

Quantum approximate optimization algorithm applied to the binary perceptron

Pietro Torta^{1,*}, Glen B. Mbeng,² Carlo Baldassi,³ Riccardo Zecchina,³ and Giuseppe E. Santoro^{1,4,5}

¹SISSA, Via Bonomea 265, I-34136 Trieste, Italy

²Universität Innsbruck, Technikerstraße 21 a, A-6020 Innsbruck, Austria

³Department of Computing Sciences, Bocconi University, I-20136 Milan, Italy

⁴International Centre for Theoretical Physics (ICTP), P.O.Box 586, I-34014 Trieste, Italy

⁵CNR-IOM, Consiglio Nazionale delle Ricerche - Istituto Officina dei Materiali, c/o SISSA Via Bonomea 265, I-34136 Trieste, Italy

 (Received 20 May 2022; revised 21 November 2022; accepted 15 December 2022; published 2 March 2023)

We apply digitized quantum annealing (QA) and quantum approximate optimization algorithm (QAOA) to a paradigmatic task of supervised learning in artificial neural networks: the optimization of synaptic weights for the binary perceptron. At variance with the usual QAOA applications to MaxCut, or to quantum spin-chains ground-state preparation, here the classical cost function is characterized by highly nonlocal multispin interactions. Yet, we provide evidence for the existence of optimal *smooth* solutions for the QAOA parameters, which are *transferable* among typical instances of the same problem, and we prove numerically an enhanced performance of QAOA over traditional QA. We also investigate on the role of the classical cost-function landscape geometry in this problem. By artificially breaking this geometrical structure, we show that the detrimental effect of a gap-closing transition, encountered in QA, is also negatively affecting the performance of our QAOA implementation.

DOI: [10.1103/PhysRevB.107.094202](https://doi.org/10.1103/PhysRevB.107.094202)

I. INTRODUCTION

Quantum optimization is a very active branch of quantum computation [1], which started with quantum annealing (QA) [2–6] and adiabatic quantum computation (AQC) [7,8], by now well established and implemented in analog dedicated hardware [9].

New research directions are currently pursuing the design of parameterized quantum circuits and variational quantum algorithms [10]. Among the early proposals in this class are the variational quantum eigensolver (VQE) [11] and the quantum approximate optimization algorithm (QAOA) [12], hybrid quantum-classical variational optimization schemes [13] designed for quantum ground state preparation, and classical combinatorial optimization, respectively. In particular, QAOA has rapidly gained popularity, with new theoretical understanding [14–18] and successful implementations on experimental platforms [19].

From a broader perspective, a number of variational quantum algorithms have been proposed and applied beyond classical combinatorial optimization [20,21], conceptually generalizing QAOA [22] or VQE [23], or focusing on realistic quantum hardware setups [24,25]. These algorithms constitute a rich playground for short-term implementations on existing and near-future noisy intermediate-scale quantum devices [26], aiming for some type of quantum speedup [27].

The quantum Hamiltonian of a physical system typically consists of a sum of local terms, each involving only a small subset of quantum degrees of freedom. On the contrary, in order to apply QA or QAOA to a classical optimization problem, an *embedding* into a quantum setup [28] is first required.

In this second case, the standard strategy is to start from a classical cost (or energy) function $E(\sigma_1, \dots, \sigma_N)$ depending on N binary variables $\sigma_j = \pm 1$, and to map them 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 [1],

$$E(\sigma_1, \dots, \sigma_N) \rightarrow \hat{H}_z(\hat{\sigma}_1^z, \dots, \hat{\sigma}_N^z). \quad (1)$$

Nevertheless, in contrast to the Hamiltonians of physical quantum systems, the embedding Hamiltonian \hat{H}_z can be arbitrarily difficult to implement, possibly including nonlocal multispin interactions, thus proving intrinsically hard to simulate exactly [29] on a quantum device.

Remarkably, some specific combinatorial optimization problems admit a natural embedding into a 2-local Ising spin-glass Hamiltonian, such as MaxCut on random graphs [12], or a few other model-specific applications [30]. As a matter of fact, many interesting classical optimization problems do not admit such a simple reformulation in terms of a 2-local (or k -local, for any $k \ll N$) Hamiltonian.

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 [31], which is, however, nonlocal in terms of its variables (network weights and biases) due to nonlinear 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 on their effectiveness in classification tasks. From a complementary perspective, recent years have witnessed the rise of an emerging field known as quantum machine learning [32] and different proposals

*Corresponding author: ptorta@sisssa.it

have been put forward to perform machine learning tasks by developing quantum versions of neural networks: interestingly enough, parameterized quantum circuits themselves can be regarded as alternative machine learning models [33].

In this paper, we aim at investigating the potential applicability of variational quantum algorithms such as QAOA in the realm of hard nonconvex classical optimization problems, yielding highly nonlocal Hamiltonians, beyond the usual 2-local models. As a working example, we focus on the problem of learning random patterns in a single-layer neural network with binary weights, the so-called binary perceptron. Our work stems from the results obtained in Ref. [34], where the authors provided analytic and simulation evidence of exponential speed-up of quantum annealing 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 quantum annealing 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 [35]. 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 quantum annealing and 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. [34], these results are expected to generalize for more complex ANN architectures. Moreover, we show the emergence of *smooth* optimal QAOA parameters [16,18,36], 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. In fact, one of the most promising research lines in variational quantum algorithms deals with reducing or removing the need for a classical optimization loop [37], by leveraging *concentration effects* in the variational energy landscape for different problem instances of the same class [38–40], or for different sizes N of the same instance [41,42]. In our paper, we give numerical evidence of similar findings, beyond the usual 2-local models previously analyzed.

Finally, we enquire into the role of the classical cost-function landscape geometry [43] in the effectiveness of digitized-QA vs QAOA for our model. This is done by artificially permuting the classical energies associated to each spin configuration: despite the spectrum and the number of classical solutions being the same, a gap closure on the adiabatic path appears [34], 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.

The rest of the article is organized as follows: In Sec. II we review the formulation of digitized-QA and QAOA, and introduce the model Hamiltonian for the binary perceptron. In Sec. III we discuss our numerical results on the comparison of these two algorithms, along with the aforementioned qualitative features of the optimal solutions, such as smooth-

ness and mild dependence on the training set. In Sec. IV we analyze the role of the classical landscape geometry, while in Sec. V we discuss possible extensions and generalizations of our paper.

II. PROBLEM AND METHODS

A. Digitized-QA and QAOA

In this section we summarize the main ingredients of digitized-QA [44,45] and QAOA, as a variational quantum algorithm [16].

In standard QA/AQC framework [8] one constructs an interpolating Hamiltonian $\hat{H}(s) = s\hat{H}_{\text{targ}} + (1-s)\hat{H}_x$, where \hat{H}_{targ} is the problem (or target) Hamiltonian—which depends on spin-1/2 Pauli operators and whose ground state we wish to find—while $\hat{H}_x = -\Gamma_0 \sum_j \hat{\sigma}_j^x$ is a transverse field term, allowing for quantum fluctuations. For a classical minimization problem, as in the case we are investigating, $\hat{H}_{\text{targ}} = \hat{H}_z$ is simply built from $\hat{\sigma}^z$ terms as in Eq. (1), thus it is diagonal in the computational basis [1], and its ground states encode the classical solutions; in contrast, for the case of quantum state preparation, \hat{H}_{targ} contains further nondiagonal quantum fluctuation terms. An adiabatic dynamics is then pursued by slowly increasing $s(t)$ from $s(0) = 0$ to $s(\tau) = 1$ in a large total annealing time τ , starting from the easily-prepared ground state of \hat{H}_x ,

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

where $|\uparrow\rangle$ and $|\downarrow\rangle$ denote the spin up/down eigenstates of $\hat{\sigma}^z$.

The difficulty in the QA/AQC scheme is usually associated with the growing values of the annealing time τ required to adiabatically follow the instantaneous ground state of $\hat{H}(s)$, possibly diverging in the thermodynamic limit, if the system crosses a critical point or, even worse, a first-order phase transition [46]. While in principle the annealing schedule $s(t)$ can be chosen with some freedom, in practice optimizing the $s(t)$, by “slowing down” close to points where the spectral gap of $\hat{H}(s)$ closes, requires knowing such spectral information, a notoriously difficult problem [47]. Very often, a *linear* schedule $s(t) = t/\tau$ is assumed.

The digitalization of the continuous-time QA/AQC dynamics is a very natural idea, conceptually relying on the Trotter split-up of noncommuting exponentials. For the case of a classical combinatorial optimization problem, $\hat{H}_{\text{targ}} = \hat{H}_z$, we would simply write

$$e^{-i\frac{\Delta t}{\hbar}\hat{H}(s)} = e^{-i\beta\hat{H}_x}e^{-i\gamma\hat{H}_z} + \mathcal{O}((\Delta t)^2), \quad (3)$$

with $\beta = (1-s)\Delta t/\hbar$ and $\gamma = s\Delta t/\hbar$ to lowest order in the Trotter splitting. Similarly, if \hat{H}_{targ} is a combination of a $\hat{\sigma}^z$ -part \hat{H}_z and a $\hat{\sigma}^x$ -part \hat{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 β and γ obtained by the Trotter split-up.

With the standard QA/AQC assumption of a linear annealing schedule $s(t) = t/\tau$, we would perform a digitized-QA by simply setting $s_m = m/P$ for $m = 1 \dots P$, where P is the number of *Trotter steps* and $\Delta t_m = \Delta t$, hence a total annealing time $\tau = P\Delta t$. This amounts to setting, in the case

of a classical optimization problem, for all $m = 1 \cdots P$,

$$\beta_m = (1 - s_m) \frac{\Delta t}{\hbar}, \quad \gamma_m = s_m \frac{\Delta t}{\hbar}, \quad s_m = \frac{m}{P}. \quad (4)$$

The digitized-QA unitary evolution would hence be given by

$$|\psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \hat{U}(\beta_P, \gamma_P) \cdots \hat{U}(\beta_1, \gamma_1) |+\rangle^{\otimes N}, \quad (5)$$

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

$$\hat{U}(\beta_m, \gamma_m) = e^{-i\beta_m \hat{H}_x} e^{-i\gamma_m \hat{H}_z}. \quad (6)$$

This digitized-QA scheme allows for a single variational parameter, Δt in Eq. (4), which we can optimize so as to minimize the variational energy,

$$E_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \langle \psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) | \hat{H}_{\text{targ}} | \psi_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) \rangle, \quad (7)$$

with $\hat{H}_{\text{targ}} = \hat{H}_z$ for a classical optimization. Indeed, as discussed in Ref. [45], a too small value of Δt produces small Trotter errors but also a short annealing time, while a too large Δt is associated to large Trotter errors that make the final state rather inaccurate. Consequently, there is an optimal value of Δt for performing such a digitized-QA dynamics [45].

The quantum approximate optimization algorithm (QAOA) by Farhi *et al.* [12] and more general variational quantum algorithms rely on the basic variational principle of Quantum Mechanics: a trial parameterized wavefunction is defined—usually in terms of a quantum circuit—with an objective function given by the target Hamiltonian average energy on the state. One regards the quantum wavefunction parameters as *variational parameters* of the objective function in Eq. (7), which we seek to minimize by some appropriate *classical* optimization scheme: hence the hybrid quantum-classical nature of such algorithms.

In the QAOA case, to which we shall restrict our considerations from now on, the trial wavefunction has the same form as in Eqs. (5) and (6), where now $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are promoted to variational parameters for the quantum state, rather than fixed by a Trotter split-up as in digitized QA. The *pro* of such a variational scheme is that the optimal energy at the global minimum $E_P(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$ is certainly a monotonically decreasing function of P , which systematically improves on any digitized-QA approach with the same P . The *cons* is that determining the global minimum ($\boldsymbol{\beta}^*$, $\boldsymbol{\gamma}^*$) is in general a nontrivial 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* [48] can make the gradients of the objective function exponentially small in the number of spin variables.

Let us finally stress that while the implementation of $e^{-i\beta_m \hat{H}_x}$ requires a single layer of one-qubit gates, the gate decomposition (and thus the depth of the resulting quantum circuit) for the unitary $e^{-i\gamma_m \hat{H}_z}$ is strongly problem dependent, and usually represents the true bottleneck for an actual implementation of this computational paradigm. Let us mention, in this respect, that there is a whole active field—sometimes referred to as *quantum neural networks*—where parameterized quantum circuits [33] with fixed gates but free variational parameters are studied, both from the point of view of their expressive power—the “expressivity” of a parameterized state

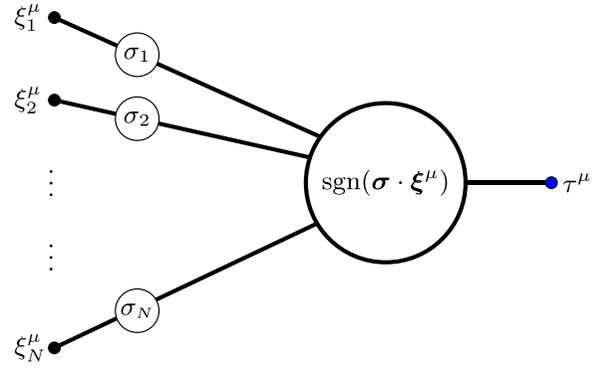


FIG. 1. Scheme of a perceptron. Binary synaptic weights σ_j have to be adjusted such that for given binary values ξ_j^μ in the N input neurons, the scalar product $\boldsymbol{\sigma} \cdot \boldsymbol{\xi}^\mu = \sum_j \sigma_j \xi_j^\mu$ has a prescribed output $\text{sgn} \tau^\mu$. Here $\mu = 1 \cdots M$, with $M = \alpha N$, labels the various input-output patterns.

being the states set of the Hilbert space that it is able to represent—as well as from the ease in finding good parameters, the so-called “trainability”. In these cases, the target Hamiltonian \hat{H}_{targ} (or the \hat{H}_z of the classical optimization problem) needs not be directly used in the construction of the unitary gates.

B. Binary perceptron model

The perceptron represents the prototypical example of a single-layer binary classifier, first introduced decades ago by Rosenblatt [49]. It is still a subject of active research, both as the fundamental unit of classical artificial neural networks [50] and as a potential candidate for simple realizations of quantum neural networks [51,52].

Following Ref. [34], 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 $\text{sgn}(\boldsymbol{\sigma} \cdot \boldsymbol{\xi}^\mu) = \tau^\mu$, see sketch in Fig. 1.

During the learning phase, a given labeled 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. Since, by hypothesis, both the components of the patterns $\boldsymbol{\xi}^\mu$ and the labels τ^μ are independent identically distributed (unbiased) Bernoulli random variables—results of a fair coin flip—we can assume without loss of generality the labels τ^μ to be all equal +1.

The search problem is immediately reformulated as a minimization problem of 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 zero-energy configurations $\boldsymbol{\sigma}^*$ of the cost function. Let us define

$$m_\mu = \frac{1}{\sqrt{N}} \sum_{j=1}^N \sigma_j \xi_j^\mu \quad (8)$$

to be the overlap between the spin configuration σ and the μ th pattern ξ^μ , 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

$$E_{n_c}(\{\sigma_j\}) := \sum_{\mu=1}^M |m_\mu|^{n_c} \Theta(-m_\mu), \quad (9)$$

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

Finding optimal solutions σ^* where $E_{n_c}(\sigma^*) = 0$ is a hard optimization problem for either choice of $n_c = 0, 1$: it has been shown that the energy landscape, in the limit of large N and for $M = \alpha N$ with $\alpha < \alpha_c \approx 0.83$ [53], 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) [54], 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 [43]. Very schematically, there exist exponentially rare regions characterized by dense ground states: 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 the training is carried out, since they are less likely to fit noise (small generalization error).

It has been conjectured and shown in Ref. [34]—with analytical and numerical evidence—that quantum fluctuations, as encoded by a path-integral Monte Carlo simulated QA [5], are particularly effective in exploring these “dense” regions. One of the aims of this paper is to provide numerical evidence of enhanced effectiveness of QAOA over QA for small-size perceptron instances, where unitary evolutions are computationally feasible to directly compare QA and QAOA.

The standard quantum mapping of the binary synaptic weights consists in promoting the classical spins σ_j to quantum spin-1/2 Pauli operators $\hat{\sigma}_j^z$, as a special case of the procedure schematized in Eq. (1). Let us remark that this standard mapping is not the only possibility to encode classical bits into a quantum setup. A recent paper [51] has implemented a quantum version of the perceptron model within 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, but it pays the price of an exponentially large number of quantum gates necessary for the state preparation [51]. 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 encoding $\sigma_j \rightarrow \hat{\sigma}_j^z$.

The target Hamiltonian associated to the perceptron is then given by

$$\hat{H}_{\text{targ}} = E_{n_c}(\{\hat{\sigma}_j^z\}). \quad (10)$$

This Hamiltonian has a complicated expression in terms of the quantum spin variables $\{\hat{\sigma}_j^z\}$, due to the Heaviside step function: in principle, it involves all possible interactions among spins, up to N -body terms. Nonetheless, the expectation value in Eq. (7) can still be estimated by a sample mean over a set of measurements in the computational basis, by repeatedly preparing the variational state in Eqs. (5) and (6) for fixed variational parameters. Each evaluation of the classical cost function still has $\mathcal{O}(1)$ complexity, independently of its expansion in terms of binary spins. On the other hand, the operator $\hat{H}_{\text{targ}} = \hat{H}_z$ is also used in building up the QAOA variational state in Eqs. (5) and (6). This structure is shared by other relevant optimization problems, such as the financial crash models considered in Ref. [55], where the target Hamiltonian also involves Heaviside functions. Whereas an exact gate decomposition of the unitaries $e^{-iy_m \hat{H}_z}$ is always possible, it would require impractical resources, i.e., an exponential number of elementary gates in the number of qubits. However, as anticipated above, the QAOA-like *ansatz* allows for some flexibility in the choice of the \hat{H}_z term appearing in the variational state, with the possibility of replacing it with a simpler set of quantum gates, a point that will be further discussed in Sec. V.

III. RESULTS

In order 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 (see Supporting Information in Ref. [34]). 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 $N \rightarrow \infty$, beyond which zero-energy solutions may no longer exist.

Following Ref. [34], 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 nonconvex 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 [34], 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*.

A. Optimal digitized-QA

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

$$\varepsilon_P(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{N} E_P(\boldsymbol{\beta}, \boldsymbol{\gamma}), \quad (11)$$

where P is fixed. For QAOA, we aim at minimizing this quantity with respect to the independent free parameters $\boldsymbol{\beta}$ and

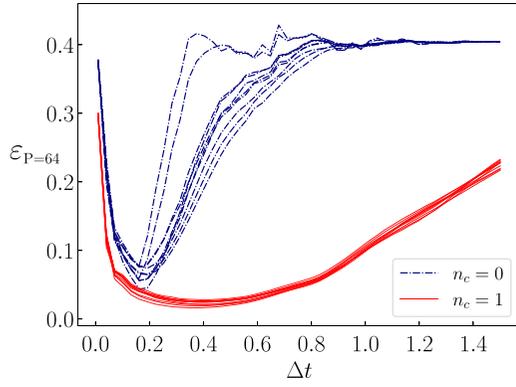


FIG. 2. The one-dimensional landscape of the variational energy density for digitized-QA, where the parameters β and γ depend only on Δt , see Eq. (4). All samples examined and both choices of $n_c = 0, 1$ are shown. The qualitative features of the landscape, and in particular the position of global minima, show mild sample-to-sample variability.

γ ; on the contrary, for digitized-QA (dQA), these are fixed as in Eq. (4) and contain Δt as a single free parameter, which we can optimize accordingly.

Focusing on dQA, for any given P we obtain a unique well-defined global minimum for the variational energy density vs Δt . This is shown, e.g., in Fig. 2 for $P = 64$ and for the two different definitions of the cost-function, $n_c = 0$ and $n_c = 1$.

As previously discussed, see also [45], 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 $\tau = P\Delta t$; 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 a first hint of general qualitative features of the model, independent of the specific sample under consideration. Additional numerical results on digitized-QA are reported in Appendix B.

B. Smooth QAOA solutions

In practical implementations of QAOA, both the choice of the classical optimization algorithm—local (gradient-based), vs global—and of the starting point for the optimization routine can be relevant, particularly when the dimensionality $2P$ of the search space grows, making the optimization harder.

The simplest approach would be to use a local gradient-based algorithm starting from a random initialization of the

parameters, but in practice this is often ineffective, due to the presence of many low-quality local minima. Moreover, recent evidence proved the ubiquitous existence of barren plateaus for variational quantum algorithms [48,56]—i.e., gradients of the objective function that are exponentially vanishing in the system size N , when the variational energy landscape is sampled at random points—a phenomenon that calls for effective warm-start strategies, further discouraging random initialization: we shall not adopt it here.

Effective heuristic warm-start strategies have been proposed, which are based on iterative procedures empirically yielding far-better quality results than a random start. These strategies adopt an iterative optimization with an increasing number of variational parameters, recursively updating a warm-start yielded by the previous step. In practice, this is done by adopting Fourier-based algorithms [18], or by interpolating from previous smaller P solutions [16,18]. Heuristically, at each step, the current optimal solution provides a very good warm-start for the next step, e.g., from $P \rightarrow P'$ (with $P' > P$, such as $P' = P + 1$). Moreover, such methods have been observed to yield high-quality optimal or quasi-optimal solutions (β, γ) that appear to be *smooth* when plotted versus the Trotter number m , albeit sometimes strongly departing from the linear choice of Eq. (4). In principle, smoothness in m of optimal or quasi-optimal solutions, and their transferability, e.g., from $P \rightarrow P' = P + 1$, are two different aspects. However, in practice, optimal or quasi-optimal nonsmooth solutions often fail in providing a good warm-start [42].

In this paper we adopt the following strategy. Let us denote by $(\beta^{\text{dQA}}, \gamma^{\text{dQA}})$ the optimal linear-choice that a digitized-QA provides, as discussed above. Using $(\beta^{\text{dQA}}, \gamma^{\text{dQA}})$ as a starting point for a Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm [57,58], we find a minimum, denoted by $(\beta^{(1)}, \gamma^{(1)})$, 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. [59–61], 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 $(\beta^{(1)}, \gamma^{(1)})$ and restarting a second BFGS optimization. This leads to a final solution $(\beta^{(2)}, \gamma^{(2)})$ that is found to be *smooth*, and to provide a systematically better variational minimum compared to the spurious minimum $(\beta^{(1)}, \gamma^{(1)})$. Schematically, here is the procedure adopted:

$$(\beta^{\text{dQA}}, \gamma^{\text{dQA}}) \rightarrow \boxed{\text{BFGS optim.}} \xrightarrow{\text{QAOA-1}} (\beta^{(1)}, \gamma^{(1)}) \rightarrow \boxed{\text{Smoothing + BFGS optim.}} \xrightarrow{\text{QAOA-2}} (\beta^{(2)}, \gamma^{(2)}). \quad (12)$$

The smoothing procedure can be carried out either by hand or by locally interpolating the curve $(\beta^{(1)}, \gamma^{(1)})$ 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 $(\beta^{(2)}, \gamma^{(2)})$.

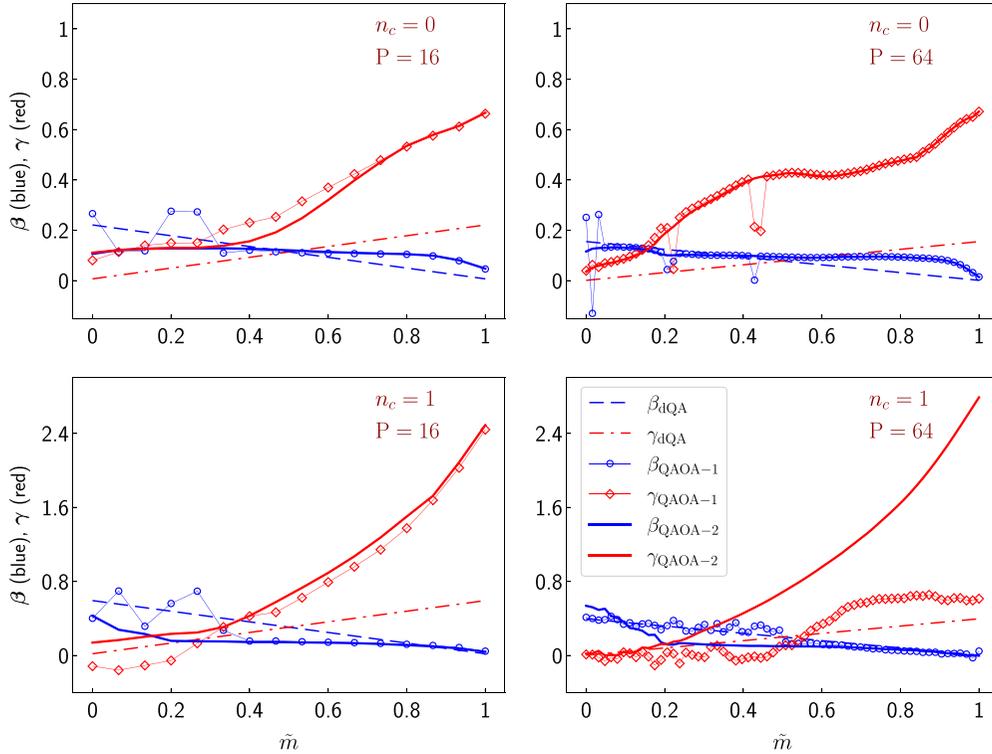


FIG. 3. Results for the optimal digitized-QA protocol (dashed and dash-dotted straight lines), QAOA-1 (open symbols with dotted lines), and QAOA-2 (solid lines), for $P = 16, 64$ (left to right) and for $n_c = 0, 1$ (top to bottom). We adopt a uniform x axis scale in terms of $\tilde{m} = (m - 1)/(P - 1) \in [0, 1]$.

We now move to illustrate our results in more detail. We performed digitized-QA and QAOA classical simulations for both $n_c = 0, 1$: in this framework, we computed the exact QAOA final state in Eq. 5 and the corresponding variational energy in Eq. (7) 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.

Figure 3 illustrates the results obtained for this sample. We show two representative 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 $(\beta^{\text{dQA}}, \gamma^{\text{dQA}})$ solutions. The empty symbols denote the optimal $(\beta^{(1)}, \gamma^{(1)})$ solutions obtained by a BFGS minimization starting from $(\beta^{\text{dQA}}, \gamma^{\text{dQA}})$: our “first shot of QAOA”, labeled 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. (12). On the other hand, for $n_c = 1$, irregularities of QAOA-1 solutions are more diffuse, and the procedure was slightly modified: We run the second BFGS local minimization from a warm-start point, obtained by interpolation [16,18] from a smoothed $P = 16$ solution, in power-of-two steps, hence from $P = 16 \rightarrow 32 \rightarrow 64 \dots$. In both cases, the resulting smooth solutions $(\beta^{(2)}, \gamma^{(2)})$ are labeled as QAOA-2 and denoted by solid lines. In Appendix A, we summarize 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. 4 we compare the obtained minima of ε_P , Eq. (11), for the optimal digitized-QA, QAOA-1 and QAOA-2 protocols. These results show a striking gain by applying QAOA, both for $n_c = 0$ 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, see Appendix A for details. 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 true for all randomly generated training set samples.

C. Transferability of a smooth ansatz

The procedure described above to obtain smooth solutions is elaborate and time-consuming. Luckily enough, it does not need to be repeated for each different training set sample.

Indeed, a *leitmotif* of recent literature on QAOA applications are *concentration* effects of the variational energy landscape: for any given P , typical instances of the same problem often yield a similar QAOA energy landscape. When this is the case, QAOA optimal parameters computed for the first instance often serve as an excellent warm-start for a local

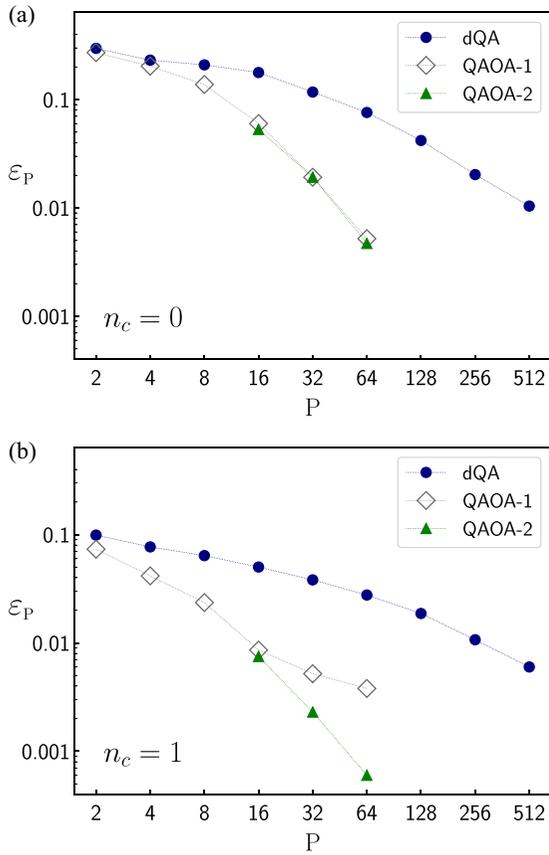


FIG. 4. Comparison of variational energy density minima for digitized-QA, QAOA-1 and QAOA-2 for a specific sample of random patterns, with $n_c = 0$ (a) and $n_c = 1$ (b). QAOA-1 outperforms digitized-QA, especially for large values of P . This gain can be further enhanced with the smooth QAOA-2 solution (see main text).

optimization for other instances, with a significant reduction of the computational cost.

While this result can be formally understood, sometimes, in terms of light-cone spreading of quantum correlations, e.g., for MaxCut problems on regular graphs with $P \ll N$ [38,40], a growing body of numerical evidence hints at concentration effects and parameter reusability in different regimes (large P) and for different models, often without a complete formal understanding. A remarkable exception is offered by [39], where the authors prove concentration for the Sherrington-Kirkpatrick (SK) model in the large- N limit, although the infinite range of two-body interactions hinders an intuitive understanding in terms of light-cone spreading of correlations.

In the following, we show numerically that similar concepts also apply to the perceptron model, which does not even admit a k -local cost function: QAOA smooth solutions are transferable among different instances of the perceptron model (i.e., for different training set samples). To show this, we adopt the following strategy: For any fixed value of P and for $n_c = 0, 1$, separately, we consider QAOA-2 optimal angles $(\beta^{(2)}, \gamma^{(2)})$ for our first sample and take them as a *smooth model-dependent ansatz* $(\beta^{\text{Ansatz}}, \gamma^{\text{Ansatz}})$ used as initial point for a BFGS minimization of a *different* sample. In this way, we are able to find *smooth* optimal solutions for all other samples, as illustrated in Fig. 5. Remarkably, these smooth solutions are

qualitatively coincident with the $(\beta^{(2)}, \gamma^{(2)})$ solutions that one would construct by adopting the QAOA-2 procedure previously outlined.

From a practical standpoint, by starting from the smooth *ansatz* $(\beta^{\text{Ansatz}}, \gamma^{\text{Ansatz}})$, 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 $(\beta^{\text{Ansatz}}, \gamma^{\text{Ansatz}})$ 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 classical hardness or large number of zero-energy configurations (see Appendix C).

IV. THE ROLE OF THE COST-FUNCTION LANDSCAPE GEOMETRY

As discussed in Ref. [34], quantum fluctuations are particularly efficient in exploring exponentially rare *dense* regions of solutions in the classical cost-function landscape, defined by Eq. (9). These dense regions are characterized by a large number of classical solutions clustering within relatively small Hamming distance. This geometrical structure of the landscape is linked to the good QA performance: The instantaneous spectral gap “seen” by the QA dynamics only closes when approaching the end of the protocol $s \rightarrow 1$, where $\hat{H}(s) \rightarrow \hat{H}_{\text{target}}$, due to the degeneracy of classical solutions.

We can scramble this geometrical structure by permuting the classical energies associated to each configuration; while keeping the spectrum unchanged, this procedure yields an unstructured problem, as previously done also in Refs. [62,63]. This scrambling was shown to be detrimental to QA in a general setting [62] and for our specific case [34]: 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 a scrambling of the cost-function landscape geometry and with 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 be a signal of an “optimal adiabatic schedule” [16], and this might suggest a worse performance.

To answer such a question, we have adopted the following strategy. For each sample considered, we generate a corresponding *randomized sample*, by permuting the classical energies associated to each configuration σ , so as to retain the same classical energy spectrum, while completely 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 same full spectra of the original samples.

We then proceed along the same lines of Sec. III: 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. By following the same scheme outlined in Eq. (12), it is again

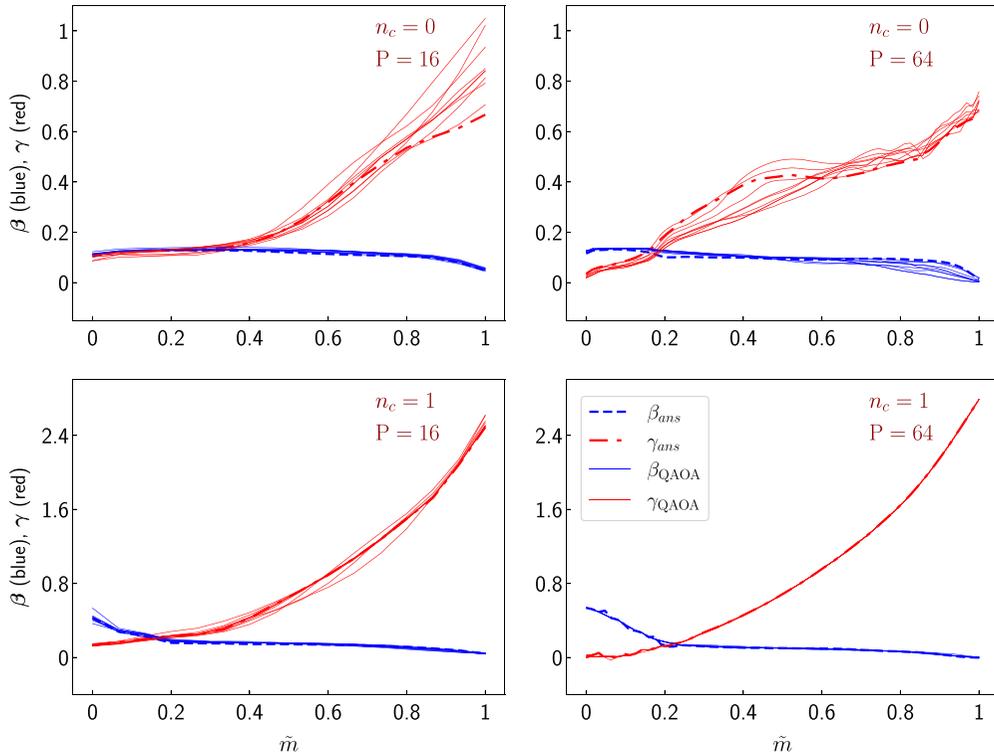


FIG. 5. QAOA optimal protocols for all tested samples, as obtained by a BFGS minimization starting from the same *smooth ansatz* ($\beta^{\text{Ansatz}}, \gamma^{\text{Ansatz}}$), corresponding to smooth optimal schedules for the first sample (see main text). Data shown for $P = 16, 64$ (left to right) and for $n_c = 0, 1$ (top to bottom). We adopt a uniform x -axis scale in terms of $\tilde{m} = (m - 1)/(P - 1) \in [0, 1]$. 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.

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). In Appendix B, we carry out a closer comparison between digitized-QA optimal- Δt values for the two cases of original vs randomized samples.

Figure 6 compares the minimized variational energy density obtained from digitized-QA and QAOA-2 for the randomized version of the sample reported in Fig. 4; the original digitized-QA and QAOA-2 results are also reported, for a direct 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 and (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 comments apply to all samples examined.

To better understand the basic mechanism behind such a degradation of performance, we plot in Fig. 7 the QAOA-2 smooth protocols for $P = 64$, reexpressed in terms of $s_m = \gamma_m/(\gamma_m + \beta_m)$, a parameter, which in digitized-QA linearly

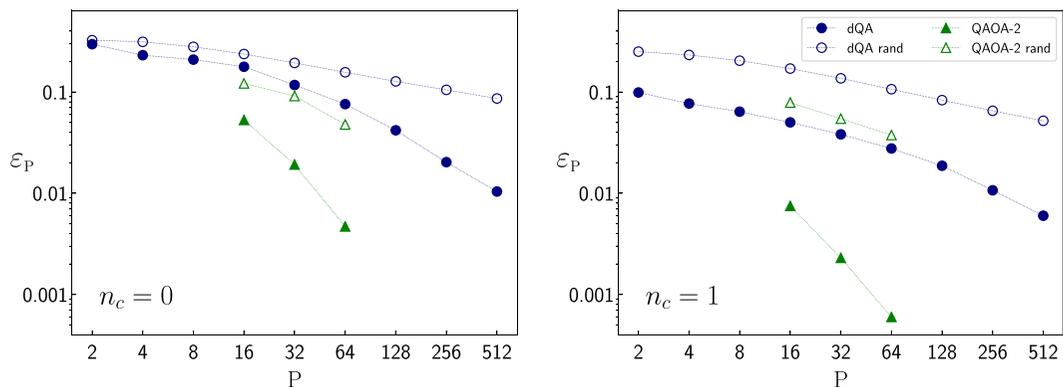


FIG. 6. Variational energy density minima for digitized-QA and QAOA-2, comparing results for the original sample (the same as in Fig. 4, full symbols) with those obtained by a *randomization* of its energy spectrum (empty symbols), for both $n_c = 0, 1$.

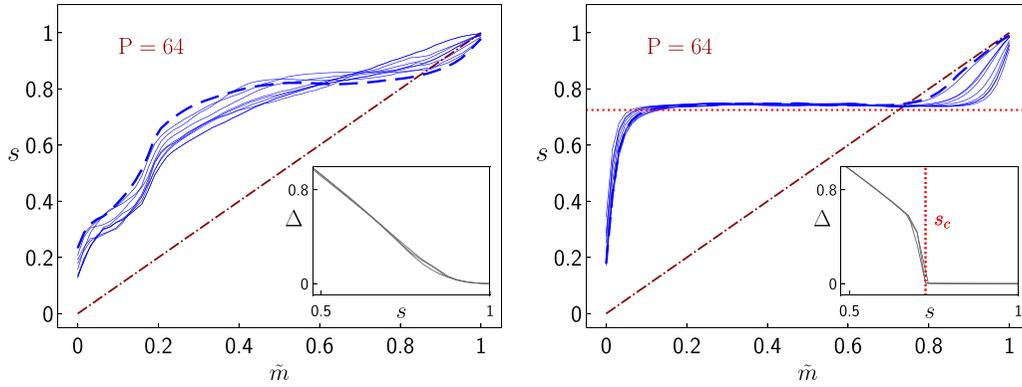


FIG. 7. Plot of the QAOA-2 smooth protocols for $P = 64$, in terms of $s_m = \gamma_m / (\gamma_m + \beta_m)$ for both the original (left) and the randomized samples (right), in the case $n_c = 0$. In both figures, the thick-dashed line represents the transferable smooth *ansatz* obtained by a detailed study of the first sample, while data for other samples is 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 $\Delta(s) = E_{\text{ex}}(s) - E_{\text{gs}}(s)$: the gap starts at $\Delta(s = 0) = 2$ (single spin-flip excitation of \hat{H}_x) and it vanishes for $s = 1$, due to the degeneracy of \hat{H}_x , 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 $n_c = 1$.

interpolates from $s = 0$ to $s = 1$ during the annealing process, see Eq. (4). 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 (see Supporting Information in Ref. [34]), an avoided level-crossing close to the numerical value $s_c = 0.725$. Correspondingly, close to this value, 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” [16], unfortunately unable to fully overcome the basic limitations of the adiabatic mechanism. Moreover, these QAOA-2 smooth solutions show striking similarity with adiabatic schedules obtained for unstructured (Grover) search in Refs. [64,65].

However, nothing guarantees that our QAOA-2 smooth solutions are the true global minimum in 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 [18]. The STA strategy goes beyond the adiabatic paradigm and is a promising framework to overcome shortcomings of the adiabatic mechanism. References [66–68] 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 [69] to modify the linear guess inspired by dQA to specifically target STA protocols.

V. DISCUSSION AND CONCLUSIONS

In our paper we provided encouraging evidence for the potential applicability and effectiveness of digitized-QA and QAOA in the realm of hard classical optimization problems

with highly nonlocal Hamiltonians, well-beyond the usual 2-local models considered both in MaxCut and quantum spin chains ground-state preparation.

Moreover, we devised an optimization scheme that leverages the transferability of optimal *smooth* QAOA schedules among different instances of the same problem (different training sets). These findings are reminiscent of previous results on parameter concentration (or instance independence) and might be further investigated, possibly leading to analytical results in the large N limit, as in Refs. [39,70].

Inspired by the results of Ref. [34], we also analyzed the role of the geometry of the classical cost-function landscape, by artificially permuting the energy spectrum. This leads to a vanishing spectral gap along the annealing path: While still providing some advantage vs an optimized- Δt linear schedule QA, QAOA seems to perform a kind of “optimal adiabatic schedule”, unable to fully overcome the basic limitations of the adiabatic mechanism.

Concerning future developments, it is an open research line to tailor smart approximation schemes, with the goal of efficiently implementing on actual quantum hardware unitaries generated by a highly nonlocal classical Hamiltonian—paving the way to experiments beyond classical simulation capabilities. Nevertheless, we highlight that similar nontrivial tasks need to be solved also for the implementation of such models on a standard quantum annealer [34]. In this respect, one possibility is a further extension of QAOA, proposed in Ref. [22]. The main idea is that, by looking back at Eqs. (5) and (6), one can redefine the variational energy in Eq. (7) as follows: (1) keep the same exact classical $\hat{H}_{\text{targ}} = \hat{H}_z$ in the objective function expectation value and (2) redefine the diagonal unitary (which is quantum-computationally hard) generated by \hat{H}_z —e.g., by using a simplified version of it, encoding some minimal information on the problem. As an additional possibility, one can even redefine the *mixing* unitary generated by \hat{H}_x , replacing it with another operator inducing tailored quantum fluctuations. One would then simply repeat the

QAOA prescription, but with a different—and possibly much more efficient—quantum gate decomposition. In contrast, note that the expectation value, in a realistic quantum device, is simply estimated by computing the sample mean of $\hat{H}_{\text{arg}} = \hat{H}_z$ for a given number of shots, each requiring only a measurement on the computational basis: each evaluation of \hat{H}_{arg} for a classical configuration σ is a trivial task, regardless of its nonlocality.

Another route, in principle, would be to use a parameterized quantum circuit agnostic about the problem Hamiltonian, tailored on the resources of available quantum devices (e.g., native gates and qubit connectivity) [24,25]. This approach, while potentially beneficial on the implementation side, might significantly suffer from well-known limitations such as local minima and barren plateaus. An efficient implementation of an effective parameterized quantum circuit might pose new interesting questions, concerning not only the presence of optimal smooth protocols and their transferability, but also their robustness to shot-noise (due to finite-sample mean estimates of the variational energy) or gate errors, effects worth studying only once an efficient gate implementation is found. A complementary approach relies on performing *efficient* classical simulations of quantum optimization methods, e.g., relying on tensor networks techniques [71].

A simple gate-efficient version of QAOA, which we have tested, is the following. Rather than using $\hat{H}_{\text{arg}} = \hat{H}_z$ in the quantum gates $e^{-i\gamma_m \hat{H}_z}$, we use the Sherrington-Kirkpatrick (SK) model Hamiltonian, which derives from taking the quadratic approximation ($|m_\mu| \Theta(-m_\mu) \rightarrow -m_\mu + m_\mu^2$ in the $n_c = 1$ cost function in Eq. (9), which, upon using Eq. (8), leads to

$$\hat{H}_z = - \sum_{j=1}^N h_j \hat{\sigma}_j^z + \sum_{j \neq j'} J_{jj'} \hat{\sigma}_j^z \hat{\sigma}_{j'}^z, \quad (13)$$

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 [50]. Unfortunately, such a choice of \hat{H}_z appears to dramatically decrease the performance of QAOA. A possibly smarter choice to improve performance, while maintaining a Hamiltonian with only two-body interactions, could be to search for an optimal SK model, either by iteratively identifying the optimal J and h parameters or by defining an appropriate inverse Ising model [72]. These options will be the subject of future research.

A complementary and promising route for future research is the extension of this work to the training of more complex discrete-weights ANNs, as well as considering training sets with correlations (data with a structure). This would allow to better inquire into the role of the classical cost-function landscape geometry, which is expected to be crucial for the effectiveness of QAOA or similar quantum optimization schemes. Incidentally, we remark that alternative cost-function definitions (i.e., different cost for a misclassified pattern) can be adopted, leading to a change of the whole classical energy landscape but zero-energy solutions. It might prove insightful to assess whether the classical cost-function

definition can be chosen to enhance the effectiveness of quantum optimization.

Interestingly, a recent paper [73] has shown that purely classical strategies using recurrent neural networks—denoted as “variational neural annealing”—can improve significantly on Simulated Annealing and simulated-QA on standard benchmark problems. Our study confirms that the perceptron model appears to belong to a class of problems where quantum effects efficiently exploit the geometry of the classical energy landscape, much better than, e.g., simulated annealing: it is therefore an ideal candidate to test and benchmark the variational neural annealing strategies against competing quantum algorithms.

ACKNOWLEDGMENTS

We thank M. Collura and G. Lami for stimulating discussions. The research was partly supported by EU Horizon 2020 under ERC-ULTRADISS, Grant Agreement No. 834402. G.E.S. acknowledges that his research has been conducted within the framework of the Trieste Institute for Theoretical Quantum Technologies (TQT).

APPENDIX A: QAOA-2 PRACTICAL IMPLEMENTATION

As mentioned in the main text, two slightly different procedures are adopted in order to single out smooth optimal QAOA-2 solutions.

In fact, some qualitative differences arise between QAOA-1 results for $n_c = 0$ and $n_c = 1$, which are visible in Fig. 3 (empty symbols) for the first training set sample, but are present for all the other samples in exam. Concisely, we observe that for $n_c = 0$ the QAOA-1 optimal parameters ($\beta^{(1)}, \gamma^{(1)}$) are noticeably different from ($\beta^{\text{dQA}}, \gamma^{\text{dQA}}$) 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. (12) (smoothing, second BFGS

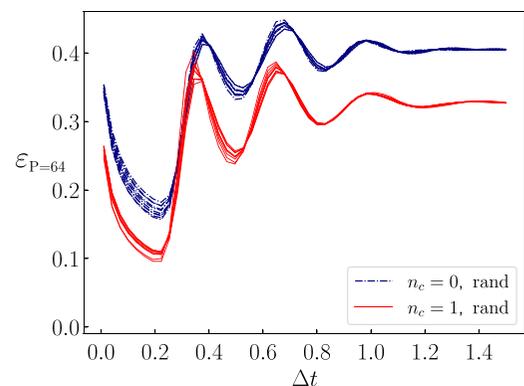


FIG. 8. The one-dimensional landscape of the variational energy density [Eq. (11)] for digitized-QA, where the parameters β and γ depend only on Δt , see Eq. (4). All *randomized* samples and both choices of $n_c = 0, 1$ are shown. We remark that the landscape, and in particular the position of global minima, show mild sample-to-sample variability.

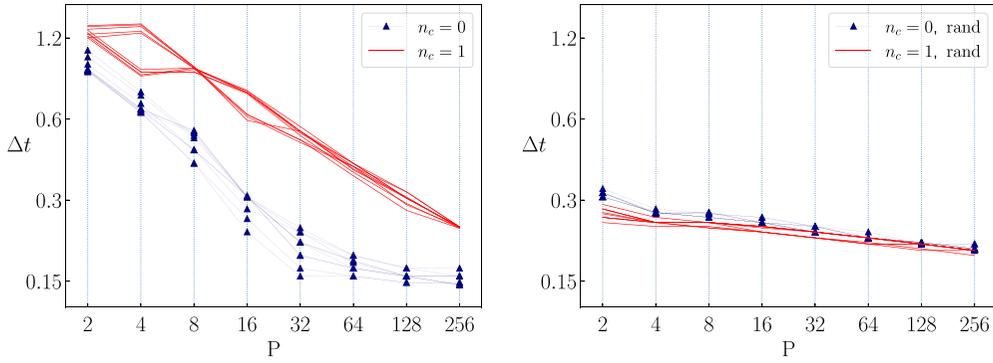


FIG. 9. Optimal digitized-QA Δt values for increasing P , with logarithmic scales on both axes. The original samples case (left) compared to the randomized samples (right). The presence of evident clusters shows mild sample-to-sample variability, with few exceptions (as $P = 4$ for $n_c = 1$, in the left panel) due to an almost-flat energy landscape in that range.

minimization), which is applied straightforwardly, e.g., for $P = 16, 64$, yielding the smooth ($\beta^{(2)}, \gamma^{(2)}$) protocols in Fig. 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 to get “trapped” in a neighborhood of ($\beta^{\text{dQA}}, \gamma^{\text{dQA}}$), also displaying more extended high-frequency oscillations in the optimal parameters ($\beta^{(1)}, \gamma^{(1)}$). This numerical evidence calls for a slightly different approach: we simply apply Eq. (12) 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., by means of a linear increment in P at each iteration.

Consistently with our intuition, one can check in Fig. 4 that 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. 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.

In conclusion, we wish to underline that—in light of the discussion upon the transferability of the ansatz (see Fig. 5)—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 *ansatz* for all other randomly generated training sets, yielding an effective unique procedure to find smooth QAOA solutions outperforming optimal digitized-QA, valid for both $n_c = 0, 1$.

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.

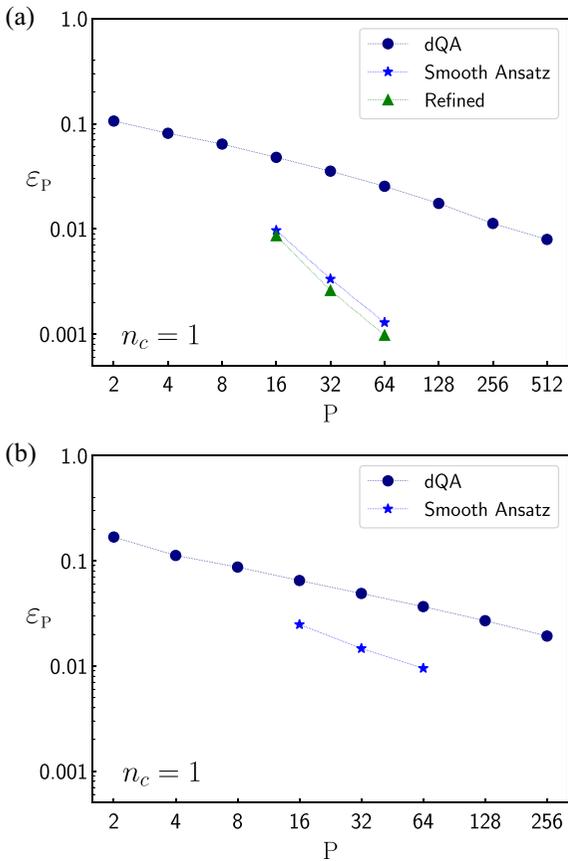


FIG. 10. (a), Variational energy density averaged over the same set of samples analyzed in Fig. 5. We plot the variational energy density for optimal dQA, for the transferred *smooth ansatz* and after a refinement optimization. (b) Variational energy density averaged over a new batch of 50 random samples. The transferred *smooth ansatz* still outperforms an optimal dQA. The same results are found for $n_c = 0$ (data not shown).

APPENDIX B: ADDITIONAL RESULTS ON DIGITIZED-QA

In this section, we report additional numerical results on optimal- Δt digitized-QA, in particular by drawing a comparison between ordered and randomized samples.

In Fig. 8 we show that, also in the randomized scenario, the Δt landscape and the position of minima are almost identical for all samples in exam: we show data for $P = 64$ and both definitions of the cost function $n_c = 0, 1$, to be compared with Fig. 2 for the original samples.

The validity of these results naturally extends to different values of P , as summarized in Fig. 9, where we display the optimal values of Δt vs P for the original samples (left panel) as well as for 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.

APPENDIX C: ADDITIONAL RESULTS ON SMOOTH ANSATZ TRANSFERABILITY

Following on the discussion on the transferability of a *smooth ansatz* among different training sets, in Fig. 10

(top 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. 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 reoptimizing the QAOA variational parameters for the new training set. The same results hold true also for $n_c = 0$. As visible in Fig. 5, for $n_c = 1$ the reoptimized 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 on the hardness for a classical optimization. This is shown in Fig. 10 (bottom panel): also in this case a transferred *smooth ansatz* outperforms an optimal dQA, without the need of a further optimization. The same results are verified for $n_c = 0$. As 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 [34].

-
- [1] M. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000)
 - [2] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Quantum annealing: A new method for minimizing multidimensional functions, *Chem. Phys. Lett.* **219**, 343 (1994).
 - [3] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse Ising model, *Phys. Rev. E* **58**, 5355 (1998).
 - [4] J. Brooke, D. Bitko, T. F. Rosenbaum, and G. Aeppli, Quantum annealing of a disordered magnet, *Science* **284**, 779 (1999).
 - [5] G. E. Santoro, R. Martoňák, E. Tosatti, and R. Car, Theory of quantum annealing of an Ising spin glass, *Science* **295**, 2427 (2002).
 - [6] G. E. Santoro and E. Tosatti, Optimization using quantum mechanics: Quantum annealing through adiabatic evolution, *J. Phys. A: Math. Gen.* **39**, R393 (2006).
 - [7] 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).
 - [8] T. Albash and D. A. Lidar, Adiabatic quantum computation, *Rev. Mod. Phys.* **90**, 015002 (2018).
 - [9] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, *et al.*, Quantum annealing with manufactured spins, *Nature (London)* **473**, 194 (2011).
 - [10] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
 - [11] 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, *Nat. Commun.* **5**, 4213 (2014).
 - [12] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028).
 - [13] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
 - [14] S. Lloyd, Quantum approximate optimization is computationally universal, [arXiv:1812.11075](https://arxiv.org/abs/1812.11075).
 - [15] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás, On the universality of the quantum approximate optimization algorithm, *Quant. Info. Proc.* **19**, 291 (2020).
 - [16] G. Bigan Mbeng, R. Fazio, and G. Santoro, Quantum annealing: A journey through digitalization, control, and hybrid quantum variational schemes, [arXiv:1906.08948](https://arxiv.org/abs/1906.08948).
 - [17] L. T. Brady, L. Kocia, P. Bienias, A. Bapat, Y. Kharkov, and A. V. Gorshkov, Behavior of analog quantum algorithms, [arXiv:2107.01218](https://arxiv.org/abs/2107.01218).
 - [18] 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).
 - [19] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, *et al.*, Quantum approximate optimization with a trapped-ion

- quantum simulator, *Proc. Natl. Acad. Sci. USA* **117**, 25396 (2020).
- [20] W. W. Ho and T. H. Hsieh, Efficient variational simulation of non-trivial quantum states, *SciPost Phys.* **6**, 029 (2019).
- [21] 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).
- [22] 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).
- [23] X. Liu, A. Angone, R. Shaydulin, I. Safro, Y. Alexeev, and L. Cincio, Layer VQE: A variational approach for combinatorial optimization on noisy quantum computers, *IEEE Trans. Quantum Eng.* **3**, 1 (2022).
- [24] E. Farhi, J. Goldstone, S. Gutmann, and H. Neven, Quantum algorithms for fixed qubit architectures, [arXiv:1703.06199](https://arxiv.org/abs/1703.06199).
- [25] 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 (London)* **549**, 242 (2017).
- [26] J. Preskill, Quantum Computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [27] 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).
- [28] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5 (2014).
- [29] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [30] 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).
- [31] 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, Amsterdam, 2007), pp. 3–24.
- [32] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature (London)* **549**, 195 (2017).
- [33] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, *Quantum Sci. Technol.* **4**, 043001 (2019).
- [34] C. Baldassi and R. Zecchina, Efficiency of quantum vs. classical annealing in nonconvex learning problems, *Proc. Natl. Acad. Sci. USA* **115**, 1457 (2018).
- [35] 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).
- [36] G. B. Mbeng, R. Fazio, and G. E. Santoro, Optimal quantum control with digitized quantum annealing, [arXiv:1911.12259](https://arxiv.org/abs/1911.12259).
- [37] M. Streif and M. Leib, Training the quantum approximate optimization algorithm without access to a quantum processing unit, [arXiv:1908.08862](https://arxiv.org/abs/1908.08862).
- [38] 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, [arXiv:1812.04170](https://arxiv.org/abs/1812.04170).
- [39] 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).
- [40] A. Galda, X. Liu, D. Lykov, Y. Alexeev, and I. Safro, Transferability of optimal QAOA parameters between random graphs, [arXiv:2106.07531](https://arxiv.org/abs/2106.07531).
- [41] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte, Parameter concentrations in quantum approximate optimization, *Phys. Rev. A* **104**, L010401 (2021).
- [42] 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).
- [43] 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, *Proc. Natl. Acad. Sci. USA* **113**, E7655 (2016).
- [44] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. L. Heras, R. Babbush, A. G. Fowler, B. Campbell, Y. Chen, *et al.*, Digitized adiabatic quantum computing with a superconducting circuit, *Nature (London)* **534**, 222 (2016).
- [45] G. B. Mbeng, L. Arceci, and G. E. Santoro, Optimal working point in digitized quantum annealing, *Phys. Rev. B* **100**, 224201 (2019).
- [46] 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, *Phys. Rep.* **523**, 127 (2013).
- [47] T. S. Cubitt, D. Perez-Garcia, and M. M. Wolf, Undecidability of the spectral gap, *Nature (London)* **528**, 207 (2015).
- [48] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
- [49] F. Rosenblatt, *The Perceptron - A Perceiving and Recognizing Automaton*, Tech. Rep. 85-460-1 (Cornell Aeronautical Laboratory, Ithaca, 1957).
- [50] J. A. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (CRC Press, Boca Raton, 2018).
- [51] F. Tacchino, C. Macchiavello, D. Gerace, and D. Bajoni, An artificial neuron implemented on an actual quantum processor, *npj Quantum Inf.* **5**, 26 (2019).
- [52] M. Schuld, I. Sinayskiy, and F. Petruccione, Simulating a perceptron on a quantum computer, *Phys. Lett. A* **379**, 660 (2015).
- [53] W. Krauth and M. Mézard, Storage capacity of memory networks with binary couplings, *J. Phys. France* **50**, 3057 (1989).
- [54] H. Horner, Dynamics of learning for the binary perceptron problem, *Z. Phys. B* **86**, 291 (1992).
- [55] R. Orús, S. Múgel, and E. Lizaso, Forecasting financial crashes with quantum computing, *Phys. Rev. A* **99**, 060301(R) (2019).
- [56] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).

- [57] J. Nocedal and S. Wright, *Numerical Optimization* (Springer Science & Business Media, New York, 2006).
- [58] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* **17**, 261 (2020).
- [59] E. R. Anschuetz and B. T. Kiani, Beyond barren plateaus: Quantum variational algorithms are swamped with traps, *Nat. Commun.* **13**, 7760 (2022).
- [60] 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 (PMLR, 2021), Vol. 139, pp. 12144–12155.
- [61] E. R. Anschuetz, Critical points in quantum generative models, in *International Conference on Learning Representations* (2022).
- [62] E. Farhi, J. Goldstone, S. Gutmann, and D. Nagaj, How to make the quantum adiabatic algorithm fail, [arXiv:quant-ph/0512159](https://arxiv.org/abs/quant-ph/0512159).
- [63] A. Callison, N. Chancellor, F. Mintert, and V. Kendon, Finding spin glass ground states using quantum walks, *New J. Phys.* **21**, 123022 (2019).
- [64] J. Roland and N. J. Cerf, Quantum search by local adiabatic evolution, *Phys. Rev. A* **65**, 042308 (2002).
- [65] 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).
- [66] 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. Res.* **4**, 013141 (2022).
- [67] J. Wurtz and P. J. Love, Counterdiabaticity and the quantum approximate optimization algorithm, *Quantum* **6**, 635 (2022).
- [68] 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).
- [69] G. B. Mbeng and W. Lechner, Rotated ansatz for approximate counterdiabatic driving, [arXiv:2207.03553](https://arxiv.org/abs/2207.03553).
- [70] 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).
- [71] G. Lami, P. Torta, G. E. Santoro, and M. Collura, Quantum Annealing for Neural Network optimization problems: A new approach via Tensor Network simulations, [arXiv:2208.14468](https://arxiv.org/abs/2208.14468).
- [72] H. C. Nguyen, R. Zecchina, and J. Berg, Inverse statistical problems: From the inverse Ising problem to data science, *Adv. Phys.* **66**, 197 (2017).
- [73] M. Hibat-Allah, E. M. Inack, R. Wiersema, R. G. Melko, and J. Carrasquilla, Variational neural annealing, *Nat. Mach. Intell.* **3**, 952 (2021).