound on the QAOA performance up to P = 12. This argument goes beyond a "reverse causal cone" approach since the all-to-all connectivity implies that, already for P = 1, each term in the variational energy sum "sees" the whole graph.

Moreover, the authors prove a stronger concentration result for the SK model at $N \to \infty$. Not only does the QAOA energy landscape concentrate for almost every instance in the thermodynamic limit (concentration over instances), but also concentration of measurements is proven. In a nutshell, for any point $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in the energy landscape, and for almost every instance, by sampling the QAOA state one gets classical strings that concentrate with probability tending to one at the calculated value $E_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma})$. Once the global minimum is classically computed, one would need a quantum computer only to prepare the corresponding QAOA state: in light of the concentration of measurements, only one shot (or very few shots) would be needed to finally obtain the candidate optimal string.

This work has been recently generalized in Ref. [120], which initially tackles unweighted MaxCut on large-girth D-regular graphs [14]. By providing an iterative algorithm to compute analytically the QAOA variational energy for any value of P and girth larger than 2P + 1, the authors can classically compute the global minimum. In the limit of large D, which implies the thermodynamic limit, they prove a promising performance, even compared to the best classical algorithms [120]. Remarkably, the analytical expression found for $E_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma})$ for any finite P and $D \to \infty$, is *equal* to the ensemble-averaged expression found in Ref. [100] for the SK model in the $N \to \infty$ limit. However, this new formula is found by looking at a single tree subgraph and has an enhanced computational efficiency $\mathcal{O}(\mathrm{P}^2 2^{2\mathrm{P}})$, allowing the authors to reach P = 20.

Quite recently, Refs. [102, 103] analytically proved parameter concentration for models beyond 2-local in the large-N limit. Ref. [102] proves both concentration over instances and concentration of measurements for a 1-layer QAOA applied to generic classical optimization problems involving random multi-spin interactions. In particular, the authors consider random combinatorial optimization problems in the form of Eq. 1.12, where each of $\{h_i\}$, $\{J_{ij}\}$, $\{K_{ijk}\}$, etc. is a set of IID random variables drawn from symmetric distributions with variance depending only on the

---

[14]The girth of a graph is the length of its shortest cycle. If a graph is acyclic, its girth is considered to be infinite.

number of bodies involved. Due to their fully-connected nature, these models are also known as mixed-spin Sherrington-Kirkpatrick models or mixed p-spin models [15].

These results were even extended in Ref. [103] to generic constant depth-P QAOA, applied to ensembles of random combinatorial optimization problems — not necessarily fully connected — in the thermodynamic limit. Moreover, building on Ref. [120], the authors generalize the aforementioned equivalence between Max-Cut on large-girth D-regular graphs (for $D \to \infty$) and the ensemble-averaged SK model (for $N \to \infty$): an equivalence result is proven (in the same limits) between Max-p-XORSAT (a generalization of MaxCut) on sparse random hypergraphs and the fully connected pure p-spin model.

In both these works beyond 2-body models, the authors leverage the methodology of Ref. [100], which already involved a great deal of complicated combinatorics. A deep formal understanding of even a single one of the impressive Refs. [100, 102, 103, 120] would require many months of ascetic spiritual retreat away from modern society, which I have not indulged myself in. However, the main open challenge along this route is manifest: all these results were obtained *for fixed depth* P in the large N limit. The cases where P scales e.g. as $\mathcal{O}(\log(N))$, $\mathcal{O}(\text{poly}(\log(N)))$, or even $\mathcal{O}(\text{poly}(N))$ are still in the class of polynomial-time algorithms. A formal understanding of these regimes would require new mathematical instruments and may prove extremely challenging even for top-class mathematicians. However, this seems the leading avenue to provide formal proof of (or exclude) quantum advantage over the best-known classical algorithms for classical optimization problems. A more physically oriented open path is to study use cases of realistic binary optimization problems, which may include higher-order interactions as in Eq. 1.12, but do not satisfy the ideal hypothesis of IID coefficients drawn from symmetric distributions. A natural framework would be the supervised learning of Binary Neural Networks: our first exploratory work is described in Chapter 2.

On a practical level, transferability may open up an interesting opportunity. Indeed, already in the seminal work of Brandao et el. [98], the authors proposed to leverage concentration effects *not only* to transfer optimal QAOA parameters for depth P among typical instances with the same size N, but also *from a small to a larger system*. Clearly, this approach would be invaluable as a warm-start procedure for a real experiment on a quantum device: a classically computed solution could seed a large-scale experiment for a large qubit count N, beyond the reach of classical simulations. In this framework, the quantum device could be employed for a less expensive refinement optimization or, alternatively, only to prepare the optimal QAOA state with transferred parameters and measure it to sample optimal strings. A systematic study of optimal parameter transferability from small to large

---

[15]In contrast, a pure p-spin model — or simply a p-spin model — involves only p-bodies all-to-all terms. If all coefficients are equal, this is an integrable model, due to the conservation of the total angular momentum.

instances can be found in Ref. [121] for P = 1 QAOA and unweighted MaxCut on d-regular graphs. Remarkably, transferability holds among d_1-regular graphs and d_2-regular graphs only provided that d_1 and d_2 have the same parity. When this is the case, optimal P = 1 QAOA parameters found for a graph with $N = 8$ nodes can be transferred successfully up to a graph with $N = 64$ nodes, with a negligible performance loss compared to a direct optimization on the large model. The validity of this approach has been proven analytically, with a scaling correction in inverse powers of N, for the specific case of a Grover search with QAOA, first in Ref. [122] and then in Ref. [123] [16]. Transferability from small to large instances of a problem class is also observed in Ref. [124], which proposes a meta-learning scheme for the optimal VQA parameters using classical learning, specifically Recurrent Neural Networks.

1.3.4. Patterns in optimal QAOA parameters

Another common trend in QAOA, which may or may not be related to concentration and transferability depending on the context, is the following: optimal QAOA parameters often appear to have a pattern. As we progress through the QAOA circuit, the optimal parameters γ_m appear to increase (nearly) monotonically, and the parameters β_m appear to decrease similarly. Informally, the overall trend appears to be quite regular, or "smooth": if we plot the two discrete sets of optimal angles against a rescaled index $\tilde{m} = (m-1)/(P-1) \in [0,1]$, these seem to be sampled from two continuous functions $\gamma(x), \beta(x)$ defined in the same interval. These qualitative features are reminiscent of linear digitized-QA in Eq. (1.20). Although the optimal QAOA parameters often depart significantly from dQA, exploiting this similarity for a warm-start procedure may sometimes be useful, as shown in Ref. [125] for the usual MaxCut problem. This approach adds to the toolbox of warm-start procedures for QAOA, and it has been adapted to our tasks in Chapter 2 and in Chapter 4.

In Ref. [100], this regular pattern is observed in the global minimum of QAOA with a fixed depth P in the infinite-size SK model. As stated above, this work has been generalized in Ref. [120] to unweighted MaxCut on large-girth D-regular graphs: for large D, also in this case (and in Max-p-XORSAT) the *global* optimum is obtained in terms of smooth sequences of γ_m and β_m. This set of results is quite remarkable, as the global optimality of smooth solutions is proven analytically.

---

[16] A sign of the ubiquitous confusion on concentration and transferability is that Ref. [123] is often mistakenly regarded as a general proof of transferability for QAOA, as stated e.g. in the QAOA review in Ref. [104]. In reality, by choosing $\widehat{H}_z = |t\rangle\langle t|$, i.e. a projector on a classical state, this is nothing but the Grover search in disguise. Not only the proof of concentration (for low P) is valid only in this case (which was already derived in Ref. [122]), but also, this operator is global, and it does not admit any practical gate decomposition.

Analogous results are obtained numerically in Ref. [74] for unweighted and weighted 3-regular MaxCut instances: also in this case, the authors numerically "prove" that smooth solutions correspond to global minima. To leverage this pattern in optimal parameters, two iterative approaches to QAOA optimization are devised, which serve as a tool to avoid local minimum traps. At each step, the QAOA ansatz acquires new variational parameters (e.g. new layers), and the new optimization is seeded with a warm start yielded by the previous step. In practice, this is done by adopting Fourier-based algorithms [74], or by interpolating from previous smaller P solutions [74, 75]. More explicitly, in the latter case, the current optimal solution at each step heuristically provides a good warm-start for the next step, e.g. from $P \rightarrow P'$ (with $P' > P$, such as $P' = P+1$). Similar tools are utilized, with some modifications, in all subsequent Chapters, both in the context of classical optimization and ground-state preparation.

Smooth solutions are commonly found numerically by leveraging these iterative techniques without proving their global optimality, a task that is usually unfeasible beyond shallow circuits and small systems. A few examples include Ref. [126] for MaxCut on random graphs, and Ref. [75] for the antiferromagnetic Transverse-Field Ising Model (TFIM) on a chain. Significant evidence of the existence of smooth QAOA solutions is provided by Ref. [95], where the authors focus on the long-range Ising model with power-law decaying interactions, both quantum (with a transverse field term) and classical. Remarkably, the QAOA algorithm is implemented experimentally on a trapped-ion quantum simulator with up to 40 qubits. The smooth solutions are not proven to be a global minimum, but turn out to be very effective. They are found to be transferable from a small-N to a large instance of the same model, and show similarities even in different points of the phase diagram (e.g. by modifying the transverse field strength) and for different values of the power-law decay coefficient. Note that the focus here is on the transferability of the smooth optimal solution: there is neither numerical nor analytical evidence of parameter concentration in the whole variational energy landscape, and due to long-range interactions, a "reverse causal cone" argument does not apply. Smooth optimal parameters are also observed up to large values of P in CD-QAOA, a generalization of QAOA including counter-diabatic (CD) terms [127], and for standard QAOA applied to the challenging task of Max K-Cut (for small values of P). In most cases, smooth QAOA solutions serve as good-quality local minima without guaranteeing global optimality. Indeed, there can exist better or equal-quality solutions that are non-smooth, as proven in Ref. [75] for the TFIM. In practice, even distinct smooth solutions for the same problem instance, e.g. implementing shortcuts to adiabaticity, can be found in certain settings: this last point is exemplified in a paper in preparation, which focuses on the interplay between quantum optimal control techniques and QAOA for hard instances of the weighted MaxCut problem.

A natural question is whether there is some physical insight or theoretical expla-

nation for this pattern in optimal QAOA solutions. By utilizing optimal control, a significant advancement is made in Ref. [112]. As previously discussed at the end of Sec. 1.3.2, the authors describe the typical behavior of an *optimal analog protocol* for fixed time τ: this often has a bang-anneal-bang structure with a central annealing region characterized by a continuous trend and a superposed oscillatory pattern. These qualitative features turn out to be closely related to the regular pattern observed in optimal QAOA schedules. Before summarizing these results, let us clear up a potential ambiguity: the "smoothness" of the central annealing region — which is a continuous-time analog evolution — is an entirely distinct concept from the "smoothness" of the QAOA optimal parameters plotted against a rescaled layer index $\tilde{m}$. Indeed, these always implement bang-bang dynamics, and the regular pattern concerns the time evolutions of this discontinuous schedule, which are precisely the values of β_m and γ_m.

To understand the connection between features of an analog optimal schedule and patterns in QAOA solutions, let us go back to Eq. 1.20, obtained as a result of time-discretization and first-order Trotter splitting for a linear annealing schedule. If we rewrite it for a generic step function attaining values s_m for small time intervals Δt_m, we get [17]

$$\begin{cases} \beta_m = (1 - s_m)\Delta t_m \\ \gamma_m = s_m \Delta t_m \end{cases} \tag{1.26}$$

By interpreting the angles β_m and γ_m as free independent variational parameters, as done in QAOA, we can invert the previous relation, finding

$$\begin{cases} s_m = \dfrac{\gamma_m}{\gamma_m + \beta_m} \\ \Delta t_m = \gamma_m + \beta_m \end{cases} \tag{1.27}$$

By reversing the argument of Sec. 1.3.1, one could think of recomposing the optimal bang-bang QAOA protocol into a discretized time-evolution:

$$e^{-i\beta_P \widehat{H}_x} e^{-i\gamma_P \widehat{H}_z} \cdots e^{-i\beta_1 \widehat{H}_x} e^{-i\gamma_1 \widehat{H}_z} \Rightarrow e^{-i\Delta t_P \widehat{H}(s_P)} \cdots e^{-i\Delta t_1 \widehat{H}(s_1)} , \tag{1.28}$$

with $\widehat{H}(s) = s\widehat{H}_z + (1-s)\widehat{H}_x$ as in Eq. 1.14. However, this approximation would be valid only for small β_m and γ_m, leading to a small Trotter error. Finally, one could even rephrase this discretized dynamics into a continuous-time analog evolution (see Eq. 1.17), driven by a control field $s(t)$ obtained by an interpolation of the optimized step function in Eq. 1.27:

$$e^{-i\Delta t_P \widehat{H}(s_P)} \cdots e^{-i\Delta t_1 \widehat{H}(s_1)} \Rightarrow \mathcal{T} \exp\left(-i \int_0^t \widehat{H}(s(t'))dt'\right) . \tag{1.29}$$

This tempting approach would be justified in the large P limit for vanishing optimal QAOA angles, leading to small Trotter and time-discretization errors. This

---

[17]We set $\hbar = 1$ for simplicity of notation.

is the case for dQA with a small time step Δt and large P, leading to small angles in Eq. 1.20. However, as noted in Ref. [112], this is not the case for QAOA: the optimal solutions may fall along certain smooth curves $\gamma(x)$, $\beta(x)$ for increasing circuit depth, as discussed above. However, these asymptotic curves prescribe angles of constant order, invalidating the previous argument [18].

As demonstrated in Ref. [112], the central annealing region of an optimal analog protocol does have some connection with Eq. 1.27 [19]. In summary, numerical results show two main features, when comparing a QAOA solution with the optimal bang-anneal-bang curve characterized by the same value of τ (see Eq. 1.25). Semi-analytical arguments also back these results. As previously stated, the central region of the bang-anneal-bang schedule oscillates about some base curve. The period of such oscillations matches up with the length of the QAOA layers, namely Δt_m in Eq. 1.27, and the number of oscillations appears to be equal to P, the number of QAOA layers. Moreover, the underlying annealing curve, obtained by filtering out the oscillations, matches with the s_m values. We remark that the bang-anneal-bang optimal curve approaches an adiabatic limit, characterized by the shrinking of the initial and final bangs and the damping of intermediate oscillations. Hence, if the QAOA optimal schedule still matches this optimal protocol, the regular pattern (smoothness) of QAOA optimal angles may follow as a consequence.

Despite being heuristically valid, this argument does not seem conclusive. As previously happened for the alleged optimality of the bang-bang scheme [114], the validity of some assumptions and semi-analytical arguments in Ref [112] might be less general than expected. Also, the match with QAOA optimal angles is observed on a limited set of problems, for a very small number of qubits, and it is not analytically proven. Finally, a rigorous optimal control strategy can be applied only to very modest system sizes, e.g. $N = 8$, and requires exact classical simulations. To address this shortcoming, a modified variational scheme, including the main features of these optimal protocols, is proposed in Ref. [112]. However, its effectiveness should be ultimately verified empirically and a mixed bang-anneal-bang protocol may prove particularly challenging to implement on real quantum hardware.

1.4. Outline of the Thesis

Having introduced the theoretical framework and provided a summary of the most relevant topics, we now present the outline of the next Chapters.

---

[18]On the other hand, for small enough values of P, smooth optimal curves for QAOA can be found by leveraging a linear warm start, given by an optimal-Δt dQA. This is the thesis of Ref. [125], as previously mentioned. This topic is subtle and perhaps not yet completely understood.

[19]Note that the authors of Refs. [111, 112] use a different convention for $\widehat{H}(s)$, by swapping $\widehat{H}_z$ and $\widehat{H}_x$, so that $s(0) = 1$ and $s(\tau) = 0$.

In Chapter 2, we adapt the toolbox of warm-start QAOA to a classical optimization task that involves non-local multi-spin interactions, namely the supervised learning of a binary perceptron for an unstructured dataset. We show the existence of smooth optimal solutions that are transferable among different instances, each characterized by a specific training set.

In Chapter 3, we extend this heuristic approach to ground state preparation through the HVA in Eq. 1.24, focusing on many-body spin models. We study the Heisenberg XYZ model [128] and the antiferromagnetic Longitudinal-Transverse-Field Ising Model (LTFIM) [129], two ubiquitous lattice systems with interesting phase diagrams. Although their ground state preparation through the HVA is affected by BPs [55, 57], we avoid this issue by transferring an optimal *smooth* solution, obtained for small system size via iterative techniques, to a large model, where a random-start optimization would fail due to BPs. Remarkably, other equal-quality non-smooth solutions for the small system fail to provide a useful warm start.

In Chapter 4, we adopt similar strategies to prepare and characterize the ground state of a $\mathbb{Z}_2$ lattice gauge theory model. Our approach leads to the detection of a topological quantum phase transition and proves accurate even in the deconfined (topologically non-trivial) regime.

Finally, in Chapter 5, we present a summary of some promising research directions and open questions in the field, also stemming from the discussion and results reported in this Thesis.

2. QAOA for the binary perceptron

> *It expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.*
>
> —Frank Rosenblatt about the perceptron, July 7, 1958.
> He was about 70 years ahead of his time.

In Chapter 1, we introduced digitized Quantum Annealing (dQA) and the Quantum Approximate Optimization Algorithm (QAOA), by commenting on their connection with AQC/QA. We mentioned several applications of QAOA to classical combinatorial optimization leading to 2-local Ising (or QUBO) models in Eq. (1.11), notably MaxCut. Besides, we summarized the relevance of a few studies that explored the applicability and effectiveness of QAOA beyond this standard framework for optimization tasks in the general form of Eq. (1.12) [100–103].

An interesting example — of paramount importance in machine learning— is offered by the training process required in supervised learning for artificial neural networks (ANNs): this is naturally formulated as a minimization problem of a suitable cost function [130], which is, however, not 2-local in terms of its variables (network weights and biases) due to non-linear activation functions. An intriguing question is to explore *if* and *how* quantum computation might provide more efficient algorithms to train ANNs, while potentially offering some deeper theoretical understanding of their effectiveness in classification tasks.

Here, we apply dQA and QAOA to a paradigmatic task of supervised learning: the optimization of synaptic weights for the binary perceptron, whose cost function is expressed in terms of highly non-local multi-spin interactions.

2.1. Introduction and motivation

Our work stems from the results obtained in Ref. [131], where the authors provided analytic and simulation evidence of exponential speed-up of Quantum Annealing (QA) vs. classical Simulated Annealing for the training process of the binary perceptron. The exponential speedup arises from the geometric structure of the solution space of the problem: the presence of rare and yet dense regions of solutions allows

QA to converge efficiently despite the presence of an exponential number of local minima traps. This property appears to be present even in more complex, highly overparameterized neural networks such as the so-called deep neural networks [132]. Quantum algorithms could thus be highly efficient also for this kind of models, which define the state of the art in contemporary machine learning.

Here, we focus on a digitized version of QA and on QAOA: in particular, we provide numerical evidence on how QAOA, by efficiently exploiting optimized quantum fluctuations among classical states, systematically outperforms standard QA. As in Ref. [131], these results are expected to generalize for more complex ANN architectures. Moreover, we show the emergence of *smooth* optimal QAOA parameters, which seem to be independent of the details of the training problem. This finding allows us to develop an effective heuristic procedure to speed up the convergence of QAOA, in a similar fashion to previous results for 2-local models, by leveraging the *transferability* of optimal solutions among typical instances of the same problem.

Finally, we enquire into the role of the classical cost-function landscape geometry [65] in the effectiveness of digitized-QA vs QAOA for our model. This is done by artificially permuting the classical energies associated with each spin configuration: despite the spectrum and the number of classical solutions being the same, a gap closure on the adiabatic path appears [131], which has well-known detrimental effects on QA performance. We show that also our QAOA implementation is affected by this gap closure, even though it still offers some advantages compared to digitized QA.

2.2. Binary perceptron model

The perceptron represents the prototypical example of a single-layer binary classifier, first introduced decades ago by Rosenblatt [133]. It is still a subject of active research, both as the fundamental unit of classical artificial neural networks [134] and as a potential candidate for basic realizations of quantum neural networks [135, 136].

Following Ref. [131], we address the problem of supervised learning of $M = \alpha N$ random patterns in a perceptron with N neurons in the input layer: any configuration of the binary synaptic weights $\boldsymbol{\sigma} = \{\sigma_j\} \in \{-1, 1\}^N$ correctly classifies a randomly generated pattern $\boldsymbol{\xi}^\mu = \{\xi_j^\mu\} \in \{-1, 1\}^N$ into a prescribed binary label $\tau^\mu = \pm 1$ if $\mathrm{sgn}(\boldsymbol{\sigma} \cdot \boldsymbol{\xi}^\mu) = \tau^\mu$, see sketch in Fig. 2.1.

During the learning phase, a given training dataset $\{\boldsymbol{\xi}^\mu, \tau^\mu\}_{\mu=1}^M$ is provided, and the task consists in finding the weight configurations $\boldsymbol{\sigma}$ such that all the patterns are correctly classified. Here, we focus on an unstructured training dataset: by hypothesis, the pattern components $\boldsymbol{\xi}^\mu$ and the labels τ^μ are IID (unbiased) Bernoulli
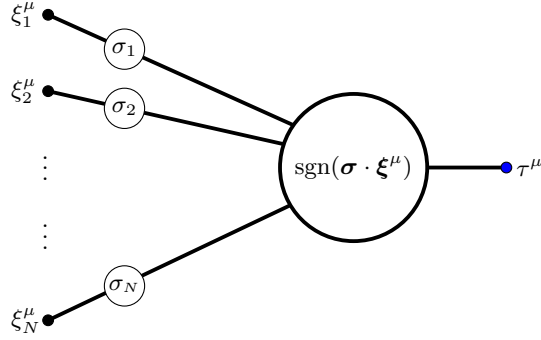
Figure 2.1.: Scheme of a perceptron. Binary synaptic weights σ_j have to be adjusted such that for given binary values ξ_j^μ in the input layer, the scalar product $\boldsymbol{\sigma} \cdot \boldsymbol{\xi}^\mu = \sum_j \sigma_j \xi_j^\mu$ has a prescribed output sign τ^μ. Here $\mu = 1 \cdots M$, with $M = \alpha N$, labels the various input-output patterns.

random variables, results of a fair coin flip. Hence, we can set the labels τ^μ all equal to $+1$ without loss of generality.

The search problem can be easily reformulated as a minimization problem for a suitable cost function. The underlying idea is to associate a positive energy cost for every pattern incorrectly classified. The exact solutions to the classification problem are the cost function's zero-energy configurations $\boldsymbol{\sigma}^*$. Let us define

$$m_\mu = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \sigma_j \xi_j^\mu \tag{2.1}$$

to be the overlap between the spin configuration $\boldsymbol{\sigma}$ and the μ-th pattern $\boldsymbol{\xi}^\mu$, normalized in such a way that, upon averaging over the random patterns, one gets $\overline{m_\mu^2} = 1$. A possible definition of the cost function is:

$$E_{\mathrm{n_c}}\left(\{\sigma_j\}\right) := \sum_{\mu=1}^{M} |m_\mu|^{\mathrm{n_c}} \, \Theta\left(-m_\mu\right) \;, \tag{2.2}$$

where $\Theta(x) = (1 + \mathrm{sgn}(x))/2$ is the Heaviside step function. The energy cost for an incorrect classification of a pattern, $m_\mu < 0$, is simply $+1$ if $\mathrm{n_c} = 0$, or proportional to the error, $|m_\mu|$, if $\mathrm{n_c} = 1$. We remark that for both values of $\mathrm{n_c} = 0, 1$, the cost function yields the same global minima $\boldsymbol{\sigma}^*$ at zero-energy (exact solutions). In contrast, the energy landscapes and the local minima are generally different.

Finding optimal solutions $\boldsymbol{\sigma}^*$ where $E_{\mathrm{n_c}}(\boldsymbol{\sigma}^*) = 0$ is a hard optimization problem for either choice of $\mathrm{n_c} = 0, 1$: it has been shown that the energy landscape, in the limit of large N and for $M = \alpha N$ with $\alpha < \alpha_c \approx 0.83$ [137], is characterized by an exponential number of zero-energy solutions and local minima. The latter play the role of metastable states for classical stochastic search algorithms, such as Simulated Annealing (SA) [138], which typically get stuck, for large N, in one of these local minima, with extensive energy costs (of the order $\mathcal{O}(N)$).

More recently, further insight has been gained about the geometrical structure of the ground states [65]. Very schematically, exponentially rare regions where the ground states are dense exist. These regions are composed of an *exponentially large* number of zero-energy solutions at extensive but relatively small Hamming distances, thus possessing very high local entropy. Despite being exponentially rare, these dense regions might be particularly well-suited for making predictions after training since they are less likely to fit noise (small generalization error).

It has been conjectured and shown in Ref. [131] — with analytical and numerical evidence — that quantum fluctuations, as encoded by a Path-Integral Monte Carlo simulated QA [78], are particularly effective in exploring these "dense" regions. Here, we will provide numerical evidence of the enhanced effectiveness of QAOA over QA for small-size perceptron instances, where unitary evolutions are computationally feasible to compare QA and QAOA directly.

The standard quantum mapping of the binary synaptic weights consists of promoting the classical spins σ_j to quantum spin-$1/2$ Pauli operators $\hat{\sigma}_j^z$, as a particular case of the procedure schematized in Eq. (1.13). Let us note that this standard mapping is not the only possibility for encoding classical bits into a quantum setup. For instance, in Ref. [135], the authors implemented a quantum version of the perceptron model by employing the so-called *amplitude encoding*: such a scheme is, in principle, very efficient in terms of memory resources, as it requires $\log_2 N$ quantum spins to represent N classical spins. Still, it pays the price of an exponentially large number of quantum gates necessary for the state preparation [135]. In our study, we focus on leveraging quantum fluctuations to train a *classical* perceptron, rather than implementing a quantum version of it. As a natural choice, in the following, we proceed with the standard base encoding $\sigma_j \rightarrow \hat{\sigma}_j^z$. The target Hamiltonian associated with the perceptron is then given by

$$\widehat{H}_{\mathrm{targ}} = E_{\mathrm{n_c}}\left(\{\hat{\sigma}_j^z\}\right) \ . \tag{2.3}$$

Multi-spin interactions

The Hamiltonian in Eq. 2.3 has a complicated expression in terms of the quantum spin variables $\{\hat{\sigma}_j^z\}$, due to the Heaviside step function: in principle, it may involve all possible multi-spin interactions, up to N-body terms. This structure is shared by other relevant optimization problems, such as the financial crash models considered in Ref. [139], where the target Hamiltonian also involves the Heaviside function. Similar problems featuring high-order multi-body terms can be addressed by employing the parity architecture, see e.g. Ref. [140] and references therein.

In the standard QAOA framework, the expectation value

$$E_{\mathrm{P}}\left(\boldsymbol{\beta}, \boldsymbol{\gamma}\right) = \langle\psi_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma})|\widehat{H}_{\mathrm{targ}}|\psi_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle \tag{2.4}$$

is estimated as the sample mean over a set of measurements on the computational basis. This is done by repeatedly preparing the variational state for fixed parameters $\boldsymbol{\beta} = (\beta_1, \cdots \beta_\mathrm{P})$, $\boldsymbol{\gamma} = (\gamma_1, \cdots \gamma_\mathrm{P})$, i.e.

$$
\begin{cases}
|\psi_\mathrm{P}(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \widehat{U}(\beta_\mathrm{P}, \gamma_\mathrm{P}) \cdots \widehat{U}(\beta_1, \gamma_1)|+\rangle^{\otimes N} \\
\widehat{U}(\beta_m, \gamma_m) = \mathrm{e}^{-i\beta_m \widehat{H}_x} \mathrm{e}^{-i\gamma_m \widehat{H}_z}
\end{cases}
, \tag{2.5}
$$

with $\widehat{H}_x = -\sum_j \hat{\sigma}_j^x$.

Therefore, the diagonal operator $\widehat{H}_\mathrm{targ} = \widehat{H}_z$ is used in building up the QAOA variational state. Whereas an exact gate decomposition of the unitaries $\mathrm{e}^{-i\gamma_m \widehat{H}_z}$ is always possible, it would require impractical resources, i.e. a number of elementary gates growing exponentially with the qubit number N. However, as mentioned in Sec. 1.3.2, the QAOA ansatz allows for some flexibility in the choice of the $\widehat{H}_z$ term appearing in the variational state, with the possibility of replacing it with a simpler set of quantum gates. This aspect is the focus of ongoing research.

Note that the target Hamiltonian also appears in the expectation value in Eq. 2.4. Nonetheless, once the QAOA state — or an approximate version of it — has been prepared, one would perform a measurement and obtain a classical string: Each evaluation of the classical cost function would require $\mathcal{O}(MN)$ complexity, independently of its expansion in terms of binary spins.

2.3. Results

To perform a fair comparison of QAOA against QA (in its digitized form), we consider a set of 10 instances of the perceptron problem for $N = 21$ spins, which were previously analyzed in [131](SI). For each instance, we aim at classifying correctly a training set of $M = 17$ patterns, corresponding to $\alpha = \frac{M}{N} \approx 0.81$. This is close to the critical value $\alpha_c \approx 0.83$, valid in the thermodynamic limit $N \to \infty$, beyond which zero-energy solutions may no longer exist.

Following Ref. [131](SI), these instances were obtained by randomly generating 450 candidate *training set samples*, and: i) keeping only those with a sufficiently large number of solutions (> 21, thus hinting at a non-convex optimization problem); ii) keeping only the instances for which SA failed to reach good approximate solutions. The rationale of selecting these instances is to mimic the typical behavior of larger system sizes [131], which cannot be tackled directly with exact classical simulations.

From here on, we shall refer to a perceptron instance characterized by a specific randomly generated training set simply as *sample*.

2.3.1. Optimal digitized-QA

A natural figure of merit to compare the performance of digitized QA and QAOA is the variational energy density

$$\varepsilon_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{N} E_{\mathrm{P}}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \,, \tag{2.6}$$

defined in terms of Eq. 2.4, with fixed P. For QAOA, we aim at minimizing this quantity w.r.t. the independent free parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$; on the contrary, for digitized-QA (dQA), these are fixed as in Eq. (1.20), hence Δt is the unique free parameter to be optimized.

Let us now focus on dQA. Interestingly, for any given P, we obtain a unique, well-defined global minimum for the variational energy density as a function of Δt. This is shown in Fig. 2.2 for the specific case of P = 64, and both cost function definitions, $n_c = 0$ and $n_c = 1$.
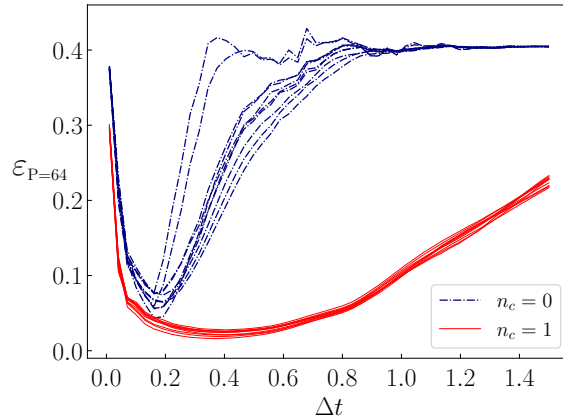
Figure 2.2.: The one-dimensional landscape of the variational energy density for digitized-QA, where the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ depend only on Δt, see Eq. (1.20). All samples examined, and both choices of $n_c = 0, 1$ are shown. The qualitative features of the landscape, and in particular the position of global minima, show mild sample-to-sample variability.

As previously discussed, see also [90], the rationale behind the presence of an optimal Δt is simple: essentially, by increasing Δt at fixed P, ε_{P} initially decreases, because we are allowing for a longer annealing time $\tau = \mathrm{P}\Delta t$; however, upon further increase of Δt, Trotter errors start spoiling the result, leading to a noise-dominated regime. Remarkably, the Δt-landscape and the optimal values depend significantly only on the cost function definition ($n_c = 0, 1$), while much smaller sample-to-sample variability is present. This is the first hint of general qualitative features of the model that are largely independent of the specific sample under consideration. The validity of these results naturally extends to different values of P, as summarized in Appendix B.

2.3.2. Smooth QAOA solutions

In practical implementations of QAOA, both the choice of the classical optimization algorithm and the starting point for the optimization routine can be relevant, particularly when the search space's dimensionality 2P grows, making the optimization harder. The most straightforward approach would be to use a local gradient-based algorithm starting from a random initialization of the parameters. Still, this is often ineffective in practice due to many low-quality local minima and Barren Plateaus, as discussed in Chapter 1.

Effective heuristic warm-start strategies have been proposed, which are based on iterative procedures empirically yielding far better quality results than a random start. Here, we adopt the following strategy. Let us denote by $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$ the optimal linear choice that a digitized QA provides, as discussed above. Using $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$ as a starting point for a Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm [141, 142], we find a minimum, denoted by $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$, which is often "close" to be a smooth curve, with occasional high-frequency localized oscillations of the optimal parameters. We associate these high-frequency oscillations with the presence of spurious minima in the variational energy landscape. As predicted in Refs. [47, 143, 144], we expect the proliferation of such spurious minima to limit the performance of the BFGS algorithm, which is unable to escape towards better solutions. To overcome this performance limitation, we enforce smoothness by applying a smoothing procedure to $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$ and restarting a second BFGS optimization. This leads to a final solution $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$ that is found to be *smooth*, and to provide a systematically better variational minimum compared to the spurious minimum $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$. Schematically, here is the procedure adopted:

$$
\begin{aligned}
(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}}) \rightarrow &\boxed{\text{BFGS optim.}} \xrightarrow{\text{QAOA}-1} (\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)}) \rightarrow \\
\rightarrow &\boxed{\text{Smoothing + BFGS optim.}} \xrightarrow{\text{QAOA}-2} (\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)}) \,.
\end{aligned}
\tag{2.7}
$$

The smoothing procedure can be carried out either by hand or by locally interpolating the curve $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$ with a low-degree polynomial. We remark that details on the smoothing procedure are not particularly relevant, since it only provides a new *educated guess* for the second BFGS optimization, eventually converging to a *smooth* optimal curve $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$.

We now move to illustrate our results in more detail. We performed digitized-QA and QAOA classical simulations for both $n_{\mathrm{c}} = 0, 1$. In this framework, we computed the exact QAOA final state in Eq. (2.5) and the corresponding variational energy in Eq. (2.4) by applying the algebra of Quantum Mechanics. While this analysis was carried out for all samples under study, we now focus, for the sake of clarity, on a single sample (or training set), with similar comments and results applying to all samples.

Fig. 2.3 illustrates the results obtained for this sample. We show two represen-
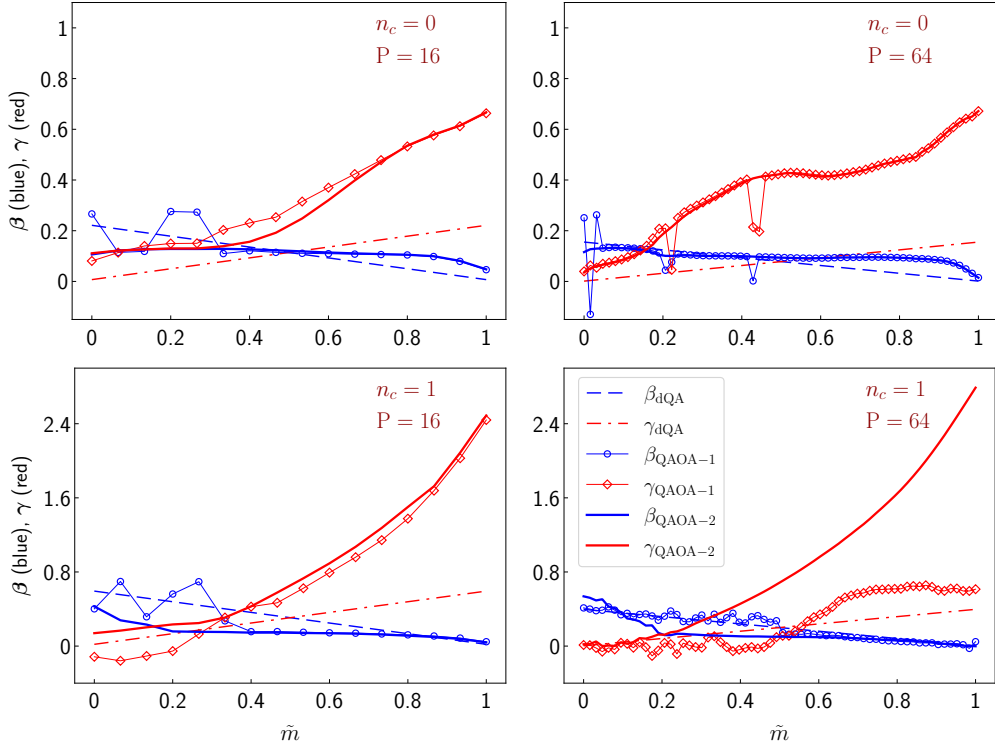
Figure 2.3.: Results for the optimal digitized-QA protocol (dashed and dash-dotted straight lines), QAOA-1 (open symbols with dotted lines), and QAOA-2 (solid lines), for $P = 16, 64$ (left to right) and for $n_c = 0, 1$ (top to bottom). We adopt a uniform x-axis scale in terms of $\tilde{m} = (m-1)/(P-1) \in [0,1]$.

tative values of $P = 16, 64$ (left to right) and both energy-cost functions $n_c = 0, 1$ (top to bottom). The dashed straight lines denote the optimal $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$ solutions. The empty symbols denote the optimal $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$ solutions obtained by a BFGS minimization starting from $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$: our "first shot of QAOA", labelled as QAOA-1. Notice the irregularities on top of an overall "smooth" behavior, particularly evident for $n_c = 0$, where they are quite localized. For $n_c = 0$, we apply a smoothing procedure, and start a "second shot" of QAOA simply as summarized in Eq. (2.7). On the other hand, for $n_c = 1$, irregularities of QAOA-1 solutions are more diffuse. The procedure was slightly modified: we run the second BFGS local minimization from a warm-start point, obtained by interpolation [74, 75] from a smoothed $P = 16$ solution, in power-of-two steps, hence from $P = 16 \rightarrow 32 \rightarrow 64 \cdots$. In both cases, the resulting smooth solutions $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$ are labelled as QAOA-2 and denoted by solid lines. In Appendix B, we summarize a few more technical details concerning these two procedures to single out QAOA-2 smooth solutions; however — as discussed in the next section — we anticipate that these are not particularly crucial: once a smooth solution for a *single* training set sample is found, there is no need to repeat the whole procedure for other samples.

In Fig. 2.4 we plot the minimum values for ε_{P}, Eq. (2.6), by comparing the optimal digitized-QA, QAOA-1 and QAOA-2 protocols. These results show a striking

gain by applying QAOA for both $n_c = 0$ and $n_c = 1$. Moreover, as anticipated, smooth QAOA-2 protocols yield systematically better results compared to QAOA-1 protocols. As expected, the gain is larger for $n_c = 1$ since our QAOA-2 implementation provides (for $P > 16$) a qualitatively different smooth optimal curve. In light of these findings, the QAOA-1 solutions can always be interpreted as spurious local minima — where the classical BFGS optimization gets trapped — systematically of lower quality than the corresponding smoothed QAOA-2 protocols. These results hold for all randomly generated training set samples.
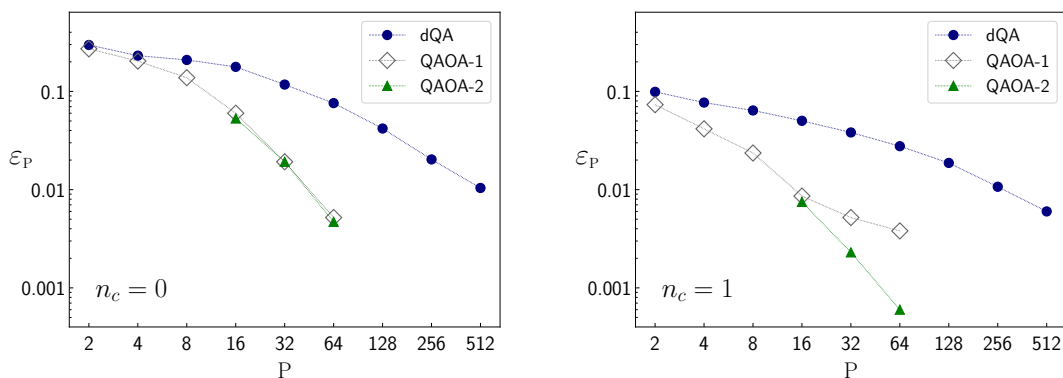
Figure 2.4.: Comparison of variational energy density minima for digitized-QA, QAOA-1 and QAOA-2 for a specific sample of random patterns, with $n_c = 0$ (left) and $n_c = 1$ (right). QAOA-1 outperforms digitized-QA, especially for large values of P. This gain can be further enhanced with the smooth QAOA-2 solution (see main text).

2.3.3. Transferability of a smooth ansatz

The procedure described in the previous section to obtain smooth solutions is elaborate and fine-tuned. Nonetheless, there is no need to repeat it for each training set sample. Indeed, we show numerically that QAOA smooth solutions are transferable among different instances (i.e. for different training set samples). This result seems quite relevant, as our model goes beyond 2-local Hamiltonians and, at the same time, does not satisfy ad-hoc hypotheses on the random coefficients, as discussed in Sec. 1.3.3 for Refs. [102, 103]. To show this, we proceed as follows: For any fixed value of P and $n_c = 0, 1$, separately, we consider QAOA-2 optimal angles $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$ for our first sample and take them as a *smooth model-dependent ansatz* $(\boldsymbol{\beta}^{\mathrm{Ansatz}}, \boldsymbol{\gamma}^{\mathrm{Ansatz}})$ used as an initial point for a BFGS-minimization of a *different* sample. In this way, we can find *smooth* optimal solutions for all other samples, as illustrated in Fig 2.5. Remarkably, these smooth solutions are qualitatively coincident with the $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$ solutions that one would construct by adopting the QAOA-2 procedure previously outlined.

From a practical standpoint, by starting from the smooth ansatz $(\boldsymbol{\beta}^{\mathrm{Ansatz}}, \boldsymbol{\gamma}^{\mathrm{Ansatz}})$,
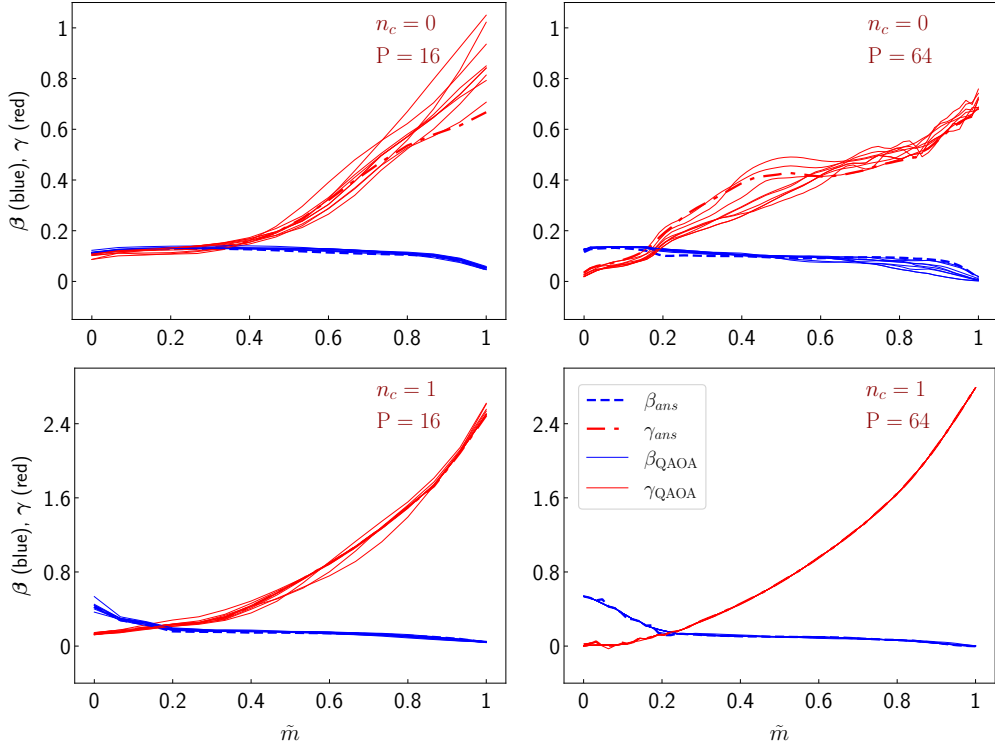
Figure 2.5.: QAOA optimal protocols for all tested samples, as obtained by a BFGS minimization starting from the same *smooth ansatz* ($\boldsymbol{\beta}^{\mathrm{Ansatz}}, \boldsymbol{\gamma}^{\mathrm{Ansatz}}$), corresponding to smooth optimal schedules for the first sample (see main text). Data shown for P = 16, 64 (left to right) and for $n_c = 0, 1$ (top to bottom). We adopt a uniform x-axis scale in terms of $\tilde{m} = (m - 1)/(\mathrm{P} - 1) \in [0, 1]$. The qualitative similarity of QAOA solutions for different samples is remarkable, particularly for $n_c = 1$, where optimal protocols for different samples are almost indistinguishable.

the convergence of the BFGS optimization is always faster. Moreover, we note that all smooth solutions have the same qualitative shape of the smooth ansatz: remarkably, even without a new BFGS-optimization, the previously found ($\boldsymbol{\beta}^{\mathrm{Ansatz}}, \boldsymbol{\gamma}^{\mathrm{Ansatz}}$) already provides a better result compared to an optimal-Δt dQA. This finding is also confirmed by additional simulations on a new (larger) set of samples, randomly drawn without any requirement on the classical hardness or a large number of zero-energy configurations (see Appendix B).

2.3.4. The role of the cost-function landscape geometry

As discussed in Ref. [131], quantum fluctuations are particularly efficient in exploring exponentially rare *dense* regions of solutions in the classical cost-function landscape, defined by Eq. (2.2) for the binary perceptron case. These dense regions are characterized by a large number of classical solutions clustering within a relatively small Hamming distance. This geometrical structure of the landscape is linked to the

good performance of a linear QA schedule, defined by $\widehat{H}(s) = s\widehat{H}_{\mathrm{targ}} + (1-s)\widehat{H}_x$, simply with $s(t) = t/\tau$. Indeed, the instantaneous spectral gap "seen" by the QA dynamics only closes when approaching the end of the protocol $s \to 1$, where $\widehat{H}(s) \to \widehat{H}_{\mathrm{targ}}$, due to the degeneracy of classical solutions.

An interesting benchmark is provided by scrambling this geometrical structure by permuting the classical energies associated with each configuration. Despite keeping the spectrum unchanged, this procedure yields an unstructured problem, previously studied in Refs. [145, 146]. This scrambling was shown to be detrimental to QA in a general setting [145] and for our specific case [131]: it causes a sharp drop of the instantaneous spectral gap at a finite $s_c < 1$, the usual bottleneck of QA, and a drastic worsening of its performance. It is natural to investigate to which extent QAOA might be able to cope with such scrambling of the cost-function landscape geometry and its associated ultra-narrow spectral gap (avoided level-crossing). After all, QAOA is based on the variational principle rather than on the adiabatic theorem. However, smooth QAOA solutions might signal an "optimal adiabatic schedule" [75], which might suggest a worse performance.

To answer such a question, we adopted the following strategy. For each sample considered, we generate a corresponding *randomized sample* by permuting the classical energies associated with each configuration $\hat{\boldsymbol{\sigma}}$, to retain the same classical energy spectrum while destroying any geometrical feature of the classical energy landscape. In summary, the randomized samples should be interpreted as a benchmark: they provide (on average) the *most difficult* optimization problem (no geometrical structure), retaining the identical full spectra of the original samples. We then proceed along the same lines of Sec. 2.3: by starting from an optimal-Δt digitized-QA solution for the randomized samples, we run QAOA and compare the minima of the variational energy density for the two methods. Incidentally, in Appendix B, we show a comparison of digitized-QA optimal-Δt values for the two cases of original vs randomized samples.

By following the same scheme outlined in Eq. (2.7), it is again possible to single out *smooth* QAOA-2 solutions, being at the same time qualitatively different compared to those for the original samples, and *transferable* among different randomized samples (data not shown). Fig. 2.6 compares the minimized variational energy density obtained from digitized-QA and QAOA-2 for the randomized version of the sample reported in Fig. 2.4; the original digitized-QA and QAOA-2 results are also reported, to ease the comparison. Two main remarks are worthwhile: 1) QAOA-2 solutions for the randomized sample considerably improve on the corresponding digitized-QA results, especially for $n_c = 0$, where they become comparable to the original sample digitized-QA results; 2) the quality of these QAOA-2 solutions is much lower compared to QAOA-2 solutions for the original sample, witnessing a degradation of performance. These same comments apply to all samples examined.
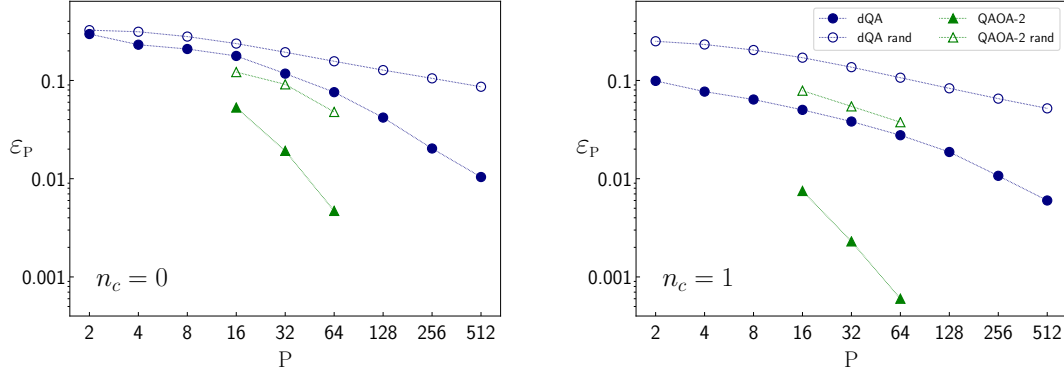
Figure 2.6.: Variational energy density minima for digitized-QA and QAOA-2, comparing results for the original sample (the same as in Fig. 2.4, full symbols) with those obtained by a *randomization* of its energy spectrum (empty symbols), for both $n_c = 0, 1$.

To better understand the basic mechanism behind such a degradation of performance, we plot in Fig. 2.7 the QAOA-2 smooth protocols for $P = 64$, re-expressed in terms of $s_m = \gamma_m/(\gamma_m + \beta_m)$, a parameter that in digitized-QA linearly interpolates from $s = 0$ to $s = 1$ during the annealing process, see Eq. (1.27) and the accompanying analysis. The two figures correspond to the original (left) and the randomized samples (right), in the case $n_c = 0$. These optimal schedules should be compared with the instantaneous spectral gap $\Delta(s) = E_{\text{ex}}(s) - E_{\text{gs}}(s)$, which is plotted in the inset for both cases. We observe that the instantaneous gap of randomized samples displays, as expected [131](SI), an avoided level-crossing close to the numerical value $s_c = 0.725$. Correspondingly, in proximity of s_c, the optimal schedule parameter s_m shows a wide marked plateau, particularly evident for large values of P: this is reminiscent of a "slowing down" of the annealing near the gap closure, a kind of "optimal adiabatic schedule" [75], unfortunately unable to fully overcome the basic limitations of the adiabatic mechanism. We note that these QAOA-2 smooth solutions show striking similarity with adiabatic schedules obtained for unstructured (Grover) search in Refs. [115, 147].

However, there is no assurance that the smooth solutions derived from our QAOA-2 approach represent the true global minimum within the 2P-dimensional variational energy landscape: in principle, there might be other better-performing QAOA protocols, e.g. similar to the "shortcut to adiabaticity" (STA) schedules found in some hard-instances of 3-MaxCut [74]. The STA strategy goes beyond the adiabatic paradigm and is a promising framework to overcome the shortcomings of the adiabatic mechanism. Refs. [127, 148, 149], have recently developed counterdiabatic-QAOA (CD-QAOA) approaches, to extend the QAOA variational ansatz by including terms that generate STAs. Since STA protocols are generally smooth, a generalization of QAOA-2 might succeed in finding them. However, this requires developing efficient strategies [150] to modify the linear guess inspired by dQA to
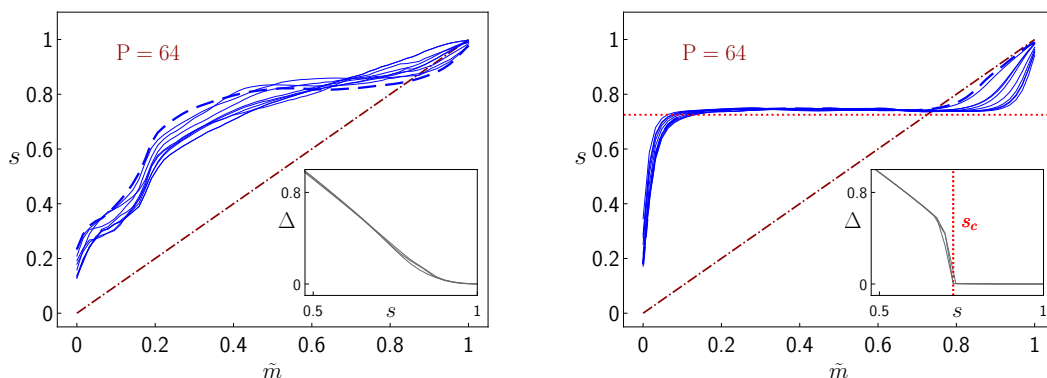
target STA protocols specifically.

Figure 2.7.: Plot of the QAOA-2 smooth protocols for P = 64, in terms of $s_m = \gamma_m/(\gamma_m + \beta_m)$ for both the original (left) and the randomized samples (right), in the case $n_c = 0$. In both figures, the thick dashed line represents the transferable smooth ansatz obtained by a detailed study of the first sample, while data for other samples are obtained by exploiting this transferability result. As a visual reference, we also plot the digitized-QA linear interpolation from $s = 0$ to $s = 1$. In the insets, we show the instantaneous spectral gap $\Delta(s) = E_{\mathrm{ex}}(s) - E_{\mathrm{gs}}(s)$: the gap starts at $\Delta(s = 0) = 2$ (single spin-flip excitation of $\widehat{H}_x$) and it vanishes for $s = 1$, due to the degeneracy of $\widehat{H}_z$, while negligible sample-to-sample variability is observed. Remarkably, $\Delta(s)$ shows a sharp drop around $s_c = 0.725$ for the randomized samples, and a wide plateau is highlighted in the corresponding smooth optimal schedules by a red dotted horizontal line. Similar results and comments apply for $n_c = 1$.

2.4. Conclusion and outlook

In this Chapter, we aimed to provide evidence for the applicability and effectiveness of digitized-QA and QAOA to solve hard classical optimization problems defined by highly non-local Hamiltonians, well beyond the usual 2-local models studied e.g. for MaxCut and ground state preparation of a quantum spin model. To do so, we employed an optimization scheme that leverages the transferability of *smooth* optimal QAOA schedules among different instances of the same problem (different training sets). Our results show similarity to previous work on parameter concentration (instance independence) and might be further investigated, possibly leading to analytic results in the large N limit, as in the papers discussed in Sec. 1.3.3 [102, 103]. Inspired by the analysis in Ref. [131], we assessed the importance of the geometrical structure of the classical cost-function landscape. An artificial permutation of the energy spectrum leads to a vanishing spectral gap along the annealing path. In this regime, QAOA still provides some advantage vs an optimized-Δt dQA; however, it seems to perform a kind of "optimal adiabatic schedule", unable to fully overcome the basic limitations of the adiabatic mechanism.

The primary objective for future developments, a.k.a. the elephant in the room, is to design smart approximation schemes, enabling an efficient implementation of unitaries generated by a highly non-local classical Hamiltonian. This is necessary to implement our proposal on actual quantum hardware and assess its effectiveness beyond classical simulation capabilities. We remark that similar challenges are also encountered if one wishes to implement an analog real-time evolution on a Quantum Annealer, as proposed in [131]. A possible step forward in this direction is to extend the standard formulation of QAOA, as proposed in Ref. [33]. Indeed, one could redefine the variational energy in Eq. (2.4) as follows: 1) keep the same exact classical $\widehat{H}_{\text{targ}} = \widehat{H}_z$ in the objective function expectation value and 2) redefine the diagonal unitary (which is quantum-computationally hard) generated by $\widehat{H}_z$ — *e.g.*, by using a simplified version of it, encoding some minimal information on the problem. This would replace QAOA with a more hardware-friendly PQC, admitting a simpler gate decomposition. Moreover, one might even redefine the *mixing* unitary generated by $\widehat{H}_x$, by replacing it with another operator inducing tailored quantum fluctuations.

An alternative strategy for developing a scalable algorithm involves employing a parameterized quantum circuit independent of the problem Hamiltonian. This design would be specifically adapted to the resources of existing quantum devices, such as native gates and qubit connectivity [151, 152]. This approach enjoys clear benefits in the ease of implementation, but it is expected to suffer well-known drawbacks such as the proliferation of local minima and Barren Plateaus. After an efficient parameterized quantum circuit is implemented, applying the techniques described above would be straightforward to assess the presence of smooth optimal protocols and their transferability. Subsequently, one could test the robustness of these protocols to shot-noise in the optimization or to mild gate errors.

Another option is to conduct *efficient* classical simulations of quantum optimization methods, e.g. relying on tensor networks techniques. These concepts have been explored in our publication in Ref. [4], but are not covered in this Thesis.

A simple gate-efficient modification of QAOA, which we have tested, is the following. Rather than using $\widehat{H}_{\text{targ}} = \widehat{H}_z$ in the quantum gates $\mathrm{e}^{-i\gamma_m \widehat{H}_z}$, we use the Sherrington-Kirkpatrick (SK) model Hamiltonian, which derives from taking the quadratic approximation $(|m_\mu|)\,\Theta(-m_\mu) \to -m_\mu + m_\mu^2$ in the $n_c = 1$ cost function in Eq. (2.2), as depicted in Fig. 2.8. This quite arbitrary approximation, upon using Eq. (2.1), leads to:

$$\widehat{H}_z = -\sum_{j=1}^{N} h_j \hat{\sigma}_j^z + \sum_{j \neq j'} J_{jj'} \hat{\sigma}_j^z \hat{\sigma}_{j'}^z \, , \tag{2.8}$$

where $h_j = \frac{1}{\sqrt{N}}\sum_{\mu=1}^{M} \xi_j^\mu$ is a local field provided by all the input patterns at site j, while $J_{jj'} = \frac{1}{N}\sum_{\mu=1}^{M} \xi_j^\mu \xi_{j'}^\mu$ is the standard Hebbian-rule coupling [134]. Unfortunately, such a choice of $\widehat{H}_z$ appears to dramatically decrease the performance of QAOA. However, recall that we are studying the supervised learning of random
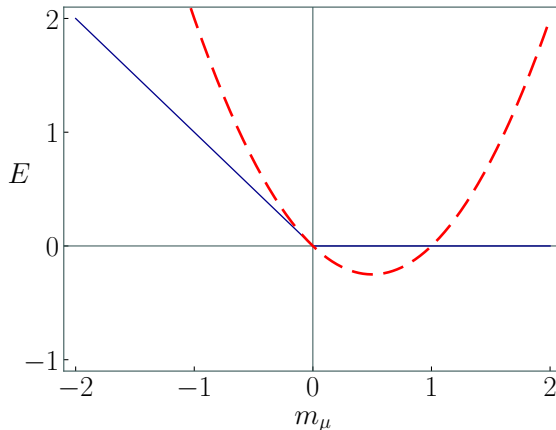
Figure 2.8.: The exact energy cost for the classification of a single pattern in the training set (solid blue line): the sum of these contributions yields $E_{\mathrm{n_c}=1}$ in Eq (2.2). This is compared to a quadratic approximation (dashed red line), see the main text.

patterns drawn as IID random variables from a given distribution. This approach is ubiquitous in the statistical physics community but leads to a dataset without any structure. In contrast, in real-world applications, supervised learning is always conducted on datasets with a structure: a standard testbed consists of datasets of hand-written digits. In this framework, preliminary results indicate that a 2-local expansion might be a valid approximation for the classical loss function of simple instances of binary neural networks [153]. A potentially more effective strategy to enhance performance involves maintaining a Hamiltonian with only two-body interactions, by searching for an optimal SK model, either by iteratively identifying the optimal J and h parameters or by defining an appropriate inverse Ising model [154].

Alongside a better account for implementability and gate decomposition, a natural extension of our work is to examine the training of more complex binary neural networks, possibly further generalizing to low-precision discrete-weights neural networks. This could be tested numerically on more realistic correlated datasets (data with a structure) and is a subject of ongoing research. This analysis would allow us to better investigate the role of the classical cost-function landscape geometry, which is expected to be crucial for the effectiveness of QAOA or analogous quantum optimization schemes. Furthermore, it would be possible to evaluate the generalization error using a test set composed of previously unseen data that follows the same distribution.

3. Barren Plateaus in the Hamiltonian Variational Ansatz

> *Simulating the behavior of 100 billion neurons of the human brain is not feasible by classical computers but quantum machine learning promises to fulfill that requirement.*
>
> —Amit Ray, Indian author and "spiritual master", 2023.
> A *slightly* optimistic take on quantum computing.

In Chapter 1, we introduced and motivated the unprecedented scientific and technological interest in Variational Quantum Algorithms (VQAs). To obtain any *practically useful* and *scalable* hybrid quantum-classical algorithm, the scientific community must not only overcome formidable experimental issues but also solve theoretical and algorithmic challenges.

The cost function optimization is known to be a difficult task [48]: only a careful choice of the ansatz is usually *expressive* enough to approximately find the ground state of $\widehat{H}_{\mathrm{targ}}$ and, at the same time, *trainable* enough for the optimization to succeed. In particular, the landscape of the cost function may not be easy to inspect for two reasons: the proliferation of low-quality local minima traps, and the exponential flattening of the landscape by increasing the number of qubits, a phenomenon dubbed *barren plateaus*, which can severely hinder the scalability of the VQA scheme beyond small system sizes amenable to classical simulations.

A review of barren plateaus and trainability issues in VQAs is provided in Chapter 1, where we also mention some promising approaches to boost trainability.

Here, we draw a new connection between smooth optimal solutions — obtained employing iterative methods — and barren plateaus, developing a novel, efficient scheme to circumvent this issue. Our procedure leverages the *transferability* of a *smooth* optimal solution, obtained for a small system size, to solve the same task with a larger number of qubits, where a direct optimization would fail due to barren plateaus. In a nutshell, the transferred smooth solution serves as an excellent warm-start with low variational energy for the large system, and a subsequent refinement optimization is observed to be free of the barren plateau issue. Remarkably, even though other (non-smooth) solutions for the small system can be obtained by standard random-start local optimization, they do not provide any useful warm-start

for larger systems and, crucially, a refinement optimization still suffers from barren plateaus in their neighborhood.

3.1. Problem definition

For definiteness, we focus on the Heisenberg XYZ model [128] and the antiferromagnetic Longitudinal-Transverse-Field Ising Model (LTFIM) [129], two ubiquitous models in quantum physics with rich phase diagrams, whose ground state preparation with VQAs is affected by barren plateaus [55, 57]. We select ansatz states in the Hamiltonian Variational Ansatz (HVA) form

$$|\psi_{\mathrm{P}}(\boldsymbol{\theta})\rangle = \prod_{m=1}^{\mathrm{P}} \mathrm{e}^{-i\theta_{m,M}\widehat{H}_M} \cdots \mathrm{e}^{-i\theta_{m,1}\widehat{H}_1}|\psi_0\rangle \,, \tag{3.1}$$

where $|\psi_0\rangle$ is a simple initial state, and $m = 1\cdots \mathrm{P}$ labels successive circuit layers, each composed by $j = 1\cdots M$ alternating unitaries generated by Hamiltonian operators $\widehat{H}_j$. The target Hamiltonian $\widehat{H}_{\mathrm{targ}}$ can be linearly decomposed in terms of the generators, which are selected to enforce model symmetries into the variational wavefunction: this procedure is detailed in Sec. 3.1.1. This way, we restrict the Hilbert space to the ground state symmetry sector, boosting trainability. Moreover, this choice leads to a reduction in the number of independent Pauli correlators needed to compute the cost function, namely the usual variational energy:

$$E_{\mathrm{P}}(\boldsymbol{\theta}) = \langle \psi_{\mathrm{P}}(\boldsymbol{\theta})|\widehat{H}_{\mathrm{targ}}|\psi_{\mathrm{P}}(\boldsymbol{\theta})\rangle \,. \tag{3.2}$$

The first class of models we consider is the spin-1/2 XYZ [128, 155] Hamiltonian:

$$\widehat{H}_{\mathrm{XYZ}} = \sum_{j=1}^{N} \left(\hat{\sigma}_j^x \hat{\sigma}_{j+1}^x + \Delta_{\mathrm{Y}} \hat{\sigma}_j^y \hat{\sigma}_{j+1}^y + \Delta_{\mathrm{Z}} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z \right) \,. \tag{3.3}$$

We restrict our considerations to the antiferromagnetic case $\Delta_{\mathrm{Y}}, \Delta_{\mathrm{Z}} > 0$. In this quadrant, the system is gapped, except at three critical half-lines/segments: $\Delta_{\mathrm{Y}} \leq 1, \Delta_{\mathrm{Z}} = 1$; $\Delta_{\mathrm{Y}} = 1, \Delta_{\mathrm{Z}} \leq 1$; $\Delta_{\mathrm{Y}} = \Delta_{\mathrm{Z}}, \Delta_{\mathrm{Z}} \geq 1$ [156, 157]. The Hamiltonian (3.3) is integrable in the whole $(\Delta_{\mathrm{Y}}, \Delta_{\mathrm{Z}})$ plane. In particular, $\Delta_{\mathrm{Y}} = 1$ corresponds to the XXZ model, while $\Delta_{\mathrm{Y}} = \Delta_{\mathrm{Z}} = 1$ corresponds to the spin-isotropic Heisenberg model. The second Hamiltonian we examine is the antiferromagnetic LTFIM [158, 159]:

$$\widehat{H}_{\mathrm{LTFIM}} = \sum_{j=1}^{N} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - g_x \sum_{j=1}^{N} \hat{\sigma}_j^x - g_z \sum_{j=1}^{N} \hat{\sigma}_j^z \,. \tag{3.4}$$

We restrict our analysis to positive local fields $g_x, g_z > 0$. The system is gapped in the whole positive quadrant, except for a line connecting the two points $(g_x = 1, g_z = 0)$ and $(g_x = 0, g_z = 2)$, obtained numerically in Ref. [158]. While for $g_z = 0$ the model

is integrable by a Jordan-Wigner transformation to free fermions [27, 75, 160–162], integrability is generically lost for $g_z \neq 0$. In Sec. 3.3.4, we specifically address the integrable Transverse Field Ising Model (TFIM) line ($g_z = 0$). For both the XYZ model and the LTFIM, we examine even values of N, and assume periodic boundary conditions.

As anticipated, our ansatz states are in the general form of Eq. (3.1) with $M = 2$ generating Hamiltonians only, defined to encode some symmetries of the model. To illustrate this idea for the XYZ case, let us split $\widehat{H}_{\mathrm{XYZ}}$ into two mutually non-commuting parts that refer to the *even* $(2j-1, 2j)$ and to the *odd* bonds $(2j, 2j+1)$, $\widehat{H}_{\mathrm{XYZ}} = \widehat{H}_{\mathrm{even}} + \widehat{H}_{\mathrm{odd}}$, with

$$\widehat{H}_{\mathrm{even}} = \sum_{j=1}^{N/2} \left(\hat{\sigma}_{2j-1}^x \hat{\sigma}_{2j}^x + \Delta_{\mathrm{Y}} \hat{\sigma}_{2j-1}^y \hat{\sigma}_{2j}^y + \Delta_{\mathrm{Z}} \hat{\sigma}_{2j-1}^z \hat{\sigma}_{2j}^z \right) \tag{3.5}$$

and similarly for $\widehat{H}_{\mathrm{odd}}$. Next, in the spirit of AQC [81], imagine an *interpolating Hamiltonian* connecting $\widehat{H}_{\mathrm{even}}$ to the full $\widehat{H}_{\mathrm{XYZ}}$:

$$\widehat{H}(s) = s\widehat{H}_{\mathrm{XYZ}} + (1-s)\widehat{H}_{\mathrm{even}} = \widehat{H}_{\mathrm{even}} + s\widehat{H}_{\mathrm{odd}} , \tag{3.6}$$

with $s \in [0, 1]$. For $s = 0$, the ground state of $\widehat{H}(0) = \widehat{H}_{\mathrm{even}}$ is a valence-bond state of singlets on the even bonds

$$|\psi_0\rangle = \prod_{j=1}^{N/2} \frac{1}{\sqrt{2}} \Big(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \Big)_{2j-1,2j} , \tag{3.7}$$

which is taken as the initial state. This suggests, in close analogy with QAOA, the following ansatz for the XYZ ground-state wavefunction:

$$|\psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}\rangle = \widehat{U}_{\mathrm{P}} \cdots \widehat{U}_2 \, \widehat{U}_1 |\psi_0\rangle . \tag{3.8}$$

Here, $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} = (\beta_1 \cdots \beta_{\mathrm{P}}, \alpha_1 \cdots \alpha_{\mathrm{P}})$ are 2P variational parameters, and the unitary operators $\widehat{U}_m = \widehat{U}(\beta_m, \alpha_m)$, for $m = 1 \cdots \mathrm{P}$, evolve the state according to $\widehat{H}_{\mathrm{even}}$ and $\widehat{H}_{\mathrm{odd}}$, in an alternating fashion:

$$\widehat{U}_m = \widehat{U}(\beta_m, \alpha_m) = \mathrm{e}^{-i\beta_m \widehat{H}_{\mathrm{even}}} \mathrm{e}^{-i\alpha_m \widehat{H}_{\mathrm{odd}}} . \tag{3.9}$$

As usual in the HVA framework, the goal is to minimize the variational energy

$$E_N (\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} = \langle \psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} | \widehat{H}_{\mathrm{targ}} | \psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} \rangle , \tag{3.10}$$

with $\widehat{H}_{\mathrm{targ}} = \widehat{H}_{\mathrm{XYZ}}$. We explicitly indicated the system size N and the QAOA depth P. The connection with AQC is restored in the $\mathrm{P} \to \infty$ limit, by setting specific values for $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}$, as prescribed by a Trotter split-up of the continuous-time AQC dynamics [75].

Before moving to a more detailed analysis, let us summarize here the main points of the discussion provided below in Sec 3.1.1. Firstly, the ansatz state lies in the same symmetry subsector of the XYZ ground states for the following symmetries: translations by *two* lattice spacings $\widehat{T}^2$ (which maps $j \to j + 2$), lattice inversion $\widehat{I}$ (which maps $j \leftrightarrow N - j + 1$) and parity $\widehat{P}_b = \prod_j \hat{\sigma}_j^b$, with $(b = x, y, z)$. Additionally, for the su(2)-invariant Heisenberg model, this holds true for the total spin $\widehat{S}_{\mathrm{tot}}^b$ ($b = x, y, z$) and $\widehat{S}_{\mathrm{tot}}^2$, while for the u(1)-invariant XXZ model only for $\widehat{S}_{\mathrm{tot}}^z$. As a result, the variational energy in Eq. (3.10) requires the evaluation of only six independent two-point correlators, which may be further reduced to four (two) by exploiting rotational symmetries in the XXZ (XXX) case.

The ansatz for the ground state preparation of the LTFIM is defined along similar lines. Indeed, it reads as in Eq. (3.8), with a single layer unitary given by:

$$\widehat{U}_m = e^{i\beta_m \widehat{H}_{\mathrm{X}}} e^{-i\alpha_m \left(\widehat{H}_{\mathrm{ZZ}} - g_z \widehat{H}_{\mathrm{Z}}\right)} , \qquad (3.11)$$

where we defined $\widehat{H}_{\mathrm{ZZ}}$, $\widehat{H}_{\mathrm{Z}}$, and $\widehat{H}_{\mathrm{X}}$ simply as the sum of nearest-neighbors interactions, Pauli-z and Pauli-x operators, respectively. In this setting, the initial state is simply the fully polarized state along-x $|\psi_0\rangle = |+\rangle^{\otimes N}$, once again bearing a direct connection with AQC state preparation for $P \to \infty$. The goal is to minimize the variational energy as in Eq. (3.10), now with $\widehat{H}_{\mathrm{targ}} = \widehat{H}_{\mathrm{LTFIM}}$. Also in this case (see Sec 3.1.1), the variational ansatz is restricted to the correct symmetry subsector of the target ground state, for both single-site translation $\widehat{T}$ (full translational invariance) and lattice inversion $\widehat{I}$.

3.1.1. Symmetries encoded in the HVA

An important aspect of the story concerns the symmetries of the ground state we need to construct and those of our ansatz wavefunction. This section describes the relevant symmetries of the XYZ models and the LTFIM. We will show that the choice of ansatz wavefunctions as in Eqs. (3.8), (3.9), (3.11) restricts the search space from the whole Hilbert space to a specific symmetry sub-sector, which is precisely the one where the target ground state belongs.

Let us start with general considerations valid for problem-inspired HVA in Eq. (3.1). Suppose we identify a set of symmetries of the target Hamiltonian $\widehat{H}_{\mathrm{targ}}$ — whose ground state we aim to prepare — and let us focus in particular on a specific symmetry (unitary) operator $\widehat{S}$.

In practice, a smart strategy can be to select the generators $\widehat{H}_1, \ldots, \widehat{H}_M$ such that they *all* commute with $\widehat{S}$. In fact, if we select as the initial state an easy-to-prepare symmetry eigenstate $\widehat{S}|\psi_0\rangle = e^{i\phi}|\psi_0\rangle$, it immediately follows that the HVA is confined to the same symmetry subsector, since:

$$\widehat{S}|\psi(\boldsymbol{\gamma})\rangle = e^{i\phi}|\psi(\boldsymbol{\gamma})\rangle , \qquad (3.12)$$

for any choice of the variational parameters. The rationale behind this procedure is simple: we should select the correct symmetry subsector, where the target ground state belongs, by properly choosing $|\psi_0\rangle$: this sector is then preserved by applying only symmetry-commuting unitaries.

For clarity, let us now restrict our discussion to the XYZ model; the LTFIM requires only minor changes, which are summarized at the end of this section. As a preliminary observation, we remark that the various bond terms appearing in $\widehat{H}_{\text{even}}$ (or equivalently, in $\widehat{H}_{\text{odd}}$) form a set of mutually commuting operators, hence the corresponding unitaries factorize; even more, since $\left[\hat{\sigma}_j^b \hat{\sigma}_{j+1}^b, \hat{\sigma}_j^{b'} \hat{\sigma}_{j+1}^{b'}\right] = 0$ for $b, b' = x, y, z$, the various unitaries also factorize in the xx, yy and zz terms. This leads to a standard parameterized quantum circuit, which can be further decomposed into a basis set of native gates (e.g. CNOT and single-qubit rotations) [5]. Let us simplify the notation for this ansatz state, evaluated in a generic point of the search space, by setting $|\psi_{\text{P}}\rangle \equiv |\psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\text{P}}\rangle$.

The initial state $|\psi_0\rangle$ in Eq. (3.7) has obvious symmetries with respect to translations by *two* lattice spacings $\widehat{T}^2$ (which sends $j \to j + 2$), lattice inversion $\widehat{I}$ (which maps $j \leftrightarrow N - j + 1$), parity $\widehat{P}_b = \prod_j \hat{\sigma}_j^b$ and total spin $\widehat{S}_{\text{tot}}^b$, with $b = x, y, z$, as well as $\widehat{S}_{\text{tot}}^2$. Clearly, $\widehat{S}_{\text{tot}}^2 |\psi_0\rangle = 0$, and $\widehat{S}_{\text{tot}}^b |\psi_0\rangle = 0$. The singlets, however, are *odd* under the exchange of the two spins and also under the application of $\widehat{P}_b$. Hence, while $\widehat{T}^2 |\psi_0\rangle = |\psi_0\rangle$, we have that $\widehat{P}_b |\psi_0\rangle = (-1)^{\frac{N}{2}} |\psi_0\rangle$ and $\widehat{I} |\psi_0\rangle = (-1)^{\frac{N}{2}} |\psi_0\rangle$.

Concerning the symmetries of the ansatz state $|\psi_{\text{P}}\rangle$, they are inherited by the symmetries of $\widehat{H}_{\text{even}}$ and $\widehat{H}_{\text{odd}}$. Hence, full spin rotational invariance is broken except for $\Delta_{\text{Y}} = \Delta_{\text{Z}} = 1$; for $\Delta_{\text{Y}} = 1$, $\widehat{S}_{\text{tot}}^z$ symmetry is preserved and $\widehat{S}_{\text{tot}}^z |\psi_{\text{P}}\rangle = 0$. Moreover, since both $\widehat{H}_{\text{even}}$ and $\widehat{H}_{\text{odd}}$ commute with $\widehat{T}^2$, $\widehat{P}_b$ and $\widehat{I}$, we immediately deduce that:

$$
\begin{aligned}
\widehat{T}^2 |\psi_{\text{P}}\rangle &= |\psi_{\text{P}}\rangle & \widehat{I} |\psi_{\text{P}}\rangle &= (-1)^{\frac{N}{2}} |\psi_{\text{P}}\rangle \\
\widehat{P}_b |\psi_{\text{P}}\rangle &= (-1)^{\frac{N}{2}} |\psi_{\text{P}}\rangle .
\end{aligned}
$$

These are precisely the quantum numbers of the ground states we want to construct for $\widehat{H}_{\text{XYZ}}$, as it can be verified numerically for small-size systems with exact diagonalization.

Restricting the variational wavefunction to the ground state symmetry subsector may foster trainability, but this is not the only practical advantage. Indeed, also the number of independent Pauli correlators needed to compute the variational energy is reduced. To prove this fact, we need to obtain an explicit formula for the variational energy in Eq. (3.10). It is useful to introduce the k-points correlation functions:

$$
C_{i,j,\ldots,k}^{b_i, b_j, \ldots, b_k} (\boldsymbol{\beta}, \boldsymbol{\alpha})_{\text{P}} := \langle \psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\text{P}} | \hat{\sigma}_i^{b_i} \hat{\sigma}_j^{b_j} \cdots \hat{\sigma}_k^{b_k} | \psi(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\text{P}} \rangle , \tag{3.13}
$$

where the lower indices enumerate the involved spins $i, j, k = 1 \cdots N$, while the upper indices assign corresponding directions $b = x, y, z$. Contrarily to quantum chemistry applications [163] or some classical optimization problems (see discussion

in Chapter [2]), our quantum spin model is 2-local, hence the expectation value in Eq. (3.10) only requires calculating two-point correlators of the type

$$C_{i,j}^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} := \langle \psi \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} | \hat{\sigma}_i^b \hat{\sigma}_j^b | \psi \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} \rangle \ . \tag{3.14}$$

In addition, $C_i^b = \langle \psi_{\mathrm{P}} | \hat{\sigma}_i^b | \psi_{\mathrm{P}} \rangle = 0$ thanks to the parity symmetry. Importantly, we can exploit ansatz symmetries to reduce the number of correlators needed: from the $\widehat{\mathrm{T}}^2$ symmetry, it immediately follows that

$$C_{(2j-1),(2j-1)+i}^b = C_{1,1+i}^b \ , \quad C_{2j,2j+i}^b = C_{2,2+i}^b \ . \tag{3.15}$$

Additionally, since only nearest-neighbors correlators are needed, the $\widehat{H}_{\mathrm{XYZ}}$ expectation value reduces to

$$\frac{2}{N} E_N \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} = \sum_{b=x,y,z} C_{1,2}^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} + C_{2,3}^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} \ , \tag{3.16}$$

involving only six independent correlators, which may be further reduced to four (two) by exploiting rotational symmetries in the XXZ (XXX) case. Moreover, a significant reduction in the number of shots to estimate expectation values in real experiments might be accomplished by a final rotation into the Bell basis at the end of the circuit, allowing direct access to the correlator statistics for $b = x, y, z$ by usual measurements in the computational basis [164].

It is relevant to notice that this symmetry-encoding procedure can be applied only to a subset of symmetries of $\widehat{H}_{\mathrm{targ}}$, and it need not be applied to all of them. Indeed, for the XYZ models, our ansatz state encodes all the aforementioned symmetries of $\widehat{H}_{\mathrm{targ}} = \widehat{H}_{\mathrm{XYZ}}$, whereas it does not encode its one-site translational symmetry. Nevertheless, the latter is almost exactly restored for optimal variational parameters, as clearly shown in Sec [3.3.1] (see Fig [3.3]).

The previous discussion extends straightforwardly to the LTFIM, with minor modifications. This variational ansatz lies in the same symmetry subsector as the target ground state: precisely, both are eigenstates with eigenvalue $+1$ of the symmetry operators $\widehat{\mathrm{T}}$ (full translational invariance) and $\widehat{\mathrm{I}}$. This fact is once again easily verified since the initial fully-polarized along-x state $|\psi_0\rangle$ is in the same symmetry sector, and both symmetries commute with the generators of the HVA wavefunction. Incidentally, note that the LTFIM parameterized quantum circuit reduces to the usual QAOA ansatz for the TFIM ($g_z = 0$), where also the parity symmetry $\widehat{\mathrm{P}}_x$ is restored, and the QAOA ansatz has the same eigenvalue of the ground state (for $g_x > 0$).

Along the same lines as the previous discussion, due to the full translational invariance, we now have:

$$\begin{aligned} C_{i,j}^z \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} &= C_{1,2}^z \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} \\ C_i^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} &= C_1^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} \ , \end{aligned} \tag{3.17}$$

for any $b = x, y, z$, leading to the following expression for the variational energy:

$$\frac{1}{N} E_N \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} = C_{1,2}^z \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} - \sum_{b=x,z} g_b \, C_1^b \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} \ . \qquad (3.18)$$

This reduction in the number of independent correlators — thanks to symmetry encoding in HVA — is handy for classical simulations but also in realistic experiments on a quantum device [53].

As a final remark concerning additional symmetries, since both the XYZ and LT-FIM Hamiltonians are real-valued matrices, and we are using ansatz wavefunctions of the generic form in Eq. (3.1), then $E_N \left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)_{\mathrm{P}} = E_N \left(-\boldsymbol{\beta}, -\boldsymbol{\alpha}\right)_{\mathrm{P}}$ (time-reversal symmetry).

3.1.2. Alternative HVA implementations

Before moving to the results, let us briefly mention alternative HVA implementations that have appeared in the recent literature. To this end, we remark that the number of variational parameters of our HVA for the XYZ models (defined by Eqs. (3.8), (3.9)) is always equal to 2P. This is at variance with a different version of the HVA for the XXZ model, studied e.g. in [56], which introduces more parameters to account for the possible spin anisotropies in the Hamiltonian: here, on the contrary, the $\widehat{H}_{\mathrm{XYZ}}$ spin anisotropies are directly accounted for by $\widehat{H}_{\mathrm{even}}$ and $\widehat{H}_{\mathrm{odd}}$ in Eq. (3.9), using only two parameters per layer. Another possible approach was tested in [165], by adopting a more general class of ansatz wavefunctions, with a number of variational parameters per layer proportional to the system size N.

Alternative HVA formulations have also been proposed for LTFIM ground state preparation, as in Ref. [55, 108]. Once more, with our definitions, we only need two variational parameters per layer, regardless of the phase diagram point under consideration.

3.2. The reverse causal cone for XYZ and LTFIM

A natural interpretation in terms of the spreading of quantum correlations emerges for both our ansatz wavefunctions. This is an example of a "reverse causal cone", as discussed in Sec. 1.3.3 to explain the concentration and transferability of optimal parameters for MaxCut, in the regime of shallow circuits and large system size. As we shall now explain, it follows that the whole variational energy landscape, once rescaled by the system size N, becomes independent of N for $N > \tilde{N}_{\mathrm{P}}$, where $\tilde{N}_{\mathrm{P}} = 4\mathrm{P} + 2$ and $\tilde{N}_{\mathrm{P}} = 2\mathrm{P} + 1$ for the XYZ and the LTFIM, respectively.
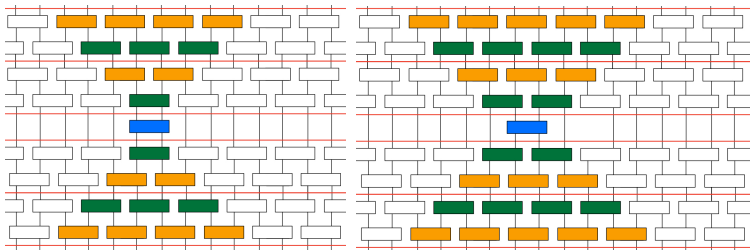
Figure 3.1.: The reverse causal cone spreading of quantum correlations for $C_{1,2}^b$ (left) and $C_{2,3}^b$ (right) for the XYZ ansatz as in Eqs 3.8, 3.9, with a given depth P (here P = 2). The blue blocks represent the observable, while the orange (green) blocks represent the relevant part of $e^{-i\beta_m \widehat{H}_{\text{even}}}$ ($e^{-i\alpha_m \widehat{H}_{\text{odd}}}$). Only a reduced spin chain of length proportional to P is involved in the calculation: the white blocks (acting on spins outside the reduced chain) can be trivially contracted. The gates are arranged according to a *brickwork* [166] architecture.

Indeed, in Sec 3.1.1 above, we derived a simplified expression of the variational energy for the XYZ model and the LTFIM, requiring the evaluation of a small set of two-points and one-point correlators, as in Eqs. (3.16) and (3.18). The calculation of these correlators admits a simple graphical interpretation in terms of a "reverse causal cone" spreading of quantum correlations, which is an immediate consequence of the locality of two-body spin interactions, reminiscent of Lieb-Robinson bounds [167].

Let us first focus on the XYZ case. Given Eq. (3.16), it is sufficient to compute only $C_{1,2}^b = \langle \psi_P | \hat{\sigma}_1^b \hat{\sigma}_2^b | \psi_P \rangle$ and $C_{2,3}^b = \langle \psi_P | \hat{\sigma}_2^b \hat{\sigma}_3^b | \psi_P \rangle$: this can be done by addressing a reduced spin chain of length proportional to P, as sketched in Fig. 3.1. Remarkably, the reduced spin chain is smaller than the whole chain — and it "does not see" the boundary conditions [75] — only for small-enough values of P, corresponding to a low-depth quantum circuit for our ansatz. When this is the case, it can be proven that $C_{1,2}^b$ and $C_{2,3}^b$ do not depend on the system size N, and neither does the rescaled variational energy in Eq. (3.16). More precisely, within this graphical interpretation, it is easy to observe that $C_{1,2}^b$ is independent of N if $4P < N$, while for $C_{2,3}^b$ the condition reads $4P + 2 < N$ (notice the minor differences between the two reverse causal cones in Fig. 3.1). Therefore, once we have fixed the depth P of the ansatz, for large-enough sizes N of the XYZ chain the whole (rescaled) variational energy landscape defined by Eq. (3.16) does not depend on N. Precisely, this holds true for any $N > \tilde{N}_P$, where $\tilde{N}_P = 4P + 2$. This analysis implies that the optimal parameters for given P, found for an XYZ chain of size $N > \tilde{N}_P$, can be *exactly* transferred to any chain of size $N' > N$.

Similarly to the XYZ case, a description of quantum correlations spreading in terms of a light cone also emerges for the LTFIM (analogously to the discussion in [75] for the TFIM). In particular, one can prove that the rescaled variational energy in Eq. (3.18) does not depend on the system size N if $N > \tilde{N}_P$, with $\tilde{N}_P =$
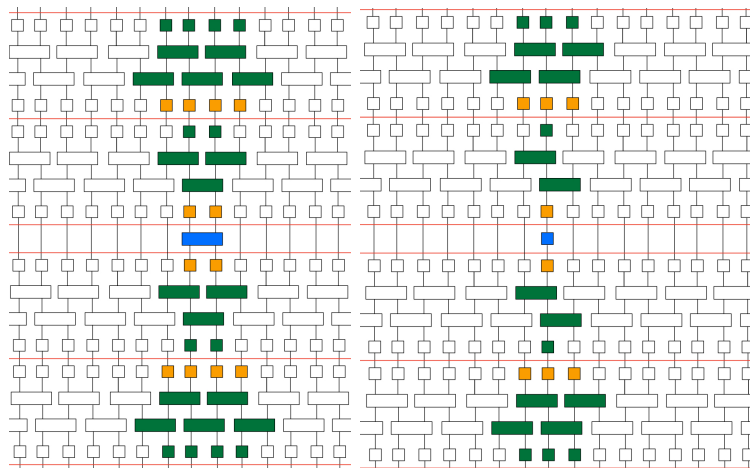
Figure 3.2.: The reverse causal cone for $C_{1,2}^{z,z}$ (left) and C_1^x, C_1^z (right) for the LTFIM ansatz as in Eqs. (3.8), (3.11) for a depth P (here P = 2). The blue blocks represent the observable, while the orange (green) blocks represent the relevant part of $\mathrm{e}^{i\beta_m \widehat{H}_X}$ ($\mathrm{e}^{-i\alpha_m(\widehat{H}_{ZZ}-g_z\widehat{H}_Z)}$). Also in this case, only a reduced spin chain of length proportional to P is involved in the calculation: the white blocks (acting on spins outside the reduced chain) can be trivially contracted.

2P + 1. This follows from the graphical interpretation in Fig. 3.2.

3.3. Results

As mentioned at the beginning of this Chapter, we seek *smooth* and transferable optimal variational parameters. With this in mind, we adopt an iterative interpolation scheme known as INTERP, first introduced in [74]. Essentially, the idea is to perform a sequence of local optimizations for increasing values of P, each starting from an educated guess that is iteratively updated by interpolating on the optimal parameters found at the previous step. A formal description of the procedure, including algorithmic details of our implementation, is reported in Appendix A.

Although INTERP was initially formulated for standard QAOA applied to classical optimization problems, here, we apply this heuristic to more general HVA wavefunctions as in Eq. (3.1), with the goal of quantum many-body ground state preparation. In the first place, we need to provide numerical evidence that both XYZ and LTFIM ground states can be efficiently prepared across their phase diagrams, reaching high fidelity values.
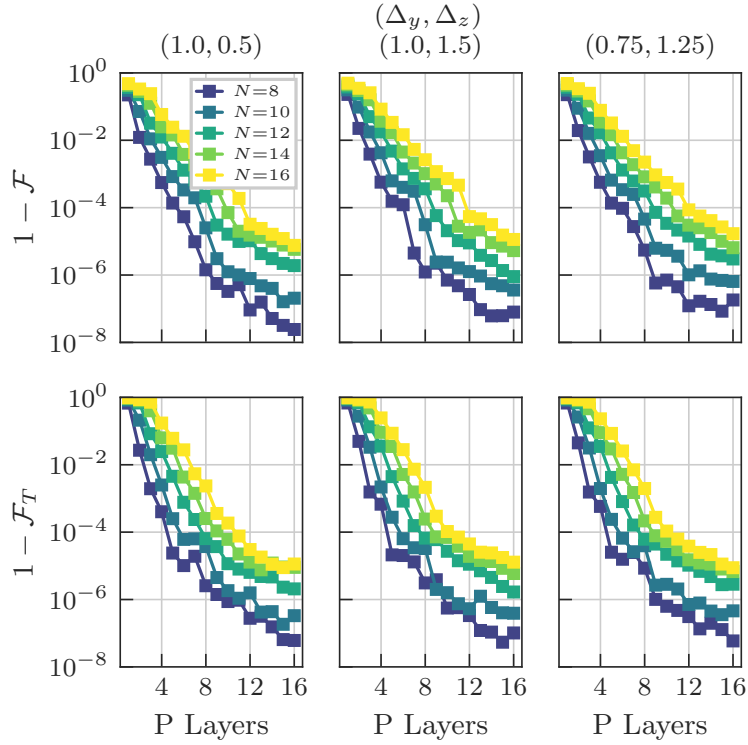
Figure 3.3.: Infidelity $1-\mathcal{F}_P$ (top graphs) and translational infidelity $\mathcal{F}_P^T$ (bottom graphs) for the optimal XYZ ansatz, as a function of increasing number of layers. Even lower (better) values can be obtained for the Heisenberg model ($\Delta_Y = \Delta_Z = 1$, data not shown). The translational infidelity follows a pattern similar to the infidelity, proving that full translational symmetry is correctly restored. The optimal parameters always lie on a smooth curve, as shown in Fig. 3.6.

3.3.1. Ground state preparation

We test the effectiveness of INTERP in providing excellent approximate ground states for our models by applying it to the ansatz wavefunctions defined, for the XYZ, by Eqs. (3.8), (3.9), and, for the LTFIM, by Eqs. (3.8), (3.11). We quantify the accuracy of our ground state approximation — namely the ansatz state evaluated at optimal parameters $(\boldsymbol{\beta}^\star, \boldsymbol{\alpha}^\star)_P$ — with the ground state fidelity

$$\mathcal{F}_P = \left| \langle \psi_{\mathrm{gs}} | \psi(\boldsymbol{\beta}^\star, \boldsymbol{\alpha}^\star)_P \rangle \right|^2 . \tag{3.19}$$

Another useful quantity for the XYZ model is the *translational fidelity* of the optimal state with its one-site translated version:

$$\mathcal{F}_P^T = \left| \langle \psi(\boldsymbol{\beta}^\star, \boldsymbol{\alpha}^\star)_P | \widehat{\mathrm{T}} | \psi(\boldsymbol{\beta}^\star, \boldsymbol{\alpha}^\star)_P \rangle \right|^2 . \tag{3.20}$$

Indeed, as stated in Sec 3.1.1, the ansatz for the XYZ model does not encode one-site translational symmetry. Nevertheless, this translational fidelity is expected to converge to one when approximating the true ground state with high fidelity, thus restoring the full translational symmetry.
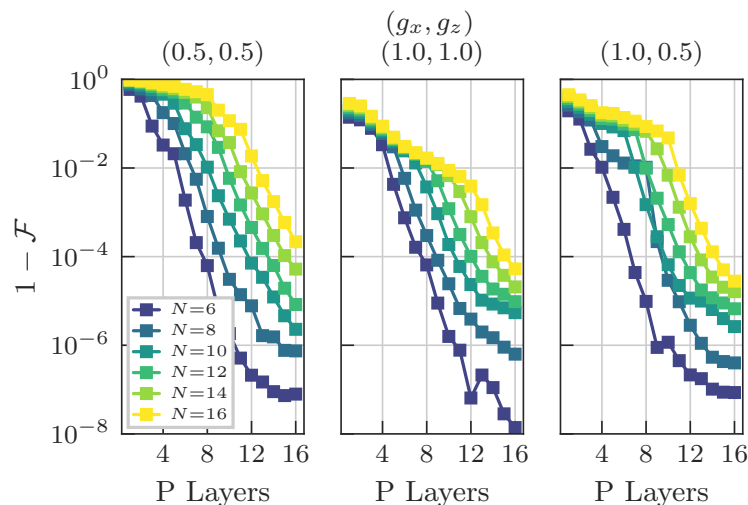
Figure 3.4.: Infidelity for the optimal LTFIM ansatz, as a function of increasing number of layers. As for the XYZ models, we observed empirically that similar infidelities can be obtained for other points of the phase diagram (different values of g_x and g_z). Once again, the optimal parameters lie on a smooth curve, as shown in Fig. 3.6.

Some illustrative results for the ground state fidelity are reported in Fig 3.3 for XYZ models (also displaying data for the translational fidelity) and in Fig 3.4 for the LTFIM. Note that, for ease of visualization, we plot infidelity (translational infidelity) values, defined as $1 - \mathcal{F}_P$ $(1 - \mathcal{F}_P^T)$. Remarkably, the INTERP method avoids low-quality local minima of the energy landscape [47], allowing the obtaining of high-fidelity values. Moreover, as shown below in Fig. 3.6, it converges to *smooth* optimal curves for the variational parameters.

The mechanism underlying the effectiveness of INTERP can be understood intuitively as follows. Let us adopt the usual definition of residual energy

$$\varepsilon_N(\boldsymbol{\beta}, \boldsymbol{\alpha})_P = \frac{E_N(\boldsymbol{\beta}, \boldsymbol{\alpha})_P - E_N^{\min}}{E_N^{\max} - E_N^{\min}} \ , \tag{3.21}$$

where $E_N(\boldsymbol{\beta}, \boldsymbol{\alpha})_P$ is the variational energy in Eq. (3.10), while $E_N^{\min}$ ($E_N^{\max}$) is the ground-state (maximum) energy of the spectrum. By looking at Fig. 3.5, it is evident how this iterative scheme provides an effective warm start at each iteration of the algorithm. Here, the residual energy is plotted for increasing values of P, both evaluated at $(\boldsymbol{\beta}, \boldsymbol{\alpha})_P^{\text{start}}$ (before the local optimization) and at $(\boldsymbol{\beta}, \boldsymbol{\alpha})_P^{\text{opt}}$ (after it).

We remark that the same ansatz (defined by Eqs. (3.8), (3.9)) successfully prepares the ground state for the Heisenberg model ($\Delta_Y = \Delta_Z = 1$), the XXZ in both phases ($\Delta_Z < 1$ and $\Delta_Z > 1$, with $\Delta_Y = 1$), and the XYZ model. Here, we show data for arbitrary values of Δ_Y, Δ_Z, but we verified that our results extend to different points in the phase diagram. The same comments apply for the LTFIM ansatz (defined by Eqs. (3.8), (3.11)), which is effective for the whole phase diagram. High fidelity values are obtained despite the finite number of iterations ($N_{\text{iter}} = 100$) set for the classical optimizer.
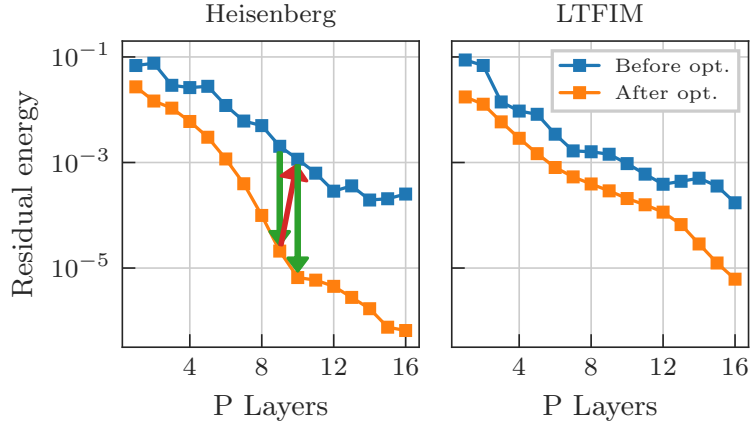
Figure 3.5.: Residual energy vs P before and after each local optimization leading from $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}^{\mathrm{start}}$ to $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}^{\mathrm{opt}}$. Data refer to $N = 16$ qubits, for both the $(\Delta_{\mathrm{Y}} = 1, \Delta_{\mathrm{Z}} = 1)$ Heisenberg model and the $(g_x = 1, g_z = 1)$ LTFIM. Remarkably, the residual energy before each optimization is already quite low, especially for large values of P. The vertical green arrows show the improvement of the local optimization, while the red arrow depicts the interpolation step from $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}^{\mathrm{opt}}$ to $(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P+1}}^{\mathrm{start}}$.

3.3.2. Smooth solutions and transferability

Usually, by adopting such iterative methods, one finds optimal angles that are smooth functions of the layer index $m = 1 \cdots \mathrm{P}$. For this reason, we dub them *smooth* solutions. This is consistently observed in all phases of our models, as shown in Fig. 3.6 at the critical point of XXZ (Heisenberg model) and close to the critical line of the LTFIM [158]. On top of that, we note that these smooth optimal curves are qualitatively similar for different system sizes. Inspired by this observation, we verify numerically that smooth optimal solutions $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$ — obtained by applying INTERP to a small-size system with dimension N_G up to a certain value of P — can be transferred to solve the same task for a larger number of qubits, thus providing an effective educated guess. In the following, we will always indicate with N_G the "guess" size used to obtain the smooth optimal solution, which will be eventually transferred to a larger system with $N > N_G$ lattice sites. Unless otherwise stated, we set $N_G = 8$. To estimate the effectiveness of our transferability protocol, we employ the residual energy for a system with N qubits, defined in Eq. 3.21, now evaluated at $\varepsilon_N(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$. In Fig. 3.7, we plot this quantity for different points in the phase diagram of our models: strikingly, smooth optimal curves obtained for a small system provide an excellent educated guess for the ground-state preparation up to $N = 24$ lattice sites.

A few comments are in order. The residual energy is not always a good proxy for the fidelity with the ground state. However, in our case, the good agreement between these two quantities is verified in Sec 3.3.3, where we compute the actual fidelity of transferred solutions with target ground states. In practice, the residual energy is a convenient choice for a figure of merit since it can be directly evaluated
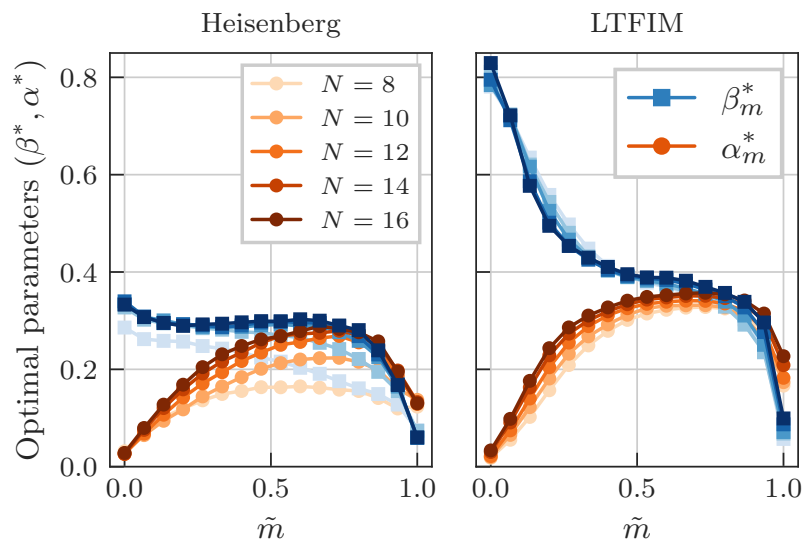
Figure 3.6.: Smooth optimal parameters $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N} = (\beta_1^* \cdots \beta_\mathrm{P}^*, \alpha_1^* \cdots \alpha_\mathrm{P}^*)$ obtained with INTERP, plotted vs the rescaled index $\tilde{m} \equiv (m-1)/(\mathrm{P}-1)$ in the x-axis range $[0, 1]$. Results are shown for the Heisenberg model ($\Delta_\mathrm{Y} = 1$, $\Delta_\mathrm{Z} = 1$) (left) and the LTFIM ($g_x = 1$, $g_z = 1$) (right) for $\mathrm{P} = 16$, and they are qualitatively similar for different sizes N. These solutions are stable by further increasing the number of layers P. Similar smooth solutions can be found for different points of the phase diagram.

on a quantum computer[1], in contrast to the ground state fidelity. Intuitively, the residual energy may roughly evaluate at ≈ 0.5 when computed at a random point in the energy landscape [2], while its values obtained via transferability are remarkably lower.

As a second crucial remark, the transferability of this class of smooth solutions found via INTERP holds for larger values of P; *in contrast*, other equal-quality *non-smooth* solutions for the small N_G-size system — obtained using random-start local optimization — do not provide any helpful guess for the ground state preparation of the same model with a larger number of qubits. The analysis of these results is also deferred to Sec 3.3.3. Additionally, we tested the existence of smooth curves and their transferability to a larger number of qubits, also for the TFIM: our results are confirmed up to much larger sizes by leveraging a standard mapping to free fermions [75, 160, 161], as reported in Sec 3.3.4.

Now, despite the good educated guess provided by the transferability of smooth solutions, one may be tempted to refine the ground state approximation for the N-size model, e.g. by aiming at a target value of fidelity such as 99.9%. However, for such large sizes, both the XYZ models and the LTFIM are affected by barren plateaus [2, 55]. Therefore, any local optimization starting from a random

---

[1] If the values of $E_N^{\max}$ and $E_N^{\min}$ are unknown, one can use the variational energy.

[2] This is approximately verified if the density of energy levels of $\widehat{H}_{\mathrm{targ}}$ is almost uniform in the spectrum.
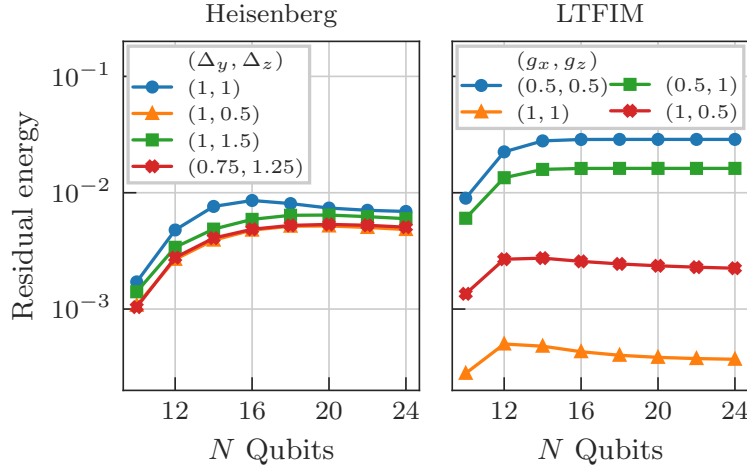
Figure 3.7.: Residual energies $\varepsilon_N(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$ for different system sizes using parameters from a small-size "guess-system" ($N_G = 8$) computed for different flavours of the two models. We show the transferability of smooth optimal solutions $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$ with $\mathrm{P} = 10$ for XYZ models (left) and the LTFIM (right).

point in the parameter space is doomed to fail on a realistic quantum device, due to vanishingly small gradients requiring an exponential scaling of resources [34, 49]. Remarkably — and this is one of the main original results of this Chapter — we find that transferred smooth optimal solutions stand out in this respect: in their neighborhood, the landscape does not suffer from small gradients, and a local optimization would succeed.

Figures 3.8 and 3.9 illustrate this important point. For conciseness, we show data for ($\Delta_{\mathrm{Y}} = 1$, $\Delta_{\mathrm{Z}} = 1$) and ($g_x = 1$, $g_z = 1$), but our results extend to other points of the phase diagrams. Specifically, in Fig. 3.8, we plot the variance of a representative gradient component of the variational energy in Eq. (3.10), as customary in studies on barren plateaus [49, 54, 56], which is sampled at random in the whole landscape. As expected, its exponential decay with the system size N confirms the presence of barren plateaus. However, if we sample the same gradient component only in a neighborhood of radius ϵ of the transferred smooth solution, its magnitude does not show any appreciable exponential decay. This result is observed for both classes of models under exam, and it is further evidenced in Fig. 3.9, showing data for a fixed value of P: the exponential decrease of the gradients in the whole search space is equivalent to that in a neighborhood of radius ϵ of any given set of angles, with the exception of the smooth transferred curve $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$. Once more, also this local landscape property does not extend to the neighborhood of other transferred non-smooth solutions, which neither provide a useful educated guess for the large system nor solve the barren plateau issue for local optimization. This is shown in Sec 3.3.3, along with data supporting the effectiveness of a refinement optimization performed classically in the neighborhood of the transferred smooth curve.

Incidentally, for each value of N, the sample variance in the whole search space
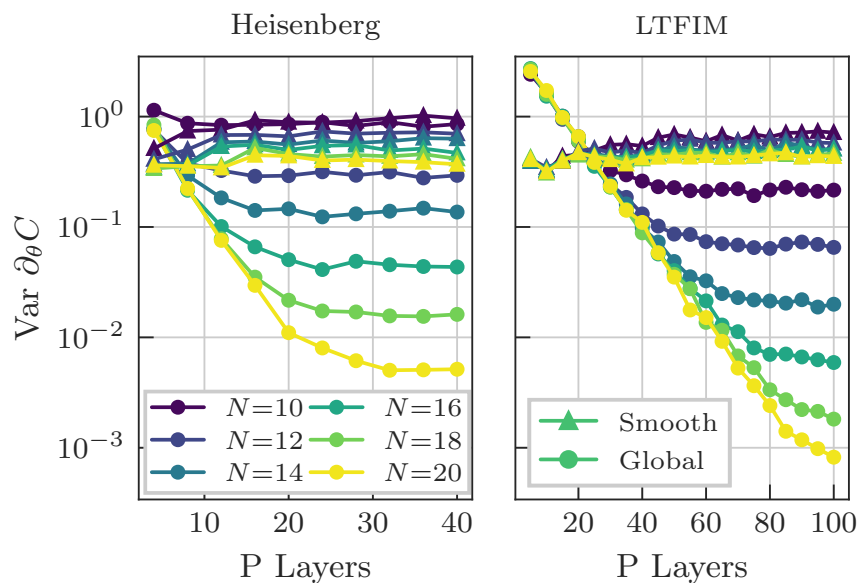
Figure 3.8.: Barren plateaus in the whole search space (data denoted as "Global" and shown as solid circles of different colors for different N, in both plots). A qualitatively different trend is observed in the ϵ-neighborhood of the transferred smooth solution $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P},N_G}$, obtained with INTERP for a small system size $N_G = 8$ (data denoted as "Smooth", and shown as solid triangles, with colors matching those of "Global" data for various N, in both plots). Here, we focus on a single partial derivative w.r.t. $\theta = \alpha_1$ (see Eqs. (3.9), (3.11)) of the variational energy in Eq. (3.10), rescaled as in Sec 3.1.1 (see Eqs. (3.16), (3.18)) and dubbed C. We plot the sample variance of the partial derivative as a function of the number P of HVA layers in the circuit. We fix $\epsilon = 0.05$ and a batch of 1000 samples for each value of P and N.
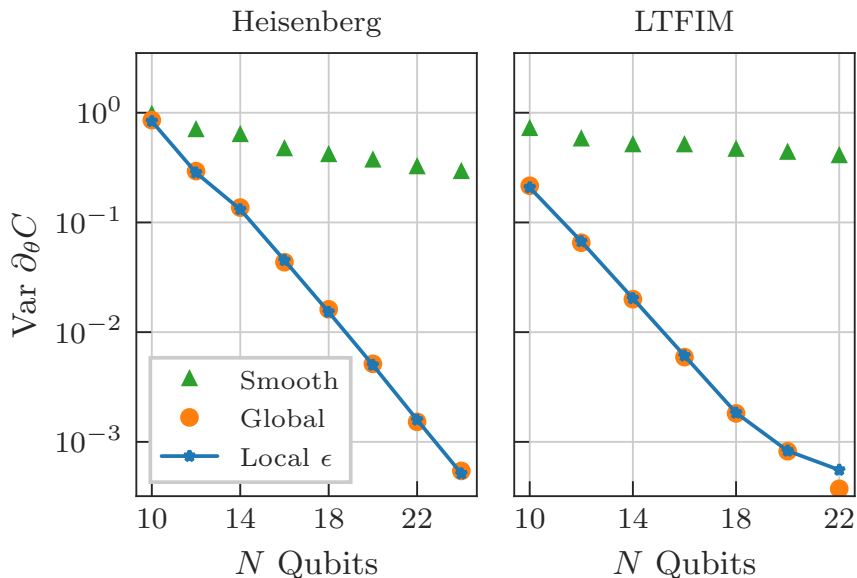
Figure 3.9.: Same quantity as in Fig. 3.8, here plotted vs the qubit number for a fixed circuit depth P = 40 (P = 100) for the Heisenberg model (LTFIM). The labels "Global" and "Smooth" refer to the same data of Fig. 3.8. We denote with "Local ϵ" a sampling performed in an ϵ-region centered around a random point. Data are averages over 20 random points (generated independently for each value of N) with $\epsilon = 0.05$. We use a constant batch size of 1000 samples, for each N and sampling region. A clear trend appears: the neighborhood of a random point exhibits the same exponential decay as the whole space, which is not present in the neighborhood of the transferred smooth solution.

saturates after a certain circuit depth P, as argued in [34, 49] and clearly shown in Fig. 3.8. This fact is usually linked to the ansatz parameterized quantum circuit approaching an approximate 2-design [52, 168, 169] on its symmetry subspace [55].

In conclusion of this section, let us highlight that our findings pave the way to an improved scheme to prepare the ground state of this class of many-body quantum systems with a large number of qubits: the smooth optimal curves can be found *classically* for a small system and then transferred to solve the same task for larger N, beyond the reach of classical simulations. In this scenario, the *quantum* device would only be needed for a refinement optimization without barren plateaus.

3.3.3. Additional results on transferability and barren plateaus

In the previous section, we described the transferability properties of a class of smooth solutions found via INTERP. However, for a small system (e.g. "guess size" $N_G = 8$) and a large value of P, one can easily find other solutions employing standard random-start local optimization, not displaying any smoothness property as a function of the layer index $m = 1 \cdots P$. Remarkably, these do not offer, in general, any useful educated guess for the ground state preparation of a larger system, as
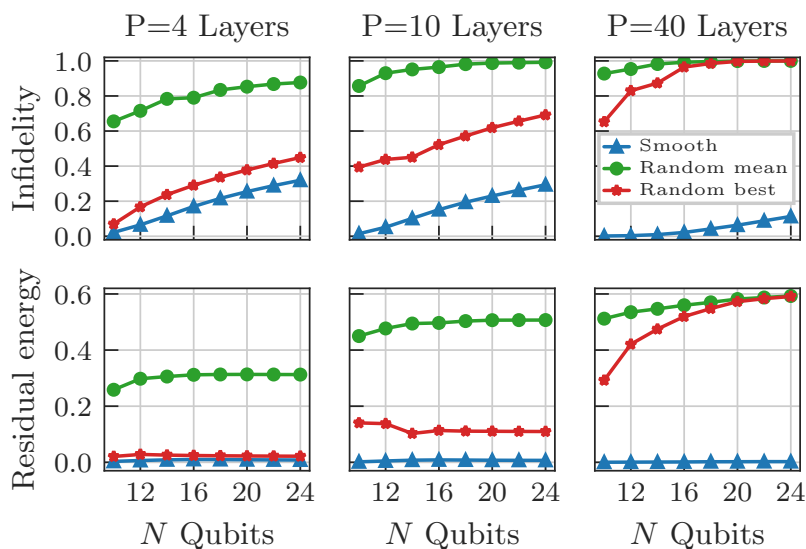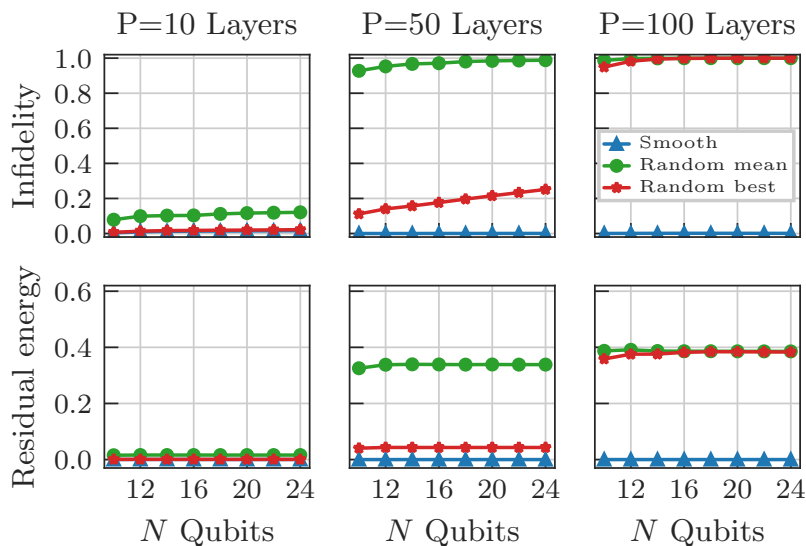
Figure 3.10.: Transferability of smooth optimal solutions (INTERP) vs random-start solutions (local optimization), both obtained for a small-size "guess" system ($N_G = 8$) and then transferred to a larger system sizes N. Data refer to the ($\Delta_Y = 1$, $\Delta_z = 1$) Heisenberg model; we plot the average and the best out of 20 random-start local optimizations, compared to the usual INTERP smooth curve $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P}, N_G}$. INTERP solutions are always observed to provide an excellent educated guess for larger systems: this is particularly evident for large values of P, where random-start solutions always fail in this respect.

Figure 3.11.: Transferability of smooth optimal solutions (INTERP) vs random-start solutions (local optimization), both obtained for a small-size "guess" system ($N_G = 8$) and then transferred to a larger system sizes N. Data refer to the ($g_x = 1$, $g_z = 1$) LTFIM and we plot the average and the best out of 20 random-start local optimizations. The same comments apply as in Fig. 3.10.

shown in Figs. 3.10 and 3.11 for the Heisenberg model and the LTFIM, both in terms of ground state infidelity and residual energy. Notice that both INTERP and any random-start solution prepare the ground state for the small system equally well
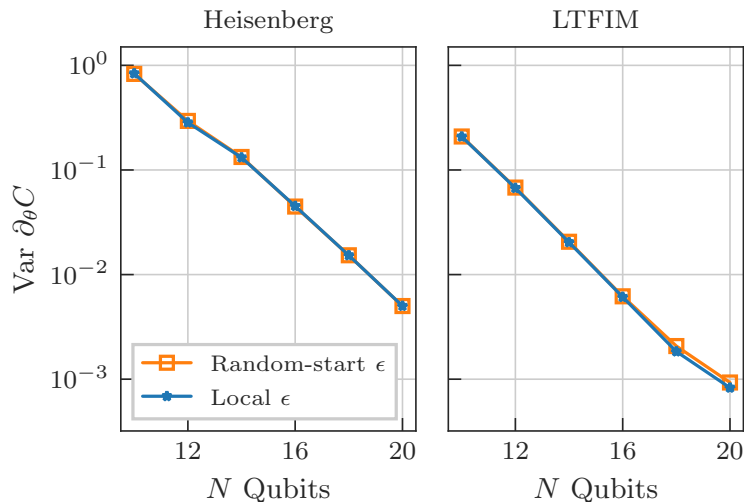
Figure 3.12.: Comparison between the local landscape of transferred random-start (non-smooth) solutions with that of random points in the search space (dubbed "Local ϵ", same data as in Fig. 3.9). In both cases, we average over 20 instances and, for each instance, we sample 1000 points in a neighborhood of radius $\epsilon = 0.05$. The exponential decay is manifest, with a striking similarity. This is at variance with the local landscape of transferred smooth solutions found via INTERP, not displaying barren plateaus (compare with Fig. 3.9).

(essentially, with zero infidelity). Nevertheless, only the former class of solutions is observed always to yield an excellent educated guess for the same task with a larger number of qubits. This fact is particularly evident for large values of P. Indeed, in this regime, transferred random-start solutions perform as poorly as the ansatz evaluated at random in an arbitrary point of the search space (i.e., almost-unit infidelity, residual energy of the order of ≈ 0.5). Moreover, smooth solutions found via INTERP also stand out concerning their favorable local landscape, where gradients do not show any appreciable exponential decay. In contrast, this is not observed in the neighborhood of transferred non-smooth solutions, where barren plateaus are as marked as in the neighborhood of a random point in the search space (or as in the global search space, see Fig. 3.9). These results are outlined in Fig 3.12. Here, and in the rest of the section, we focus again on the ($\Delta_Y = 1$, $\Delta_z = 1$) Heisenberg model and the ($g_x = 1$, $g_z = 1$) LTFIM.

Thanks to this favorable local landscape, one can effectively perform a refinement optimization for the large system to further increase the ground state fidelity above a target threshold. An example is shown in Fig. 3.13, by setting a maximum number of iterations $N_{iter} = 100$ for the classical routine performing the refinement optimization. Despite this constraint, the local optimization succeeds, significantly lowering the final infidelity. Incidentally, we verified that the refined optimal curve is still in the "basin of attraction" of the transferred smooth curve, as shown in Fig. 3.14. At this level of detail, the curves appear almost exactly overlapping.
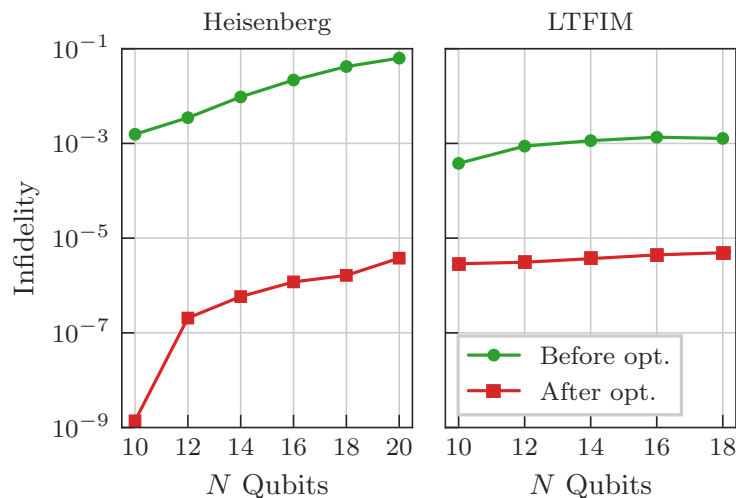
Figure 3.13.: Refinement optimization starting from the transferred smooth curve, for the Heisenberg model with P = 40 (left panel) and the LTFIM with P = 100 (right panel). Even with a fixed number of iterations $N_{\text{iter}} = 100$ for the classical optimizer, we can significantly reduce the ground state infidelity.

3.3.4. Large-scale simulations for the TFIM

In this section, we address the line ($g_z = 0$) of the LTFIM phase diagram, corresponding to the integrable Transverse Field Ising Model (TFIM), with Hamiltonian

$$\widehat{H}_{\text{TFIM}} = \sum_{i=1}^{N} \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - g_x \sum_{i=1}^{N} \hat{\sigma}_i^x \ . \tag{3.22}$$

A mapping to non-interacting fermions allows us to perform large-N VQA simulations, well beyond the usual limits of exact diagonalization techniques. In particular, we used the same ansatz as in Eqs. (3.8), (3.11), which reduces to the standard QAOA ansatz for $g_z = 0$.

As argued in [55], we verified numerically that the TFIM ground state preparation is not affected by barren plateaus. However, the possibility of large-scale simulations offers a useful benchmark on the effectiveness of INTERP in this regime, particularly concerning the existence of smooth optimal solutions and their transferability. The variational energy is given by Eq. (3.10) with $\widehat{H}_{\text{targ}} = \widehat{H}_{\text{TFIM}}$, while the residual energy reads as in Eq. (3.21).

Refs. [27, 75, 162] previously discussed how to efficiently simulate QAOA using the fermionic mapping of the TFIM. They showed that, after applying a Jordan-Wigner and a Bogoliubov transformation, a system of an even number of spins N decomposes in a direct sum of $N/2$ independent two-level systems, which are labeled by the wave-vectors $k_n = (2n-1)\pi/N$ with $n = 1 \cdots N/2$. The total residual energy
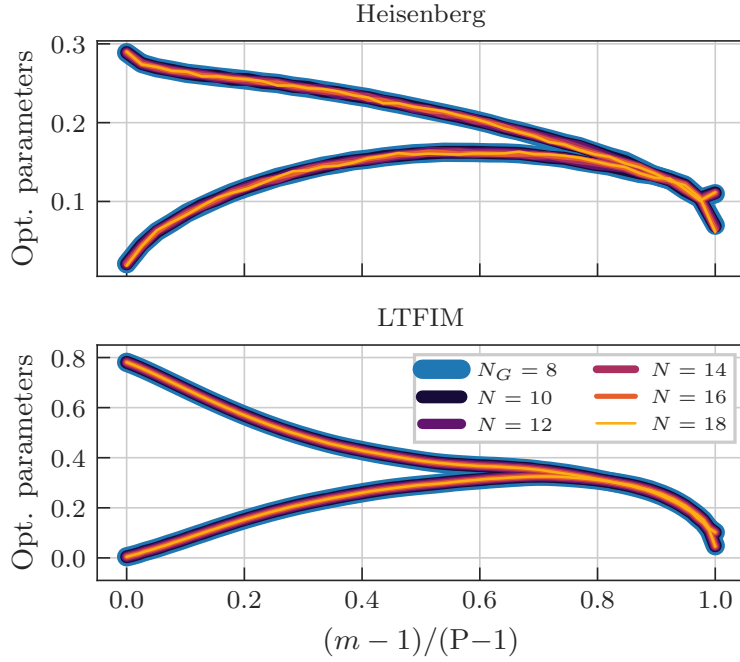
Figure 3.14.: A coarse-grained picture of the optimal parameters *after* performing a refinement optimization, compared with the initial smooth guess (obtained by INTERP) at $N_G = 8$, i.e. $(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*)|_{\mathrm{P}, N_G}$. These are the refined optimal parameters corresponding to Fig. 3.13, i.e. P = 40 (P = 100) for the Heisenberg model (LTFIM).

of the one-dimensional TFIM then reads

$$\varepsilon_N(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} = \sum_{n=1}^{N/2} \varepsilon^{(k_n)}(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} \ , \tag{3.23}$$

where $\varepsilon^{(k_n)}(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}$ are the residual energies associated to each two-level system. The analytical expression for $\epsilon^{(k_n)}(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}$, as provided in Ref. [75], is

$$\epsilon^{(k)}(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}} = \frac{1}{2} - \frac{1}{2} \boldsymbol{v}_k^T \left(\prod_{m=1}^{\leftarrow \mathrm{P}} R_{\hat{\boldsymbol{z}}}(4\beta_m) R_{\boldsymbol{b}_k}(4\alpha_m) \right) \hat{\boldsymbol{z}} \ ,$$

where $\hat{\boldsymbol{z}} = (0, 0, 1)^T$, $\boldsymbol{b}_k = (-\sin k, 0, \cos k)^T$ and $\boldsymbol{v}_k = (\boldsymbol{b}_k + g_x \hat{\boldsymbol{z}})/||\boldsymbol{b}_k + g_x \hat{\boldsymbol{z}}||$ are three-dimensional unit vectors. $R_{\hat{\boldsymbol{\omega}}}(\theta)$ is the 3×3 matrix associated with a rotation of an angle θ around the unit vector $\hat{\boldsymbol{\omega}}$, and their product is "time"-ordered from right to left for increasing $m = 1 \cdots \mathrm{P}$.

These formulas allow for efficient computation of $\varepsilon_N(\boldsymbol{\beta}, \boldsymbol{\alpha})_{\mathrm{P}}$, enabling us to numerically study the performance of QAOA for a large number of qubits. Also for the TFIM, we find smooth curves, shown in Fig. 3.15 at the critical point $g_x = 1$, up to sizes as large as $N = 128$. In the same figure, we also show that smooth curves — prepared by applying INTERP to a small system of size $N_G = 8$ — are transferable up to sizes as large as 250 qubits, i.e. they offer a good educated guess for TFIM ground state preparation. This is in stark contrast with other (non-smooth)
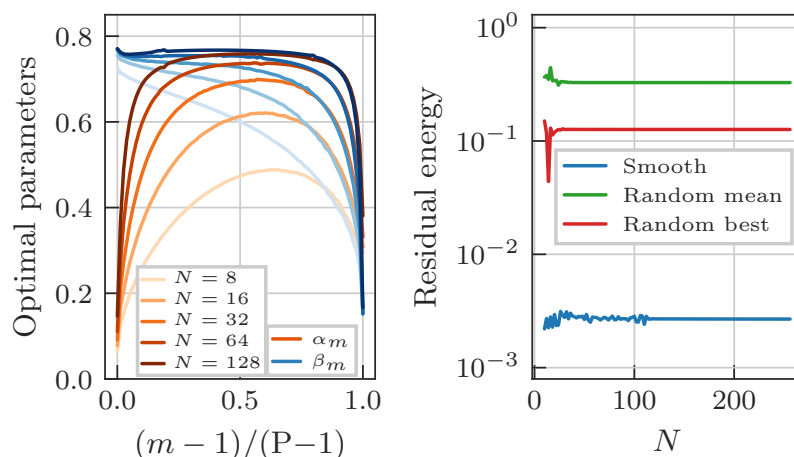
Figure 3.15.: (Left panel) Smooth optimal parameter curves for the TFIM, obtained by applying INTERP up to P = 80. Smooth curves are qualitatively similar for a wide range of system sizes. (Right panel) Numerical evidence on the transferability of smooth optimal solutions — obtained for a small-size "guess" system ($N_G = 8$) — up to very large systems. In contrast, non-smooth solutions found via random-start local optimization do not provide, on average, any useful guess for the ground state preparation of a larger system, yielding a residual energy ≈ 0.5. We averaged over 100 random-start solutions but, remarkably, not even the best of them is nearly comparable with the one obtained via INTERP. Data refer to the critical point $g_x = 1$, and the same behavior is observed in other regions of the TFIM phase diagram.

solutions, found via random-start local optimization for the small system: despite preparing the small-size ground state with perfect accuracy (the same as applying INTERP), they do not provide any useful educated guess for the large system.

3.4. Conclusion and outlook

We tackled many-body ground state preparation via problem-inspired VQAs, and provided extensive numerical evidence on the transferability of a class of optimal *smooth* solutions — obtained using iterative schemes for a small number of qubits — to solve the same task for larger system sizes. Remarkably, these solutions provide an excellent educated guess for the ground-state wavefunction, as opposed to other solutions that can be easily obtained for small systems without appropriate iterative schemes (e.g. with random-start local optimization). These results are confirmed up to larger sizes for the TFIM.

For ground state preparation of these models, our procedure overcomes the well-known difficulties related to the highly non-trivial structure of the variational energy landscape. On top of avoiding low-quality local minima traps daunting random-start local optimization [47, 74], we provided evidence of a novel feature of this class of solutions: the cost-function landscape is observed to become free of barren

plateaus in their neighborhood, potentially allowing for further effective refinement optimizations with a quantum device on a classically obtained smooth guess.

This work paves the way for many novel, exciting research directions. Our new effective way of approaching ground state preparation for larger many-body systems may allow to deal with 2D lattice models, ranging from spin systems to Hubbard-like systems, with or without disorder. Furthermore, an interesting proposal would be to test our scheme in the presence of noise and for the mitigation of noise-induced barren plateaus [59]. Theoretically, it might be interesting to prove analytically the transferability and landscape properties of smooth solutions found via INTERP. Previous numerical and analytical results on reusable optimal parameters exist, as explained in detail in Chapter 1. These "parameter concentration" results are usually valid for the whole landscape, while here, they apply specifically to smooth optimal solutions for large values of P, providing a link between solution transferability and the local absence of barren plateaus. Hence, a possible connection between this class of solutions and adiabaticity might be investigated [75, 112, 127]. Finally, our scheme could be directly tested with near-term technology on real quantum devices beyond the size limits of classical computation.

4. Ground state preparation of a Lattice Gauge Theory

In topology, you don't understand things. You just get used to them.

—James Waddell Alexander II

In this Chapter, we address the realization and characterization on near-term quantum devices of the ground states of a special class of interacting many-body systems, namely lattice gauge theories. We focus on a two-dimensional $\mathbb{Z}_2$ lattice gauge theory model on a shallow quantum circuit, a paradigmatic problem of interest for its implications in high-energy physics and the realization of topological quantum memories. Our implementation involves single and two-qubit gates compatible with present-day technologies, employing a Variational Quantum Algorithm in the class of Quantum Approximate Optimization Algorithms (QAOA).

Our results are based on systems with 18 qubits, and, despite their small size, enable us to detect the presence of a phase transition between confined and deconfined regimes, which can be detected by measuring expectation values of Wilson loop operators or the topological entropy. Moreover, if periodic boundary conditions are implemented, the same optimal solution is transferable among all four different topological sectors, without any need for further optimization of the variational parameters.

These techniques can be generalized to more complex lattice gauge theories or models with topological order, and they constitute a useful instrument to be added to the toolbox for the digital quantum simulation of gauge theories, especially for quantum computers based on arrays of superconducting qubits.

4.1. Introduction

When considering current quantum computers based on superconducting qubits, the typical platforms are constituted by qubits arranged in two-dimensional arrays [170, 171]. These systems allow us to manipulate the qubits via single and two-qubit gates and, in most cases, two-qubit gates are local, i.e. they can be ap-

plied to neighboring qubits only. These platforms are usually initialized starting from simple product states, while the efficient creation of complex entangled states is a non-trivial challenge. Despite the limitations of the noisy hardware currently available, these architectures open the path to various applications. One of the most important is the implementation of error-correcting codes and, to this purpose, considerable attention has been devoted to the realization of surface codes [172], thus to the preparation via local gates of states displaying topological order. In this context, the first experimental realization of the ground state of a surface code has been successfully achieved [173], and it allowed for the study of the main topological properties of its anyonic excitations.

In this framework, Lattice Gauge Theories (LGT) emerged as a paradigmatic research subject. They constitute the backbone of particle physics, and many of their important features display a non-perturbative nature, requiring advanced numerical techniques to be studied. Furthermore, some of the simplest two-dimensional (2D) LGTs share the same topological properties of surface codes which, indeed, can be seen as the extreme deconfined limit of systems with $\mathbb{Z}_2$ gauge symmetry. In the last decade, the application of quantum technologies to LGT became a lively field of research [174–179], progressing both on the development of several technologies and algorithms to tackle the complexity of LGTs and on the study of LGTs themselves.

One task that can be addressed through the quantum simulation of LGTs is the study of their static properties. A key step to this purpose is the efficient initialization of their ground states, which allows the investigation of their phase diagrams at low temperatures.

In the next sections, we explore the possibility of studying the ground state of a 2D pure lattice gauge theory within the framework of quantum circuits and digital quantum simulations. Indeed, the recent developments of quantum simulations provide complementary approaches to other quantum many-body approximation techniques such as Tensor Networks [176, 180–184], which are challenging to implement in 2D with current technologies, in particular for what regards quantum dynamics. We shall focus on the 2D $\mathbb{Z}_2$ LGT, which is known to display a confinement-deconfinement phase transition between a trivial (confined) phase and a topologically ordered (deconfined) phase, matching the topological features of Kitaev's toric code [185]. As we show, most of the interesting ground state properties linked to topological order, which are usually described in the thermodynamic limit, can be characterized even with small lattices. To do so, we employ QAOA to prepare the ground state at arbitrary values of the coupling and we show that the algorithm reaches high fidelity within a very small number of variational parameters, corresponding to a shallow quantum circuit. To reliably find optimal or quasi-optimal minima, we employ a two-step local optimization procedure similar to the one in Chapter 2, which provides regular schedules that can be efficiently transferred to larger systems.

Targeting the ground state in the confined phase, where there is no long-range entanglement, can always be performed efficiently and our numerical simulations suggest that QAOA can be scaled up to larger sizes without increasing the circuit depth. Concerning the preparation of states in the deconfined phase, instead, it is known that topologically ordered states cannot be obtained exactly with circuits of fixed depth for growing system size. In particular, for the ground states of the toric code, the required circuit depth scales linearly with the system width [173, 186–189]. This is also a general property of QAOA, where long-range correlation and perfect control on the system is attained only with an extensive number of layers [75, 190, 191]. As a consequence, when targeting states in different phases, we compare two strategies: either we apply directly the QAOA evolution on a trivial product state, or we first build exactly the toric code state and then apply the variational circuit from that starting point. The two approaches offer optimal results for targeting states in the confined or deconfined phase, respectively. They also display similar performances for the small system sizes we consider, except for the required overhead of the second approach.

The ground states prepared with QAOA are then used to characterize the crossover from the confined-topologically trivial phase to the deconfined-topologically ordered one. In particular, we focus on the behavior of Wilson loop operators and the topological entropy. We also discuss the possibility of exploring the ground state degeneracy when the lattice has periodic boundary conditions. Remarkably, all these indicators of a topological phase transition display very small deviations from their expected behavior in the thermodynamic limit despite the reduced size of our lattices. The successful implementation of 2D $\mathbb{Z}_2$ LGT and the correct description of this crossover — with very limited resources in terms of qubit numbers and circuit depth — provides a proof of principle of the feasibility of quantum simulations of deconfined and topological phases of lattice gauge theories in general.

4.2. $\mathbb{Z}_2$ lattice gauge theory

In this work, we consider a pure $\mathbb{Z}_2$ gauge theory model on a regular square lattice. The discretized gauge fields are represented by qubits on the links of the lattice. Using a lattice of size $L \times L$, there are $2L^2$ qubits if periodic boundary conditions are imposed. The Hamiltonian we use is the sum of two competing terms

$$\widehat{H} = \widehat{H}_E + h\,\widehat{H}_B \;, \tag{4.1}$$

which represent "electric" and "magnetic" non-commuting contributions. Their structure comes from an analogy with the QED Hamiltonian, where both space and the gauge group $U(1)$ are discretized: the real space becomes a lattice, and $U(1)$ is discretized to $\mathbb{Z}_n$. Here we focus on the smallest discrete group $n = 2$, which is

naturally encoded in terms of qubits. The electric contribution to the Hamiltonian is

$$\widehat{H}_E = \sum_l (1 - \hat{\sigma}_l^x) \, , \qquad (4.2)$$

where the index l runs over all the links in the lattice and the Pauli matrices are denoted $\hat{\sigma}_l^\alpha$, with $\alpha = x, \, y, \, z$. This specific choice of $\widehat{H}_E$ is motivated by the QED analogy since the electric field enters the Hamiltonian via $\vec{E}^2$ and our term is positive definite. A spin in the eigenstate $\hat{\sigma}_l^x |+\rangle_l = |+\rangle_l$ brings no contribution to the electric energy and corresponds to a vanishing electric field. The state $|-\rangle_l$ indicates instead the presence of a $\mathbb{Z}_2$ electric excitation on the link l, with energy cost assigned by $\widehat{H}_E$. The magnetic term reads

$$\widehat{H}_B = -\sum_p \mathcal{B}_p = -\sum_p \hat{\sigma}_{p1}^z \hat{\sigma}_{p2}^z \hat{\sigma}_{p3}^z \hat{\sigma}_{p4}^z \, , \qquad (4.3)$$

where p labels the *plaquettes* of the lattice and the *plaquette operator* $\mathcal{B}_p$ involves the product of the four spin variables $\hat{\sigma}^z$ around the four links $p_1, \cdots p_4$ of the p-th plaquette (see Fig. 4.1). In particular, $\mathcal{B}_p = -1$ represents a magnetic flux through the p-th plaquette, and the interaction term $h\widehat{H}_B$ assigns an energy $2h$ to each of these excitations. Note that $\widehat{H}_B$ is diagonal in the computational basis, whereas the electric Hamiltonian $\widehat{H}_E$ effectively provides kinetic energy to the magnetic fluxes.

The local gauge constraint is the analog of Gauss's law and it selects the physically relevant sector of the Hilbert space. For each vertex v of the lattice, physical states must be left invariant by gauge transformations, thus satisfying

$$\mathcal{A}_v |\psi\rangle_{\text{phys}} = \prod_{l \in v} \hat{\sigma}_l^x |\psi\rangle_{\text{phys}} = |\psi\rangle_{\text{phys}} \, , \qquad (4.4)$$

where the *star operator* $\mathcal{A}_v$ is the product of the spin operators $\hat{\sigma}^x$ on the four links connected to the vertex v, as represented in Fig. 4.1. It is straightforward to check that any plaquette operator $\mathcal{B}_p$ commutes with any star operator $\mathcal{A}_v$, i.e.

$$[\mathcal{B}_p, \mathcal{A}_v] = 0 \qquad \forall \, p, v \, , \qquad (4.5)$$

hence the star operators are local symmetries of the model since they commute with the Hamiltonian in Eq. (4.1) for any value of the field h.

This Hamiltonian has two well-known limits, for $h \to 0$ and $h \to \infty$. When only $\widehat{H}_E$ is present ($h \to 0$), the electric ground state is a trivial product state with all spins aligned along the x direction $|\Omega_E\rangle = \bigotimes_l |+\rangle_l$, which satisfies the local gauge constraints in Eq. (4.4) and corresponds to the absence of any electric field excitation.

In the opposite limit $h \to \infty$ only the magnetic term remains, the system behaves like a surface code [172, 185, 192] and displays topological order. In this case, the number of ground states depends on the boundary conditions. The ground states
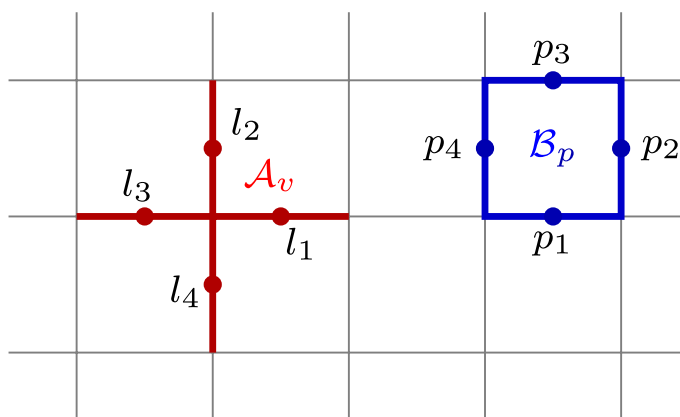
Figure 4.1.: Representation of a star operator $\mathcal{A}_v$ (in red) and a plaquette operator $\mathcal{B}_p$ (in blue), with the corresponding qubits on the links (solid circles).

of $\widehat{H}_B$ are the simultaneous eigenstates of all plaquette and star operators with eigenvalue 1 and correspond to the absence of magnetic fluxes. In the case of open and smooth boundaries [172], there is a single magnetic ground state which can be expressed as an equal amplitude linear superposition of all possible contractible electric flux loops Γ:

$$|\Omega_B\rangle = \mathcal{N} \sum_{\Gamma} \mathcal{W}_\Gamma |\Omega_E\rangle = \prod_p \left(\frac{1 + \mathcal{B}_p}{\sqrt{2}} \right) |\Omega_E\rangle \ . \tag{4.6}$$

Here $\mathcal{N}$ is a normalization factor and $\mathcal{W}_\Gamma$ is the Wilson loop operator associated to a closed path Γ, defined as the product of $\hat{\sigma}^z$ matrices on the links belonging to Γ:

$$\mathcal{W}_\Gamma = \prod_{l \in \Gamma} \hat{\sigma}_l^z \ . \tag{4.7}$$

Since $\hat{\sigma}^z|\pm\rangle = |\mp\rangle$, Wilson loops applied to the electric ground state $|\Omega_E\rangle$ create closed lines of electric field excitations. Note that any contractible Wilson loop can be uniquely written as a suitable product of plaquette operators $\mathcal{B}_p$, thus, by taking the uniform superposition in Eq. (4.6), one obtains an eigenstate with eigenvalue 1 for each $\mathcal{B}_p$. This is made explicit by the second form of $|\Omega_B\rangle$, expressed as the normalized product of the projectors on the eigenstates of each plaquette operator with eigenvalue 1.

The magnetic coupling h drives the system across a topological phase transition, occurring at h_c, between the electric and the magnetic phases, which are distinguished by different behaviors of the expectation values of the Wilson loop operators. From the definition of the limiting ground states $|\Omega\rangle_E$, $|\Omega\rangle_B$, it follows that in the two limits $h = 0$ and $h \to \infty$ we have for all paths Γ

$$\langle\Omega_E|\mathcal{W}_\Gamma|\Omega_E\rangle = 0 \quad \text{and} \quad \langle\Omega_B|\mathcal{W}_\Gamma|\Omega_B\rangle = 1 \ .$$

At a finite value of h, the expectation value of Wilson loops on the ground state decreases exponentially in the size of Γ, with a leading contribution given by [193,

194]

$$\langle \mathcal{W}_\Gamma \rangle = \mathrm{e}^{-\chi(h)A_\Gamma - \delta(h)P_\Gamma} \,, \tag{4.8}$$

where A_Γ and P_Γ are the area enclosed by the loop Γ and its perimeter, respectively, while χ and δ are two positive functions. For $h < h_c$, the system is in a phase dominated by the electric term $\widehat{H}_E$, $\chi(h) > 0$ and $\langle \mathcal{W}_\Gamma \rangle$ decays with an "area law". This means that large loops of electric excitations are strongly suppressed, which is a signature of *confinement* [195, 196]. In the opposite *deconfined* phase, where $h > h_c$ and the dominant term is $\widehat{H}_B$, $\chi(h) \to 0$, and the behavior of large Wilson loops follows a "perimeter law".

When periodic boundary conditions (PBC) are considered in both directions, the Hamiltonian acquires an extra $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry related to non-contractible 't Hooft loops. Consider a closed path $\mathcal{C}$ in the dual lattice. $\mathcal{C}$ crosses orthogonally a sequence of links of the direct lattice that we denote schematically by $\ell \cap \mathcal{C}$. The 't Hooft loop operator

$$\tau_\mathcal{C} = \prod_{l \cap \mathcal{C}} \hat{\sigma}_l^x \,, \tag{4.9}$$

commutes with the Hamiltonian (4.1) for any closed loop $\mathcal{C}$. However, if $\mathcal{C}$ is contractible, $\tau_\mathcal{C}$ can always be expressed as a product of star operators A_v, so that $[\tau_\mathcal{C}, \hat{H}] = 0$ does not provide any additional information that is not already contained in the gauge-invariance of the Hamiltonian. Considering PBC, the lattice becomes a torus and there are indeed two inequivalent non-contractible loops, whose corresponding 't Hooft operators τ_h and τ_v cannot be expressed as a product of star operators, and thus provide new symmetries. Fig. 4.2 shows examples of 't Hooft loop operators. Similarly, with PBC, two non-contractible Wilson loop operators $\mathcal{W}_h$ and $\mathcal{W}_v$ can be constructed, associated with a horizontal or vertical winding around the lattice, respectively. These two operators commute with $\widehat{H}_B$ and $\mathcal{A}_v$, but they can not be written as a product of plaquette operators. The following mixed commutation relations are easily derived:

$$\begin{aligned} [\mathcal{W}_h, \tau_h] &= [\mathcal{W}_v, \tau_v] = 0 \\ \{\mathcal{W}_h, \tau_v\} &= \{\mathcal{W}_v, \tau_h\} = 0 \,. \end{aligned} \tag{4.10}$$

In the trivial limit $h \to 0$, the expectation values of τ_h and τ_v on the electric ground state $|\Omega_E\rangle$ are $+1$. In the topological limit $h \to \infty$, when $[\mathcal{W}_h, \hat{H}] = [\mathcal{W}_v, \hat{H}] = 0$, we get four degenerate ground states characterized by different eigenvalues of the 't Hooft loops, corresponding in the basis $|\tau_h, \tau_v\rangle$ to:

$$\begin{aligned} |++\rangle &= |\Omega_B\rangle \,, \\ |+-\rangle &= \mathcal{W}_h |\Omega_B\rangle \,, \\ |-+\rangle &= \mathcal{W}_v |\Omega_B\rangle \,, \\ |--\rangle &= \mathcal{W}_h \mathcal{W}_v |\Omega_B\rangle \,. \end{aligned}$$
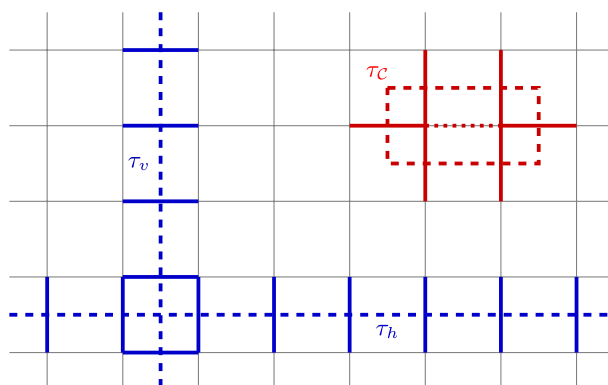
Figure 4.2.: The two non-contractible 't Hooft loops τ_h, τ_v and a simple example of how a contractible 't Hooft loop operator $\tau_{\mathcal{C}}$ is written as a product of star operators. The dotted link indicates a cancellation due to $(\hat{\sigma}_l^x)^2 = 1$.

Now, one clearly sees that τ_h and τ_v measure the winding number of the lines of electric field excitations in the vertical and horizontal directions, created by the non-contractible Wilson loops $\mathcal{W}_v$ and $\mathcal{W}_h$, respectively. A non-trivial topological ground state is signaled by eigenvalues -1 in the basis $|\tau_h, \tau_v\rangle$. This is an immediate consequence of the anticommutation relations.

At finite values of $h > h_c$, the perfect degeneracy between these four states gets lifted by an energy splitting vanishing exponentially with the linear size L. Hence, in the thermodynamic limit, the topological phase of a $\mathbb{Z}_2$ LGT with PBC is characterized by four exactly degenerate ground states in different topological sectors. The phase transition occurs at $h_c = 3.04438(2)$ [197] and it can be understood by considering the duality between the Hamiltonian in Eq. (4.1) and the 2-dimensional quantum Ising model with transverse field (2D-TFIM), valid under the following assumptions on our LGT model: the lattice has PBC, the gauge-symmetry $A_v|\psi\rangle_{phys} = |\psi\rangle_{phys}$ is imposed at each vertex, and the Hilbert space is restricted to the $\tau_h = \tau_v = +1$ sector [198]. Indeed, we can define new Pauli spin variables X_p and Z_p on the dual lattice, where p denotes the plaquette centers, by identifying

$$\begin{cases} X_p &= \mathcal{B}_p \\ Z_p Z_{p'} &= \hat{\sigma}^x_{l(p,p')} \end{cases}, \tag{4.11}$$

where p and p' are neighboring plaquettes and $l(p, p')$ is the link shared by p, p'. With the mapping in Eq. (4.11), the Hamiltonian becomes a transverse-field Ising model on the dual lattice:

$$\widehat{H} = \sum_{\langle p,p'\rangle} \left(1 - Z_p Z_{p'}\right) - h \sum_p X_p \,. \tag{4.12}$$

One can check that the algebra generated by the new operators is the same as the original one, confirming the unitary equivalence of the two models. Notice that the number of degrees of freedom is now halved: L^2 qubits (one for each plaquette),

instead of $2L^2$ (one for each link). This is an effect of the gauge symmetries, which are now automatically incorporated into the model. Finally, it is important to mention the fact that this duality fixes the global $\mathbb{Z}_2$ symmetry of the Ising model: in the original representation, the product of all plaquette operators is the identity $\prod_p \mathcal{B}_p = 1$. In the 2D-TFIM, this is reflected in the condition $\prod_p X_p|\psi\rangle = |\psi\rangle$, which means that the physical states must be invariant under a global spin flip.

In this work, we shall employ both the original formulation of Eq. (4.1) and its dual model in Eq. (4.12). The dual Ising model will be exploited to speed up our numerical analysis of the $\mathbb{Z}_2$ LGT and, in particular, to verify the scalability of QAOA between different system sizes.

4.2.1. Ground state preparation with QAOA

To prepare the ground state of the LGT Hamiltonian in Eq. (4.1), we use a version of the Quantum Approximate Optimization Algorithm (QAOA) described in Chapter 1, in the broad class of AQC-inspired variational states. Considering the two terms $\widehat{H}_B$ and $\widehat{H}_E$ in the LGT Hamiltonian, we construct the following variational ansatz

$$|\psi_P(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle = \widehat{U}(\gamma_P, \beta_P) \ldots \widehat{U}(\gamma_1, \beta_1) |\psi_0\rangle \,, \tag{4.13}$$

where $\boldsymbol{\beta} = \beta_1, \ldots, \beta_P$ and $\boldsymbol{\gamma} = \gamma_1, \ldots, \gamma_P$ are 2P free real parameters, and the unitary operators $\widehat{U}(\gamma_m, \beta_m)$, for $m = 1 \cdots P$, evolve the state according to $\widehat{H}_B$ and $\widehat{H}_E$, in an alternating fashion. More precisely, the initial state $|\psi_0\rangle$ can either be the electric ground state $|\Omega_E\rangle$ or the magnetic one $|\Omega_B\rangle = |++\rangle$, and, depending on the choice of $|\psi_0\rangle$, we define the operator $\widehat{U}_m = \widehat{U}(\gamma_m, \beta_m)$ in Eq. (4.13) as [1]

$$\widehat{U}_m = \begin{cases} e^{-i\beta_m \widehat{H}_E} e^{-i\gamma_m \widehat{H}_B} & \text{if} \quad |\psi_0\rangle = |\Omega_E\rangle \,, \\ e^{-i\gamma_m \widehat{H}_B} e^{-i\beta_m \widehat{H}_E} & \text{if} \quad |\psi_0\rangle = |\Omega_B\rangle \,. \end{cases} \tag{4.14}$$

For a given choice of the coupling h, which identifies a target Hamiltonian $\widehat{H}_{\text{target}}(h) = \widehat{H}_E + h\widehat{H}_B$, an approximation of the associated ground state is found using a classical minimization of the variational energy

$$E_P(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle \psi_P(\boldsymbol{\gamma}, \boldsymbol{\beta}) | \widehat{H}_{\text{target}}(h) | \psi_P(\boldsymbol{\gamma}, \boldsymbol{\beta})\rangle \tag{4.15}$$

in this 2P-dimensional energy landscape. This is a specific implementation of the general framework outlined in Chapter 1.

As previously discussed, precisely determining the global minimum is generally a challenging task. This difficulty arises because local optimization routines often become trapped in one of the numerous local minima. In the following, we shall

---

[1] This choice avoids adding only a global phase in the first step.

adopt an effective strategy to search for optimal (or quasi-optimal) solutions by a two-step QAOA procedure that starts from a linear schedule for the parameters, drawing inspiration from digitized Quantum Annealing. This is essentially the same procedure we dubbed QAOA-1 in Chapter 2.

4.2.2. Circuit implementation of the variational ansatz

The QAOA variational wavefunction in Eq. (4.13) is obtained by applying P layers of local unitary operators, by alternating the time evolutions generated by plaquette and electric field interactions. In what follows, we are going to describe how to implement the operations involved in each layer of the variational circuit by using only single and two-qubit gates. Since we shall focus on a single application of the unitary operations $\mathrm{e}^{-i\beta_m \widehat{H}_E}$ and $\mathrm{e}^{-i\gamma_m \widehat{H}_B}$, the index m will be dropped from the parameters.

The electric term of Eq. (4.2) is a sum of single-qubit operators, therefore the evolution it generates can be realized as a product of single-qubit rotations around the x axis by the angle β, up to an irrelevant global phase. As usual, the computational basis we adopt hereafter is the $\hat{\sigma}^z$ eigenbasis. Therefore, we employ Hadamard gates to diagonalize $\widehat{H}_E$, and we reproduce the electric evolution during a single QAOA step by simultaneously applying operators $U_p(\beta) = \mathrm{e}^{i\beta\hat{\sigma}^z}$ to all qubits, i.e. a global rotation of angle β around the z-axis [2]. A schematic representation of the single-qubit gates required is sketched in Fig. 4.3(b).

The implementation of the time evolution associated with plaquette operators is less trivial [199–201], but it can be realized in a local way as a combination of single- and two-qubit gates. Fig. 4.3(a) shows that a single-plaquette unitary operator $\mathrm{e}^{i\gamma\mathcal{B}_p}$ is obtained by a suitable combination of CNOT gates and a single-qubit rotation $U_p(\gamma)$ applied to the fourth qubit of the plaquette. The fourth qubit is the target of all CNOTs and it is restored to its initial logical state by the last three CNOTs, such that $U_p(\gamma)$ successfully applies the phase $\gamma\mathcal{B}_p$ only. An alternative technique based on ancillary qubits is presented in Refs. [202–204].

For a lattice composed of several plaquettes, the circuits in Fig. 4.3(a) cannot be simultaneously run on all of them, since two neighboring plaquettes share a qubit: as shown in Fig. 4.3(c), the qubit 4 not only acts as the target for the left plaquette but also as one of the controls for the right plaquette. The time evolution of the plaquette operators, however, can still be efficiently parallelized. For the sake of simplicity, we shall first consider systems with an even number of columns (for an even number of rows the situation is formally equivalent). In this case, we can

---

[2]In the $\hat{\sigma}^z$ eigenbasis the circuit depth is 3; however, in a realistic experimental setup, it is often possible to implement single-qubit rotations around the x axis, requiring only a single layer of gates
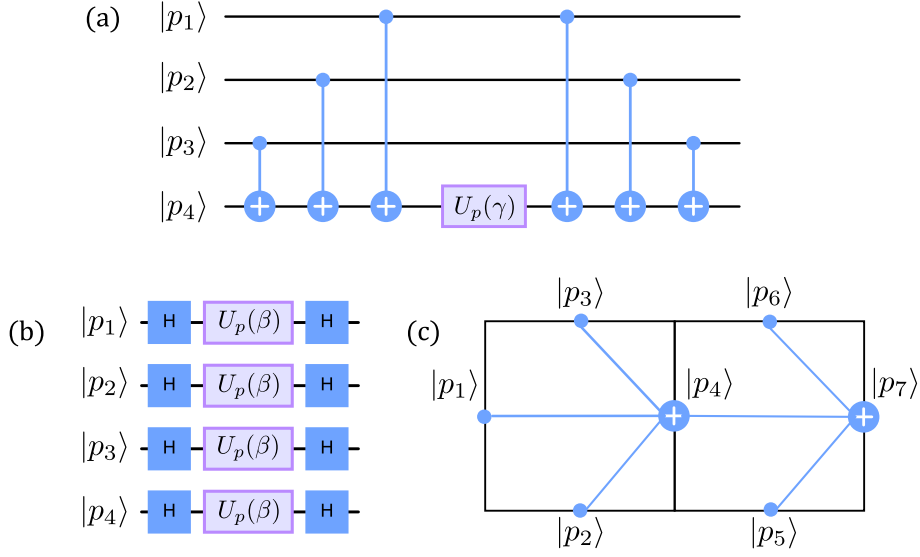
Figure 4.3.: (a) Circuit implementation of the operator $e^{i\gamma\mathcal{B}}$ acting on a single plaquette, with a target and 3 control qubits. The states $|p_i\rangle$ are expressed in the $\hat{\sigma}_z$ eigenbasis and $U_p(\gamma) = e^{i\gamma\hat{\sigma}^z}$ is a single-qubit rotation around the z axis. (b) Implementation of the single-qubit operations that describe $e^{-i\beta\hat{H}_E}$. (c) Example of the *sequential* implementation of the operator $e^{i\gamma\mathcal{B}}$ on two neighboring plaquettes. Here, the qubit $|p_4\rangle$ is used first as the target qubit for the left plaquette and afterward as a control qubit for the right plaquette.

decompose the whole lattice into sets of two neighboring horizontal plaquettes, each set having the same structure as depicted in Fig. 4.3(c). Let us focus on a single set, which corresponds to our basic unit. We show in Fig. 4.4 the corresponding quantum circuit that will be run *in parallel* for all such sets. Neighboring pairs of plaquettes share qubits at their boundary: this can be understood by ideally replicating the pair in Fig. 4.3(c), to build a lattice. For instance, the qubits 3 and 6 of our set also correspond to the qubits 2 and 5 of the set above the one in exam. Similarly, qubits 2 and 5 are homologous of qubits 3 and 6 for the set below, whereas qubit 7 matches qubit 1 of the plaquette pair lying on the right of the one depicted, and so on.

The quantum circuit defined in Fig. 4.4 performs the rotation of both plaquettes in 12 steps and it can be run in parallel for all plaquette pairs. This procedure is based on applying the phase gates $U_p(\gamma)$ on qubits 4 and 7 (thus on all the qubits of the lattice lying on the vertical links) to implement the plaquette rotation. The gates partially depicted in green are related to the simultaneous realization of the same algorithm for the neighboring plaquettes. All the qubits lying on the vertical links are required to be connected via CNOTs to four neighboring qubits, whereas the horizontal qubits are just connected to two neighbors each. As a result of the previous scheme, each of the P steps of QAOA can be realized with a circuit of depth 13 on systems with open boundaries, or systems with closed boundaries and an even number of rows or columns. In this work, we mainly focus on the numerical
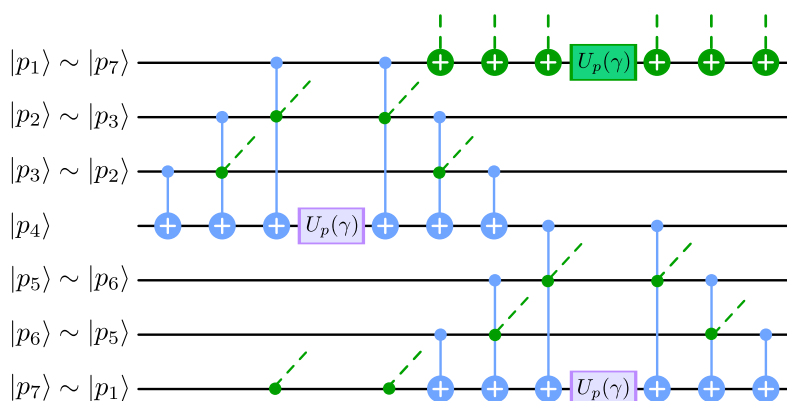
Figure 4.4.: Algorithm to implement the plaquette operator $e^{i\gamma\mathcal{B}}$ on the two neighboring plaquettes depicted in Fig. 4.3(c). The labeling of the qubit lines emphasizes that all the boundary qubits are shared with the neighboring plaquette pairs. The partially depicted green gates (dashed lines) are related to the simultaneous implementation of the same algorithm on a couple of neighboring plaquettes. They connect the displayed qubits with qubits belonging to the surrounding lattice sites, based on suitable translations of the two-plaquette unit.

investigation of a lattice with 3×3 plaquettes and periodic boundary conditions. Its implementation on actual quantum hardware requires some additional care due to the boundary conditions and, in that case, each QAOA step can be realized with a circuit of depth 18. This is explained in detail in Appendix C.

Preparation of the initial state $|\psi_0\rangle$

To estimate the computational cost for the preparation of a target ground state of the $\mathbb{Z}_2$ LGT, it is also important to consider the preparation of the initial state $|\psi_0\rangle$, which will be subsequently modified by the QAOA layers. The electric ground state $|\Omega_E\rangle$ is a trivial product state and it can be prepared by applying Hadamard gates on all qubits, to rotate them into the eigenstate $|+\rangle$ of the $\hat{\sigma}^x$ operator. The toric code (magnetic) ground state $|\Omega_B\rangle$ displays instead topological order and long-range entanglement [187]. To initialize this state, we follow the technique adopted in Ref. [189]: assuming that each qubit is initially in the eigenstate $|\uparrow\rangle$ of $\hat{\sigma}^z$ (the usual computational basis state $|0\rangle$), in each plaquette we first apply three Hadamard gates on the control qubits and then three CNOT gates targeting the fourth [3]. This procedure is similar to the first 3-CNOT sequence of the circuit depicted in Fig. 4.3(a), with the addition of Hadamard gates on the qubits $|p_1\rangle$, $|p_2\rangle$, and $|p_3\rangle$.

These operations can be performed in parallel on plaquettes belonging to a single column and then repeated L times to cover the whole lattice. Operations on dif-

---

[3] Our implementation is equivalent to the procedure in Ref. [189] up to a rotation of the basis. Hence, we have three control qubits and one target for each plaquette, instead of one control and three targets.

ferent rows, however, cannot be parallelized because in each plaquette the CNOT gates must be applied *before* using one of the control qubits as the target for the neighboring column. If we consider Fig. 4.3(c), the plaquette operations must start from the rightmost column, in such a way that the qubit $|p_4\rangle$ is used as a control before becoming the target of the plaquette on the left. Incidentally, this is the opposite order to what is described in the caption of Fig. 4.3(c) and depicted in Fig. 4.4. However, in that case, the specific order is irrelevant, since different plaquette operators trivially commute.

We remark an important difference between the circuit implementing $|\Omega_B\rangle$ and the gate decomposition of $\mathrm{e}^{-i\gamma\widehat{H}_B}$: despite the preparation of the ground state of $\widehat{H}_B$ on a single plaquette requires a smaller number of gates, for large systems the initialization of $|\Omega_B\rangle$ requires a deeper circuit than the Hamiltonian gate, which, instead, can be run in parallel on all the even or odd columns (rows). This reflects the necessity of having a circuit with depth $O(L)$ to prepare a state with long-range entanglement, such as $|\Omega_B\rangle$, which has been well studied in the literature [186–188]. In conclusion, when we compare the QAOA results with different choices of the initial state $|\psi_0\rangle$, we need to take into account the overhead required for preparing $|\Omega_B\rangle$.

Possible generalizations

The construction of the QAOA layers we presented so far was restricted to the case of an Abelian $\mathbb{Z}_2$ LGT. We stress, however, that the same procedure can be generalized to pure 2D LGTs with arbitrary discrete gauge groups. In particular, by suitably extending the Hilbert space associated with each link of the square lattice, it is possible to implement the time evolution steps of both the electric and magnetic Hamiltonian based on local unitary operators [201].

In this respect, the simplest generalization is provided by $\mathbb{Z}_n$ LGTs (see, for instance [203, 205–211]). In this case, a gauge degree of freedom is encoded into an n-dimensional Hilbert space, as common in quantum clock models with $\mathbb{Z}_n$ symmetries [212, 213]. The electric field assumes indeed n different values, which may be represented by suitable qudits (or by embedding each degree of freedom in a set of qubits). $\widehat{H}_E$ remains a local Hamiltonian, whose time evolution can be performed in a parallel way over all links.

As in the case of the $\mathbb{Z}_2$ theory, also for $\mathbb{Z}_n$ symmetries, there is a suitable unitary transformation mapping the eigenstates of the electric Hamiltonian into the eigenstates of the magnetic operators adopted to build the plaquette terms (the so-called connection operators). Such unitary transformations generalize the Hadamard gates we adopted and correspond to a quantum Fourier transform. Additionally, the plaquette term maintains the same 4-body interaction form through a suitable replace-

ment of $\hat{\sigma}_z$ with quantum clock operators. The implementation of the plaquette operator thus requires generalizing the CNOT to controlled $\mathbb{Z}_n$ clock gates. Again, the phase diagram of pure 2D $\mathbb{Z}_n$ LGT models presents a deconfined and topological phase at large h, whose topological order matches the $\mathbb{Z}_n$ generalization of the toric code [214, 215], and a confined phase whose ground state becomes a trivial product for $h = 0$.

By following the Kogut-Susskind Hamiltonian construction, a further generalization can be implemented to investigate ground states of discrete non-Abelian 2D LGTs (see, for example, Refs. [201, 216, 217]). In this case, the gauge degrees of freedom can be represented either in an eigenbasis associated with the irreducible representations of the group, which is diagonal in the electric term of the Hamiltonian, or in an eigenbasis associated to the group elements, which is diagonal in the magnetic term of the Hamiltonian. The general structure of a quantum algorithm for implementing the QAOA steps, in this case, is analogous to the previous one and can be based on the construction in Ref. [201]. Given the non-Abelian nature of the group, however, the implementation of the magnetic time evolution requires a further technical generalization. In this case, the irreducible representations are not one-dimensional and correspondingly the connection operators acquire a tensor form, therefore the gauge-invariant plaquette terms must be written in terms of their trace [195], requiring, in turn, to extend the rotation operators $U_p(\gamma)$ to more general single-link gates, which apply phases given by the traces of gauge group matrices.

4.3. Numerical results

In this section we analyze the QAOA performance on our LGT model, showing that the ground state can be prepared through shallow circuits with good fidelity both in the confined and in the topological phase. Unless otherwise stated, the numerical analysis is performed on a lattice with 3×3 plaquettes (18 qubits) and implemented through the python package Qiskit [35], using the circuit sketched in Appendix C that slightly generalized the one in Fig. 4.4. Simulations of larger systems ($L = 4,\ 5$), instead, exploit the mapping onto the 2D-TFIM to reduce the Hilbert space dimension and allow for the exact evaluation of the QAOA ansatz.

4.3.1. Energy landscape

As remarked in Chapter 1, the presence of rugged energy landscapes is a common problem that severely affects the classical optimization loop of VQAs by making it prone to remaining stuck in local minima, some of which might be far from the ground state energy. Here, we address this issue by inspecting the structure of
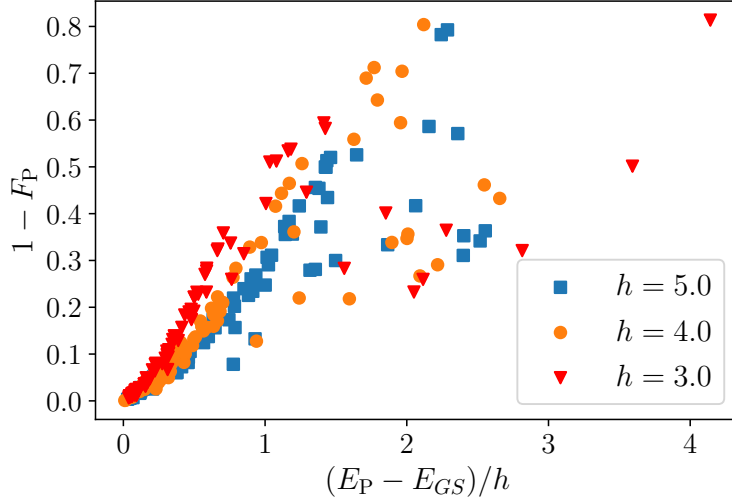
Figure 4.5.: Infidelity vs residual energy rescaled over the magnetic coupling h, for $h = 5, 4, 3$. Notice that the infinite-size value of the critical coupling is $h_c = 3.04438(2)$ [197]. Data refer to 100 local optimizations on a system with linear dimension $L = 3$, initial random guess of the QAOA parameters, and circuit depth P = 5. The initial state is $|\psi_0\rangle = |\Omega_E\rangle$.

the energy landscape, which determines whether the classical optimization of the variational parameters can be performed efficiently or not [22]. To characterize the energy landscape and to assess the quality of the optimized variational ansatz $|\psi_P(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)\rangle$, two common metrics are the residual energy and the fidelity. Given a target Hamiltonian $\widehat{H}_{\text{targ}}(h)$, we denote with $|\psi_{\text{targ}}\rangle$ its ground state and with E_{GS} the corresponding energy, both obtained with exact diagonalization. The residual energy is simply the difference between the minimized variational energy $E(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$, defined in Eq. (4.15), and E_{GS}, while the fidelity follows the usual definition

$$F_P(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*) = |\langle\psi_P(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)|\psi_{\text{targ}}\rangle|^2 . \tag{4.16}$$

Clearly, the fidelity is a precise estimate of the accuracy of the approximation of the target state, even though, in an actual experiment, it is hardly accessible because it requires an exponential number of measurements. The energy, instead, is easily estimated: the magnetic contribution is diagonal in the computational basis, while the electric contribution is obtained by applying a basis rotation on each qubit, i.e. a set of Hadamard gates, before the measurement.

In the $\mathbb{Z}_2$ LGT model, the energy landscape emerging from the QAOA ansatz is characterized by many local minima covering a wide energy interval, making random-start local optimization impractical. This effect is particularly severe if the target state is in a phase different from the initial one $|\psi_0\rangle$. This is illustrated in Fig. 4.5, where we show the residual energy and the infidelity $1 - F_P(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$ for 100 different random-start local optimizations with P = 5 QAOA layers and three different values of the magnetic coupling, around or above the topological phase

transition at $h_c = 3.04438(2)$ [197] (for $L \to \infty$). The initial state is the product state $|\Omega_E\rangle$ in the extreme confined limit ($h = 0$) and the local minimizations are performed with the BFGS algorithm [141]. Although there is a clear concentration of data in the corner corresponding to successful optimizations, where both the infidelity and the residual energy tend to zero, there are many local minima far from the ground state, suggesting that more refined optimization techniques are needed for this model.

However, the clear correlation between the residual energy and the fidelity is reassuring since the absence of low-energy minima with small projection on the ground state guarantees that any scheme that allows for a reliable minimization of the energy will also lead to a good approximation of the target state. This correlation is intuitively justified by the existence of a gap in the topological phase. Indeed, the construction in Eq. (4.13) cannot mix different topological sectors of the model and, for each topological sector, there is only one ground state. Away from the critical point, the ground state is protected by a finite gap, such that the correlation between infidelity and residual energy must hold below this energy scale. Close to the critical point, other orthogonal low-energy states may appear and spoil the correlation. This, however, seems not to be the case. This resilience is not surprising for small system sizes in which the gap does not close even at h_c. However, we observe that the correlation between infidelity and residual energy holds also when we increase the linear dimension L of the lattice, and it actually appears to be even sharper, as shown in App. C. The variational energy is therefore a reliable figure of merit for the optimization, it can be efficiently measured in experiments, and the procedure is still effective when the system size is scaled up. All these represent positive factors for the feasibility of the implementation of QAOA on the $\mathbb{Z}_2$ LGT model in near-term quantum devices.

4.3.2. Heuristic local optimization: two-step QAOA

Because of the large number of suboptimal minima present in the energy landscape, it is important to adopt an efficient strategy to reliably find a good approximation of the true ground state of the target Hamiltonian. As previously discussed, this is, indeed, a crucial task for QAOA and VQAs in general, where the classical optimization outer loop is often the main computational bottleneck and several strategies have been proposed that go beyond a local search from a random start. These strategies range from problem-specific methods to general iterative techniques, based on observed patterns in the optimal schedules, which have already been discussed in the previous Chapters. Here, we adopt a simple two-step minimization protocol inspired by a version of digitized Quantum Annealing [75, 90] that corresponds to a slow turning-on of either the magnetic or the electric Hamiltonian. We outline this procedure below, which is closely related to the QAOA-1 implementation

in Chapter [2].

The idea behind this two-step optimization is to leverage the formal analogy between QAOA and digitized Quantum Annealing discussed in Chapter [1]: for depth-P QAOA, we first optimize the total run time of a digitized linear QA [90] of the same depth, and then fine-tune the variational parameters around this schedule. This approach can be used effectively when the system is initialized either in the electric state $|\Omega_E\rangle$ or in the magnetic one $|\Omega_B\rangle$. For generic applications, the confined electric ground state $|\Omega_E\rangle$ — a uniform superposition of all possible states in the computational (magnetic) basis — is a standard choice for initializing the variational circuit, because it is easy to prepare. However, this choice is non-optimal when we target states with long-range entanglement in the deconfined/topological phase, since a circuit of local unitary gates with bounded depth cannot create states with topological order beyond a certain system size [187].

When the initial state is $|\Omega_E\rangle$ we adiabatically turn on the magnetic coupling through $\widehat{H}(t) = \widehat{H}_E + (t/\tau)h\widehat{H}_B$ in a total time τ. After a digitization by Trotter decomposition (we set $\hbar = 1$)

$$\mathrm{e}^{-i\Delta t\widehat{H}(t_m)} \approx \mathrm{e}^{-i\Delta t\widehat{H}_E}\mathrm{e}^{-i\frac{m\Delta t}{\mathrm{P}}h\widehat{H}_B} \,, \tag{4.17}$$

with $t_m/\tau = m/\mathrm{P}$, we set $\gamma_m^0 = \frac{m\Delta t}{\mathrm{P}}h$ and $\beta_m^0 = \Delta t$ in the state in Eq. (4.13). In our first step, these linear-schedule parameters are optimized by searching for the optimal digitized QA Δt^* — a simple one-dimensional minimization — which leads to setting:

$$\gamma_m^{\mathrm{dQA}} = \frac{m\,\Delta t^*}{\mathrm{P}}h \,, \quad \beta_m^{\mathrm{dQA}} = \Delta t^* \,. \tag{4.18}$$

The second step in our QAOA procedure is to perform 10 local BFGS optimizations in the 2P-dimensional parameters space, starting from $(\boldsymbol{\gamma}^{\mathrm{dQA}}, \boldsymbol{\beta}^{\mathrm{dQA}}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a small 2P-dimensional vector with random entries uniformly sampled in the interval $[-0.025, 0.025)$, keeping the best outcome out of these local optimizations. Schematically:

$$(\boldsymbol{\gamma}^{\mathrm{dQA}}, \boldsymbol{\beta}^{\mathrm{dQA}}) + \boldsymbol{\epsilon} \xrightarrow{\text{best BFGS}} (\boldsymbol{\gamma}^\star, \boldsymbol{\beta}^\star) \,. \tag{4.19}$$

The toric code ground states $|\Omega_B\rangle$, corresponding to the extreme deconfined limit $h \to \infty$, provide a better initial state $|\psi_0\rangle$ when targeting ground states in the topological phase: they can be exactly prepared with local circuits whose depth scales linearly with the system width L [173, 189]. Proceeding with an adiabatic turning-on, now of the electric part of the Hamiltonian, through $\widehat{H}(t) = h(\widehat{H}_B + (t/h\tau)\widehat{H}_E)$, leads to a Trotter decomposition ($\hbar = 1$):

$$\mathrm{e}^{-i(\Delta t/h)\widehat{H}(t_m)} \approx \mathrm{e}^{-i\Delta t\widehat{H}_B}\mathrm{e}^{-i\frac{m\Delta t}{h\mathrm{P}}\widehat{H}_E} \,, \tag{4.20}$$

with $t_m/\tau = m/\mathrm{P}$. Hence, we set $\gamma_m^0 = \Delta t$ and $\beta_m^0 = \frac{m\Delta t}{h\mathrm{P}}$ in the state in Eq. (4.13). Once again, these can be optimized by searching for the optimal digitized QA Δt^*,

which leads to:

$$\gamma_m^{\text{dQA}} = \Delta t^* \,, \qquad \beta_m^{\text{dQA}} = \frac{m \, \Delta t^*}{h \, \text{P}} \,. \tag{4.21}$$

The second step in our QAOA procedure is identical to the previous case, as schematically indicated in Eq. (4.19).

In close analogy to the results of Chapter 2 and Chapter 3, two noteworthy features of the QAOA minima obtained by applying our two-step QAOA procedure are the *smoothness* of the schedules $(\gamma^\star, \beta^\star)$ (see Appendix C), and the closely related *transferability* of such smooth schedules from a smaller to a larger $L' > L$ model, discussed in Sec. 4.3.3. We benchmarked our heuristic two-step QAOA approach against a computationally expensive global optimization, finding comparable quality results in terms of ground-state fidelity, both in the confined and in the deconfined phase: we illustrate this in Appendix C.

In the following, we will compare the performance of our two-step QAOA for systems prepared either in the electric ground state $|\Omega_E\rangle$, or in the toric code ground state $|\Omega_B\rangle = |++\rangle$. For a fair comparison, a remark is in order: while $|\Omega_E\rangle$ is trivially prepared with one layer of single-qubit Hadamard gates, for the preparation of $|\Omega_B\rangle$ one should include an overhead circuit with $O(3L^2)$ gates, organized in L layers applied sequentially. As explained in Sec. 4.2.2, although no optimization is necessary for this preliminary step, it is still required to apply 3 CNOT gates for each plaquette.

The QAOA results obtained from the initial product state $|\Omega_E\rangle$ are reported in Fig. 4.6, where we show the infidelity $1 - F_{\text{P}}(\gamma, \beta)$ as a function of the circuit depth P for several values of the magnetic coupling, both in the confined phase ($h \lesssim 3$) and in the deconfined one ($h \gtrsim 3$). As expected, the variational ansatz converges faster to states in the same phase (e.g. $h = 1, 2$) but QAOA can reach very good fidelity $1 - F_{\text{P}} < 10^{-3}$, when P ≥ 5, for all the couplings we considered.

With the "reversed" protocol, starting from the toric code ground state $|\Omega_B\rangle = |++\rangle$, we obtain an overall behavior similar to what observed for $|\psi_0\rangle = |\Omega_E\rangle$, see Fig. 4.7(a), with the important difference that now the optimization converges faster when targeting the deconfined phase. Indeed, only P = 3 QAOA layers are now needed to reach $1 - F_{\text{P}} < 10^{-3}$ when $h > h_c$, see data for $h = 4$ or $h = 5$, while confined states require more QAOA layers to reach comparable accuracy.

For both choices of initial state, we observe that the infidelity decreases exponentially with the circuit depth; the only exceptions for P = 5, 6 can be ascribed to the algorithm remaining stuck in a (high-quality) local minimum, when the target state is very close to the initial one (see Appendix C). However, if we focus on the minimal resources to approximate the target state within a certain fidelity threshold, we can further reduce the number of parameters required. Figure 4.7(b) shows a comparison of QAOA performance with the two possible choices of the initial
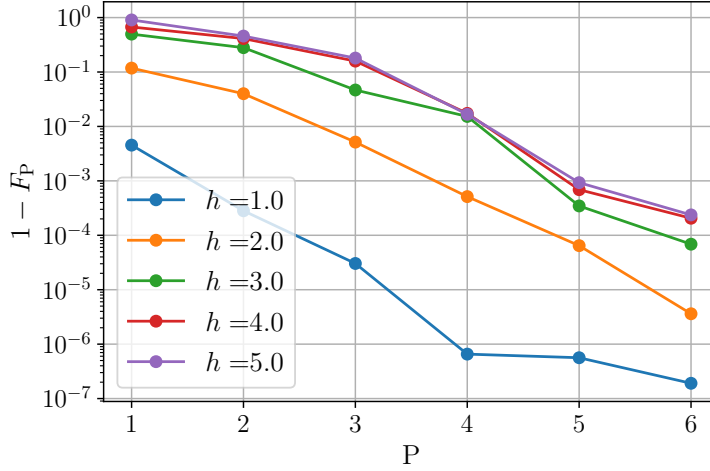
Figure 4.6.: Infidelity vs the number of QAOA layers P for $L = 3$ and several values of the magnetic coupling h. The initial state is the electric ground state $|\Omega_E\rangle$, thus the convergence towards the exact GS is faster for smaller values of h. We show the best result out of ten local BFGS minimizations following the heuristic two-step optimization method.

state, for P = 2 and P = 3, by looking at the best fidelity reached by the two-step optimization as a function of the coupling h. Remarkably, such shallow variational circuits are enough to prepare with high fidelity the ground states in the confined and deconfined phases, provided the initial state is selected in the same phase as the target ground state. Unsurprisingly, the region that requires a larger number of parameters, i.e. a deeper variational circuit, corresponds to the crossover between the two regimes, where $2 \lesssim h \lesssim 3$.

We finally observe that the choice of the initial state based on the target value of h plays a role analogous to the choice of the electric or magnetic representation of the LGT Hamiltonians applied in the quantum simulation protocols presented in Refs. [218, 219].

4.3.3. Schedule transferability

A promising route to reduce the computational cost of the outer-loop classical optimization in VQAs is the *transferability* of the optimal parameters from small to large instances of the same model. Indeed, as empirically observed or proven in specific applications of VQAs, see e.g. Chapter 3, if you consider two instances of the same model and a fixed variational circuit depth P, the optimal parameters obtained for the small system of size L may serve as a very good warm-start (or educated guess) for a local optimization for the L'-size model ($L' > L$).

Classical numerical simulations soon become unfeasible even for modest sizes, of-
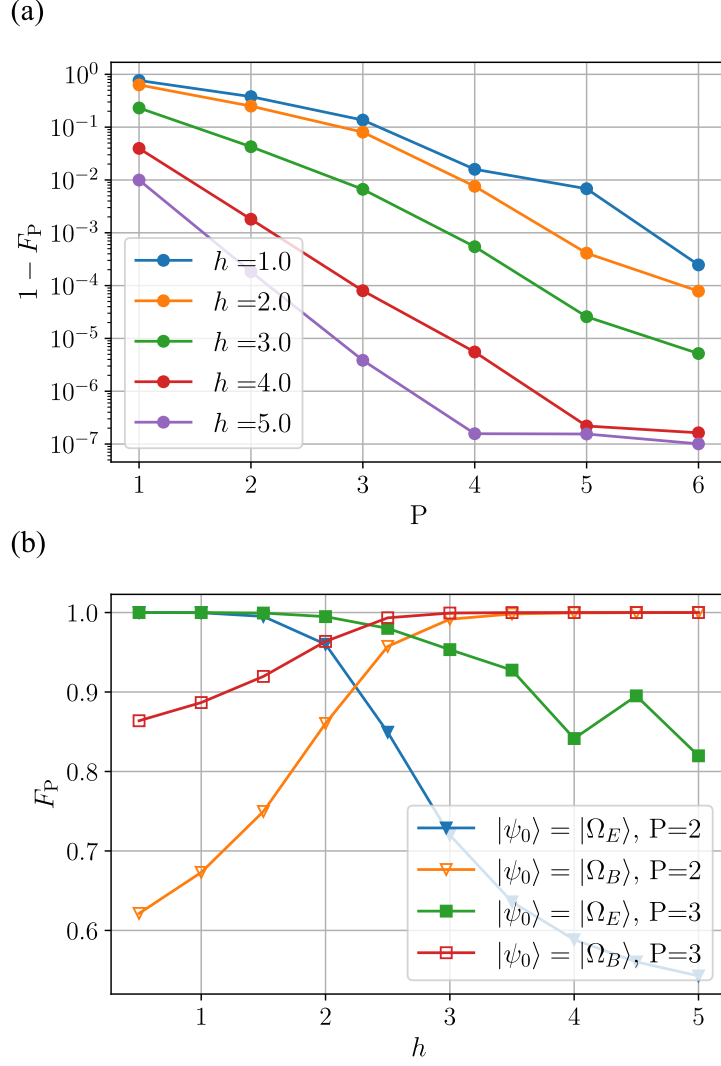
(a)

(b)

Figure 4.7.: (a): Infidelity vs the number of QAOA layers P for $L = 3$ and several values of the magnetic coupling h. Data correspond to the best out of 10 results obtained in the two-step optimization, performed on a state initially prepared in the toric code ground state $|\Omega_B\rangle$, hence the convergence is now faster for larger couplings h. (b): Comparison between two-step QAOA performance by starting from the electric and the magnetic ground states: we plot the fidelity vs magnetic coupling h at fixed values of P. Here P only takes into account the number of QAOA layers with parameterized gates, while it does not include the computational overhead for the preparation of $|\Omega_B\rangle$, compared to preparing $|\Omega_E\rangle$.

ten hindering a more systematic analysis on this issue: for our LGT model, which requires $2L^2$ quantum spins, even sizes as small as $L = 4, 5$ can be extremely challenging to simulate exactly. To partially overcome the size limitation, we exploit the mapping onto the 2D-TFIM, explained in Sec. 4.2, which involves only L^2 spins on a square lattice, by taking advantage of the restrictions imposed by the gauge constraints. This allows us to simulate exactly the variational optimization for $L = 4, 5$. To study schedule transferability, we first perform the two-step QAOA on the system with $L = 3$, as described in Sec. 4.3.2. The optimal angles $(\gamma^\star, \beta^\star)$ found for $L = 3$ are then used as warm-start points for a local optimization on larger sizes. In particular, we keep the best run out of 10 BFGS optimizations on the larger systems, each of them starting in the neighborhood of $(\gamma^\star, \beta^\star)_{L=3}$, similarly to the strategy used in the second part of the two-step QAOA protocol. This procedure is repeated for different values of the coupling.

The results obtained are reported in Fig. 4.8, where we compare the fidelity $F_\mathrm{P}(\gamma^*, \beta^*)$ vs h, for circuit depth P $= 6$ — which allows us to prepare the ground state for arbitrary h with an error $1 - F_\mathrm{P}(\gamma^*, \beta^*) < 10^{-3}$ for $L = 3$ — and both possible initial states: $|\psi_0\rangle = |\Omega_E\rangle$ (full symbols) and $|\psi_0\rangle = |\Omega_B\rangle$ (empty symbols). The transferability of the parameters is almost perfect when the initial and target states are in the same topological phase, leading to very high fidelities both in the small and large magnetic coupling regimes. Even when we target a ground state in a different phase than the initial one — for instance, $|\psi_0\rangle = |\Omega_E\rangle$ and $h > h_c \sim 3.0$ — the final fidelity is still large, allowing us to characterize the topological properties of the final state. Interestingly, the warm-start initialization provided by the $L = 3$ optimal parameters leads to a successful *local* minimum search for $L = 4, 5$, with an accuracy close to what can be achieved with a full global minimization. These additional results are discussed in Appendix C. Moreover, the number of iterations needed for the local optimization is rather small ($N_{iter} \lesssim 50$), confirming the benefit of the transferability of optimal solutions: once the $L = 3$ two-step solution is provided, only a small overhead in computation resources is required to fine-tune the parameters for larger sizes. Hence, transferability provides a speed advantage over starting from scratch a two-step optimization: even though the fidelity reached is comparable, the latter requires more runs of the quantum circuit, making it less efficient when the optimal schedule for a smaller system is already known.

This transferability evidence may be linked to the observation of the smooth schedules we found with the two-step optimization. Indeed, consistently with the discussion in Chapter 3, it is important to remark that the schedule transferability is not a general property of any minimum in the energy landscape, but it is associated with the smooth solution found with the two-step protocol. For instance, a global optimization yields slightly better results on the $L = 3$ system, but it often represents a poor choice as an educated guess to initialize a local minimum search on larger sizes, as discussed in Appendix C: this phenomenon might have some similarity to
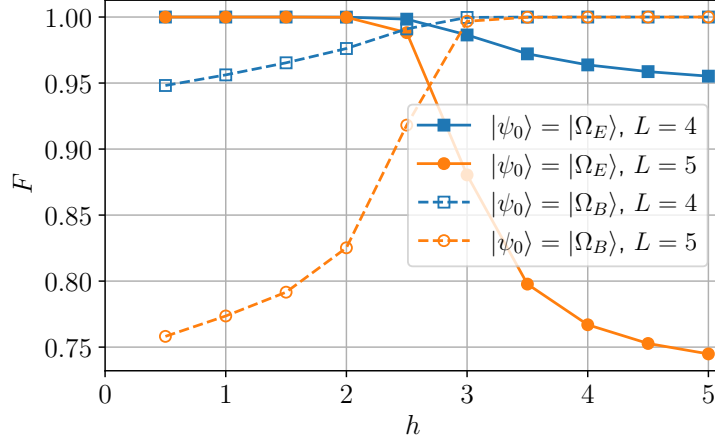
Figure 4.8.: Fidelity vs the magnetic coupling h. The data are obtained by using the two-step optimal schedules for $L = 3$, as an initial guess for 10 local BFGS optimizations performed on the larger systems $L = 4, 5$ (we fixed P = 6). Here, we report the best result out of the 10 runs. The full and empty symbols respectively correspond to $|\psi_0\rangle = |\Omega_E\rangle$ and $|\psi_0\rangle = |\Omega_B\rangle$.

overfitting in machine learning [220]. In this respect, the two-step scheme appears to outperform an extensive global search.

4.3.4. Ground state characterization

In the following, we turn our attention to the properties of the approximate ground states we prepare with QAOA. Despite the finite size limitations of our simulations, the states obtained through QAOA display most of the main features associated with the appearance of topological order and the crossover from a confined to a deconfined phase as h increases. The main observables to distinguish these two regimes are the Wilson loops, as defined in Eq. (4.7). We consider in particular Wilson loops $\mathcal{W}_{l_x, l_y}$ defined over rectangles of size $l_x \times l_y$.

As explained in Sec. 4.2, it is known that the deconfined phase is characterized by an exponential decay of $\langle \mathcal{W}_{l_x, l_y} \rangle$ with the perimeter P of the loop, whereas the confined phase displays a decay dictated by the area A of the loop [194]. In particular, the magnetic ground states $|\Omega_B\rangle$ are such that $\langle \mathcal{W}_{l_x, l_y} \rangle = 1$, while in the electric ground state $|\Omega_E\rangle$ Wilson loops always present vanishing expectation values. Recalling Eq. (4.8), the overall behavior of a Wilson loop can be approximated by $\langle \mathcal{W} \rangle \propto e^{-\chi A - \delta P}$. Indeed, if $\chi > 0$, the exponential decay with the area dominates for large loops, while if instead $\chi = 0$, the decay is dictated by the perimeter law only. To extract the information about the χ coefficient we estimate the so-called
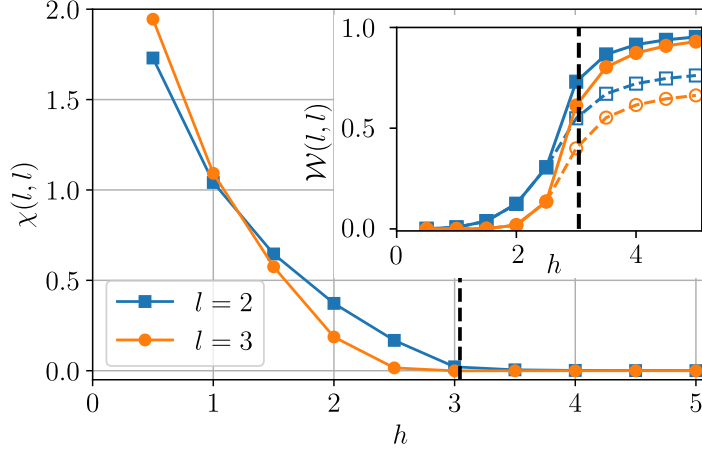
Figure 4.9.: Cruetz ratio $\chi(l, l)$, defined in Eq. (4.22), for two different loops in a system with $L = 5$. Inset: expectation value of Wilson operators $\mathcal{W}_{l,l}$, corresponding to the data in the main plot, vs the coupling strength h. The vertical dashed line indicates the critical value of the coupling h_c. All data refer to the best energy out of 20 BFGS local optimizations, with $\mathrm{P} = 6$ and $|\psi_0\rangle = |\Omega_E\rangle, |\Omega_B\rangle$, performed on a system of linear size $L = 5$ and initialized with the optimal parameters found for $L = 3$. Empty symbols in the inset are data obtained with $|\psi_0\rangle = |\Omega_E\rangle$.

Creutz ratio [196]:

$$\chi(l, l) = -\log \frac{\langle \mathcal{W}_{l,l} \rangle \langle \mathcal{W}_{l-1,l-1} \rangle}{\langle \mathcal{W}_{l,l-1} \rangle \langle \mathcal{W}_{l-1,l} \rangle}. \tag{4.22}$$

This ratio is indeed built to cancel the perimeter contribution to the decay of the observables and approximate the coefficient χ, which is recovered for large l.

Fig. 4.9 displays the Creutz ratio in a system with $L = 5$ and periodic boundary conditions for states obtained with $P = 6$ QAOA steps applied either to the state $|\Omega_E\rangle$ (for $h < 3$) or to $|\Omega_B\rangle$ (for $h \geq 3$). The optimization on the $L = 5$ systems was initialized with the best result obtained with the two-step protocol for $L = 3$, on top of which we performed 20 local minimum searches, keeping the best outcome.

Analogously to other LGT studies on small lattices [184, 221, 222], the finite size effects in our computations are strong. When considering a Wilson loop of width 3, its opposite sides lay at distance 2. This implies that what we observe in Fig. 4.9 may provide a quantitative estimate of the behavior in thermodynamic systems only if the correlation length is sufficiently smaller than this distance, thus only sufficiently far from the phase transition. Despite this limitation, the Creutz ratio $\chi(3, 3)$ presents a behavior that clearly distinguishes the confined phase ($\chi > 0$) and the deconfined phase ($\chi \to 0$) appearing for $h \gtrsim 3$, although a quantitative identification of h_c is beyond the possibilities of these small systems and loops.

The inset of Fig. 4.9 reports the expectation value of the Wilson loop opera-

tors corresponding to the Creutz ratios shown in the main plot. It clearly shows a crossover between the trivial, confined state with $\langle \mathcal{W}_\Gamma \rangle \to 0$ and the topologically ordered, deconfined limit $\langle \mathcal{W}_\Gamma \rangle \to 1$. With the chosen scheme, i.e. starting from $|\Omega_E\rangle$ or $|\Omega_B\rangle$ depending on the target state, they perfectly match the results from exact diagonalization (not shown) as expected from the high fidelity reached, see Fig. 4.8. We emphasize that the possibility of obtaining a reliable estimate of the expectation value of the Wilson loops yields further important implications: Ref. [219] shows that, in a $U(1)$ LGT, even the expectation value of the single plaquette operator can be used to extract the running coupling of the model, which is a fundamental quantity related to its renormalization.

If we chose to start always from the electric ground state, the deviation from exact results would become larger in the deconfined phase, as also expected from the fidelity drop observed in Fig. 4.8. However, the results obtained in this non-optimal case still provide an acceptable scaling of the Wilson loop for the deconfined regime (empty symbols in Fig. 4.9): even without perfect reconstruction of the target state, it is still possible to signal the existence of a deconfined phase. This is, indeed, useful for experimental investigation, where realistic setups are limited to shallow circuits and noise would decrease the quality of the approximate ground state.

Another observable that marks the onset of topological order is the topological entropy [223, 224]. Given a connected subsystem $A \cup B \cup C$ of the whole lattice, one should define its topological entropy as

$$S_{topo} = S_A + S_B + S_C - S_{AB} - S_{BC} - S_{AC} + S_{ABC} \ . \tag{4.23}$$

Here S_X is the von Neumann entanglement entropy of a generic subsystem X, obtained by tracing out all degrees of freedom in the complement of X with respect to the whole system, and $\{A, B, C\}$ is a tripartition of the region of which we compute the topological entropy. This specific choice allows for the cancellation of terms that are proportional to each separation boundary between the partition elements [223], a step conceptually similar to the Creutz ratio definition. In the toric code state $|\Omega_B\rangle$, the topological entropy of any subsystem is $S_{topo} = -\ln 2$ and the total entanglement entropy is

$$S_{ABC} = N_v \ln 2 + S_{topo} = (N_v - 1) \ln 2 \ , \tag{4.24}$$

where N_v is the number of vertex operators $\mathcal{A}_v$ cut by the edge of the bipartition X [173, 225]. In a product state, such as $|\Omega_E\rangle$, we expect both quantities to be zero, while for generic values of h the entropy should interpolate between the two limits. To compute the entropy, we choose a subsystem X with 6 qubits, as depicted in Fig. 4.10(a), and we divide it into three further regions A, B, and C with two qubits each. We compute the entanglement entropy of all the subsets used in Eq. (4.23) by tracing out explicitly their complements and obtain the data plotted in Fig. 4.10(b). Despite the small dimension of the lattice and its subsystem, our results perfectly
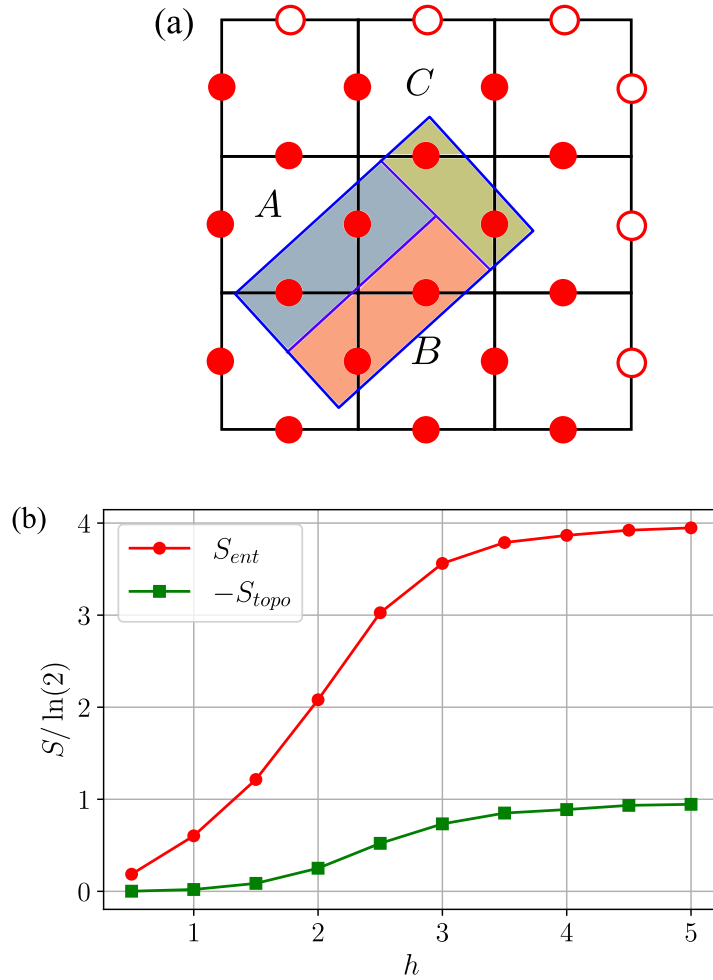
Figure 4.10.: (a): Graphical representation of the subsystem used to compute the topological and entanglement entropy, with the tripartition A, B, C highlighted. Notice that a total of $N_v = 5$ vertices are cut by the outer edge of $A \cup B \cup C$. Empty dots indicate the presence of PBC in the lattice, thus identifying the upper edge with the lower one and the right edge with the left one. (b): Entanglement and topological entropy as a function of the coupling h. Notice that we plot $-S_{topo}$ to make it positive.

agree with the theoretical prediction: in the deconfined phase, the total entanglement entropy is $S_{\text{ent}} = 4\ln 2$, since the partition ABC cuts five vertices and the topological entropy approaches $S_{\text{topo}} = -\ln 2$.

Finally, we would like to show that it is possible to manipulate the state constructed with QAOA to change its symmetry sector when the system has PBC. Let $|++\rangle_{\text{P}}$ denote the approximate ground state constructed with a QAOA circuit of P layers. We then construct approximate candidate ground states in the other topological sectors by applying non-contractible Wilson loops $\mathcal{W}_{v/h}$, i.e., $|+-\rangle_{\text{P}} = \mathcal{W}_h|++\rangle_{\text{P}}$, $|-+\rangle_{\text{P}} = \mathcal{W}_v|++\rangle_{\text{P}}$, $|--\rangle_{\text{P}} = \mathcal{W}_h\mathcal{W}_v|++\rangle_{\text{P}}$. The subscript label P is here used to distinguish the states obtained via QAOA from the exact eigenstates of the Hamiltonian. Non-contractible Wilson loops $\mathcal{W}_{v/h}$ are readily implemented via L single-qubit gates $\hat{\sigma}^z$ acting on a vertical or horizontal line. By doing so, however, we introduce an extra error on top of the finite accuracy of the QAOA state: indeed, the choice of a specific vertical or horizontal Wilson loop to change the symmetry sector of the system breaks the translational invariance of the constructed state, producing a small excitation. This effect is visible in Fig. 4.11, where we show the energies of the state approximated with QAOA, denoted by $|++\rangle_{\text{P}}$, and of the other three states obtained by applying $\mathcal{W}_h$ and $\mathcal{W}_v$ on $|++\rangle_{\text{P}}$. For comparison, we also plot the low eigenvalues obtained by exact diagonalization (drawn with solid blue lines). For large h, the four lowest energy levels should be almost degenerate, and, indeed, the exact diagonalization results are almost indistinguishable for $h \geq 4$. In the same region, the excess energy of the approximate states $|\tau_h, \tau_v\rangle$ is instead clearly visible, although well below the topological gap with the first proper excited state.

An alternative procedure to explore the different topological sectors in the deconfined regime, is to apply *first* the relevant Wilson loop on $|\Omega_B\rangle$ and then the QAOA unitaries. In such a way, the initial state $\mathcal{W}_{h/v}|\Omega_B\rangle$ is *exactly* degenerate with $|\Omega_B\rangle$. We find that the optimal schedule $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$ used to prepare the state $|++\rangle_{\text{P}}$ minimizes also the expectation value of the energy in the other topological sectors, so no further optimization is required. However, the picture presented in Fig. 4.11 remains valid, and small excitations are created in the other topological sectors. In other words, by inverting the order of application of the operators $\mathcal{W}_{h/v}$ and $\widehat{U}(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$, we observe nearly irrelevant changes in the expectation value of the energy; $\widehat{U}(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$ is the QAOA evolution operator with optimal parameters for the state $|++\rangle_{\text{P}}$.

The expectation value of the 't Hooft loops τ_h and τ_v, which distinguish the different topological sectors, is perfectly reconstructed by the algorithm. This last feature is, however, independent from the specific values of h and P, since the QAOA evolution respects the global $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry ($\tau_{h/v}$ commute with both $\widehat{H}_E$ and $\widehat{H}_B$) and $\tau_{h/v}$ always anticommute with $\mathcal{W}_{v/h}$.
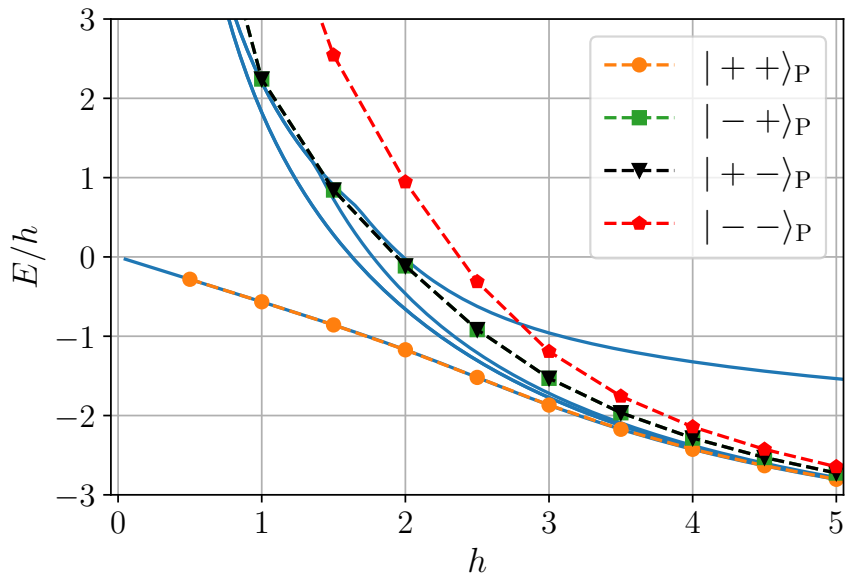
Figure 4.11.: The solid lines correspond to the 5 lowest eigenvalues obtained with exact diagonalization: The 5th eigenvalue, corresponding to the first excited state, is shown to highlight the topological gap in the deconfined phase. Only four lines are visible because $|+-\rangle$ and $|-+\rangle$ are exactly degenerate. The symbols indicate the four lowest energy states in different symmetry sectors, as approximated by QAOA. All data refer to the best solution found for $L = 3$, P $= 6$, and $|\psi_0\rangle = |\Omega_E\rangle$. States different from $|++\rangle_P$ are obtained by acting with non-contractible Wilson loops after the unitary evolution, a procedure yielding a small excitation energy. Note that also the approximate states $|+-\rangle_P$ and $|-+\rangle_P$ are degenerate.

4.4. Conclusion and outlook

In this Chapter, we studied the ground-state properties of a two-dimensional $\mathbb{Z}_2$ lattice gauge theory using a version of QAOA, leading to good-quality variational approximations while keeping circuits with a low depth. Hence, this allows us to prepare the target state with a number of standard single-qubit rotations and CNOT gates comparable with the realistic expectations for near-term quantum technologies.

We focused on the minimal resources needed for an accurate description of the ground state in a quantum circuit setup, to show that interesting physics can be observed despite the small system size. In particular, we showed that both the behavior of Wilson loops and entanglement entropy clearly distinguish the trivial and the topological phase, also characterizing the confinement/deconfinement transition. To reliably find good approximations of the ground state, we adopted a two-step protocol for QAOA, which produces regular optimal schedules that can be successfully transferred to larger sizes. In this respect, the two-step protocol outperforms a resource-costly global minimum search, as well as other local optimization strategies that are prone to remaining trapped in low-quality local minima.

However, the role of noise brought by measurements and gates has been neglected, even though it will inevitably appear in a realistic implementation of our proposal. In general, the effect of noise on VQAs is still an important open question [22], and error mitigation techniques, or even quantum error correction, could be integrated into our scheme in the near-term future. In the worst-case scenario, where noise prevents an accurate reconstruction of the cost function $E_\mathrm{P}(\boldsymbol{\gamma}, \boldsymbol{\beta})$ for optimization purposes in a large system, one might still use a "simulated" QAOA to infer good variational parameters for a smaller system in the same phase, which might be transferred to the quantum device for measuring physical properties or for a refinement optimization on the large system.

We emphasize that the QAOA technique we propose can be easily combined to extend several proposals for the study of 2D LGTs through digital quantum simulations [201–203, 217, 219, 226–229]. Digital quantum simulations of LGTs on small systems have already been implemented in trapped ion experiments [106, 230] and superconducting qubit platforms [231–234]. These experiments inspired several theoretical studies aimed at investigating the dynamics of the most important LGT excitations [222, 235–238]. Our results provide a tool to efficiently initialize the ground states of LGTs, which, in turn, make it possible to engineer in a controlled way several of the excited states studied to explore the dynamical and topological properties of LGTs, including, for example, flux excitations and mesons. The system we considered can be regarded as a surface code perturbed by onsite interactions that provide kinetic energy to its plaquette excitations [239, 240]. Hence, the study of its dynamics delivers information on the resilience of topological quantum memories in which anyons acquire a non-trivial dispersion. Furthermore, the topological order of

the $\mathbb{Z}_2$ LGT is the same as the most common topological quantum spin liquids, and our QAOA approach can be extended, for instance, to the study of quantum dimer models based on plaquette interactions, such as the Rokhsar Kivelson model [241], which displays this kind of topological phases and transitions on suitable lattices [198].

More in general, our variational quantum optimization successfully enables exploration of the properties of Hamiltonians with non-trivial four-body interactions, which represent not only an essential element for designing topological phases but also a useful tool for encoding classical optimization problems [199, 200, 242, 243]. Such interactions are compatible with the native geometry and qubit gate connectivity of several recently developed quantum computation platforms, encompassing both two-dimensional superconducting architectures, such as the Google Sycamore array [170, 173], and programmable arrays of Rydberg atoms [244–247]. By matching model interactions with the hardware connectivity (coupling map) of a physical platform, no additional overhead would be needed to map virtual into physical qubits and measurements in the computational basis would give direct information on the addressed models, as in the case of the $\mathbb{Z}_2$ LGT.

In conclusion, the combination of QAOA, initialization of the excitations, and digital quantum simulation of their time evolution opens the path to study many aspects of the dynamics of the confined and deconfined phases in LGTs as well as the anyonic excitations appearing in topologically ordered phases.

5. Conclusions

> *Do the best you can until you know better. Then when you know better, do better.*
>
> —Maya Angelou

In this Thesis, I have explored new applications of hybrid quantum-classical schemes that can be implemented on near-term quantum devices.

In Chapter 1, I briefly described the profoundly different approach of VQAs and near-term quantum computing, compared to fault-tolerant algorithms. I summarized the main algorithmic challenges for VQAs, focusing on trainability issues, namely local minimum traps and barren plateaus.

After a quick summary of Adiabatic Quantum Computation / Quantum Annealing, I switched to a gate-based framework and thoroughly reviewed the Quantum Approximate Optimization Algorithm (QAOA), a version of which has been implemented across all our case studies. Instead of examining the vast array of recent literature on QAOA applications, I concentrated on the principal theoretical findings and certain ubiquitous, yet not fully understood, phenomena. I specifically addressed landscape concentration, optimal parameter reusability across different instances of the same problem or from small to large systems, and patterns in QAOA optimal solutions. These concepts have been observed and generalized in the next Chapters, both for classical optimization problems beyond a 2-local Hamiltonian and for many-body ground state preparation.

In Chapter 2, I tackled the supervised learning of a binary perceptron for the highly idealized case of an unstructured dataset, as customary in the statistical physics community. This work showed promising evidence, yet is quite preliminary, in particular for what concerns its implementation on realist quantum hardware. Nonetheless, it has contributed to identifying a challenging research direction: designing problem-inspired variational quantum algorithms for classical optimization problems that involve multi-spin interactions, as illustrated in Eq. 1.12. A natural generalization of our work concerns the supervised learning of more general Binary Neural Networks, tested on a realistic dataset with a structure. Promising results have been showcased in Ref. [153], and we are currently working on this topic. The trade-off between trainability and gate decomposition is particularly subtle for this problem class, and it could represent a valid benchmark for many proposals of ansatz

design, see e.g. Ref. [248].

A related question regards the connection between the geometric structure of the *classical* cost function landscape — for instance, the occurrence of clusters of optimal strings proximate in Hamming distance — and the effectiveness of quantum optimization. This was investigated in Ref. [131] and in this Thesis, but I am not aware of other similar results. Introducing a structured training dataset seems the first relevant step forward in this direction. Moreover, analytical tools may yield formal results in the large N limit for QAOA on Binary Neural Networks, at least for unstructured datasets. One could envisage a generalization of analytical tools developed in Refs. [102, 103, 120], potentially proving concentration results or performance bounds in typical cases.

Next, I moved to ground state preparation of quantum many-body spin systems on a lattice. This is a crucial task for quantum simulation, and it could enable a direct study of quantum many-body systems beyond the reach of classical techniques. In this framework, the study of two-dimensional systems is particularly relevant, since standard tools like Matrix Product States are unsuccessful. An optimized strategy would be to tailor the details of an algorithmic implementation to the available hardware, employing a decomposition into native gates. Moreover, it is necessary to develop efficient schemes to train a parameterized Quantum Circuit (PQC), e.g. the Hamiltonian Variational ansatz that generalizes QAOA.

In Chapter 3, I described how to successfully tackle the ground state preparation for the one-dimensional Heisenberg XYZ model and the longitudinal-and-transverse field Ising model. We employ a problem-inspired ansatz that embodies some of the symmetries of the model, boosting its trainability. We circumvent local minimum traps and barren plateaus through the transferability of a specific class of optimal schedules. Indeed, our numerical results reveal that a smooth optimal solution — obtained through iterative interpolation techniques for a smaller system — serves as an excellent warm start for preparing the ground state of the identical model at the same point in the phase diagram, but for a system of considerably larger size. One may further refine the transferred solution via a local optimization, in a favorable landscape free of barren plateaus. These findings have been confirmed also for the simpler TFIM, by leveraging a free-fermions mapping and reaching larger system sizes.

Besides a natural extension of this scheme to other quantum many-body models, it would be fascinating to theoretically characterize this class of smooth solutions, elucidating potential relations with adiabaticity or shortcuts to adiabaticity. Contrarily to the QAOA for a binary perceptron, our proposal here can be directly tested on existing hardware. This involves classically simulating a small system and then transferring the solution to a quantum-classical setup, potentially beyond the limits of classical simulation capabilities.

In Chapter 4, I focused on a $\mathbb{Z}_2$ lattice gauge theory model, a problem of interest due to its implications in high-energy physics and its role in quantum error correction and surface codes. Despite small system sizes and a limited number of gates, allowing for a potentially straightforward implementation on NISQ devices, it was possible to detect the presence of a topological phase transition between confined and deconfined regimes. These techniques can be adapted to more complex lattice gauge theories or other models with topological order. Digital simulation of lattice gauge theory is a very active research field, and an ambitious goal is to simulate excited states and dynamical properties of increasingly realistic models with local symmetries and topological properties. During my PhD, I have only touched upon this interdisciplinary field. However, it appears an ideal playground to evaluate the effectiveness of VQAs and simultaneously test advanced physical theories.

From a broader perspective, the field of variational quantum computing is evolving rapidly, with significant challenges and best practices that transcend any single application area. Although not the focus of this Thesis, error mitigation and the first implementations of logical (error-corrected) qubits [23] are essential tools for the progress of near-term quantum computing. A major open question concerns the ultimate trainability of hybrid quantum-classical algorithms. The comparison with classical learning systems seems discouraging: Artificial Neural Networks feature billions (or trillions) of free parameters, which are routinely trained with standard algorithms, based on noisy versions of gradient descent. The identification of a benign overparameterization for VQAs, namely a regime characterized by a simple variational landscape, would constitute an undisputed leap forward for quantum computing. As discussed in Chapter 1, another crucial question relates to the classical resources that are necessary to simulate a VQA. The hybrid quantum-classical loop relies on iterative state preparations and measurements on the PQC: if the samples can be efficiently generated classically (*weak* simulability) there is no room for quantum speed up. The identification of new classes of wavefunctions that can be classically simulated seems to be a constitutional requirement for guiding ansatz design in VQAs.

On a practical level, it would be ideal if practical algorithmic proposals were to include, at a minimum, simulated shot noise and intermediate system sizes (e.g. $N > 14$ qubits). This would allow practitioners to seriously tackle trainability issues, such as the proliferation of local minimum traps and barren plateaus, that are hidden under the carpet for small-scale exact simulations. In particular, the computation of gradients of the VQA cost function should not be performed exactly, as in any realistic setup it would rely on samples and thus inherit a statistical noise, even in the idealized scenario of a noise-free quantum computer without coherent or incoherent errors.

A comprehensive analysis of QAOA theory and the development of a conceptual framework that synthesizes parameter concentration, the transferability of optimal

schedules, and patterns in optimal solutions remain incomplete. Specifically, exploring connections with adiabaticity and counter-diabatic effects could be advantageous. Interestingly, as shown in Chapter 3, these techniques are likely to prove beneficial also for quantum ground state preparation.

Finally, as duly noted in Ref. [66], the advent of VQAs has democratized the "quantum" research field. Nonetheless, the comparison with traditional quantum computing, even considering e.g. the basic Deutsch-Jozsa algorithm, seems both perplexing and somewhat embarrassing: We have moved from a meticulously crafted design, which delicately manipulates quantum information, to, literally, a bunch of quantum gates, with some variational parameters to be optimized. From a skeptical standpoint, encountering technical hurdles with this approach should not be overly surprising. Perhaps it would be beneficial to delve into the theory of quantum computing once again and understand how to manipulate qubits in a more structured fashion, while ensuring resource scaling that aligns with the capabilities of near-term quantum devices.

A. Algorithmics details

A.1. Interpolation algorithm (INTERP)

In this section, we describe the INTERP (interpolation) procedure, specifying technical information on our implementation for Chapter 3. The INTERP strategy, introduced in [74], is an algorithm devised for an iterative optimization of variational parameters of the cost function, originally applied in the context of QAOA for classical combinatorial optimization tasks. The INTERP strategy works as follows:

1. The optimization starts from a guess of the initial parameters at $P = P_{min}$ (e.g. $P_{min} = 1$), namely $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P=P_{min}}$.

2. We run a local optimization starting from $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P=P_{min}}$, using a classical local optimization routine, in order to minimize the cost function and obtain new optimized angles $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{opt}}_{P=P_{min}}$.

3. We run the following instructions on a loop up to $P = P_{max}$:

   a) Given the optimal parameters at step P, $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{opt}}_{P}$, we set the initial parameters at step $P + 1$, $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P+1}$, using the interpolation formula of Ref. [74] for $i = 1, 2, \ldots, P + 1$:

   $$\left[\boldsymbol{\beta}^{\text{start}}_{P+1}\right]_i = \frac{i-1}{P}\left[\boldsymbol{\beta}^{\text{opt}}_{P}\right]_{i-1} + \frac{P-i+1}{P}\left[\boldsymbol{\beta}^{\text{opt}}_{P}\right]_i,$$

   where $\boldsymbol{\beta}^{\text{opt}}_{P}$ is a P-dimension vector. Note that it is not required to define values of $\left[\boldsymbol{\beta}^{\text{opt}}_{P}\right]_0$ and $\left[\boldsymbol{\beta}^{\text{opt}}_{P}\right]_{P+1}$, since they are multiplied by null coefficients in the formula. The same rule applies to $\boldsymbol{\alpha}$ angles.

   b) We run a new local optimization starting from $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P+1}$, yielding a new set of angles $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{opt}}_{P+1}$.

   c) We increment the value of P by one unit: $P \rightarrow P + 1$.

As a visual support, in Fig. A.1 we sketch the INTERP procedure for a simple case, starting from the optimized parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{opt}}_{P=3}$, and finding (in sequence) the angles $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P=4}$, $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{opt}}_{P=4}$, $(\boldsymbol{\beta}, \boldsymbol{\alpha})^{\text{start}}_{P=5}$.

In practice, this is the specific version of INTERP we used in this paper, involving a single-unit increase of the value of P at each iteration, starting from $P = 1$.
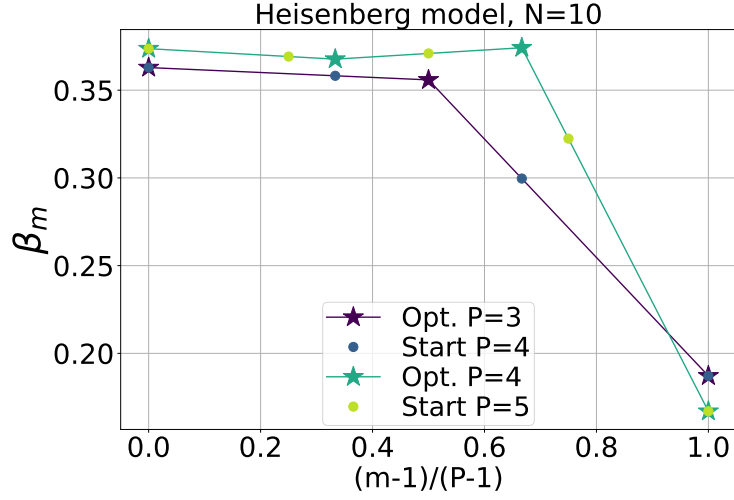
Figure A.1.: Visual example of the interpolation strategy, showing few initial steps of the iterative algorithm that is described in the main text.

Nevertheless, several minor modifications can be made to this scheme, and whole other iterative methods have also been developed [249]. Concerning the initial guess at P = 1, we always set $(\beta_1, \alpha_1)_{P=1}^{\text{start}} = (1/10, 1/10)$, as a starting point to run the first preliminary optimization. Albeit this choice might be arbitrary, in practice we verified that this preliminary optimization always converges to a well-defined minimum in the (P = 1) search space. Moreover, this minimum is close to the origin, which might provide a useful bias toward short total coherence times in the iterative construction of the smooth optimal curve.

The code for numerical simulations is written with *Qiskit* [35] (using as classical optimizer the L-BFGS-B algorithm [250]). We test INTERP algorithm by artificially fixing a maximum number of iterations for the classical optimizer: throughout Chapter 3 we set $N_{\text{iter}} = 100$, but our results are qualitatively unaffected by moderately reducing (or increasing) this value. This fixed maximum number of iterations sets an upper bound on the computational resources of the algorithm. In practice, by stopping the optimization loop, we find quasi-optimal schedules, which, however, are good enough to obtain almost-unit fidelity with the exact ground state.

B. Additional results on QAOA for the binary perceptron

In this Appendix, we include supplementary numerical results and technical details on Chapter 2. Let us begin with additional evidence on optimal-Δt digitized-QA, in particular by drawing a comparison between ordered and randomized samples. The randomization procedure and its geometrical significance are described in Chapter 2. Here we show that, also in the randomized scenario, the Δt-landscape and the position of minima are almost identical for all samples in exam. This is depicted in Fig. B.1: for conciseness, we plot data for P = 64 and both definitions of the cost function $n_c = 0, 1$, to be compared with Fig. 2.2 for the original samples.

As anticipated in Sec. 2.3.1, these results are qualitatively valid also for different values of P. This is illustrated in Fig B.2, where we plot the optimal values of Δt vs P for the original samples (left panel) and randomized samples (right panel). In the latter case, we notice that the sample-to-sample variability of the optimal values of Δt is even smaller, and also the differences — for any fixed value of P — between $n_c = 0, 1$ are negligible (especially for large values of P). Apparently, by scrambling the classical cost-function landscape geometry, the initial specification of the cost function becomes less relevant. In contrast, we remark that the optimal values of Δt differ significantly between any original sample and its randomized version.

Let us now remark on some details of the procedure outlined by Eq. (4.19). In practice, two slightly different procedures are adopted in order to single out smooth optimal QAOA-2 solutions. This choice is related to some qualitative differences between QAOA-1 results for $n_c = 0$ and $n_c = 1$, which are visible in Fig. 2.3 (empty symbols) for the first training set sample, but are present for all the samples in exam. Concisely, we observe that for $n_c = 0$ the QAOA-1 optimal parameters $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$ are noticeably different from $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$ for all values of P, and the high-frequency oscillations are either completely absent or well-localized on top of the smooth solutions. This observation motivates the original QAOA-2 procedure outlined in Eq. (4.19) (smoothing, second BFGS minimization), which is applied straightforwardly e.g. for $P = 16, 64$, yielding the smooth $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\gamma}^{(2)})$ protocols in Fig. 2.3 (solid curves).

On the contrary, for $n_c = 1$, we observe the same qualitative features e.g. for $P = 16$, whereas for larger values such as $P = 32, 64$ the QAOA-1 solutions seem
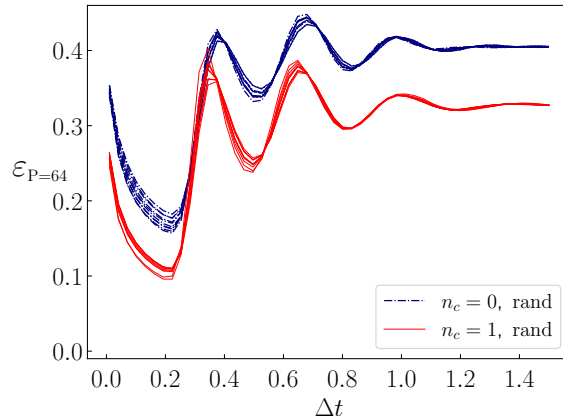
Figure B.1.: The one-dimensional landscape of the variational energy density (Eq. (2.6)) for digitized-QA, where the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ depend only on Δt, see Eq. (1.20). All *randomized* samples and both choices of $n_c = 0, 1$ are shown. We remark that the landscape, and in particular the position of global minima, show mild sample-to-sample variability.

to get "trapped" in a neighborhood of $(\boldsymbol{\beta}^{\mathrm{dQA}}, \boldsymbol{\gamma}^{\mathrm{dQA}})$, also displaying more extended high-frequency oscillations in the optimal parameters $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}^{(1)})$. This numerical evidence calls for a slightly different approach: we simply apply Eq. (4.19) prescription only for P = 16, and we find smooth solutions for larger values P$'$ using an iterative procedure: for each P$'$ > 16, we determine the new starting point for BFGS minimization by interpolating on the smooth optimal curve found for the previous value of P. We implement this procedure in power-of-two steps, hence from P = 16 $\rightarrow$ 32 $\rightarrow$ 64, but we expect to obtain similar results e.g. using a linear increment in P at each iteration. Consistently with our intuition, and as depicted in Fig. 2.4, QAOA-2 offers a noticeable improvement for $n_c = 1$, since the solutions for $P = 32, 64$ have now "escaped" the digitized-QA qualitative shape.

We remark that these details — concerning only the technical implementation of our QAOA-2 framework — do not affect our central message, as reported in the main text: for each sample in exam, QAOA hints at the presence of a smooth solution that systematically outperforms (optimal-Δt) digitized-QA, as shown in Fig. 2.4. This QAOA-1 solution is sometimes affected by the presence of high-frequency oscillations which can be smoothed out without spoiling the result: on the contrary, QAOA-2 is systematically (albeit sometimes negligibly) outperforming QAOA-1.

Let us remark that — in light of the discussion on transferability in Sec. 2.3.3 — the specific procedure adopted to obtain QAOA-2 solutions becomes less relevant: once a detailed study of a single sample is carried out, its optimized smooth solutions serve as an excellent warm start for all other randomly generated training sets, yielding an effective unique procedure to find smooth QAOA solutions outperforming optimal digitized-QA, valid for both $n_c = 0, 1$.
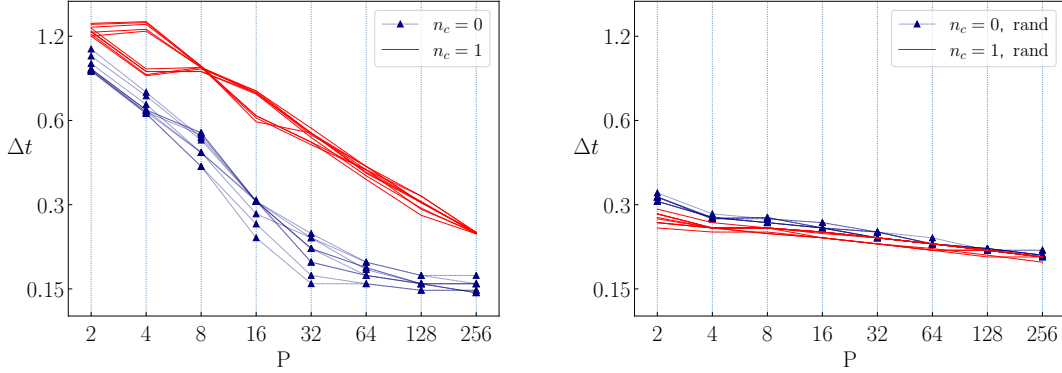
Figure B.2.: Optimal digitized-QA Δt values for increasing P, with logarithmic scales on both axes. The original samples (left) are compared to the randomized samples (right). The presence of evident clusters shows mild sample-to-sample variability, with few exceptions (as P = 4 for $n_c = 1$, in the left panel) due to an almost-flat energy landscape in that range.

Concerning our study on randomized samples, we proceeded with the same iterative interpolation strategy starting from $P = 16$, for both $n_c = 0, 1$. Once a smooth QAOA-2 solution is obtained for the first sample, the transferability of the ansatz yields a well-defined strategy to apply QAOA on all the other randomized samples.
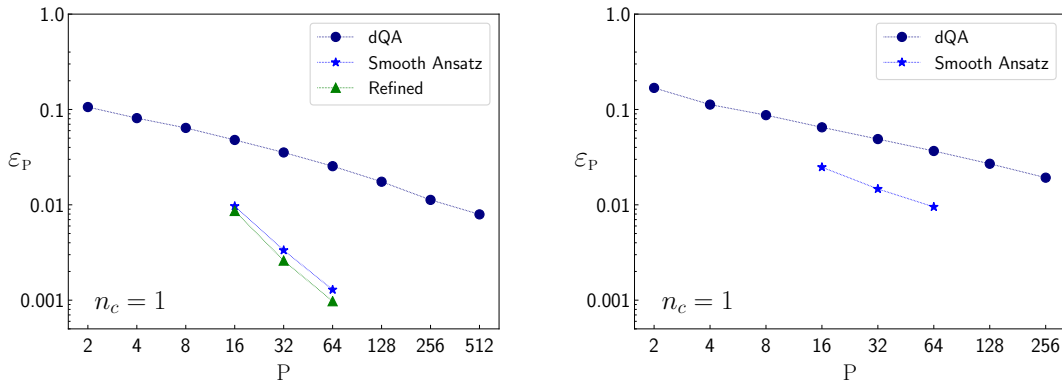
Figure B.3.: (Left panel) Variational energy density averaged over the same set of samples analyzed in Fig. 2.5. We plot data for optimal dQA, and for the transferred *smooth ansatz*, before and after a refinement optimization. (Right panel) Variational energy density averaged over a new batch of 50 random samples. The transferred *smooth ansatz* still outperforms an optimal dQA. The same results are found for $n_c = 0$ (data not shown).

In conclusion, we briefly discuss some additional evidence on transferability. In Fig. B.3 (left panel), we plot the variational energy density obtained with an optimal-Δt dQA, compared to the one of the *smooth ansatz*, before and after a refinement optimization. These data correspond to the protocols shown in Fig. 2.5: here, we show results averaged over the training set samples for $n_c = 1$. Remarkably, the transferred solution significantly outperforms an optimal dQA, even without re-

optimizing the QAOA variational parameters for the new training set. The same results are valid for $n_c = 0$. As depicted in Fig. 2.5, for $n_c = 1$ the re-optimized parameters are almost indistinguishable from the *ansatz*: a new BFGS-optimization provides only a minor increase in performance.

The effectiveness of a transferred QAOA solution is confirmed on a new batch of 50 training sets, generated randomly without any a posteriori selection on the number of zero-energy solutions or the hardness for a classical optimization. This is shown in Fig. B.3 (right panel): also here, a transferred *smooth ansatz* outperforms the optimal dQA, even without a further optimization. The same results are verified for $n_c = 0$. On a side-note, we notice that the average variational energy obtained with the *smooth ansatz* for the original 10 samples is lower than the values obtained here for the new simulations: this may be due to the selection of samples with a large number of classical solutions, on which quantum methods are expected to perform particularly well [131].

C. Additional results on ground state preparation of a Lattice Gauge Theory

C.1. Realization of the plaquette rotation in the 3×3 torus

When considering a system with periodic boundary conditions and an odd number of rows and columns, some further details must be considered when implementing the two-plaquette rotation, by modifying the picture shown in Fig. 4.4

In Fig. C.1 we depict a circuit that applies to the case of the 3×3 system considered throughout most of the paper. To this purpose, we consider a stripe of three plaquettes, as shown in panel (a). The operator $\mathrm{e}^{i\beta \mathcal{B}_p}$ is applied to all three plaquettes through the quantum circuit displayed in panel (b), which has depth 17.

When considering larger systems with periodic boundaries and an odd number of rows and columns, a suitable combination of the schemes presented in Fig. 4.4 and C.1 allows us to perform each QAOA step with a circuit of depth 18, involving only CNOTs between neighboring qubits and single-qubit rotations.

C.2. Energy landscape

In the $\mathbb{Z}_2$ LGT model, the energy landscape associated to the QAOA ansatz is characterized by the presence of many local minima, as discussed in Sec. 4.3.1. This makes the employment of a clever optimization strategy, such as the two-step protocol or schedule transferability, a necessity to target reliably low energy minima. However, for general variational problems, there might exist a deep minimum in the energy landscape, associated with a state with small or no overlap with the target one. This is not the case for the problem under investigation, where there is a clear correlation between the fidelity and the variational energy for the minima in the energy landscape, see Fig.4.5 in the main text.
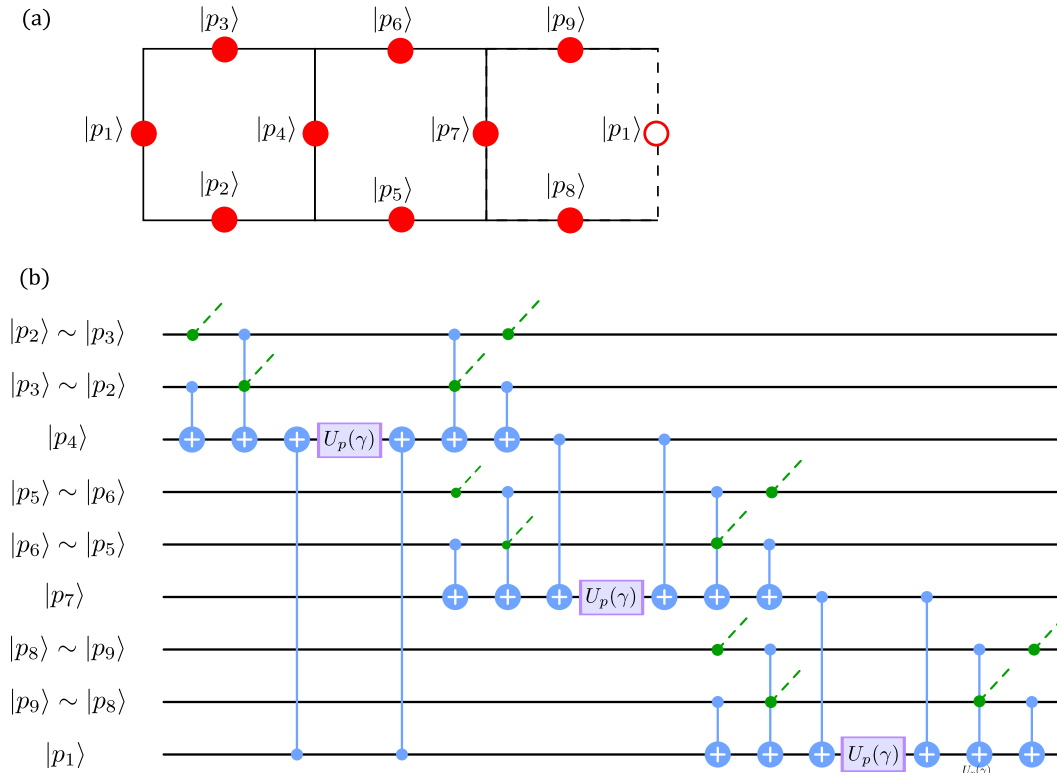
Figure C.1.: (a) A stripe of the 3×3 lattice with periodic boundary conditions (b) Quantum circuit to implement the plaquette operator $\mathrm{e}^{i\gamma\mathcal{B}}$ on all three plaquettes. The labeling of the qubit lines emphasizes that all the boundary qubits are shared with the plaquette stripes above and below. The partially depicted green CNOTS are related to the simultaneous implementation of the same algorithm on the neighboring stripes: qubits 2,3,5,6,8,9 act as controls also for the circuit in neighboring plaquettes.

Here, we show that this correlation holds also for the larger systems considered in this paper, $L = 4$ and $L = 5$, corresponding to 32 and 50 qubits respectively. We repeat the analysis of Sec. 4.3.1: focusing on $|\psi_0\rangle = |\Omega_E\rangle$, we perform 100 QAOA runs with random initial parameters, targeting states in the deconfined phase $h \geq 3$. In Fig. C.2 we plot the infidelity $1 - F_P$ vs the residual energy of the minima found with this procedure, for both $L = 4$ (panel (a)) and $L = 5$ (panel (b)) and a circuit depth of P = 6. Interestingly, it appears that increasing the system size leads to a sharper correlation between energy and fidelity, compared to the data presented in Fig. 4.5.

C.3. Details on the QAOA implementation

In the following, we discuss additional details regarding our two-step implementation of QAOA for the problem under investigation. Firstly, we focus on a benchmark of our heuristic approach against a global minimum search, which, remarkably, yields similar-quality results for both phases, in terms of ground-state fidelity, offering a convincing numerical validation of our scheme. The global search is performed with the basin hopping method [251] from the SciPy Python library. Secondly, we comment on the transferability of the optimal schedules, obtained by either a two-step optimization or basin hopping, to larger system sizes, a strategy that would provide an educated guess to lower the computational cost for a new optimization. Finally, we observe some patterns for optimal QAOA variational parameters obtained with the two-step scheme, in particular their smoothness as a function of $m = 1 \cdots P$, similar to results showcased in Chapter 2 and Chapter 3.

C.3.1. Global optimization vs two-step scheme

In order to prove the effectiveness of the two-step optimization protocol, we compare it with a global minimum search, based on the basin hopping method. In this case, we run up to 500 local minimizations, each of them initialized in the proximity of a previously found local minimum. The parameter space is explored with an effective temperature chosen to allow jumps between typical low-energy minima. To reliably determine the absolute minimum we run the basin hopping optimization 100 times and select the best result.

Figure C.3 shows a comparison between the fidelity obtained with the global and the two-step optimization procedures, as a function of h for fixed P = 6. The initial state is $|\psi_0\rangle = |\Omega_E\rangle$. For $h < 3$ the two-step approach matches the global optimization performance and it yields the same results, while for $h \geq 3$ it finds a sub-optimal local minimum. However, we stress that, even in this case, the final fidelity is almost one: the difference in the accuracy between the two methods is
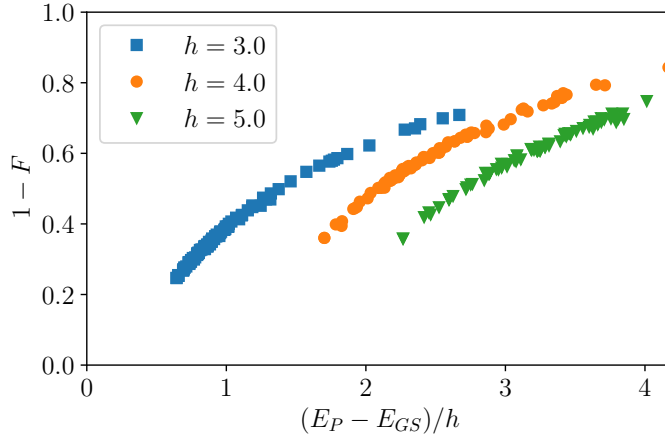
(a)
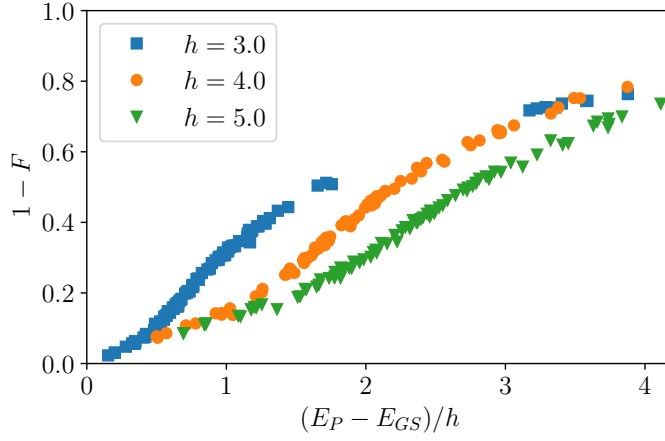
(b)

Figure C.2.: Infidelity vs residual energy for 100 minima in the energy landscape found with random initialization of local BFGS searches, with $|\psi_0\rangle = |\Omega_E\rangle$ and P = 6. Panel (a) corresponds to the lattice with 4×4 plaquettes, panel (b) to 5×5 plaquettes. We only show data that falls in the interval $(E_P - E_{GS})/h \in [0, 4.2]$.
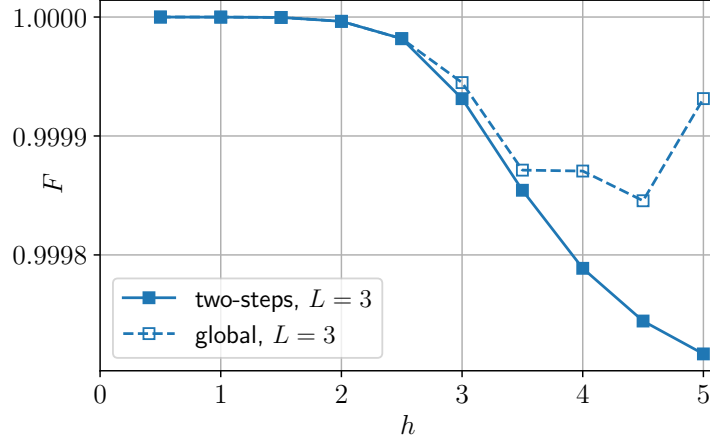
Figure C.3.: Comparison of the accuracy with respect to the magnetic coupling h between global optimization (empty squares) and the two-step approach (full squares) described in the text. The data refer to $P = 6$ and $|\psi_0\rangle = |\Omega_E\rangle$.

much lower than a realistic experimental resolution. Moreover, the two-step protocol has the clear advantage of requiring only a single local optimization — on top of a modest computational overhead for the one-dimensional optimal Δt grid search — to be compared with 500×100 local optimizations for the basin hopping method. Consequently, the two-step heuristics certainly requires drastically fewer function evaluations and it is, therefore, a better candidate to be implemented on a realistic quantum device and also much faster to simulate on a classical computer.

Concerning the transferability of the optimal schedules to larger system sizes, we use the optimal angles obtained for $L = 3$, either with basin hopping or with a two-step optimization, as an initial guess for local optimizations of larger $L = 4, 5$. Specifically, for each value of h, we compare the best fidelity out of 10 BFGS local search runs, each of them initialized with the optimal 2P parameters previously found for $L = 3$, plus a small noise to facilitate the exploration of the energy landscape. Our results are reported in Fig. C.4(a), where we compare local minimizations starting from the $L = 3$ two-step optimal schedules (star symbols), local searches starting from the $L = 3$ global minimum (circles), and the two-step process applied directly on the larger system (squares). We find that the optimal angles returned by the two-step algorithm provide a better guess for larger systems, resulting in higher fidelity than a local search initialized with the global minimum for $L = 3$. This fact is linked to the existence of some patterns in the optimal parameters found with the two-step scheme, in particular, their smoothness as a function of $m = 1 \cdots P$, consistently with the results presented in Chapter 3. The performance of the two-step optimization applied directly on the target system $L = 4$ or $L = 5$ is instead comparable with the transfer of the schedule from $L = 3$, although the latter is slightly better for large magnetic fields.
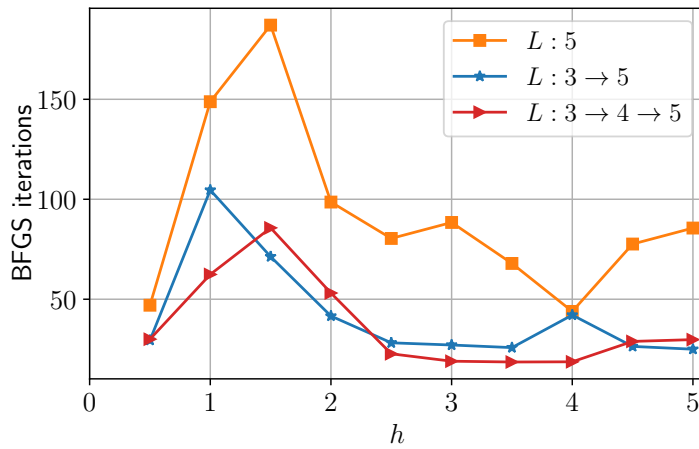
(a)
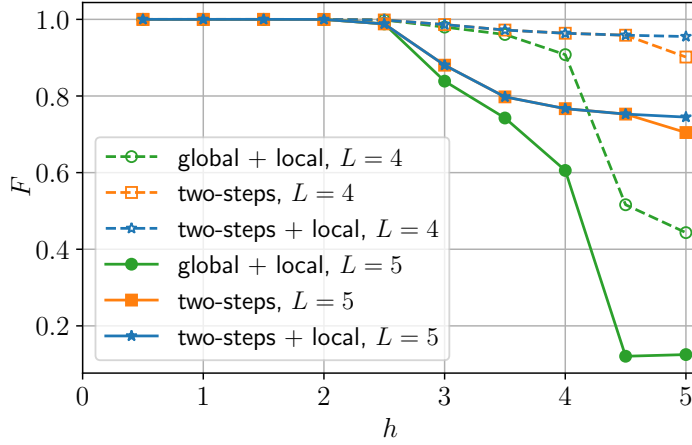
(b)

Figure C.4.: (a) Comparison of the two-step vs global schedules for $L = 3$ as an educated guess for optimization on larger system sizes $L = 4, 5$. The data refer to the best out of 10 local BFGS minimizations, run by starting close to the optimal two-step schedule (blue stars) or optimal global schedule (green circles). The plot also shows the fidelity of the optimal two-step schedule (orange squares) run for $L = 4, 5$. (b) Number of iterations required for the convergence of the final BFGS optimization; comparison between transferring the schedule from $L = 3$ to $L = 5$ (blue stars), transferring from $L = 4$ to $L = 5$ (red triangles), and two-step optimization directly on $L = 5$ (orange squares). All data refer to P $= 6$ and $|\psi_0\rangle = |\Omega_E\rangle$.
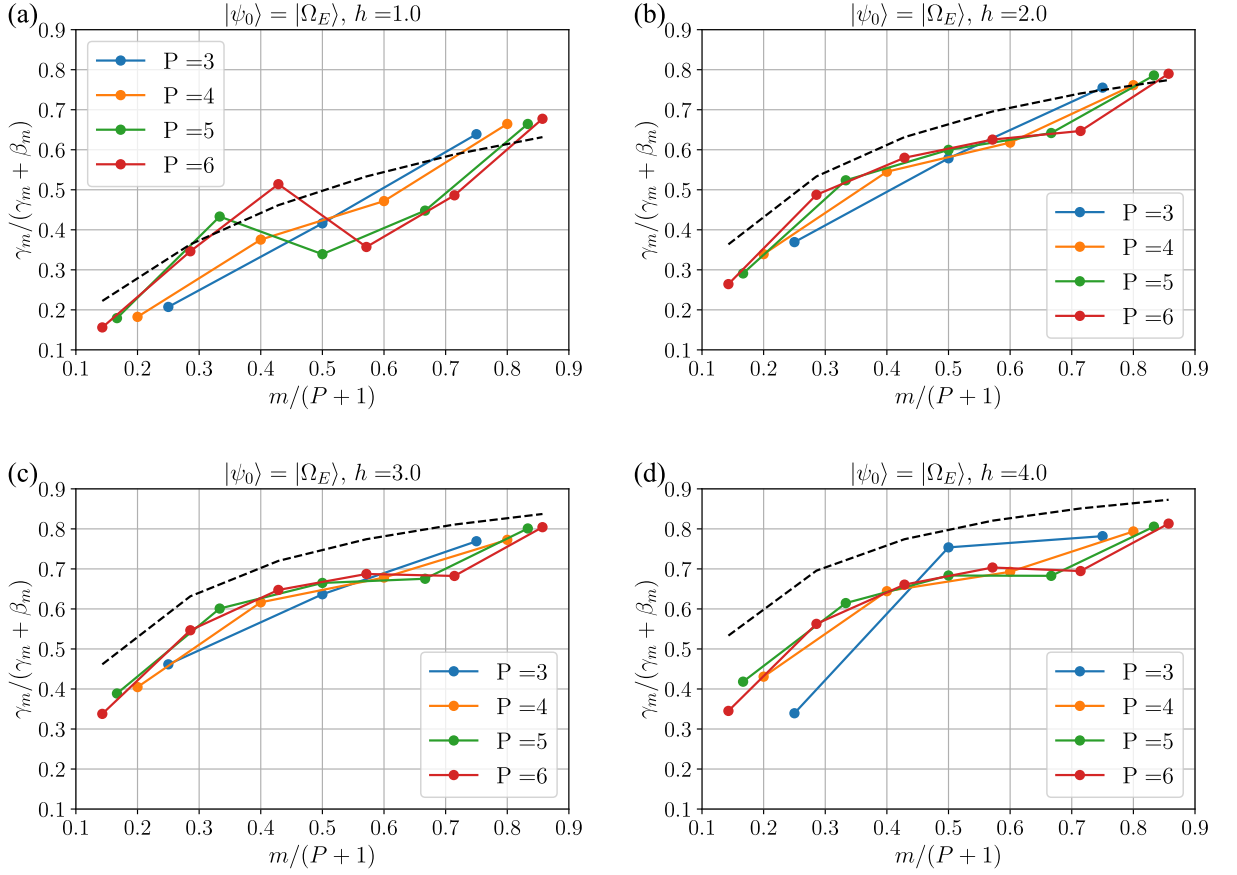
Figure C.5.: Variational parameters associated with the best result out of 10 local minimum searches, as prescribed by the two-step optimization procedure. In each panel, the black dashed line corresponds, through Eq. (C.5), to the linear annealing schedule defined by Eq.(4.18), with the corresponding value of h and P = 6 steps. The quantum circuit is initially prepared in the electric ground state $|\Omega_E\rangle$.

However, once the optimal schedule for a given system size is known, it is convenient to leverage that result to initialize the QAOA search for larger systems, instead of running a new two-step optimization from scratch. In fact, although the performance in terms of final fidelity is similar, the schedule transfer requires fewer iterations than the two-step optimization. This is shown in Fig. C.4(b), where we compare the number of BFGS iterations required in the final local minimum search on a system with $L = 5$ for different optimization strategies: schedule transferring from $L = 3$ to $L = 5$, from $L = 4$ to $L = 5$, and the two-steps protocol directly on $L = 5$. The latter requires in general a larger number of iterations and its overall cost must be added to the resources required for the optimization of the time step Δt.

C.3.2. Smooth schedules

For the $\mathbb{Z}_2$ LGT model, we found that the two-step optimization scheme produces smooth protocols for the optimal variational parameters more easily than other heuristic methods present in the literature, such as the application of iterative schemes based on parameter interpolation or Fourier component optimization [74, 75]. Moreover, it provides a minimum $(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^*)$ for a chosen circuit depth P without requiring the solution for shallower circuits with $\mathrm{P}' < \mathrm{P}$, contrarily to both the iterative approaches just mentioned. However, the effectiveness of any similar heuristic method seems model-dependent: for instance, in Chapter 3, the iterative interpolation scheme resulted in a winning strategy.

The presence of regular patterns in the optimal parameters suggests a comparison with a dQA scheme, such as the ones adopted to initialize the two-step QAOA. In Sec. 4.3.2 we defined two possible annealing protocols, depending on the choice of the initial state. If $|\psi_0\rangle = |\Omega_E\rangle$, we construct the time-dependent Hamiltonian

$$\widehat{H}(t) = \widehat{H}_E + h\frac{t}{\tau}\widehat{H}_B \,, \tag{C.1}$$

while if $|\psi_0\rangle = |\Omega_B\rangle$ we use

$$\widehat{H}(t) = h\left(\frac{t}{h\tau}\widehat{H}_E + \widehat{H}_B\right) \,. \tag{C.2}$$

In both cases, $t \in [0, \tau]$ and at the end of the protocol $\widehat{H}(t = \tau) = \widehat{H}_{\text{target}}$. The corresponding parameters γ_m and β_m of a dQA are reported in Eq. (4.18) and Eq. (4.21), respectively.

For a graphical representation of smooth optimal two-step schedules and a direct comparison with dQA, it is useful to consider the following more general protocol, as customary in AQC:

$$\widehat{H}(s) = (1 - s)\widehat{H}_E + s\widehat{H}_B \,, \tag{C.3}$$

where $s(t) \in [0, 1]$ is a time-dependent parameter that interpolates between $\widehat{H}_E$ and $\widehat{H}_B$. Note that this time-dependent Hamiltonian is in the same form introduced in Chapter 1. With this notation, we can identify Eq. (C.1) with a process starting from $s(0) = 0$ and ending in $s(\tau) = s_{\text{f}}$, with $s_{\text{f}} = \frac{h}{h+1}$; Eq. (C.2), instead, corresponds to a process with $s(0) = 1$, ending again in $s(\tau) = s_{\text{f}}$ (both identifications are valid modulo an overall multiplicative factor).

As described in Chapter 1, dQA requires a discretization of the time interval $[0, \tau]$ into P small time steps Δt_m, such that $\sum_m \Delta t_m = \tau$. Correspondingly, the continuous schedule $s(t)$ is discretized into a sequence of short-time evolutions generated by $\widehat{H}(s_m)$, where

$$s_m = s_0 + (s_{\text{f}} - s_0)\frac{m}{\mathrm{P}} \,, \tag{C.4}$$

with $m = 1 \cdots \mathrm{P}$. The resulting expression can be further simplified with a first order Trotter split-up, neglecting quadratic terms in Δt_m. Thus, the final state is written as the variational ansatz in Eq. (4.13), with *fixed* parameters given by

$$
\begin{cases}
s_m = \dfrac{\gamma_m}{\gamma_m + \beta_m} \ , \\
\Delta t_m = \gamma_m + \beta_m \ ,
\end{cases}
\tag{C.5}
$$

with $\hbar = 1$. Along the same lines of Sec. 1.3.4, once we have found optimal smooth QAOA parameters $\boldsymbol{\gamma}^\star$, $\boldsymbol{\beta}^\star$ with our two-step QAOA scheme discussed in Sec. 4.3.2, we can extract the corresponding digitized schedule s_m^* and compare it with the linear dQA protocol s_m^{dQA}, which we used as an educated guess for the local minimization.

As examples of typical smooth QAOA optimal parameters, we report in Fig. C.5(a)-(d) the schedules s_m^* corresponding to four different values of the coupling h, both below and above the "topological transition", with $\mathrm{P} \geq 3$ and initial state $|\psi_0\rangle = |\Omega_E\rangle$. The dashed black lines correspond to the linear annealing schedule of Eq.(4.18) that we used as a starting point for the local minimizations, with $\mathrm{P} = 6$. In all four cases, it appears evident that as P increases, the parameters gradually approach a smooth continuous behavior, with the possible exception of a single localized irregularity, which seems to appear in Fig. C.5(a) for $\mathrm{P} = 5, 6$. This is not surprising, however, since we are preparing a state very close to the initial one. Thus, a large value of P could "overfit" the target state, and many different parameter choices, usually non-smooth, could yield similar accuracy. A comparison with Fig. 4.6 for the case $h = 1$, clearly shows a degradation of performance (almost-flat curve) of the infidelity vs P, exactly for $\mathrm{P} = 5, 6$: this irregularity can thus be interpreted as a local lower-quality minimum or a saturation of the numerical precision of the algorithm. For larger values of the coupling h, instead, we observe a clear continuity in the optimal schedule s_m^*, as we change both P and h. This leads to the interesting consequence that the optimal schedule for a given $\widehat{H}_{\mathrm{target}}$ and circuit depth P could be used as a seed to initialize the optimization for different values of h, requiring only a small fine-tuning of the parameters to adapt the schedule to the new target ground state.

A similarly smooth pattern is observed when we initialize the system in the magnetic ground state $|\Omega_B\rangle$, as reported in Fig. C.6(a)-(d). The dashed black lines correspond here to the schedules s_m^{dQA} extracted from Eq.(4.21), with $\mathrm{P} = 6$. The main difference is that the smoothness now is more easily lost when targeting the deconfined phase, see panel(d), which is closer to the initial state. Similar comments on this irregularity apply as to the previous case, by comparing with Fig. 4.7(a). On a side note, we notice that the evident irregularity in panel(d) for $\mathrm{P} = 5, 6$ involves a single point with a numerical value smaller than 0.4: this is not a significant feature, and it could easily be eliminated by an appropriate smoothing procedure with a likely improvement in performance.
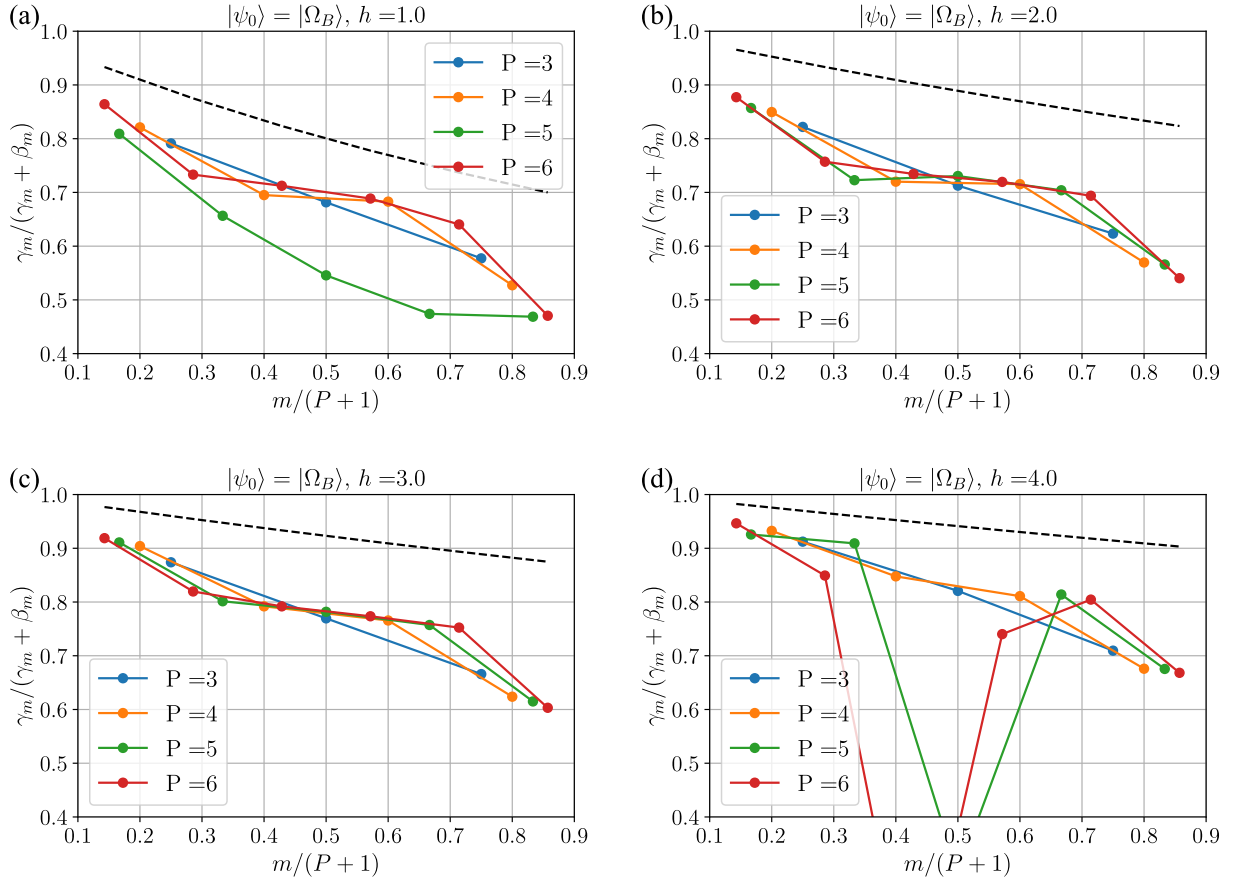
Figure C.6.: Variational parameters associated with the best result out of 10 local minimum searches, as prescribed by the two-step optimization procedure. In each panel, the black dashed line corresponds, through Eq. (C.5), to the linear annealing schedule defined by Eq.(4.21), with the corresponding value of h, and P = 6 steps. The quantum circuit is initially prepared in the magnetic ground state $|\Omega_B\rangle$.

Unsurprisingly, the two-step optimization might get trapped in a (high-quality) local minimum even when we target the opposite phase: this is seen, e.g., for the outlier set of $h = 1$ and P $= 5$ in Fig. C.6(a), which might be associated to a suboptimal minimum. This observation is once again consistent with the corresponding data in Fig. 4.7(a), where the curve for $h = 1$ shows a small spike in correspondence to P $= 5$.

Regarding the comparison with the linear dQA protocol s_m^{dQA} (dashed black lines), in both Figs. C.5 and C.6, the overall monotonicity of optimized s_m^* is the same of the original schedule, i.e., an increasing function of m when $|\psi_0\rangle = |\Omega_E\rangle$, and a decreasing function when $|\psi_0\rangle = |\Omega_B\rangle$. However, when targeting states in a phase that differs from the initial one, the optimal schedule deviates more and more from the original ansatz, highlighting the importance of the local optimization of the parameters.

Bibliography

[1] P. Torta, G. B. Mbeng, C. Baldassi, R. Zecchina, and G. E. Santoro, *Quantum approximate optimization algorithm applied to the binary perceptron*, Phys. Rev. B **107**, 094202 (2023).

[2] A. A. Mele, G. B. Mbeng, G. E. Santoro, M. Collura, and P. Torta, *Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz*, Phys. Rev. A **106**, L060401 (2022).

[3] L. Lumia, P. Torta, G. B. Mbeng, G. E. Santoro, E. Ercolessi, M. Burrello, and M. M. Wauters, *Two-Dimensional $\mathbb{Z}_2$ Lattice Gauge Theory on a Near-Term Quantum Simulator: Variational Quantum Optimization, Confinement, and Topological Order*, PRX Quantum **3**, 020320 (2022).

[4] G. Lami, P. Torta, G. E. Santoro, and M. Collura, *Quantum annealing for neural network optimization problems: A new approach via tensor network simulations*, SciPost Phys. **14**, 117 (2023).

[5] M. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[6] J. Preskill, *Quantum Computing in the NISQ era and beyond*, Quantum **2**, 79 (2018).

[7] J. Preskill, Quantum computing 40 years later, (2023), arXiv:2106.10522 [quant-ph] .

[8] R. P. Feynman, *Simulating physics with computers*, Int J Theor Phys **21**, 467 (1982).

[9] S. Lloyd, *Universal Quantum Simulators*, Science **273**, 1073 (1996).

[10] D. Deutsch, *Quantum theory, the church-turing principle and the universal quantum computer*, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences (1985).

[11] P. W. Shor, *Polynomial-time algorithm for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comp. **26**, 1484 (1997).

[12] P. W. Shor, *Scheme for reducing decoherence in quantum computer memory,* Phys. Rev. A **52**, R2493 (1995).

[13] A. M. Steane, *Error Correcting Codes in Quantum Theory,* Phys. Rev. Lett. **77**, 793 (1996).

[14] P. Shor, *Fault-tolerant quantum computation,* in *Proceedings of 37th Conference on Foundations of Computer Science* (1996) pp. 56–65.

[15] A. W. Harrow, A. Hassidim, and S. Lloyd, *Quantum Algorithm for Linear Systems of Equations,* Phys. Rev. Lett. **103**, 150502 (2009).

[16] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, *Quantum computational chemistry,* Rev. Mod. Phys. **92**, 015003 (2020).

[17] D. Ferraro, M. Campisi, G. M. Andolina, V. Pellegrini, and M. Polini, *High-Power Collective Charging of a Solid-State Quantum Battery,* Phys. Rev. Lett. **120**, 117702 (2018).

[18] L. K. Grover, *Quantum mechanics helps in searching for a needle in a haystack,* Phys. Rev. Lett. **79**, 325 (1997).

[19] A. Acin, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm, *The quantum technologies roadmap: a European community view,* New Journal of Physics **20**, 080201 (2018).

[20] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, *Defining and detecting quantum speedup,* Science **345**, 420 (2014).

[21] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, *Noisy intermediate-scale quantum algorithms,* Rev. Mod. Phys. **94**, 015004 (2022).

[22] M. Cerezo, A. Arrasmith, R. Babbush, S. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. Coles, *Variational Quantum Algorithms,* Nature Reviews Physics (2021).

[23] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin, *Logical quantum processor based on reconfigurable atom arrays,* Nature **626**, 58 (2023).

[24] G. E. Santoro, *Introduction to Quantum Computation and Information – Lecture Notes* (2023).

[25] S. Olivares, *Lecture Notes on quantum computing* (2021).

[26] W. W. Ho and T. H. Hsieh, *Efficient variational simulation of non-trivial quantum states*, SciPost Phys. **6**, 29 (2019).

[27] D. Wierichs, C. Gogolin, and M. Kastoryano, *Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer*, Phys. Rev. Research **2**, 043246 (2020).

[28] E. Farhi, J. Goldstone, and S. Gutmann, A Quantum Approximate Optimization Algorithm, (2014), arXiv:1411.4028 [quant-ph] .

[29] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles, Variational Quantum Linear Solver, (2019), arXiv:1909.05820 [quant-ph] .

[30] K. C. Tan and T. Volkoff, *Variational quantum algorithms to estimate rank, quantum entropies, fidelity, and Fisher information via purity minimization*, Phys. Rev. Research **3**, 033251 (2021).

[31] E. Tang, *Dequantizing algorithms to understand quantum advantage in machine learning*, Nature Reviews Physics **4**, 692 (2022).

[32] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *A variational eigenvalue solver on a photonic quantum processor*, Nature Communications **5**, 4213 (2014).

[33] S. Hadfield, Z. Wang, B. O'Gorman, E. Rieffel, D. Venturelli, and R. Biswas, *From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz*, Algorithms **12**, 34 (2019).

[34] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, *Cost function dependent barren plateaus in shallow parametrized quantum circuits*, Nature Communications **12** (2021).

[35] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim, D. Bucher, F. J. Cabrera-Hernández, J. Carballo-Franquis, A. Chen, C.-F. Chen, and et al., *Qiskit: An Open-source Framework for Quantum Computing*, (2019).

[36] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, PennyLane: Automatic differentiation of hybrid quantum-classical computations, (2020), arXiv:1811.04968 [quant-ph] .

[37] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Evaluating analytic gradients on quantum hardware,* Physical Review A **99** (2019).

[38] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, *General parameter-shift rules for quantum gradients,* Quantum **6**, 677 (2022).

[39] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, *Quantum Natural Gradient,* Quantum **4**, 269 (2020).

[40] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, *Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information,* Quantum **5**, 567 (2021).

[41] M. V. den Nest, Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond, (2009), arXiv:0811.0898 [quant-ph] .

[42] J. Biamonte and V. Bergholm, Tensor Networks in a Nutshell, (2017), arXiv:1708.00006 [quant-ph] .

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

[44] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, *Evidence for the utility of quantum computing before fault tolerance,* Nature **618**, 500 (2023).

[45] T. Begušić and G. K.-L. Chan, Fast classical simulation of evidence for the utility of quantum computing before fault tolerance, (2023), arXiv:2306.16372 [quant-ph] .

[46] D. Failde, J. D. Viqueira, M. Mussa Juane, and A. Gomez, *Using Differential Evolution to avoid local minima in Variational Quantum Algorithms,* Scientific Reports **13** (2023).

[47] E. R. Anschuetz and B. T. Kiani, *Quantum variational algorithms are swamped with traps,* Nature Communications **13** (2022).

[48] L. Bittel and M. Kliesch, *Training Variational Quantum Algorithms Is NP-Hard,* Phys. Rev. Lett. **127**, 120502 (2021).

[49] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, *Barren plateaus in quantum neural network training landscapes,* Nature Communications **9** (2018).

[50] A. W. Harrow and R. A. Low, *Random Quantum Circuits are Approximate 2-designs,* Communications in Mathematical Physics **291**, 257 (2009).

[51] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Exact and approximate unitary 2-designs and their application to fidelity estimation,* Physical Review A **80** (2009).

[52] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: On the structure of unitary designs,* Journal of Mathematical Physics **48**, 052104 (2007).

[53] M. Kliesch and I. Roth, *Theory of Quantum System Certification,* PRX Quantum **2** (2021).

[54] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, *Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus,* PRX Quantum **3**, 010313 (2022).

[55] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, *Diagnosing Barren Plateaus with Tools from Quantum Optimal Control,* Quantum **6**, 824 (2022).

[56] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, *Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz,* PRX Quantum **1**, 020319 (2020).

[57] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, *Exploiting Symmetry in Variational Quantum Machine Learning,* PRX Quantum **4** (2023).

[58] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, *Entanglement-Induced Barren Plateaus,* PRX Quantum **2**, 040316 (2021).

[59] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, *Noise-induced barren plateaus in variational quantum algorithms,* Nature Communications **12** (2021).

[60] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, *Equivalence of quantum barren plateaus to cost concentration and narrow gorges,* Quantum Science and Technology **7**, 045015 (2022).

[61] M. Cerezo and P. J. Coles, *Higher order derivatives of quantum neural networks with barren plateaus,* Quantum Science and Technology **6**, 035006 (2021).

[62] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, *Effect of barren plateaus on gradient-free optimization,* Quantum **5**, 558 (2021).

[63] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A Unified Theory of Barren Plateaus for Deep Parametrized Quantum Circuits, (2023), arXiv:2309.09342 [quant-ph] .

[64] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, The Adjoint Is All You Need: Characterizing Barren Plateaus in Quantum Ansätze, (2023), arXiv:2309.07902 [quant-ph] .

[65] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina, *Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes,* Proceedings of the National Academy of Sciences **113**, E7655 (2016).

[66] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing, (2023), arXiv:2312.09121 [quant-ph] .

[67] J. Dborin, F. Barratt, V. Wimalaweera, L. Wright, and A. G. Green, *Matrix product state pre-training for quantum machine learning,* Quantum Science and Technology **7**, 035014 (2022).

[68] M. S. Rudolph, J. Miller, D. Motlagh, J. Chen, A. Acharya, and A. Perdomo-Ortiz, *Synergistic pretraining of parametrized quantum circuits via tensor networks,* Nature Communications **14**, 8367 (2023).

[69] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, *Layerwise learning for quantum neural networks,* Quantum Machine Intelligence **3**, 5 (2021).

[70] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, *An initialization strategy for addressing barren plateaus in parametrized quantum circuits,* Quantum **3**, 214 (2019).

[71] S. H. Sack, R. A. Medina, A. A. Michailidis, R. Kueng, and M. Serbyn, *Avoiding Barren Plateaus Using Classical Shadows,* PRX Quantum **3**, 020365 (2022).

[72] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, *Absence of Barren Plateaus in Quantum Convolutional Neural Networks,* Physical Review X **11** (2021).

[73] E. Cervero Martín, K. Plekhanov, and M. Lubasch, *Barren plateaus in quantum tensor network optimization,* Quantum **7**, 974 (2023).

[74] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, *Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices,* Phys. Rev. X **10**, 021067 (2020).

[75] G. B. Mbeng, R. Fazio, and G. Santoro, Quantum Annealing: a journey through Digitalization, Control, and hybrid Quantum Variational schemes, (2019), arXiv:1906.08948 [quant-ph] .

[76] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Quantum annealing: A new method for minimizing multidimensional functions,* Chem. Phys. Lett. **219**, 343 (1994).

[77] T. Kadowaki and H. Nishimori, *Quantum annealing in the transverse Ising model,* Phys. Rev. E **58**, 5355 (1998).

[78] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, *Theory of quantum annealing of an Ising spin glass,* Science **295**, 2427 (2002).

[79] G. E. Santoro and E. Tosatti, *Optimization using quantum mechanics: Quantum annealing through adiabatic evolution,* J. Phys. A: Math. Gen. **39**, R393 (2006).

[80] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem,* Science **292**, 472 (2001).

[81] T. Albash and D. A. Lidar, *Adiabatic quantum computation,* Rev. Mod. Phys. **90**, 015002 (2018).

[82] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, *Quantum annealing of a disordered magnet,* Science **284**, 779 (1999).

[83] A. Lucas, *Ising formulations of many NP problems,* Frontiers in Physics **2**, 5 (2014).

[84] F. Glover, G. Kochenberger, and Y. Du, A Tutorial on Formulating and Using QUBO Models, (2019), arXiv:1811.11538 [cs.DS] .

[85] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, *The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective,* Physics Reports **523**, 127 (2013).

[86] T. S. Cubitt, D. Perez-Garcia, and M. M. Wolf, *Undecidability of the spectral gap,* Nature **528**, 207 (2015).

[87] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, *Adiabatic quantum computation is equivalent to standard quantum computation,* in *45th Annual IEEE Symposium on Foundations of Computer Science* (2004) pp. 42–51.

[88] S. Yarkoni, E. Raponi, T. Bäck, and S. Schmitt, *Quantum annealing for industry applications: introduction and review,* Reports on Progress in Physics **85**, 104001 (2022).

[89] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. L. Heras, R. Babbush, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, E. Solano, H. Neven, and J. M. Martinis, *Digitized adiabatic quantum computing with a superconducting circuit,* Nature **534**, 222 (2016).

[90] G. B. Mbeng, L. Arceci, and G. E. Santoro, *Optimal working point in digitized quantum annealing,* Phys. Rev. B **100**, 224201 (2019).

[91] C. Yi, *Success of digital adiabatic simulation with large Trotter step,* Phys. Rev. A **104**, 052603 (2021).

[92] S. Lloyd, *Quantum approximate optimization is computationally universal,* arXiv e-prints , arXiv:1812.11075 (2018), arXiv:1812.11075 [quant-ph] .

[93] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás, *On the universality of the quantum approximate optimization algorithm,* Quantum Information Processing **19** (2020).

[94] E. Farhi and A. W. Harrow, Quantum Supremacy through the Quantum Approximate Optimization Algorithm, (2019), arXiv:1602.07674 [quant-ph] .

[95] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe, *Quantum approximate optimization of the long-range Ising model with a trapped-ion quantum simulator,* Proceedings of the National Academy of Sciences **117**, 25396 (2020), https://www.pnas.org/content/117/41/25396.full.pdf .

[96] M. Alam, A. Ash-Saki, and S. Ghosh, Analysis of Quantum Approximate Optimization Algorithm under Realistic Noise in Superconducting Qubits, (2019), arXiv:1907.09631 [quant-ph] .

[97] A. H. Karamlou, W. A. Simon, A. Katabarwa, T. L. Scholten, B. Peropadre, and Y. Cao, *Analyzing the performance of variational quantum factoring on a superconducting quantum processor,* npj Quantum Information **7**, 156 (2021).

[98] F. G. S. L. Brandao, M. Broughton, E. Farhi, S. Gutmann, and H. Neven, For Fixed Control Parameters the Quantum Approximate Optimization Algorithm's Objective Function Value Concentrates for Typical Instances, (2018), arXiv:1812.04170 [quant-ph] .

[99] G. G. Guerreschi and A. Y. Matsuura, *QAOA for Max-Cut requires hundreds of qubits for quantum speed-up,* Scientific Reports **9** (2019).

[100] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, *The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size,* Quantum **6**, 759 (2022).

[101] M. M. Wauters, G. B. Mbeng, and G. E. Santoro, *Polynomial scaling of the quantum approximate optimization algorithm for ground-state preparation of the fully connected p-spin ferromagnet in a transverse field,* Phys. Rev. A **102**, 062404 (2020).

[102] J. Claes and W. v. Dam, *Instance Independence of Single Layer Quantum Approximate Optimization Algorithm on Mixed-Spin Models at Infinite Size,* Quantum **5**, 542 (2021).

[103] J. Basso, D. Gamarnik, S. Mei, and L. Zhou, *Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models,* in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022).

[104] K. Blekos, D. Brand, A. Ceschini, C.-H. Chou, R.-H. Li, K. Pandya, and A. Summer, A Review on Quantum Approximate Optimization Algorithm and its Variants, (2023), arXiv:2306.09198 [quant-ph] .

[105] D. Wecker, M. B. Hastings, and M. Troyer, *Progress towards practical quantum variational algorithms,* Phys. Rev. A **92**, 042303 (2015).

[106] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and et al., *Self-verifying variational quantum simulation of lattice models,* Nature **569**, 355 (2019).

[107] W. W. Ho, C. Jonay, and T. H. Hsieh, *Ultrafast variational simulation of nontrivial quantum states with long-range interactions,* Physical Review A **99** (2019).

[108] G. Matos, S. Johri, and Z. Papić, *Quantifying the Efficiency of State Preparation via Quantum Variational Eigensolvers,* PRX Quantum **2**, 010309 (2021).

[109] C.-Y. Park, Efficient ground state preparation in variational quantum eigensolver with symmetry breaking layers, (2021), arXiv:2106.02509 [quant-ph] .

[110] N. Astrakhantsev, G. Mazzola, I. Tavernelli, and G. Carleo, Algorithmic Phases in Variational Quantum Ground-State Preparation, (2022), arXiv:2205.06278 [quant-ph] .

[111] L. T. Brady, C. L. Baldwin, A. Bapat, Y. Kharkov, and A. V. Gorshkov, *Optimal Protocols in Quantum Annealing and Quantum Approximate Optimization Algorithm Problems,* Phys. Rev. Lett. **126**, 070505 (2021).

[112] L. T. Brady, L. Kocia, P. Bienias, A. Bapat, Y. Kharkov, and A. V. Gorshkov, *Behavior of Analog Quantum Algorithms*, (2021), arXiv:2107.01218 [quant-ph] .

[113] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, *Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe,* EPJ Quantum Technology **9** (2022).

[114] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, *Optimizing Variational Quantum Algorithms Using Pontryagin's Minimum Principle,* Phys. Rev. X **7**, 021027 (2017).

[115] J. Roland and N. J. Cerf, *Quantum search by local adiabatic evolution,* Phys. Rev. A **65**, 042308 (2002).

[116] M. Streif and M. Leib, *Training the quantum approximate optimization algorithm without access to a quantum processing unit,* Quantum Science and Technology **5**, 034008 (2020).

[117] R. Shaydulin, P. C. Lotshaw, J. Larson, J. Ostrowski, and T. S. Humble, *Parameter Transfer for Quantum Approximate Optimization of Weighted MaxCut,* ACM Transactions on Quantum Computing **4**, 1 (2023).

[118] S. H. Sureshbabu, D. Herman, R. Shaydulin, J. Basso, S. Chakrabarti, Y. Sun, and M. Pistoia, *Parameter Setting in Quantum Approximate Optimization of Weighted Problems,* Quantum **8**, 1231 (2024).

[119] R. Shaydulin, I. Safro, and J. Larson, *Multistart Methods for Quantum Approximate optimization,* in *2019 IEEE High Performance Extreme Computing Conference (HPEC)* (IEEE, 2019).

[120] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, *The Quantum Approximate Optimization Algorithm at High Depth for MaxCut on Large-Girth Regular Graphs and the Sherrington-Kirkpatrick Model,* (Schloss Dagstuhl - Leibniz - Zentrum für Informatik, 2022).

[121] A. Galda, X. Liu, D. Lykov, Y. Alexeev, and I. Safro, *Transferability of optimal qaoa parameters between random graphs,* in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2021) pp. 171–180.

[122] Z. Jiang, E. G. Rieffel, and Z. Wang, *Near-optimal quantum circuit for Grover's unstructured search using a transverse field,* Physical Review A **95** (2017).

[123] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte, *Parameter concentrations in quantum approximate optimization,* Phys. Rev. A **104**, L010401 (2021).

[124] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, Learning to learn with quantum neural networks via classical neural networks, (2019), arXiv:1907.05415 .

[125] S. H. Sack and M. Serbyn, *Quantum annealing initialization of the quantum approximate optimization algorithm,* Quantum **5**, 491 (2021).

[126] G. E. Crooks, Performance of the Quantum Approximate Optimization Algorithm on the Maximum Cut Problem, (2018), arXiv:1811.08419 [quant-ph] .

[127] J. Wurtz and P. J. Love, *Counterdiabaticity and the quantum approximate optimization algorithm,* Quantum **6**, 635 (2022).

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

[129] A. A. Ovchinnikov, D. V. Dmitriev, V. Y. Krivnov, and V. O. Cheranovskii, *Antiferromagnetic ising chain in a mixed transverse and longitudinal magnetic field,* Phys. Rev. B **68**, 214406 (2003).

[130] S. B. Kotsiantis, *Supervised Machine Learning: A Review of Classification Techniques,* in *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies* (IOS Press, NLD, 2007) pp. 3–24.

[131] C. Baldassi and R. Zecchina, *Efficiency of quantum vs. classical annealing in nonconvex learning problems,* Proceedings of the National Academy of Sciences **115**, 1457 (2018).

[132] C. Baldassi, C. Lauditi, E. M. Malatesta, R. Pacelli, G. Perugini, and R. Zecchina, *Learning through atypical phase transitions in overparameterized neural networks,* Phys. Rev. E **106**, 014116 (2022).

[133] F. Rosenblatt, *The perceptron - A perceiving and recognizing automaton,* Tech. Rep. 85-460-1 (Cornell Aeronautical Laboratory, Ithaca, New York, 1957).

[134] J. A. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation* (CRC Press, 2018).

[135] F. Tacchino, C. Macchiavello, D. Gerace, and D. Bajoni, *An artificial neuron implemented on an actual quantum processor,* npj Quantum Information **5**, 26 (2019).

Bibliography

[136] M. Schuld, I. Sinayskiy, and F. Petruccione, *Simulating a perceptron on a quantum computer,* Physics Letters A **379**, 660 (2015).

[137] Krauth, Werner and Mézard, Marc, *Storage capacity of memory networks with binary couplings,* J. Phys. France **50**, 3057 (1989).

[138] H. Horner, *Dynamics of learning for the binary perceptron problem,* Zeitschrift für Physik B Condensed Matter **86**, 291 (1992).

[139] R. Orús, S. Mugel, and E. Lizaso, *Forecasting financial crashes with quantum computing,* Phys. Rev. A **99**, 060301 (2019).

[140] M. Fellner, K. Ender, R. ter Hoeven, and W. Lechner, *Parity Quantum Optimization: Benchmarks,* Quantum **7**, 952 (2023).

[141] J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, 2006).

[142] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,* Nature Methods , 261 (2020).

[143] X. You and X. Wu, *Exponentially Many Local Minima in Quantum Neural Networks,* in *Proceedings of the 38th International Conference on Machine Learning,* Proceedings of Machine Learning Research, Vol. 139 (PMLR, 2021) pp. 12144–12155.

[144] E. R. Anschuetz, *Critical Points in Quantum Generative Models,* in *International Conference on Learning Representations* (2022).

[145] E. Farhi, J. Goldstone, S. Gutmann, and D. Nagaj, How to make the quantum adiabatic algorithm fail, (2005), arXiv:0512159 [quant-ph] .

[146] A. Callison, N. Chancellor, F. Mintert, and V. Kendon, *Finding spin glass ground states using quantum walks,* New Journal of Physics **21**, 123022 (2019).

[147] J. G. Morley, N. Chancellor, S. Bose, and V. Kendon, *Quantum search with hybrid adiabatic–quantum-walk algorithms and realistic noise,* Phys. Rev. A **99**, 022339 (2019).

[148] P. Chandarana, N. N. Hegade, K. Paul, F. Albarrán-Arriagada, E. Solano, A. del Campo, and X. Chen, *Digitized-counterdiabatic quantum approximate optimization algorithm,* Phys. Rev. Research **4**, 013141 (2022).

[149] J. Yao, L. Lin, and M. Bukov, *Reinforcement Learning for Many-Body Ground-State Preparation Inspired by Counterdiabatic Driving,* Phys. Rev. X **11**, 031070 (2021).

[150] G. B. Mbeng and W. Lechner, Rotated ansatz for approximate counterdiabatic driving, (2022), arXiv:2207.03553 .

[151] E. Farhi, J. Goldstone, S. Gutmann, and H. Neven, Quantum Algorithms for Fixed Qubit Architectures, (2017), arXiv:1703.06199 [quant-ph] .

[152] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,* Nature **549**, 242 (2017).

[153] J. Carrasquilla, M. Hibat-Allah, E. Inack, A. Makhzani, K. Neklyudov, G. W. Taylor, and G. Torlai, Quantum HyperNetworks: Training Binary Neural Networks in Quantum Superposition, (2023), arXiv:2301.08292 [quant-ph] .

[154] H. C. Nguyen, R. Zecchina, and J. Berg, *Inverse statistical problems: from the inverse Ising problem to data science,* Advances in Physics **66**, 197 (2017).

[155] R. J. Baxter, *One-dimensional anisotropic Heisenberg chain,* Annals of Physics **70**, 323 (1972).

[156] M. P. M. den Nijs, *Derivation of extended scaling relations between critical exponents in two-dimensional models from the one-dimensional Luttinger model,* Phys. Rev. B **23**, 6111 (1981).

[157] E. Ercolessi, S. Evangelisti, F. Franchini, and F. Ravanini, *Essential singularity in the Renyi entanglement entropy of the one-dimensional XYZ spin-$\frac{1}{2}$ chain,* Phys. Rev. B **83**, 012402 (2011).

[158] A. A. Ovchinnikov, D. V. Dmitriev, V. Y. Krivnov, and V. O. Cheranovskii, *Antiferromagnetic Ising chain in a mixed transverse and longitudinal magnetic field,* Phys. Rev. B **68**, 214406 (2003).

[159] P. Sen, *Quantum-fluctuation-induced spatial stochastic resonance at zero temperature,* Phys. Rev. E **63**, 040101 (2001).

[160] P. Jordan and E. Wigner, *Über das paulische äquivalenzverbot,* Zeitschrift für Physik **47**, 631 (1928).

[161] E. Lieb, T. Schultz, and D. Mattis, *Two soluble models of an antiferromagnetic chain,* Annals of Physics **16**, 407 (1961).

[162] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, *Quantum approximate optimization algorithm for MaxCut: A fermionic view,* Phys. Rev. A **97**, 022304 (2018).

[163] D. A. Fedorov, B. Peng, N. Govind, and Y. Alexeev, VQE method: A short survey and recent developments, (2021), arXiv:2103.08505 [quant-ph] .

[164] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families, (2019), arXiv:1907.13623 [quant-ph] .

[165] C. Lyu, V. Montenegro, and A. Bayat, *Accelerated variational algorithms for digital quantum simulation of many-body ground states,* Quantum **4**, 324 (2020).

[166] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Y. Halpern, *Linear growth of quantum circuit complexity,* Nature Physics **18**, 528 (2022).

[167] E. H. Lieb and D. W. Robinson, *The finite group velocity of quantum spin systems,* Communications in Mathematical Physics **28**, 251 (1972).

[168] F. G. S. L. Brandão, A. W. Harrow, and M. Horodecki, *Local Random Quantum Circuits are Approximate Polynomial-Designs,* Communications in Mathematical Physics **346**, 397 (2016).

[169] J. Haferkamp, Random quantum circuits are approximate unitary t-designs in depth $o\left(nt^{5+o(1)}\right)$, (2022), arXiv:2203.16571 [quant-ph] .

[170] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, *Quantum supremacy using a programmable superconducting processor,* Nature **574**, 505 (2019).

[171] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K. X. Wei, C. J. Wood, J.-B. Yau, E. J. Zhang, O. E. Dial, J. M. Chow, and J. M. Gambetta, *Demonstration of quantum volume*

64 on a superconducting quantum computing system, Quantum Science and Technology **6**, 025020 (2021).

[172] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, *Surface codes: Towards practical large-scale quantum computation,* Phys. Rev. A **86**, 032324 (2012).

[173] K. J. Satzinger *et al.*, *Realizing topologically ordered states on a quantum processor,* Science **374**, 1237 (2021).

[174] E. Zohar, J. I. Cirac, and B. Reznik, *Quantum simulations of lattice gauge theories using ultracold atoms in optical lattices,* Reports on Progress in Physics **79**, 014401 (2015).

[175] M. Dalmonte and S. Montangero, *Lattice gauge theory simulations in the quantum information era,* Contemporary Physics **57**, 388 (2016).

[176] M. C. Bañuls, R. Blatt, J. Catani, A. Celi, J. I. Cirac, M. Dalmonte, L. Fallani, K. Jansen, M. Lewenstein, S. Montangero, C. A. Muschik, B. Reznik, E. Rico, L. Tagliacozzo, K. Van Acoleyen, F. Verstraete, U.-J. Wiese, M. Wingate, J. Zakrzewski, and P. Zoller, *Simulating lattice gauge theories within quantum technologies,* The European Physical Journal D **74**, 165 (2020).

[177] M. C. Bañuls and K. Cichy, *Review on novel methods for lattice gauge theories,* Reports on Progress in Physics **83**, 024401 (2020).

[178] E. Zohar, *Quantum simulation of lattice gauge theories in more than one space dimension-requirements, challenges and methods,* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **380**, 20210069 (2022).

[179] M. Aidelsburger, L. Barbiero, A. Bermudez, T. Chanda, A. Dauphin, D. González-Cuadra, P. R. Grzybowski, S. Hands, F. Jendrzejewski, J. Jünemann, G. Juzeliunas, V. Kasper, A. Piga, S.-J. Ran, M. Rizzi, G. Sierra, L. Tagliacozzo, E. Tirrito, T. V. Zache, J. Zakrzewski, E. Zohar, and M. Lewenstein, *Cold atoms meet lattice gauge theory,* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **380**, 20210064 (2022).

[180] L. Tagliacozzo, A. Celi, and M. Lewenstein, *Tensor Networks for Lattice Gauge Theories with Continuous Groups,* Phys. Rev. X **4**, 041024 (2014).

[181] P. Silvi, E. Rico, T. Calarco, and S. Montangero, *Lattice gauge tensor networks,* New J. Phys. **16**, 103015 (2014).

[182] A. Celi, B. Vermersch, O. Viyuela, H. Pichler, M. D. Lukin, and P. Zoller, *Emerging Two-Dimensional Gauge Theories in Rydberg Configurable Arrays,* Phys. Rev. X **10**, 021057 (2020).

[183] T. Felser, P. Silvi, M. Collura, and S. Montangero, *Two-Dimensional Quantum-Link Lattice Quantum Electrodynamics at Finite Density,* Phys. Rev. X **10**, 041040 (2020).

[184] E. Zohar, *Wilson loops and area laws in lattice gauge theory tensor networks,* Phys. Rev. Research **3**, 033179 (2021).

[185] A. Kitaev, *Fault-tolerant quantum computation by anyons,* Annals of Physics **303**, 2 (2003).

[186] S. Bravyi, M. B. Hastings, and F. Verstraete, *Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order,* Phys. Rev. Lett. **97**, 050401 (2006).

[187] X. Chen, Z.-C. Gu, and X.-G. Wen, *Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order,* Phys. Rev. B **82**, 155138 (2010).

[188] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, *Obstacles to Variational Quantum Optimization from Symmetry Protection,* Phys. Rev. Lett. **125**, 260505 (2020).

[189] Y.-J. Liu, K. Shtengel, A. Smith, and F. Pollmann, Methods for simulating string-net states and anyons on a digital quantum computer, (2021), arXiv:2110.02020 [quant-ph] .

[190] E. Farhi, D. Gamarnik, and S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: A typical case, (2020), arXiv:2004.09002 [quant-ph] .

[191] E. Farhi, D. Gamarnik, and S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples, (2020), arXiv:2005.08747 [quant-ph] .

[192] S. B. Bravyi and A. Y. Kitaev, Quantum codes on a lattice with boundary, (1998).

[193] E. Fradkin and L. Susskind, *Order and disorder in gauge systems and magnets,* Phys. Rev. D **17**, 2637 (1978).

[194] J. Greensite, *The confinement problem in lattice gauge theory,* Progr. Part. Nucl. Phys. **51**, 1 (2003).

[195] J. B. Kogut, *An introduction to lattice gauge theory and spin systems,* Rev. Mod. Phys. **51**, 659 (1979).

[196] M. Creutz, *Asymptotic-Freedom Scales,* Phys. Rev. Lett. **45**, 313 (1980).

[197] H. W. J. Blöte and Y. Deng, *Cluster Monte Carlo simulation of the transverse Ising model,* Phys. Rev. E **66**, 066110 (2002).

[198] E. Fradkin, *Field Theories of Condensed Matter Physics*, 2nd ed. (Cambridge University Press, 2013).

[199] C. Dlaska, K. Ender, G. B. Mbeng, A. Kruckenhauser, W. Lechner, and R. van Bijnen, *Quantum Optimization via Four-Body Rydberg Gates,* Phys. Rev. Lett. **128**, 120503 (2022).

[200] W. Lechner, *Quantum Approximate Optimization With Parallelizable Gates,* IEEE Transactions on Quantum Engineering **1**, 1 (2020).

[201] H. Lamm, S. Lawrence, and Y. Yamauchi (NuQS Collaboration), *General methods for digital quantum simulation of gauge theories,* Phys. Rev. D **100**, 034518 (2019).

[202] E. Zohar, A. Farace, B. Reznik, and J. I. Cirac, *Digital Quantum Simulation of $\mathbb{Z}_2$ Lattice Gauge Theories with Dynamical Fermionic Matter,* Phys. Rev. Lett. **118**, 070501 (2017).

[203] E. Zohar, A. Farace, B. Reznik, and J. I. Cirac, *Digital lattice gauge theories,* Phys. Rev. A **95**, 023604 (2017).

[204] T. Armon, S. Ashkenazi, G. García-Moreno, A. González-Tudela, and E. Zohar, *Photon-Mediated Stroboscopic Quantum Simulation of a $\mathbb{Z}_2$ Lattice Gauge Theory,* Phys. Rev. Lett. **127**, 250501 (2021).

[205] D. Horn, M. Weinstein, and S. Yankielowicz, *Hamiltonian approach to $Z(N)$ lattice gauge theories,* Phys. Rev. D **19**, 3715 (1979).

[206] E. Zohar, J. I. Cirac, and B. Reznik, *Quantum simulations of gauge theories with ultracold atoms: Local gauge invariance from angular-momentum conservation,* Phys. Rev. A **88**, 023617 (2013).

[207] S. Notarnicola, E. Ercolessi, P. Facchi, G. Marmo, S. Pascazio, and F. V. Pepe, *Discrete Abelian gauge theories for quantum simulations of QED,* J. Phys. A: Math. Theor. **48**, 30FT01 (2015).

[208] E. Ercolessi, P. Facchi, G. Magnifico, S. Pascazio, and F. V. Pepe, *Phase transitions in $\mathbb{Z}_n$ gauge models: Towards quantum simulations of the Schwinger-Weyl QED,* Phys. Rev. D **98**, 074503 (2018).

[209] P. Emonts, M. C. Bañuls, I. Cirac, and E. Zohar, *Variational Monte Carlo simulation with tensor networks of a pure $\mathbb{Z}_3$ gauge theory in $(2+1)$D,* Phys. Rev. D **102**, 074501 (2020).

[210] J. Nyhegn, C.-M. Chung, and M. Burrello, $\mathbb{Z}_N$ *lattice gauge theory in a ladder geometry,* Phys. Rev. Research **3**, 013133 (2021).

[211] D. Robaina, M. C. Bañuls, and J. I. Cirac, *Simulating* $(2+1)$D $\mathbb{Z}_3$ *Lattice Gauge Theory with an Infinite Projected Entangled-Pair State,* Phys. Rev. Lett. **126**, 050401 (2021).

[212] E. Fradkin and L. P. Kadanoff, *Disorder variables and para-fermions in two-dimensional statistical mechanics,* Nuclear Physics B **170**, 1 (1980).

[213] G. Ortiz, E. Cobanera, and Z. Nussinov, *Dualities and the phase diagram of the p-clock model,* Nucl. Phys. B **854**, 780 (2012).

[214] S. S. Bullock and G. K. Brennen, *Qudit surface codes and gauge theory with finite cyclic groups,* J. Phys. A: Math. Theor. **40**, 3481 (2007).

[215] M. D. Schulz, S. Dusuel, R. Orús, J. Vidal, and K. P. Schmidt, *Breakdown of a perturbed* $\mathbb{Z}_N$ *topological phase,* New J. Phys. **14**, 025005 (2012).

[216] E. Zohar and M. Burrello, *Formulation of lattice gauge theories for quantum simulations,* Phys. Rev. D **91**, 054506 (2015).

[217] J. Bender, E. Zohar, A. Farace, and J. I. Cirac, *Digital quantum simulation of lattice gauge theories in three spatial dimensions,* New Journal of Physics **20**, 093001 (2018).

[218] J. F. Haase, L. Dellantonio, A. Celi, D. Paulson, A. Kan, K. Jansen, and C. A. Muschik, *A resource efficient approach for quantum and classical simulations of gauge theories in particle physics,* Quantum **5**, 393 (2021).

[219] D. Paulson, L. Dellantonio, J. F. Haase, A. Celi, A. Kan, A. Jena, C. Kokail, R. van Bijnen, K. Jansen, P. Zoller, and C. A. Muschik, *Simulating 2D effects in lattice gauge theories on a quantum computer,* PRX Quantum **2**, 030334 (2021).

[220] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016).

[221] E. Zohar, M. Burrello, T. B. Wahl, and J. I. Cirac, *Fermionic projected entangled pair states and local U (1) gauge theories,* Annals of Physics **363**, 385 (2015).

[222] J. Bender, P. Emonts, E. Zohar, and J. I. Cirac, *Real-time dynamics in* $2+1d$ *compact QED using complex periodic Gaussian states,* Phys. Rev. Research **2**, 043145 (2020).

[223] A. Kitaev and J. Preskill, *Topological Entanglement Entropy,* Phys. Rev. Lett. **96**, 110404 (2006).

[224] M. Levin and X.-G. Wen, *Detecting Topological Order in a Ground State Wave Function,* Phys. Rev. Lett. **96**, 110405 (2006).

[225] A. Hamma, R. Ionicioiu, and P. Zanardi, *Bipartite entanglement and entropic boundary law in lattice spin systems,* Phys. Rev. A **71**, 022315 (2005).

[226] L. Tagliacozzo, A. Celi, A. Zamora, and M. Lewenstein, *Optical Abelian lattice gauge theories,* Annals of Physics **330**, 160 (2013).

[227] L. Tagliacozzo, A. Celi, P. Orland, M. W. Mitchell, and M. Lewenstein, *Simulation of non-Abelian gauge theories with optical lattices,* Nat. Comm. **4**, 1 (2013).

[228] C. Muschik, M. Heyl, E. Martinez, T. Monz, P. Schindler, B. Vogell, M. Dalmonte, P. Hauke, R. Blatt, and P. Zoller, *U(1) Wilson lattice gauge theories in digital quantum simulators,* New J. Phys. **19**, 103020 (2017).

[229] J. C. Halimeh, H. Lang, J. Mildenberger, Z. Jiang, and P. Hauke, *Gauge-Symmetry Protection Using Single-Body Terms,* PRX Quantum **2**, 040311 (2021).

[230] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, *Real-time dynamics of lattice gauge theories with a few-qubit quantum computer,* Nature **534**, 516 (2016).

[231] N. Klco, E. F. Dumitrescu, A. J. McCaskey, T. D. Morris, R. C. Pooser, M. Sanz, E. Solano, P. Lougovski, and M. J. Savage, *Quantum-classical computation of Schwinger model dynamics using quantum computers,* Phys. Rev. A **98**, 032331 (2018).

[232] N. Klco, M. J. Savage, and J. R. Stryker, *SU(2) non-abelian gauge field theory in one dimension on digital quantum computers,* Phys. Rev. D **101**, 074512 (2020).

[233] Y. Y. Atas, J. Zhang, R. Lewis, A. Jahanpour, J. F. Haase, and C. A. Muschik, *SU(2) hadrons on a quantum computer via a variational approach,* Nat. Comm. **12** (2021).

[234] A. Ciavarella, N. Klco, and M. J. Savage, *Trailhead for quantum simulation of SU(3) Yang-Mills lattice gauge theory in the local multiplet basis,* Phys. Rev. D **103**, 094501 (2021).

[235] T. Pichler, M. Dalmonte, E. Rico, P. Zoller, and S. Montangero, *Real-Time Dynamics in U(1) Lattice Gauge Theories with Tensor Networks,* Phys. Rev. X **6**, 011023 (2016).

[236] F. M. Surace, P. P. Mazza, G. Giudici, A. Lerose, A. Gambassi, and M. Dalmonte, *Lattice Gauge Theories and String Dynamics in Rydberg Atom Quantum Simulators,* Phys. Rev. X **10**, 021041 (2020).

[237] F. M. Surace and A. Lerose, *Scattering of mesons in quantum simulators,* New Journal of Physics **23**, 062001 (2021).

[238] M. Rigobello, S. Notarnicola, G. Magnifico, and S. Montangero, *Entanglement generation in* $(1+1)$D *QED scattering processes,* Phys. Rev. D **104**, 114501 (2021).

[239] J. Vidal, S. Dusuel, and K. P. Schmidt, *Low-energy effective theory of the toric code model in a parallel magnetic field,* Phys. Rev. B **79**, 033109 (2009).

[240] J. R. Wootton and J. K. Pachos, *Bringing Order through Disorder: Localization of Errors in Topological Quantum Memories,* Phys. Rev. Lett. **107**, 030503 (2011).

[241] D. S. Rokhsar and S. A. Kivelson, *Superconductivity and the Quantum Hard-Core Dimer Gas,* Phys. Rev. Lett. **61**, 2376 (1988).

[242] W. Lechner, P. Hauke, and P. Zoller, *A quantum annealing architecture with all-to-all connectivity from local interactions,* Science Advances **1** (2015).

[243] K. Ender, R. ter Hoeven, B. E. Niehoff, M. Drieb-SchÃ¶n, and W. Lechner, Parity Quantum Optimization: Compiler, (2021).

[244] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, *Quantum phases of matter on a 256-atom programmable quantum simulator,* Nature **595**, 227 (2021).

[245] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, *Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms,* Nature **595**, 233 (2021).

[246] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J. Liu, R. Samajdar, X.-Z. Luo, B. Nash, X. Gao, B. Barak, E. Farhi, S. Sachdev, N. Gemelke, L. Zhou, S. Choi, H. Pichler, S. Wang, M. Greiner, V. Vuletic, and M. D. Lukin, Quantum Optimization of Maximum Independent Set using Rydberg Atom Arrays, (2022).

[247] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, V. Vuletic, and M. D. Lukin, A quantum processor based on coherent transport of entangled atom arrays, (2021).

[248] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, *An adaptive variational algorithm for exact molecular simulations on a quantum computer,* Nature Communications **10**, 3007 (2019).

[249] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The Variational Quantum Eigensolver: a review of methods and best practices, (2021), arXiv:2111.05176 [quant-ph] .

[250] C. Zhu, R. H. Byrd, and J. Nocedal., *L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization.* ACM Transactions on Mathematical Software **23**, 550 (1997).

[251] D. J. Wales and J. P. K. Doye, *Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms,* The Journal of Physical Chemistry A **101**, 5111 (1997).