communications physics

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42005-025-02423-4

Efficient online quantum circuit learning with no upfront training

Tom O'Leary [1,2,9] ✉, Piotr Czarnik [3,4,9] ✉, Elijah Pelofske [5], Andrew T. Sornborger [6,7], Michael McKerns [6,8] & Lukasz Cincio [1,7]

Optimization is a promising candidate for studying the utility of variational quantum algorithms (VQAs). However, evaluating cost functions using quantum hardware introduces runtime overheads that limit exploration. Surrogate-based methods can reduce calls to a quantum computer, yet existing approaches require hyperparameter pre-training and have been tested only on small problems. Here, we show that surrogate-based methods can enable successful optimization at scale, without pre-training, by using radial basis function interpolation (RBF) to construct an adaptive, hyperparameter-free surrogate. Using the surrogate as an acquisition function drives hardware queries to the vicinity of the true optima. For 16-qubit random 3-regular Max-Cut instances with the Quantum Approximate Optimization Algorithm (QAOA), our method outperforms state-of-the-art approaches, without considering their upfront training costs. Furthermore, we successfully optimize QAOA circuits for 127-qubit random Ising models on an IBM processor using 10^4-10^5 measurements. Strong empirical performance demonstrates the promise of automated surrogate-based learning for large-scale VQA applications.

The prospect of a quantum computer providing a computational advantage over a classical computer is of broad interest[1–4]. Quantum processors have developed to the point of approaching the limit of classical simulability[5–13]. However, decoherence, cross-talk error, imperfect gate calibration, and measurement error, collectively referred to as noise, place limitations on current quantum devices. Variational quantum algorithms (VQAs) are a widely studied approach to utilizing pre-fault-tolerant noisy quantum computers[14], which we expect to be of use in the fault-tolerant era. VQAs are generally formulated as optimization problems, with problem solutions encoded into the minima of a cost function. Parameterized quantum circuits (PQCs) are iteratively used to prepare candidate solutions until a cost function minimum (ideally, a global minimum) is found. A classical computer updates the quantum circuit parameters at each iteration based on previously obtained cost function values. In order to minimize the effects of noise, VQAs use low-depth quantum circuits, typically scaling at most logarithmically in system size for near-term quantum computers.

There are several challenges to obtaining a quantum advantage with this approach. In particular, for a broad range of expressive PQCs, or PQCs affected by unital noise, one encounters cost function landscapes which concentrate to their average value exponentially fast in the problem size[15–20]. These cost function barren plateaus require the cost function to be sampled a prohibitive number of times to be optimized. Even noise-free cost functions with non-vanishing gradients present a challenging optimization problem, due to the rapidly increasing number of local minima with PQC depth[21,22]. Nevertheless, PQCs with a provable absence of barren plateaus and special optimization initialization strategies have been proposed to mitigate these issues[23–34]. Furthermore, it has recently been shown that cost function landscape regions without barren plateaus can be efficiently approximated classically, if access to quantum cost function data in these regions is available[13]. Moreover, it has been shown that such data can be obtained efficiently, with a shot cost scaling at most polynomially with the number of PQC parameters[35,36]. In other words, a classical surrogate of the quantum cost function can be constructed efficiently.

The promise of surrogate-based methods lies in their ability to minimize the number of evaluations of a computationally expensive cost function, C, by constructing an inexpensive surrogate cost function, C^{surr}, to serve as a lightweight approximation of C. If evaluations of the surrogate are computationally negligible, then the computational cost of optimization strongly depends on how efficiently the surrogate method can produce an accurate C^{surr} with a minimal number of calls to C. When C^{surr} has a form that

[1]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. [2]Department of Physics, Clarendon Laboratory, University of Oxford, Oxford, UK. [3]Institute of Theoretical Physics, Jagiellonian University, Kraków, Poland. [4]Mark Kac Center for Complex Systems Research, Jagiellonian University, Kraków, Poland. [5]Information Systems & Modeling, Los Alamos National Laboratory, Los Alamos, NM, USA. [6]Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA. [7]Quantum Science Center, Oak Ridge, TN, USA. [8]The Uncertainty Quantification Foundation, Wilmington, DE, USA. [9]These authors contributed equally: Tom O'Leary, Piotr Czarnik. ✉e-mail: thomas.oleary@lmh.ox.ac.uk; piotr.j.czarnik@gmail.com

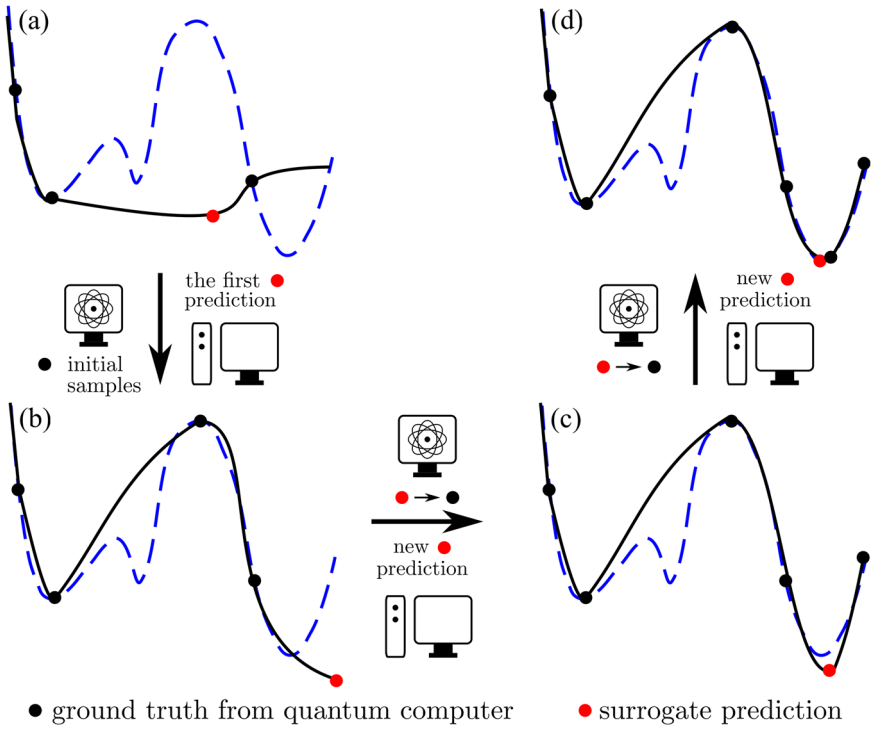

Fig. 1 | Surrogate-based optimization of parametrized quantum circuits. We use an inexpensive classical surrogate C^{surr} (black curve) to approximate the true cost function C (dashed blue curve) that is sampled using a quantum computer (black dots). **a** we fit the first surrogate using randomly sampled points from the quantum computer. We then find the surrogate minimum to determine the next point to sample on the quantum hardware (red dot). **b** we use the new point to refine our surrogate. **c**, **d** this procedure is iterated until a good approximation of the true cost minimum is found.

approximates C sufficiently well, surrogate-based optimizations of C can require orders of magnitude fewer evaluations of the true cost function, while yielding an accurate solution[37,38]. Generally, a surrogate is selected for a particular optimization problem and trained on a dataset produced by sampling the expensive cost function a large number of times, until the surrogate accurately reproduces existing results.

There is a growing body of work using classical surrogate models in VQAs[39–44] and Quantum Annealing[45,46]. Most of these approaches use Bayesian Optimization[37,38] with Gaussian Process models[47] as surrogates. The surrogate parameters are typically selected after performing upfront training, then the pre-trained surrogate is used to determine points to be evaluated on quantum hardware[39,41–44]. Some previous work also uses trust regions to constrain the parameter space for which the surrogate is valid[40,41,43]. Another approach is to use a surrogate obtained exclusively through an approximate numerical simulation of C[35,48], however, without using quantum cost data, its generality is limited[35].

From an optimization theory perspective, most modern surrogate-based approaches use a sampling method that relies on uncertainty reduction to select the new training points used to refine the surrogate. Both active learning[49] and Bayesian Optimization[37,38] commonly choose Gaussian Process[47] models as surrogates, as a Gaussian Process model can generate a mean and a variance, therefore yielding an uncertainty estimate. Information about the true response function (from the training data) is iteratively folded into the surrogate through Bayes's rule, where the surrogate can be represented by the mean of the distribution in the Gaussian Process model, and is updated each iteration when the posterior is used to inform the prior for the next iteration. Active learning and Bayesian Optimization perform optimization on an acquisition function within each iteration, to identify new training points which are used to inform the prior in the next iteration. Surrogate-based approaches generally use bound-constrained local optimization algorithms and lean heavily on incorporating information about the true response surface by iteratively informing the prior.

Motivated by these developments, we propose a surrogate-based approach that extends Bayesian optimization[37,38] to leverage recent advances in optimizer-driven sampling[50]. This approach accelerates progress toward robust automated on-device optimization of PQCs for challenging problems. Crucially, our approach does not require the computationally expensive generation of what is typically a large amount of training data for pre-training. Furthermore, by adaptively constructing a surrogate cost function, our approach is robust to noisy variable training data. We seek to use this method to minimize the total number of shots required to optimize a PQC. Such an approach is of particular importance for tackling problems that can not be simulated classically on physical devices with time-variable noise. In such a case, no prior knowledge about the device cost function landscape may be available. The method is described in detail in the Section "Methods".

While prior work has also used adaptive sampling, there are fundamental differences in our approach. We have previously shown that by applying known information about the problem through a scientific coordinate transform (essentially an operator transforming coordinate space to a constrained coordinate space), machine learning produces a surrogate whose form better approximates truth (i.e,. ground-truth on experimental data), and thus requires less training data to produce a surrogate that fits the training data well[51–53]. Notably, most implementations of Gaussian Process model-based approaches rely on uncertainty-based acquisition functions, making the implementation of the coordinate transforms challenging[47]. Our method is built to leverage these transforms in a constrained pseudo-global optimization of the surrogate. More specifically, our acquisition function directly seeks N extrema of the surrogate during each iteration, as opposed to indirectly through surrogate uncertainty reduction. By incorporating optimizer-driven sampling into our acquisition function, our method utilizes an initial exploratory step to determine the starting point of each member of an ensemble of optimizers, then proceeds with an exploitative step where the optimizers solve for the surrogate extrema. The exploratory step uses a random sampling method, such as sampling from a distribution or with an acquisition function, which essentially makes our initial step identical to Bayesian Optimization. The additional step of using an optimizer to search for the surrogate extrema serves to focus the sampling near the extrema of truth. We take the surrogate extrema as candidates, and sample the corresponding points on truth, where the error in prediction between the surrogate and truth at each iteration is used to drive the surrogate closer to truth, as shown schematically in Fig. 1. Unlike Gaussian Process model-based surrogates, our approach does not require hyperparameters specifying the correlations to the existing data points sampled

from truth. This enables more direct surrogate construction methods, such as radial basis function interpolation, that also do not require pre-training to produce good quality surrogates.

We target the Quantum Approximate Optimization Algorithm (QAOA)[1] as an example use-case of our surrogate-based optimization protocol. We apply QAOA to the Max-Cut combinatorial optimization problem[54–59] and finding low energy states of random-coefficient classical Ising models. Max-Cut problem instances are encoded as binary discrete polynomials known as quadratic unconstrained binary optimizations (QUBOs). The Max-Cut costs, or equivalently Ising model energies, are optimized with respect to parameters of the QAOA PQCs to find low-energy solutions.

For the Max-Cut problems, we numerically simulate surrogate-based optimization for 16-qubit problem instances using measurement shot numbers typically available with cloud-based quantum computers. For this setup, we find that our approach outperforms a state-of-the-art VQA method, double adaptive-region Bayesian optimization (DARBO)[43], which is described in detail in Section "Methods". We choose graph instances for which DARBO has been benchmarked against popular approaches to VQA optimization[43] and for which numerical studies[43,60] have found that the proliferation of low-quality local minima with increasing QAOA depth makes optimization challenging.

Next, we use numerical tensor network methods to simulate surrogate-based optimization of a shallow QAOA PQC for 127-qubit instances of a random Ising model on the heavy hex graph used by current IBM quantum processors. We utilize a warm-start initialization by parameter transfer from solutions of smaller problem instances found in previous work. We demonstrate a consistent improvement in solution quality over the parameter transfer solutions.

We build on this by using the `ibm_torino` quantum computer to apply our method to solve a 127-qubit Ising model instance. Again, we systematically improve upon previous parameter transfer results and show that the solutions obtained with quantum hardware give improved initial states for 133- and 156-qubit problem instances. To the best of our knowledge, this implementation is the largest on-device QAOA optimization in the literature, showcasing the potential of the surrogate-based approach. This performance gain is made possible by an extensive usage of C^{surr}, as we evaluate it three orders of magnitude more frequently than the on-device cost, C.

Results

In this section, we present the results of a numerical investigation into the effectiveness of the proposed approach. We simulated finite-shot QAOA optimization for Max-Cut problems on random graphs and random Ising models on a heavy-hex lattice. We compared our Max-Cut results to the state-of-the-art QAOA optimization method, DARBO. In Section "QAOA" we provide an overview of QAOA. In Section "QAOA for Max-Cut on 3-regular Weighted Graphs" we present the Max-Cut results, and in Section "QAOA for Random Ising Models on the 127-qubit Heavy-Hex Hardware Graph" the Ising model results.

QAOA

In QAOA, we start with a classical Hamiltonian that typically describes a combinatorial optimization problem given by a cost function, $C^{\text{classical}}(\boldsymbol{z})$, where $\boldsymbol{z} = (z_1, z_2, \ldots, z_n) \in \{-1, 1\}^n$ is a vector of decision variables. QAOA encodes the decision variables as computational basis states such that n-qubit $|\boldsymbol{z}\rangle$, with $z_i = \langle \boldsymbol{z}|Z_i|\boldsymbol{z}\rangle$, encodes $\boldsymbol{z}$. Here, Z_i is a Pauli operator acting at qubit i.

QAOA consists of five algorithmic components:

- An initial state, $|\psi_0\rangle$.
- A phase separating Hamiltonian, H_C, related to the optimization problem by $C^{\text{classical}}(\boldsymbol{z}) = \langle \boldsymbol{z}|H_C|\boldsymbol{z}\rangle$. H_C is a classical Ising Hamiltonian, which usually encodes a quadratic unconstrained binary optimization problem.
- A mixing Hamiltonian, H_M.
- An integer, $p \geq 1$, which is typically called the number of QAOA rounds.
- A total of $2p$ real-valued numbers that parameterize the time evolution with H_C and H_M for each round. These parameters are referred to as *angles*. For a round number, i, we designate the angles as γ_i and β_i, respectively. We introduce a notation $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$.

The standard QAOA protocol consists of alternating simulation of the two non-commuting Hamiltonians, H_C and H_M, acting on the initial state, $|\psi_0\rangle$, resulting in the evolved state

$$|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle = e^{-i\beta_p H_M} e^{-i\gamma_p H_C} \ldots \\ \ldots e^{-i\beta_2 H_M} e^{-i\gamma_2 H_C} e^{-i\beta_1 H_M} e^{-i\gamma_1 H_C} |\psi_0\rangle. \tag{1}$$

Applications of H_C separate computational basis states by phases $e^{-i\gamma_i C^{\text{classical}}(\boldsymbol{z})}$, while H_M produces parameterized interference between states. Sampling $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ we obtain the computational basis eigenstates $|\boldsymbol{z}\rangle$ that correspond to the classical bitstrings $\boldsymbol{z}$ from the solution space of the classical optimization problem.

If p is sufficiently large and the parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, are sufficiently well-tuned, this algorithm will generate, with high probability, low energy eigenstates of the underlying classical Hamiltonian H_C, which correspond to the bitstrings with close to the optimal $C^{\text{classical}}(\boldsymbol{z})$. The goal is to sample high-quality approximate solutions of the classical optimization problem, ideally finding a globally optimal solution or solutions. Learning optimal $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ values is computationally challenging in general, but optimal angles are not necessary for QAOA to perform well as the sampling protocol. Nonetheless, learning good parameters is important to get the best results possible, especially on noisy quantum computer hardware.

QAOA can be treated as an example of a variational quantum algorithm. In this approach, the state in Eq. (1) is compiled to a parameterized quantum circuit with parameters

$$\boldsymbol{\theta} = (\gamma_1, \gamma_2, \ldots, \gamma_p, \beta_1, \beta_2, \ldots, \beta_p), \tag{2}$$

and an optimized cost function

$$C(\boldsymbol{\theta}) \equiv C(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \langle \boldsymbol{\gamma}, \boldsymbol{\beta}|H_C|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle. \tag{3}$$

For large enough p, the angles resulting in optimal $C(\boldsymbol{\gamma}, \boldsymbol{\beta})$ correspond to a high probability of sampling from $|\boldsymbol{\gamma}, \boldsymbol{\beta}\rangle$ states $|\boldsymbol{z}\rangle$ with close to optimal $C^{\text{classical}}$, and therefore result in high-quality approximate solutions of the classical optimization problem.

QAOA for Max-Cut on 3-regular Weighted Graphs

We begin with the results for finite-shot QAOA applied to the Max-Cut problem. We considered Max-Cut defined for weighted graphs $G = (V, E)$ with a classical Hamiltonian

$$H_C = \sum_{ij} w_{ij} Z_i Z_j. \tag{4}$$

Here i, j index graph vertices ($i, j \in V$) which are connected by edges $\{i, j\} \in E$, $w_{ij} \in [0, 1]$ is the weight associated with an edge and Z_i is a Pauli operator acting at vertex i. We constructed graph instances by taking G to be 3-regular, with each vertex having three edges and choosing a set of random edge weights.

More precisely, we used 5 graph instances with $n = |V| = 16$ vertices that were used to benchmark DARBO against popular optimizers in ref. 43 (These instances can be found in Supplemental Information of ref. 43). We note that for the 16-qubits 3-regular weighted graphs a numerical study[60] found that to find a global optimum starting from random initialization standard optimizers require a number of runs that scales at least exponentially with p even in the infinite shot limit. Hence, despite moderate n the

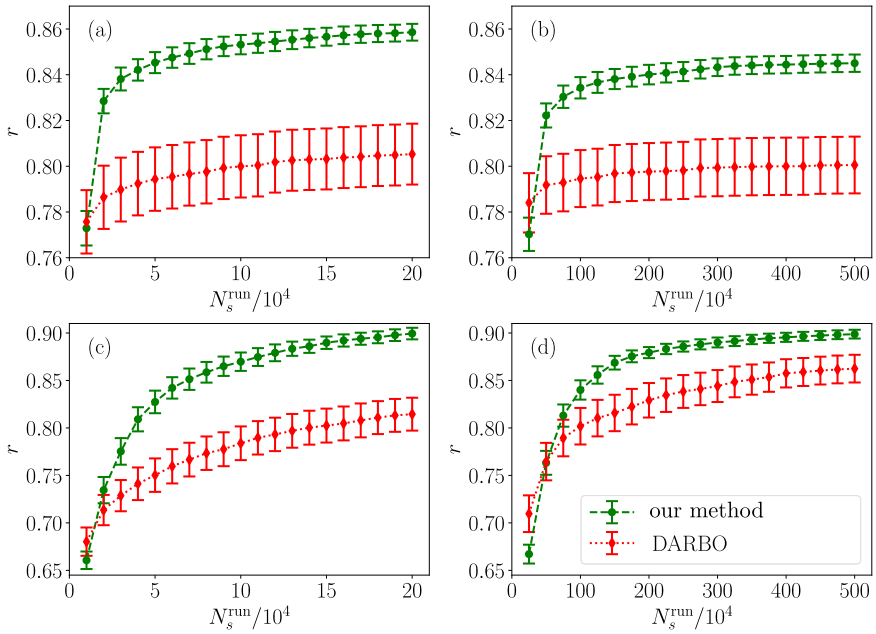

Fig. 2 | A comparison of surrogate-based optimization with DARBO for 3-regular 16-qubit Max-Cut problems. Approximation ratios, Eq. (5), of the best ground truth cost function values found by our method (green) and DARBO (red) runs within the first N_s^{run} shots of the run for p QAOA rounds. In **a** and **b** we show results for $p = 2$ and in **c** and **d** for $p = 10$. The optimization was performed in the presence of numerically simulated shot noise with $N_s = 200$ shots per cost function evaluation for **a** and **c** and $N_s = 5000$ for **b** and **d**. The markers represent r averaged over 5 random graphs and 20 randomly initialized runs per graph. The error bars are the mean standard deviations multiplied by a factor of 2, and correspond to a 95% confidence interval.

optimization landscape properties make our setup very challenging from the optimization standpoint.

We constructed the QAOA ansatz by taking $H_{\mathcal{M}} = \sum_i X_i$, for Pauli operators X_i, and using an equal weight superposition of computational basis states as an initial state, $|\psi_0\rangle$, defined by $\forall_i X_i|\psi_0\rangle = |\psi_0\rangle$. Furthermore, we chose $p \in \{2, 4, 6, 8, 10\}$. We numerically simulated the optimization of Eq. (3) in the absence of hardware noise. This matches the reference simulation setup[43].

To quantify optimization quality we used an approximation ratio defined for a cost function value, $C(\boldsymbol{\theta})$, as

$$r = \frac{C_{\max} - C(\boldsymbol{\theta})}{C_{\max} - C_{\min}}, \tag{5}$$

with $C_{\max}$ and $C_{\min}$ being the maximal and the minimal cost function values, respectively. Then, for $C = C_{\min}$, we have $r = 1$, and, for $C = C_{\max}$, we have $r = 0$. In our case, $C = C_{\min}$ ($C = C_{\max}$) for the angles that prepare the ground state of H_C ($-H_C$), respectively.

We compared the approximation ratio of an average optimization run with DARBO. The DARBO results were obtained from a repository[61], which contains the data published in ref. 43. We evaluated $C(\boldsymbol{\theta})$ using $N_s \in \{200, 5000\}$ shots. Furthermore, we set the total number of evaluations per run, $N_{\mathrm{init}} + N_{\mathrm{it}} = 1000$, with $N_{\mathrm{init}} = 50$. Consequently, the total number of shots per run was the same for our method and DARBO. We chose the β_i range to be the full parameter range after taking into account symmetries of the problem[60], i.e., $\forall_i \beta_i \in [-\pi/4, \pi/4]$. Here, we have $\forall_i \gamma_i \in [-\pi/2, \pi/2]$, the same as for the restricted search space in DARBO[60]. The initial points were obtained by random uniform sampling of the parameter ranges. For each graph and value of p we conducted 20 runs, starting each with a different random initialization. The same number of randomly initialized runs was used to generate the reference results.

In Fig. 2, we compare results for $p = 2$ and $p = 10$. These are the smallest and largest p considered in ref. 43. We present the approximation ratio of the best cost function evaluated so far, $\min\{C(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$, versus a run shot cost

$$N_s^{\mathrm{run}} = N_s(N_{\mathrm{init}} + i). \tag{6}$$

Here, i indexes optimization iteration number, as described in detail in the "Methods" section. In Fig. 2, we plot the approximation ratio, r averaged over all runs per graph and graph instances. The error bars are the mean standard deviation multiplied by a factor of 2. We find that, apart from the first iterations, our approach outperforms DARBO, resulting in higher r for $N_s^{\mathrm{run}} > 10^4$ ($N_s^{\mathrm{run}} > 5 \cdot 10^5$), when $N_s = 200$ ($N_s = 5000$), respectively. The performance gap is especially pronounced for $N_s = 200$, where for $N_s^{\mathrm{run}} = 10^5$ and $p = 2$ our approach obtained $r = 0.859(2)$ versus the $r = 0.784(7)$ achieved by DARBO. Similarly, for $p = 10$ and the same N_s^{run}, our method and DARBO achieved $r = 0.899(3)$ and $r = 0.795(9)$ respectively. We note that the shot budgets considered here ($N_s^{\mathrm{run}} < 5 \cdot 10^6$) match what is currently available through cloud access to quantum computers. We note that our method uses modest classical resources as a typical $p = 10$ optimization run from Fig. 2 takes approximately 5 hours on a single-core of an Intel Xeon Gold 5320 processor.

Next, we analyze the quality of our surrogate-based optimization method as a function of the number of QAOA rounds, comparing $p \in \{2, 4, 6, 8, 10\}$ runs. We present these results in Fig. 3, obtained using the same 5 graphs, initialization method, optimization hyperparameters, and number of runs per graph as for the DARBO comparison. We find with $N_s = 5000$ that, for large enough $N_s^{\mathrm{tot}} \geq 2 \cdot 10^6$, the results systematically improve with increasing p. Therefore, despite using the limited shot budget available with current quantum devices, the method is capable of harnessing the increase in QAOA ansatz expressibility with increasing p.

Additionally, in Supplementary Note 1 A we present the values of $C(\boldsymbol{\theta})$ computed with infinite shots and the best angles found after the finite-shot optimization runs. This is to both understand the effect of finite shot noise on the estimates of $C(\boldsymbol{\theta})$ used in Figs. 2, 3, and to more robustly quantify the quality of the angles found using finite-shot optimization. In this case, we find that our approach's advantage over DARBO, and the convergence with increasing p, remain very similar to the results shown in Figs. 2, 3.

QAOA for Random Ising Models on the 127-qubit Heavy-Hex Hardware Graph

Next, we consider optimization of the QAOA ansatz defined in Eq. (1) with $p = 3$ for a 127-qubit graph compatible with the 133-qubit `ibm_torino` quantum processor connectivity. `ibm_torino` is a superconducting qubit quantum computer that has a sparse 2-qubit gate interaction graph, generally referred to as heavy-hex connectivity[62]. We focus on 5 instances of a random Ising model with a problem Hamiltonian

$$H_C = \sum_i d_i Z_i + \sum_{\langle i,j \rangle} d_{i,j}^{(2)} Z_i Z_j + \sum_i d_{i,j,k}^{(3)} Z_i Z_j Z_k. \tag{7}$$

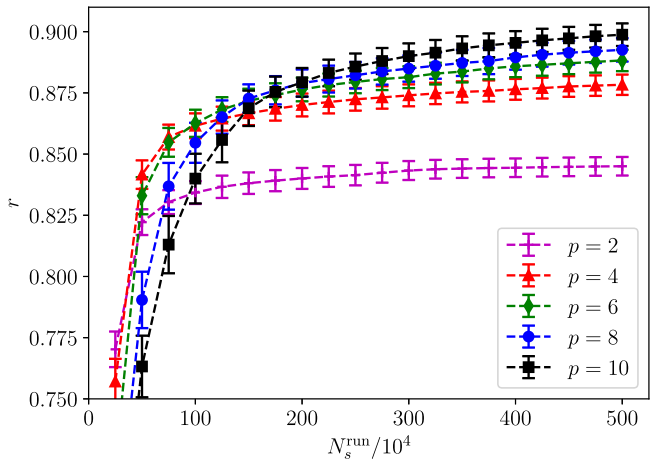

Fig. 3 | Surrogate optimization performance with increasing QAOA rounds for 3-regular 16-qubit Max-Cut problems. The approximation ratio of the best cost, $C(\boldsymbol{\theta})$, found during the first N_s^{run} shots of an optimization run is plotted for multiple QAOA round numbers $p = 2$ (magenta), $p = 4$ (red triangles), $p = 6$ (green diamonds), $p = 8$ (blue circles), $p = 10$ (black squares). Here, the number of shots per $C(\boldsymbol{\theta})$ evaluation was $N_s = 5000$. The choice of graphs, optimization initialization, and hyperparameters were the same as in Fig. 2. The markers denote the mean values obtained by averaging over 5 graphs and 20 runs per graph, and the error bars were computed in the same way as in Fig. 2.

Here, in the second term, we sum over edges of the graph, and the three-body terms occur for a subset of graph nodes indexed by i that have exactly two nearest-neighbors j, k, see refs. 63–65 for details. The coefficients $d_i, d_{i,j}^{(2)}, d_{i,j,k}^{(3)} \in \{-1, 1\}$ are chosen uniformly at random with probability 0.5.

This class of optimization problems has been designed to be highly hardware-compatible with the IBM processors that share this heavy-hex graph structure. In particular, despite including the three-body terms in H_C, a single QAOA round can be implemented using a circuit with a CNOT gate depth of 6. The graph sparsity also means that these optimization problems are relatively easy to sample using heuristic methods such as simulated annealing. Moreover, they can be solved exactly using optimization software such as CPLEX[63–65]. QAOA performance for these problems has been investigated using 127-qubit IBM quantum processors[65] using angles learned with a technique known as parameter transfer. In a variety of combinatorial optimization problem instances, it has been observed that QAOA angles seem to perform similarly for many different problem instances that are all from a similar class of problem type (i.e., similar graph structure)[66–71]. This property is typically called *parameter concentration* or *parameter transfer*. ref. 65 used this strategy to train QAOA on a 16-qubit heavy-hex Ising model instance with Hamiltonian in Eq. (7), and to apply the learned angles to larger problem instances with the same Hamiltonian. In this section, we investigate the capability of our surrogate-based optimization approach to improve on the parameter transfer angles found in ref. 65. We will refer to these parameter transfer angles as heuristic angles.

We numerically simulate finite-shot optimization for 5 random Hamiltonian instances, $p = 3$, and with $N_s = 1000$ shots per $C(\boldsymbol{\theta})$ evaluation using a Matrix Product State (MPS) simulator[65,72]. In our simulations we assume no hardware noise. The set of initial parameters Θ_{init} is built of the parameter transfer angles $\boldsymbol{\theta}_{\mathrm{heur}}^{p=3}$, and 19 random vectors of parameter $\boldsymbol{\theta}(2)$ choices. Each random choice is obtained by uniform random sampling of a full parameter range for each of $2p$ $\boldsymbol{\theta}$ elements. We remark that our initialization can be viewed as a special case of warm start VQA initialization[32]. As the cost function in Eq. (3) is invariant under shifts of γ_i and β_i by π, we choose the full parameter ranges to be $\gamma_i \in [-\pi/2, \pi/2)$, $\beta_i \in [-\pi/2, \pi/2)$.

The parameter transfer angles were found in ref. 65 by numerically optimizing small 16 qubit instances of the random Ising model defined in Eq. (7) on a heavy-hex graph, and are listed in Table 1. Furthermore, we use

Table 1 | Heuristic (parameter transfer) QAOA angles

$\boldsymbol{\beta}_{\mathrm{heur}}^{p=3}$	(0.50502, 0.35713, 0.19264)
$\boldsymbol{\gamma}_{\mathrm{heur}}^{p=3}$	(−0.14264, −0.26589, −0.34195)
$\boldsymbol{\beta}_{\mathrm{heur}}^{p=4}$	(0.54321, 0.41806, 0.28615, 0.16041)
$\boldsymbol{\gamma}_{\mathrm{heur}}^{p=4}$	(−0.12077, −0.22360, −0.29902, −0.35329)
$\boldsymbol{\beta}_{\mathrm{heur}}^{p=5}$	(0.53822, 0.44776, 0.32923, 0.23056, 0.12587)
$\boldsymbol{\gamma}_{\mathrm{heur}}^{p=5}$	(−0.11764, −0.19946, −0.268736, −0.321586, −0.34583)

QAOA ansatz parameters $\boldsymbol{\theta}_{\mathrm{heur}} = (\boldsymbol{\gamma}_{\mathrm{heur}}, \boldsymbol{\beta}_{\mathrm{heur}})$ for $p \in \{3, 4, 5\}$ rounds on random-coefficient Ising Hamiltonians with heavy-hex connectivity. Obtained by a numerical optimization on a 16-qubit graph in ref. 65. The heuristic angles are included in sets of initial angles Θ_{init} for the surrogate-based optimization of the 127-qubit instances.

$N_{\mathrm{it}} = 80$ optimization iterations during training. Our hyperparameter choices resulted in a total run shot cost of 10^5, which is accessible with cloud-based access to current IBM quantum computers.

For multiple types of classical problem Hamiltonians, it has been shown that the parameter transfer angles found by optimization of similar problems work very well for most other problem instances in that class[66–71]. In particular, ref. 65 has shown this for the heuristic angles and the heavy-hex Ising model problem instances considered here. Moreover, it has been shown that, for low-enough p, QAOA cost function landscapes frequently have multiple suboptimal local minima, which limits ansatz trainability[73]. Therefore, we expect that the inclusion of $\boldsymbol{\theta}_{\mathrm{heur}}^{p=3}$ among the initial angles may speed up the convergence of our approach towards angles with low-cost function values. Furthermore, we expect that for this case, a fully random initialization might reduce optimization shot efficiency, since many calls to the true cost function may be required to learn a complex cost landscape from scratch.

in Fig. 4, we show the best cost function found, $\min\{c(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \theta\}$, versus a run shot cost n_s^{run} of eq. (6), for each of the 5 Hamiltonian instances. The cost was averaged over 32 runs, each started with different choices of random angles in θ_{init}. Furthermore, we compare the optimized cost function to its value at the parameter transfer angles. As expected, for all Hamiltonian instances, we find that the randomly sampled initial points do not improve upon $c(\boldsymbol{\theta}_{\mathrm{heur}}^{p=3})$, as shown by the first evaluated points on the learning curve having the same energy as the parameter transfer angles for each problem instance. in contrast, the surrogate-based optimization systematically improves upon the parameter transfer angles. This improvement is clearly visible for all instances, even for n_s^{run} as small as $3 \cdot 10^4$.

To provide further evidence for this conclusion, we present infinite-shot cost function values in Supplementary Note 1 B. This is in order to more robustly quantify the cost improvement for the angles found by our method. That is, we determined the best angles, $\mathrm{argmin}\{C(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$, using $N_s = 1000$ cost function estimates from the optimization runs (the same as in the main text). Subsequently, we computed infinite-shot $C(\boldsymbol{\theta})$ for these angles. The observed behavior of $C(\boldsymbol{\theta})$ is very similar to the finite shot case, as shown in Fig. 4. Additionally, in Supplementary Note 1 B we show that these cost function estimates appear to be converged in the MPS bond dimension, D, the refinement parameter of the numerical simulations.

Hardware Implementation. We performed surrogate-based optimization for an instance of a 127-qubit heavy-hex Ising model (instance 0 from Fig. 4) with a $p \in \{3, 4, 5\}$ QAOA ansatz on the `ibm_torino` quantum computer, using $N_s = 1000$ shots for each cost function evaluation. We chose Θ_{init}, and the angle bounds in the same way as for the MPS optimization. More precisely, Θ_{init} was the union of 19 random angles and $\boldsymbol{\theta}_{\mathrm{heur}}$ from Table 1. During optimization, the QAOA angles were bound to lie within $[-\pi/2, \pi/2)$. We set the iteration number $N_{\mathrm{it}} = 380$. The optimization hyperparameters are presented in more detail in Supplementary Note 3. For each value of p, we conducted 3 or 4 optimization runs, and plot the best device cost function estimates versus shot number Eq. (6) in Fig. 5. Additionally, to benchmark angle

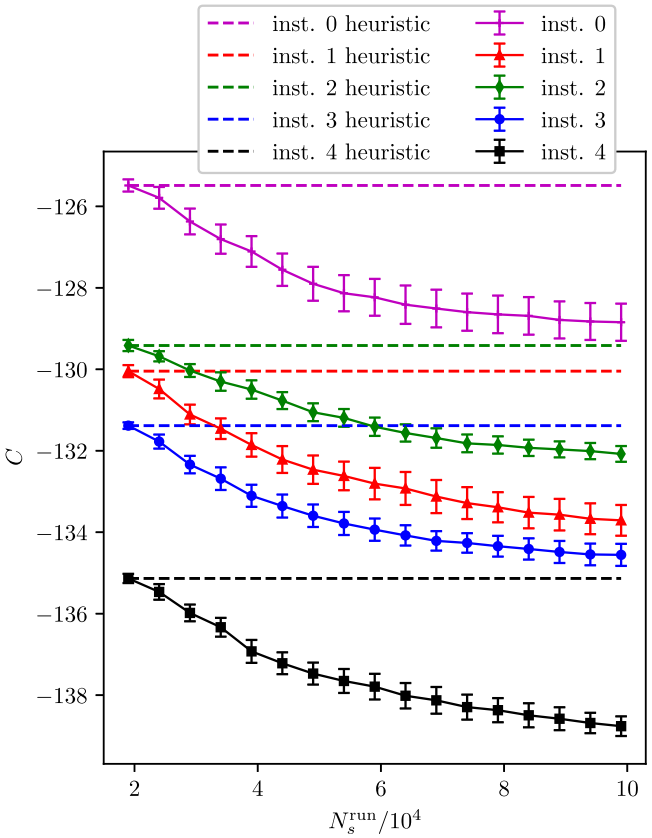

Fig. 4 | Numerical Matrix Product State optimization for $p = 3$ QAOA and 127-qubit heavy-hex Ising models. Here we plot the best cost $C(\boldsymbol{\theta})$ (energy) found during the first N_s^{run} run shots (solid lines) for 5 instances of the model coefficients Eq. (7) (shown by color) and $N_s = 1000$ shots per $C(\boldsymbol{\theta})$ evaluation. The dashed horizontal lines indicate the cost function for the parameter transfer heuristic angles $\boldsymbol{\theta}^{p=3}_{\mathrm{heur}}$, which are included as one of $N_{\mathrm{init}} = 20$ initial angles. The remaining initial parameters are chosen randomly. The markers show the mean values obtained by averaging over 32 runs started with different choices of the random initial angles. The error bars are the mean standard deviations multiplied by a factor of 2, which correspond to 95% confidence intervals.

improvement, we evaluated MPS infinite-shot cost values for the angles with the best device $C(\boldsymbol{\theta})$, and show the results in the insets of Fig. 5. The MPS cost values were not affected by the hardware noise, and were converged in the MPS bond dimension D, as discussed in Supplementary Note 2. Therefore, the MPS cost values reliably quantify optimized angle quality.

We observe that our surrogate-based approach systematically improves on the initial cost function values. For all runs, a significant part of the total improvement occurred $N_s^{\mathrm{tot}} < 5 \cdot 10^4$. Furthermore, for $p \in \{3, 4\}$, the vast majority of the gain happened for $N_s^{\mathrm{tot}} < 2 \cdot 10^5$. This demonstrates that the shot efficiency of the surrogate-based approach enables on-device training with limited, cloud-based access to the quantum hardware.

The total algorithmic runtime of a $p = 5$ optimization run, including classical computation and QPU time, was 23.5 minutes. The 1000-shot cost function evaluation for $p = 5$ took approximately 0.8 seconds of QPU time on `ibm_torino`. The total quantum runtime for the $p = 5$ angle learning was then around 320 seconds of QPU time. A single $p = 5$ surrogate update and minimization took on average 2.9 seconds on a MacBook Pro M1 notebook. This gives 1090 seconds spent on surrogate construction and optimization within one run. We note that the surrogate update and minimization time was not optimized, as the QPU queue time – median 6.73 hours across runs – took up the vast majority of the total runtime. For these problem instances an infinite-shot MPS cost evaluation took approximately 3.4 hours for $p = 5$ and MPS bond dimension $D = 2048$ using Ian ntel Xeon Gold 5120 processor.

The impact of the hardware noise is clearly demonstrated by values $C(\boldsymbol{\theta}^{p=3}_{\mathrm{heur}})$ obtained from the device, $\{-92.7, -91.4, -89.6\}$, that are larger than the MPS value of -125.4. Furthermore, for $p = 4$, we see a large discrepancy of the device cost values $C(\boldsymbol{\theta}^{p=4}_{\mathrm{heur}}) \in \{-102, -101, -44.5\}$, which is unlikely to be caused by finite N_s effects. A more likely explanation is the time variability of hardware noise. We see a similar effect for the $p = 5$ runs. In our MPS benchmarks, for $p = 3$ we observe that all 3 runs converged to the cost function values typical for MPS optimization, see Fig. 4 for comparison, improving upon the heuristic MPS cost function, and showcasing the noise-resilience of the QAOA optimization for shallow-enough circuits. In contrast, for the deeper $p = 4$ and $p = 5$ circuits, the final MPS cost C_{opt} only improved on the parameter transfer angle cost for 2 out of 3 and 2 out of 4 runs, respectively. Therefore, for $p = 4$ and $p = 5$, the noise resilience of the on-device QAOA optimization is significantly reduced. Overall, the results demonstrate that despite significant hardware noise and device

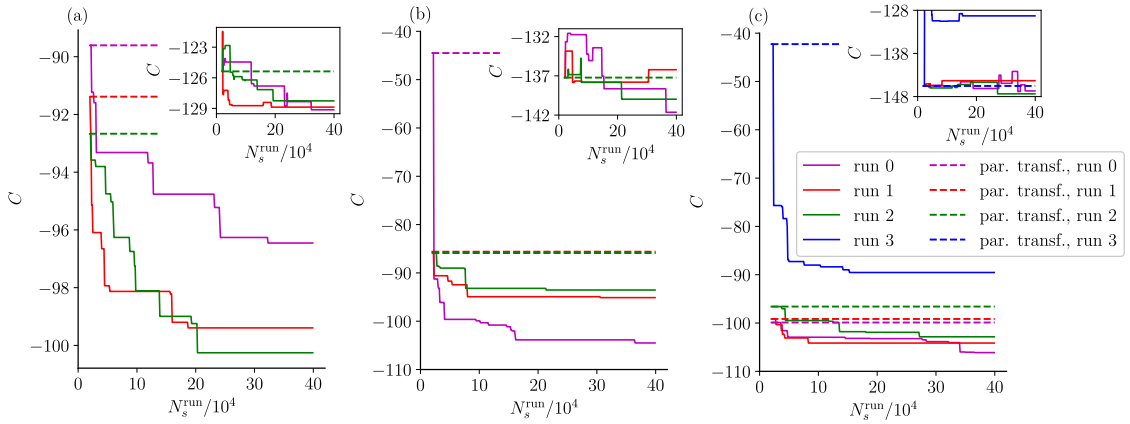

Fig. 5 | Surrogate-based optimization on the `ibm_torino` quantum computer for p QAOA rounds applied to an instance of a 127-qubit heavy-hex random Ising model. Here, **a** depicts $p = 3$, **b** $p = 4$, and **c** $p = 5$ optimization results. In the main plots, the solid lines are the best device cost function estimates $C(\boldsymbol{\theta})$, for $N_s = 1000$, found within the first N_s^{run} shots for several runs (shown by color). The solid lines in the insets show the infinite-shot Matrix Product State simulator cost function, which is not affected by hardware noise, for the same angles as in the main plots. The dashed horizontal lines shown in both cases are the cost function estimates for the parameter transfer angles from ref. 65. The parameter transfer angles are included in the runs' initial angles Θ_{init}.

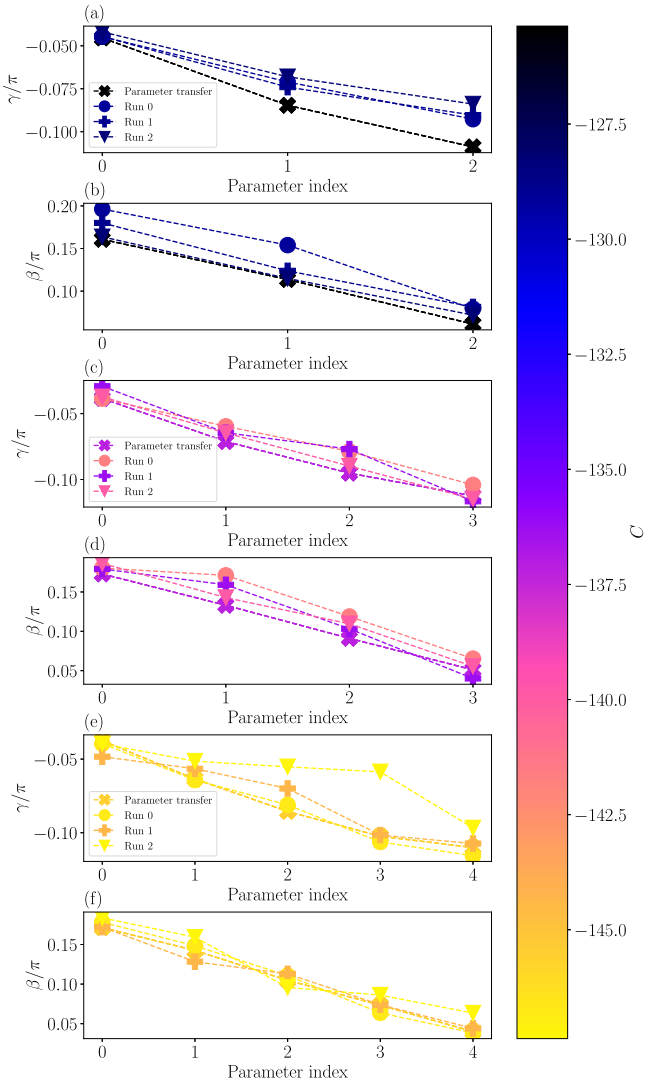

Fig. 6 | QAOA angle evolution. Comparison of the initial parameter transfer angles θ_{heur} with the angles learned from noisy device optimization, θ_{opt}, having been initialized with the parameter transfer angles. The plots show the final angles from runs 0 (circles), 1 (pluses), 2 (triangles) in Fig. 5, and the parameter transfer runs from Table 1. **a** and **b** show results for $p = 3$ QAOA rounds, **c** and **d** plot $p = 4$ results, and **e**, **f** contain $p = 5$ angles. Parameter index i, see (1), specifies to which QAOA layer the angles belong to. The color bar is the QAOA cost function evaluated at the angles shown using a Matrix Product State simulator in the infinite-shot limit.

instability, our approach is able to find good QAOA angles and improve on already known QAOA angles.

An overview of the optimized QAOA parameters $\boldsymbol{\theta}_{opt}$, which are obtained at the end of each run, is shown in Fig. 6. This plot compares $\boldsymbol{\theta}_{\text{opt}}$ and $\boldsymbol{\theta}_{\text{heur}}$, correlating the parameter values with their corresponding infinite-shot MPS cost values. We observe that for all p, $|\gamma_i|$ increases with increasing parameter index i, while $|\beta_i|$ decreases. At $p = 3$, the newly learned values of γ_i and β_i are typically larger than for the parameter transfer angles. For noise-resilience-breaking $p \in \{4, 5\}$ it is less clear whether there are patterns distinguishing $\boldsymbol{\theta}_{\text{opt}}$ and $\boldsymbol{\theta}_{\text{heur}}$.

Finally, we investigated whether the improved QAOA angles learned by our surrogate-based method could be successfully transferred to larger problem instances. This was done for $p = 3$, by comparing cost values for the QAOA parameters learned after 127-qubit device optimization $C(\boldsymbol{\theta}_{\text{opt}})$ with $C(\boldsymbol{\theta}^{p=3}_{\text{heur}})$, for 133 and 156-qubit random Ising model instances defined using Eq. (7). The cost function for both instances was evaluated using 8192 shots on the 156-qubit `ibm_marrakesh` processor and in the infinite-shot limit

Table 2 | Parameter transfer of device-learned parameters to larger problems

n	133			156		
Eval. nb.	0	1	2	0	1	2
Device $C(\boldsymbol{\theta}^{p=3}_{\text{heur}})$	−94.8	−95.2	−94.6	−115	−110	−109
MPS $C(\boldsymbol{\theta}^{p=3}_{\text{heur}})$	−139	–	–	−163	–	–
Device $C(\boldsymbol{\theta}_{\text{opt}})$	−97.7	−97.4	−97.3	−121	−116	−119
MPS $C(\boldsymbol{\theta}_{\text{opt}})$	−143	−144	−143	−173	−171	−169

Parameter transfer of the `ibm_torino` quantum computer $n = 127$, $p = 3$ optimization runs' final angles $\boldsymbol{\theta}_{\text{opt}}$ to larger (unseen) 133, and 156-qubit random heavy-hex Ising model instances of Eq. (7). Here, the QAOA cost $C(\boldsymbol{\theta})$ was evaluated on the `ibm_marrakesh` quantum computer with $N_s = 8192$ shots. Furthermore, it was also evaluated with $D = 2048$ Matrix Product States in the infinite shot limit. We compare the cost obtained for the optimized angles $C(\boldsymbol{\theta}_{\text{opt}})$ with the cost for the heuristic $\boldsymbol{\theta}^{p=3}_{\text{heur}}$ angles from Table 1. In the case of $C(\boldsymbol{\theta}_{\text{opt}})$, the evaluation number from this table is the same as the optimization run number, used in Figs. 5, 6, of a run that produced $\boldsymbol{\theta}_{\text{opt}}$. In the case of $C(\boldsymbol{\theta}^{p=3}_{\text{heur}})$ and the device cost estimates, the evaluation number indexes independent evaluations of the cost function for the same angles. As the MPS cost function was evaluated in the infinite-shot limit, and was not subject to hardware noise, it was evaluated once for each graph instance.

using an MPS simulator. The cost values are presented in Table 2. We see that the angles learned using a noisy quantum device provide a consistent improvement over the previously known angles for these new, larger problem instances. These results further indicate that our method is capable of improving estimates of good QAOA angles.

Discussion

We expect that the shot efficiency of the proposed approach can be enhanced further. In particular, in this work, we terminate the adaptive sampling after N_{it} iterations. State-of-the-art surrogate methods frequently use composite termination criteria instead, for example, taking into account the cost function change in recent iterations. Such criteria allow for a faster termination of the optimization if the cost does not change significantly. Moreover, imposing physical constraints on the surrogates, like periodicity of the cost as a function of QAOA angles, may further reduce the number of hardware calls by constructing a surrogate that more closely reflects the true cost.

Another direction to improve result quality, and ultimately the efficiency of the approach, is sampling strategies that target the exploration of sparsely sampled parts of the parameter space. Our algorithm attempts to efficiently locate the best local optimum detected out of a set whose presence is inferred from the initial data points. However, this approach may not be efficient enough in the presence of many local optima, especially when no heuristic information about their approximate location is available. In such circumstances, we expect that strategies involving multiple optimization runs, with new runs starting in the least probed regions of parameter space, may be necessary.

A related issue emerges when considering parts of the cost landscape with vanishing gradients. In the presence of barren plateaus, cost function regions with significant variance occupy vanishingly small portions of parameter space[18]. Nevertheless, in special cases, heuristic strategies to locate some of such regions have been proposed, enabling optimization[32]. However, even with such heuristic information, the presence of barren plateaus will have a negative impact on the efficiency of sampling strategies that explore the whole parameter space. Therefore, to preserve the efficiency of our method, modifications to these strategies to target the non-vanishing gradient regions may be very beneficial.

To correctly identify the barren plateau regions, one needs to handle cost variations caused by finite shot noise. In the region of a barren plateau, these variations obscure cost changes due to varying parameters. Our current approach does not take into account finite shot effects on cost estimates.

Consequently, directly accounting for shot noise is an interesting topic to explore. This development could improve method efficiency even in the absence of barren plateaus.

While this work focuses on continuous parameter optimization with a fixed ansatz, variational approaches that incorporate discrete parameters to modify the ansatz structure have been proposed[74]. An extension of this work to such cases would be of interest. Moreover, as we target QAOA as our proof-of-principle application, a natural continuation of this research is applying the method to other VQAs, for example variational quantum eigensolvers targeting strongly-correlated systems in quantum chemistry[75], or warm-started QAOA ansaetze[76–78]. Another natural direction of research would be a systematic comparison with other VQA optimization approaches across varying optimization landscapes and initialization strategies.

One potential use of our surrogate-based approach is improving the efficiency of PQC optimization with numerically-expensive and accurate classical approaches like state-of-the-art tensor network[9,79], Pauli propagation[8,80], or symmetry-exploiting approaches[12,81]. Our approach can be applied to minimize the number of cost evaluations by such numerical methods to extend the range of optimization problems to which they are applicable. Another future research direction is the usage of such numerical methods to supplement device-derived data by providing accurate results for special classically simulable parameter values, like QAOA angles corresponding to near-Clifford circuits[82].

Importantly, the optimization problems considered in this study are test cases, in particular the IBM quantum hardware compatible Ising models, which, as combinatorial optimization problems, are relatively easy to solve using classical methods[83]. These problems were chosen as the objective of this study so as to demonstrate the algorithmic capability of large-scale on-device variational learning. Future research should examine QUBO optimization problems with denser graph connectivity that are harder to solve using classical methods, but still compatible with near-term quantum computer hardware.

Finally, a future research direction is combining the surrogate-based approach to on-device learning with error mitigation methods and strategies to reduce the effects of hardware-drift[57,84,85]. A pertinent question is to what extent these methods can improve the noise-resilience of optimization. In particular, such improvements may enable successful and reliable on-device QAOA optimization for $p > 3$, for which we observe the breakdown of noise resilience. Another relevant question is to what extent error mitigation methods are sufficient to reliably judge the noise resilience of optimization[86], as for deep enough ansätze, the MPS numerical methods applied in this study are expected to fail.

Methods

Adaptive surrogate construction using thin-plate radial basis functions

Our goal is to optimize $C(\boldsymbol{\theta})$ with respect to a set of quantum circuit parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$, keeping the number of $C(\boldsymbol{\theta})$ evaluations as small as possible. We assume that the parameters are bounded. We start with an initial coarse sampling of the full parameter space to obtain N_{init} evaluations of $C(\boldsymbol{\theta})$. In other words, we evaluate the cost function for an initial set of parameters $\Theta_{\text{init}} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{N_{\text{init}}}\}$. If a heuristic choice of $\boldsymbol{\theta}$ is available, in particular when $C(\boldsymbol{\theta})$ is known to be a good parameter choice, one may include these parameters in Θ_{init} to speed up optimization convergence. Otherwise, all initial parameters are obtained by random sampling. In the "QAOA for Max-Cut on 3-regular Weighted Graphs" Section, we use uniform random sampling of the parameter range to generate Θ_{init}, and supplement these random choices with heuristic parameters in Sections "QAOA for Random Ising Models on the 127-qubit Heavy-Hex Hardware Graph" and "Hardware Implementation".

Next, we iterate the following steps until a termination criterion is met.

1. We build a classical surrogate cost function $C^{\text{surr}}(\boldsymbol{\theta})$, performing radial basis function interpolation using all $C(\boldsymbol{\theta})$ evaluations performed so far.

More explicitly, we interpolate a set of pairs

$$\{(\boldsymbol{\theta}, C(\boldsymbol{\theta})) \,|\, \boldsymbol{\theta} \in \Theta\},$$

with $\Theta = \Theta_{\text{init}}$ for the first iteration ($i = 1$), and updated later as described below.

Following ref. 50, for $\boldsymbol{\theta}$ we construct the surrogate cost as

$$C^{\text{surr}}(\boldsymbol{\theta}) = \sum_i \beta_i \phi(||\boldsymbol{\theta} - \boldsymbol{\theta}_i||_2), \tag{8}$$

where

$$\phi(r) = r^2 \ln(r), \tag{9}$$

is a 'thin-plate' radial basis function (RBF), and $||.||_2$ is a vector 2-norm also called the Euclidean distance. Here, the index i numbers all $C(\boldsymbol{\theta})$ evaluations. The coefficients β_i are obtained by solving a system of linear equations

$$M\boldsymbol{\beta} = \boldsymbol{Y},\ M_{ij} = \phi(||\boldsymbol{\theta}_i - \boldsymbol{\theta}_j||_2),\ Y_i = C(\boldsymbol{\theta}_i). \tag{10}$$

The interpolation is performed with the mystic scientific machine learning package[87,88]. We provide more details on the interpolation implementation in Supplementary Note 3.

2. We identify N_{opt} randomly chosen new points in the parameter space, and use these points as starting values for local minimizations (or maximizations) of $C^{\text{surr}}(\boldsymbol{\theta})$. We identify the best of the optimization solutions as a candidate optimum, $\boldsymbol{\theta}^{(i)}_{\text{cand}}$.
3. We evaluate $C(\boldsymbol{\theta}^{(i)}_{\text{cand}})$, and check the termination condition. After this step $\Theta \leftarrow \Theta \cup \{\boldsymbol{\theta}^{(i)}_{\text{cand}}\}$, and $i \leftarrow i + 1$.

In this work, to demonstrate the method in a simple setup, we terminate the loop after N_{it} iterations. We defer investigation of the more sophisticated termination conditions commonly used in classical surrogate approaches[50] to future work. Our approach is depicted schematically in Fig. 1.

We choose thin-plate RBFs for interpolation as they have a track record of performing well in challenging global optimization problems[50,89] and do not involve hyperparameters, removing the need for upfront training before interpolation. Furthermore, when used for interpolation, the thin-plate basis produces a surface with minimal curvature while also passing through each data point. This implicitly regularizes the surrogate cost surface, potentially smoothing out poor-quality local minima not supported by sample data. Therefore, this approach may be an effective heuristic for circumventing the local minima problem in VQAs if the global minimum can be captured sufficiently well[22].

Our procedure outputs the best cost function value found on the noisy quantum computer, C_{opt}, and the corresponding parameters, $\boldsymbol{\theta}_{\text{opt}}$, where

$$\boldsymbol{\theta}_{\text{opt}} = \text{argmin}\,(\text{argmax})\,\{C(\boldsymbol{\theta}) \,|\, \theta \in \Theta\}, \tag{11}$$

$$C_{\text{opt}} = C(\boldsymbol{\theta}_{\text{opt}}), \tag{12}$$

for minimization (maximization). Furthermore, in Sections "Results" and "Hardware Implementation", to improve the method's performance, we repeat the optimization runs multiple times, starting from different Θ_{init} each time. Then, the outcomes of the best run are accepted as the optimization output. Such a strategy is commonly used by state-of-the-art optimization algorithms in the case of cost function landscapes with multiple local optima[43].

Reference method: DARBO

DARBO[43] is a Bayesian optimization algorithm designed for QAOA. It uses a Gaussian process with a parametrized Matérn 5/2 kernel as a surrogate function to approximate the cost function by fitting previously evaluated points. The kernel parameters are determined by pre-training. Rather than

fitting all previous points, it fits only those points inside an intersection of two regions in the parameter space. The first, called the adaptive trust region, is a hypercube centered on the current best set of parameters. When the cost function is evaluated for new parameters, this hypercube grows if the value of the cost function consistently improves (consecutive successes) and shrinks when it consistently does not (consecutive failures). The second, called the adaptive search region, switches between allowing parameter selection across the whole domain or a subset of it whenever consecutive failures are detected. DARBO samples at the maxima of a parametrized acquisition function within the intersection of the trust and search regions. The acquisition function is chosen heuristically to balance sampling in the vicinity of the surrogate minima and in high-uncertainty parameter space regions.

DARBO has been benchmarked against standard VQA optimization techniques for a Max-Cut problem defined on weighted 3-regular graphs with $n = 10 - 18$ vertices[43]. In this benchmark, the method was shown to outperform other black-box optimizers, including Adam, L-BFGS-B, Nelder-Mead, COBYLA, SPSA, global Bayesian optimization, and TURBO. DARBO's advantage was especially pronounced when smaller shot numbers were used to evaluate $C(\boldsymbol{\theta})$, and has been observed to increase with growing p and n. Therefore, we treat it as a state-of-the-art reference method.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code that supports the findings of this study is available from the corresponding author upon reasonable request.

Received: 9 April 2025; Accepted: 5 November 2025;
Published online: 20 November 2025

References

1. Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm, https://arxiv.org/abs/1411.4028 arXiv preprint arXiv:1411.4028 (2014).
2. Huang, Hsin-Yuan, Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020).
3. Caro, M. C. et al. Generalization in quantum machine learning from few training data. *Nat. Commun.* **13**, 4919 (2022).
4. Childs, A. M., Maslov, D., Nam, Y., Ross, N. J. & Su, Y. Toward the first quantum simulation with quantum speedup. *Proc. Natl Acad. Sci.* **115**, 9456–9461 (2018).
5. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
6. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023).
7. Aharonov, D., Gao, X., Landau, Z., Liu, Y. & Vazirani, U. A polynomial-time classical algorithm for noisy random circuit sampling, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* 945–957 (2023).
8. Rudolph, M. S., Fontana, E., Holmes, Z. & Cincio, L. Classical surrogate simulation of quantum systems with LOWESA, https://arxiv.org/abs/2308.09109 (2023).
9. Tindall, J., Fishman, M., Stoudenmire, E. M. & Sels, D. Efficient tensor network simulation of IBM's Eagle kicked Ising experiment. *PRX Quantum* **5**, 010308 (2024).
10. Schuster, T., Yin, C., Gao, X. & Yao, N. Y. A polynomial-time classical algorithm for noisy quantum circuits, https://doi.org/10.48550/arXiv.2407.12768 (2024).
11. Begušić, T., Gray, J. & Chan, G. K.-L. Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, https://doi.org/10.1126/sciadv.adk4321 *Sci. Adv.* **10** (2024).
12. Anschuetz, E. R., Bauer, A., Kiani, B. T. & Lloyd, S. Efficient classical algorithms for simulating symmetric quantum systems. *Quantum* **7**, 1189 (2023).
13. Cerezo, M. et al. Does provable absence of barren plateaus imply classical simulability? *Nat. Commun.* **16**, 7907 (2025).
14. Cerezo, M. et al. Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
15. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 1–6 (2018).
16. Wang, S. et al. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **12**, 1–11 (2021).
17. Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **12**, 1–12 (2021).
18. Arrasmith, A., Holmes, Zoë, Cerezo, M. & Coles, P. J. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Sci. Technol.* **7**, 045015 (2022).
19. Holmes, Zoë, Sharma, K., Cerezo, M. & Coles, P. J. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum* **3**, 010313 (2022).
20. Martín, EnriqueCervero, Plekhanov, K. & Lubasch, M. Barren plateaus in quantum tensor network optimization. *Quantum* **7**, 974 (2023).
21. Bittel, L. & Kliesch, M. Training variational quantum algorithms is NP-hard. *Phys. Rev. Lett.* **127**, 120502 (2021).
22. Anschuetz, E. R. & Kiani, B. T. Beyond barren plateaus: Quantum variational algorithms are swamped with traps. *Nat. Commun.* **13**, 7760 (2022).
23. Pesah, A. et al. Absence of barren plateaus in quantum convolutional neural networks. *Phys. Rev. X* **11**, 041011 (2021).
24. Sack, S. H., Medina, R. A., Michailidis, A. A., Kueng, R. & Serbyn, M. Avoiding barren plateaus using classical shadows. *PRX Quantum* **3**, 020365 (2022).
25. Schatzki, L. et al. Theoretical guarantees for permutation-equivariant quantum neural networks. *npj Quantum Inf.* **10**, 12 (2024).
26. Zhang, Hao-Kai, Liu, S. & Zhang, Shi-Xin Absence of barren plateaus in finite local-depth circuits with long-range entanglement. *Phys. Rev. Lett.* **132**, 150603 (2024).
27. Zhang, K., Liu, L., Hsieh, M.-H. & Tao, D. Escaping from the barren plateau via Gaussian initializations in deep variational quantum circuits, in *Advances in Neural Information Processing Systems*. https://openreview.net/forum?id=jXgbJdQ2Yly.
28. Park, Chae-Yeun & Killoran, N. Hamiltonian variational ansatz without barren plateaus. *Quantum* **8**, 1239 (2024).
29. Wang, Y., Qi, B., Ferrie, C. & Dong, D. Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization. *Phys. Rev. Appl.* **22**, 054005 (2024).
30. Park, C.-Y., Kang, M. & Huh, J. Hardware-efficient ansatz without barren plateaus in any depth, https://arxiv.org/abs/2403.04844 arXiv preprint arXiv:2403.04844 (2024).
31. Shi, X. and Shang, Y. Avoiding barren plateaus via Gaussian mixture model, https://arxiv.org/abs/2402.13501 arXiv preprint arXiv:2402.13501 (2024).
32. Puig, R., Drudis, M., Thanasilp, S. & Holmes, Zoë Variational quantum simulation: A case study for understanding warm starts. *PRX Quantum* **6**, 010317 (2025).
33. Cao, C., Zhou, Y., Tannu, S., Shannon, N. & Joynt, R. Exploiting many-body localization for scalable variational quantum simulation, https://arxiv.org/abs/2404.17560 arXiv preprint arXiv:2404.17560 (2024).
34. Mele, A. A., Mbeng, G. B., Santoro, G. E., Collura, M. & Torta, P. Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz. *Phys. Rev. A* **106**, L060401 (2022).

35. Lerch, S. et al. Efficient quantum-enhanced classical simulation for patches of quantum landscapes, https://doi.org/10.48550/arXiv.2411.19896 (2024).
36. Bermejo, P. et al. Quantum convolutional neural networks are (effectively) classically simulable, https://arxiv.org/abs/2408.12739 arXiv preprint arXiv:2408.12739 (2024).
37. Mockus, J. On the Bayes methods for seeking the extremal point. *IFAC Proc. Vol.* **8**, 428–431 (1975).
38. Mockus, J. B. & Mockus, L. J. Bayesian approach to global optimization and application to multiobjective and constrained problems. *J. Optim. theory Appl.* **70**, 157–172 (1991).
39. Otterbach, J. S. et al. Unsupervised machine learning on a hybrid quantum computer, https://arxiv.org/abs/1712.05771 arXiv preprint arXiv:1712.05771 (2017).
40. Sung, K. J. et al. Using models to improve optimizers for variational quantum algorithms. *Quantum Sci. Technol.* **5**, 044008 (2020).
41. Shaffer, R., Kocia, L. & Sarovar, M. Surrogate-based optimization for variational quantum algorithms. *Phys. Rev. A* **107**, 032415 (2023).
42. Müller, J., Lavrijsen, W., Iancu, C. & de Jong, W. Accelerating noisy VQE optimization with Gaussian processes, in https://doi.org/10.1109/QCE53715.2022.00041 *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2022) pp. pages 215–225.
43. Cheng, L., Chen, Yu-Qin, Zhang, Shi-Xin & Zhang, S. Quantum approximate optimization via learning-based adaptive optimization. *Commun. Phys.* **7**, 83 (2024).
44. Kim, J. E. and Wang, Y. Quantum approximate Bayesian optimization algorithms with two mixers and uncertainty quantification, https://doi.org/10.1109/TQE.2023.3327055 *IEEE Trans. Quant. Eng.* (2023).
45. Pelofske, E., Hahn, G. & Djidjev, H. N. Advanced anneal paths for improved quantum annealing, in https://doi.org/10.1109/QCE49297.2020.00040 *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2020) pp. 256–266
46. Finžgar, JernejRudi, Schuetz, MartinJ. A., Brubaker, J. K., Nishimori, H. & Katzgraber, H. G. Designing quantum annealing schedules using bayesian optimization. *Phys. Rev. Res.* **6**, 023063 (2024).
47. Rasmussen, C. E. Gaussian processes in machine learning, in https://doi.org/10.1007/978-3-540-28650-9_4 *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*, edited by Bousquet, O., U. von Luxburg, and G. Rätsch (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) pp. 63–71.
48. Gustafson, E. J. et al. Surrogate optimization of variational quantum circuits, https://doi.org/10.48550/arXiv.2404.02951 arXiv preprint arXiv:2404.02951 (2024).
49. Cohn, D., Atlas, L. & Ladner, R. Improving generalization with active learning. *Mach. Learn.* **15**, 201–221 (1994).
50. Diaw, A., McKerns, M., Sagert, I., Stanton, L. G. & Murillo, M. S. Efficient learning of accurate surrogates for simulations of complex systems. *Nat. Mach. Intell.* **6**, 568 (2024).
51. Czarnik, P., McKerns, M., Sornborger, A. T. & Cincio, L. Improving the efficiency of learning-based error mitigation. *Quantum* **9**, 1727 (2025).
52. McKerns, M., Alexander, F., Hickmann, K., Sullivan, T. J. & Vaughn, D. https://arxiv.org/pdf/2009.06626*Advanced Analysis Solutions for Leading Experimental Techniques*, edited by K. Kleese van Dam et al. (World Scientific, 2019).
53. McKerns, M. Is automated materials design and discovery possible? https://link.springer.com/chapter/10.1007/978-3-319-99465-9_2 Materials Discovery and Design: By Means of Data Science and Optimal Learning, 15–58 (2018).
54. Crooks, G. E. Performance of the quantum approximate optimization algorithm on the maximum cut problem, https://arxiv.org/abs/1811.08419 arXiv preprint arXiv:1811.08419 (2018).
55. Boulebnane, S. & Montanaro, A. Predicting parameters for the quantum approximate optimization algorithm for MAX-CUT from the infinite-size limit, https://arxiv.org/abs/2110.10685 arXiv preprint arXiv:2110.10685 (2021).
56. Wurtz, J. & Love, P. MaxCut quantum approximate optimization algorithm performance guarantees for $p>1$. *Phys. Rev. A* **103**, 042612 (2021).
57. Sack, S. H. & Egger, D. J. Large-scale quantum approximate optimization on nonplanar graphs with machine learning noise mitigation. *Phys. Rev. Res.* **6**, 013223 (2024).
58. Santra, GopalChandra, Jendrzejewski, F., Hauke, P. & Egger, D. J. Squeezing and quantum approximate optimization. *Phys. Rev. A* **109**, 012413 (2024).
59. Maciejewski, F. B. et al. A multilevel approach for solving large-scale QUBO problems with noisy hybrid quantum approximate optimization, https://arxiv.org/abs/2408.07793 (2024).
60. Zhou, L., Wang, Sheng-Tao, Choi, S., Pichler, H. & Lukin, M. D. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Phys. Rev. X* **10**, 021067 (2020).
61. Cheng, L. https://github.com/sherrylixuecheng/EMQAOA-DARBO EMQAOA-DARBO.
62. Chamberland, C. et al. Topological and subsystem codes on low-degree graphs with flag qubits, https://doi.org/10.1103/physrevx.10.011022 *Phys. Rev.* X **10** (2020)
63. Pelofske, E., Bärtschi, A. & Eidenbenz, S. Short-depth QAOA circuits and quantum annealing on higher-order Ising models. *npj Quantum Inf.* **10**, 30 (2024).
64. Pelofske, E., Bärtschi, A. & Eidenbenz, S. Quantum annealing vs. QAOA: 127 Qubit Higher-Order Ising Problems on NISQ Computers, in https://doi.org/10.1007/978-3-031-32041-5_13 *International Conference on High Performance Computing ISC HPC'23* (2023) pp. 240–258.
65. Pelofske, E. et al. Scaling whole-chip QAOA for higher-order Ising spin glass models on heavy-hex graphs, https://doi.org/10.1038/s41534-024-00906-w npj Quantum Information (2024).
66. Galda, A., Liu, X., Lykov, D., Alexeev, Y. & Safro, I. Transferability of optimal QAOA parameters between random graphs, in https://doi.org/10.1109/QCE52317.2021.00034 *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, year 2021) pp. 171–180.
67. Akshay, V., Rabinovich, D., Campos, E. & Biamonte, J. Parameter concentrations in quantum approximate optimization. *Phys. Rev. A* **104**, L010401 (2021).
68. Brandao, F. G., Broughton, M., Farhi, E., Gutmann, S. & Neven, H. For fixed control parameters the quantum approximate optimization algorithm's objective function value concentrates for typical instances, https://arxiv.org/abs/1812.04170 arXiv preprint arXiv:1812.04170 (2018).
69. Farhi, E., Goldstone, J., Gutmann, S. & Zhou, L. The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size. *Quantum* **6**, 759 (2022).
70. Galda, A. et al. Similarity-based parameter transferability in the quantum approximate optimization algorithm. *Front. Quantum Sci. Technol.* **2**, 1–16 (2023).
71. Shaydulin, R., Lotshaw, P. C., Larson, J., Ostrowski, J. & Humble, T. S. Parameter transfer for quantum approximate optimization of weighted MaxCut. *ACM Trans. Quantum Comput.* **4**, 1–15 (2023).
72. Fannes, M., Nachtergaele, B. & Werner, R. Finitely correlated states on quantum spin chains. *Comm. Math. Phys.* **144**, 443–490 (1992).
73. Rajakumar, J., Golden, J., Bärtschi, A. & Eidenbenz, S. Trainability barriers in low-depth qaoa landscapes, in https://doi.org/10.1145/3649153.3649204 *Proceedings of the 21st ACM International Conference on Computing Frontiers*, CF '24 (Association for Computing Machinery, New York, NY, USA, 2024) p. 199–206.

74. Grimsley, H. R., Economou, S. E., Barnes, E. & Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **10**, 1–9 (2019).
75. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
76. Egger, D. J., Mareček, J. & Woerner, S. Warm-starting quantum optimization. *Quantum* **5**, 479 (2021).
77. Tate, R., Moondra, J., Gard, B., Mohler, G. & Gupta, S. Warm-Started QAOA with custom mixers provably converges and computationally beats Goemans-Williamson's Max-Cut at Low Circuit Depths. *Quantum* **7**, 1121 (2023).
78. do Carmo, R. S. et al. https://arxiv.org/abs/2504.19934 Warm-starting qaoa with xy mixers: A novel approach for quantum-enhanced vehicle routing optimization, (2025).
79. Patra, S., Jahromi, S. S., Singh, S. & Orus, R., Efficient tensor network simulation of IBM's largest quantum processors, arXiv preprint arXiv:2309.15642 https://doi.org/10.1103/PhysRevResearch.6.013326 (2023).
80. Nemkov, N. A., Kiktenko, E. O. & Fedorov, A. K. Fourier expansion in variational quantum algorithms. *Phys. Rev. A* **108**, 032406 (2023).
81. Goh, M. L., Larocca, M., Cincio, L., Cerezo, M. & Sauvage, Frédéric Lie-algebraic classical simulations for quantum computing. *Phys. Rev. Res.* **7**, 033266 (2025).
82. Begušić, T., Hejazi, K. & Chan, G. K. Simulating quantum circuit expectation values by Clifford perturbation theory, https://doi.org/10.1063/5.0269149 *J. Chem. Phys.* **162** (2025).
83. Pelofske, E., Bärtschi, A. & Eidenbenz, S. Classical combinatorial optimization scaling for random ising models on 2d heavy-hex graphs, arXiv preprint arXiv:2412.15572 (2024).
84. Cai, Z. et al. Quantum error mitigation. *Rev. Mod. Phys.* **95**, 045005 (2023).
85. Barron, S. V. et al. Provable bounds for noise-free expectation values computed from noisy samples, https://doi.org/10.1038/s43588-024-00709-1 *Nat. Comput. Sci.* 1–11 (2024).
86. Wang, S. et al. Can error mitigation improve trainability of noisy variational quantum algorithms? *Quantum* **8**, 1287 (2024).
87. McKerns, M., Strand, L., Sullivan, T., Fang, A. & Aivazis, M.A.G. Building a framework for predictive science, https://doi.org/10.48550/arXiv.1202.1056 *Proceedings of the 10th Python in Science Conference* (2011).
88. McKerns, M., Hung, P. & Aivazis, M. https://uqfoundation.github.io/project/mystic mystic: highly-constrained non-convex optimization and UQ, (2009).
89. Gutmann, H.-M. A radial basis function method for global optimization. *J. Glob. Optim.* **19**, 201–227 (2001).

Acknowledgements

We thank Marco Cerezo for helpful conversations. This work was supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001). The research for this publication was supported by a grant from the Priority Research Area DigiWorld under the Strategic Programme Excellence Initiative at Jagiellonian University. PC acknowledges support by the National Science Centre (NCN), Poland, under project 2022/47/D/ST2/03393. TO acknowledges partial support by the EPSRC through an EPSRC iCASE studentship award. LC was partially supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Contract No. DE-AC05-00OR22725 through the Accelerated Research in Quantum Computing Program MACH-Q project. This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Science Center (QSC). ATS acknowledges initial support from the LANL ASC Beyond Moore's Law project and subsequent support from the QSC. MM acknowledges support by the Uncertainty Quantification Foundation under the Statistical Learning program. The Uncertainty Quantification Foundation is a nonprofit dedicated to the advancement of predictive science through research, education, and the development and dissemination of advanced technologies. EP acknowledges support from the NNSA's ASC Beyond Moore's Law Program at LANL. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

Author contributions

T.O., P.C., E.P., A.T.S., M.M., and L.C. conceived the project, contributed to the experimental design, and wrote the manuscript. P.C. performed the experiments using Matrix Product State simulation. T.O. performed the experiments using the IBM quantum computers. M.M. is the primary author of the 'mystic' Python library used in the experiments and implemented the optimization procedures used in this study. E.P. provided code to generate the random heavy-hex lattice Ising models and the QAOA circuit details.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-025-02423-4.

Correspondence and requests for materials should be addressed to Tom O'Leary or Piotr Czarnik.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.