




Scalability challenges in variational quantum optimization under stochastic noise

Adelina Bärligea ^{1,2,*}, Benedikt Poggel ¹ and Jeanette Miriam Lorenz ^{1,3}

¹*Fraunhofer Institute for Cognitive Systems IKS, Munich, Germany*

²*Physics Department, Technical University of Munich, Garching, Germany*

³*Faculty of Physics, Ludwig-Maximilians-University Munich, Munich, Germany*



(Received 13 March 2025; accepted 29 July 2025; published 3 September 2025)

With rapid advances in quantum hardware, a central question is whether quantum devices with or without full error correction can outperform classical computers on practically relevant problems. Variational quantum algorithms (VQAs) have gained significant attention as promising candidates in this pursuit, particularly for combinatorial optimization problems. While reports of their challenges and limitations continue to accumulate, many studies remain optimistic based on small-scale, idealized testing setups, leaving doubt about the scalability of VQAs for large-scale problems. We systematically investigate this scaling behavior by analyzing how classical optimizers minimize variational quantum loss functions for random Quadratic Unconstrained Binary Optimization instances in the presence of uncertainty, modeled as effective Gaussian noise. We find that the critical noise threshold for successful optimization decreases rapidly as system size grows. This decline exceeds what can be explained solely by shrinking loss variance, confirming deeper, fundamental limitations in the loss landscapes of VQAs beyond barren plateaus. Translating these thresholds into required measurement shots reveals that achieving sufficient precision in the evaluated loss values quickly becomes impractical, even for moderately sized problems. Our findings demonstrate serious scalability challenges for VQAs in optimization stemming from mere uncertainty, indicating potential barriers to achieving practical quantum advantage with current hybrid approaches.

DOI: [10.1103/rgyh-8xw8](https://doi.org/10.1103/rgyh-8xw8)

I. INTRODUCTION

While quantum hardware has advanced rapidly [1–3], scalable and fault-tolerant quantum computing may still be a long way off. John Preskill [4] coined the term NISQ (noisy intermediate-scale quantum) to describe a generation of quantum devices, which operate with limited resources and significant noise levels. An open question of the NISQ era is whether these devices can achieve a practically relevant quantum advantage. In this context, variational quantum algorithms (VQAs) have emerged as potentially promising candidates for realizing near-term quantum advantage in a wide range of applications [5–7]. One of the most anticipated use cases, particularly in industry, is combinatorial optimization, where current quantum approaches mainly rely on approximation algorithms and heuristics such as VQAs [8].

VQAs are hybrid schemes that leverage quantum computers to evaluate a loss function while relying on classical optimization methods for loss minimization. They employ shallow, parameterized quantum circuits (ansatzes), making them well suited to the resource constraints of NISQ

devices. Yet a growing body of research highlights significant challenges and limitations of VQAs: One major issue is the barren plateau phenomenon [9,10], where the loss values and gradients vanish exponentially with system size in most of the proposed quantum ansatzes, making parameter optimization intractable. Conversely, structural properties leading to the provable absence of barren plateaus might enable efficient classical simulations [11,12]. Additionally, the classical optimization problem within VQAs has been proven to be NP-hard [13], with loss landscapes plagued by numerous suboptimal local minima [14], further complicating the convergence to high-quality solutions. These challenges are amplified by noise, which not only leads to unfavorable resource scaling and unpredictable behavior but may also eliminate any potential quantum advantage altogether [15–21]. Crucially, these limitations persist regardless of whether the effective noise comes from hardware imperfections or finite sampling errors inherent to quantum measurement. Consequently, despite ongoing efforts and widespread optimism, no practically meaningful quantum advantage has been demonstrated for VQAs on near-term devices to date.

Thus, the critical question remains: can variational quantum optimization become practical in the presence of unavoidable stochastic errors? While some studies directly address this question (see Sec. II), most research focuses on developing new heuristic approaches, typically tested on small-scale and idealized setups. Such benchmarks often yield promising insights but rarely capture how well algorithms

*Contact author: adelina.baerligea@tum.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

scale under realistic noise conditions and increasing problem sizes, potentially painting an overly optimistic picture of their viability. More importantly, many studies overlook the role of the classical optimization process—the central computational bottleneck of VQAs—which may ultimately determine whether a problem remains solvable as system size grows. Closing this gap is essential for distinguishing genuinely promising advancements from appealing but unproven claims in quantum optimization and related disciplines.

This work investigates the viability of VQAs for solving classical optimization problems in the Quadratic Unconstrained Binary Optimization (QUBO) formulation, with a specific focus on the classical optimization routine as a key driver of performance. We systematically analyze how state-of-the-art classical optimizers handle variational quantum loss functions under increasing levels of Gaussian noise, providing an empirical assessment of practical scalability limits in variational quantum optimization. The primary contributions of this study are the following:

(1) *Performance evaluation framework*: We propose a systematic methodology to assess the scaling behavior of hybrid variational quantum optimization methods under stochastic noise, providing a robust foundation for understanding their potential and limitations, even in fault tolerance.

(2) *Scaling analysis of classical optimizers*: This study presents, to the best of our knowledge, the first in-depth scaling analysis of classical optimizer performance on noisy quantum loss functions for random QUBO problem instances. By examining two essential factors—noise level (σ) and system size (n)—we identify a clear threshold $\sigma^*(n)$, beyond which reliable optimization becomes exponentially challenging. Via extrapolation to regions not (yet) accessible numerically we address the feasibility of these algorithms for larger-scale problems.

(3) *Insights on practical quantum advantage*: We quantify how uncertainty in the loss values, particularly arising from finite measurement sampling, can significantly impact the practical scalability of VQAs. These results highlight potential barriers to achieving quantum advantage in variational quantum optimization under realistic noise scenarios.

Our empirical findings offer clear benchmarks and valuable insights into the practical limits of VQAs under noise. By fostering a more realistic understanding of the potential of quantum optimization on near-term devices and beyond, these insights help to guide future research towards more promising algorithms and applications in quantum computing.

II. RELATED WORK

The trainability of VQAs has been extensively explored in the literature. Several key obstacles impairing successful parameter optimization have been identified, including reachability deficits [22], inherent NP-hardness of optimization [13], complex loss landscapes with numerous suboptimal local minima [14], and barren plateaus [9,10], i.e., the phenomenon of exponentially vanishing loss values and gradients. Specifically, recent work indicates that all sources of barren plateaus, such as high ansatz expressibility [9], global observables [23], or hardware-induced noise [16], are fundamentally connected to the scaling behavior of the dynamical

Lie algebra (DLA) associated with the chosen ansatz [23–26]. Our study explicitly circumvents these issues by considering only ansatz circuits characterized by only polynomially growing DLA dimensions along with strictly local observables, ensuring that the underlying quantum loss functions are provably barren-plateau-free and remain “trainable” in theory.

A significant body of research has also addressed the detrimental effects of hardware noise on VQA performance. Numerous studies have demonstrated that even minimal noise levels can rapidly propagate through quantum circuits, resulting in output distribution that become effectively simulable classically [17,19–21,27–29], thereby negating potential quantum advantage. In contrast, our analysis assumes a fault-tolerant scenario, simulating uncertainty exclusively via Gaussian noise applied as a postprocessing step to exact quantum loss function evaluations. This setup effectively models an idealized, fully error-corrected scenario with finite sampling noise as the dominating source of uncertainty. This setting not only circumvents previously identified noise-induced limitations but also aligns more closely with realistic conditions under which quantum advantage for (approximate) optimization can be expected to first emerge.

Given the empirical and experimental nature of our study and its focus on the classical optimization routine, our findings naturally compare to existing numerical assessments of classical optimizer performance within VQAs [30–38]. However, as already noted in Ref. [39], these studies often evaluate performance under fixed conditions using static setups primarily focusing on comparing optimizer effectiveness. Consequently, they offer limited insights into scalability constraints and the quantitative limits of classical optimizer performance in hybrid quantum algorithms with respect to noise thresholds or system size.

Most closely aligned with our work in terms of scope and metrics is the more recent study by Scriva *et al.* [18], which employs similar performance metrics and comes to similar conclusions. However, methodological differences exist: Scriva *et al.* [18] focus exclusively on shot noise and evaluate iteration complexity relative to fixed success probabilities. Conversely, our approach systematically investigates the effects of varying levels of Gaussian noise, allowing for a broader and more generalizable analysis of the precision requirements for successful optimization. A complementary theoretical investigation into the convergence properties of classical optimizers within a similar context is provided by Kungurtsev *et al.* [40].

III. THEORETICAL BACKGROUND

Optimization has long been regarded as a promising application of quantum computing, particularly for combinatorial optimization problems [41], where the number of possible solutions grows exponentially with problem size. Many such problems are NP-hard [42,43], making them intractable for classical algorithms in the worst case. Quantum algorithms offer several potential approaches to tackling these challenges [8], including Grover search [44], quantum phase estimation [45], adiabatic quantum computing [46], and heuristic approximation algorithms. Among these, variational quantum algorithms [6] have gained particular attention due to

their adaptability to near-term quantum devices. Although extensively studied [47–51], no practically relevant quantum advantage has been demonstrated in quantum optimization yet. While promising results exist for small-scale or noise-free scenarios, it remains an open question whether these approaches can ultimately outperform classical heuristics, which are already highly optimized for practical applications.

A. Quadratic unconstrained binary optimization problems

A QUBO problem is a widely used mathematical framework for formulating combinatorial optimization problems, defined as

$$\min_{\mathbf{x}} \mathbf{x}^\top \mathbf{Q} \mathbf{x}, \quad (1)$$

where $\mathbf{x} \in \{0, 1\}^n$ is a vector of binary decision variables, and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is a real-valued cost matrix that fully characterizes the problem. Due to its versatility, the QUBO formulation can encode a broad range of important optimization problems [52–55]. Beyond its flexibility, QUBO is closely linked to statistical physics, specifically to Ising spin glasses [56]. This connection allows any QUBO problem to be reformulated as an Ising Hamiltonian [57] (cf. Appendix A). This equivalence provides a bridge between classical optimization and quantum physics motivating the application of quantum computing to combinatorial optimization problems.

The Ising Hamiltonian of an n -spin system is given by

$$\mathbf{C} = \sum_{i=1}^n C_{ii} \mathbf{Z}_i + \sum_{i < j} C_{ij} \mathbf{Z}_i \mathbf{Z}_j, \quad (2)$$

where \mathbf{Z}_i is the Pauli- \mathbf{Z} operator acting on the i th spin, with eigenvalues ± 1 , C_{ij} (for $i \neq j$) represents the interaction energy between spins i and j , and C_{ii} describes the interaction between each spin with its external magnetic field. Given a computational basis state $|\mathbf{q}\rangle = \bigotimes_{i=1}^n |q_i\rangle$ with $q_i \in \{0, 1\}$, its total energy $\langle \mathbf{C} \rangle := \langle \mathbf{q} | \mathbf{C} | \mathbf{q} \rangle$ is given by

$$\langle \mathbf{C} \rangle = \sum_{i=1}^n (-1)^{q_i} C_{ii} + \sum_{i < j} (-1)^{q_i + q_j} C_{ij}. \quad (3)$$

Thus, finding the ground state $|\mathbf{q}^*\rangle$ of the system corresponds to minimizing the expectation value of Eq. (3) with respect to $|\mathbf{q}\rangle$, which is known as the Ising problem, a direct equivalent of the QUBO problem in the optimization context.

B. Variational quantum optimization

Until fault-tolerant quantum computing becomes viable, variational quantum algorithms have emerged as a leading heuristic approach for tackling computationally hard problems, such as combinatorial optimization and other complex applications like molecular simulations [5–7].

VQAs are hybrid quantum-classical algorithms where a quantum computer evaluates the expectation value of a Hermitian cost operator \mathbf{C} , for example, given by Eq. (2), with respect to a candidate state

$$|\Psi(\theta)\rangle = \mathbf{U}(\theta)|\Psi_0\rangle, \quad (4)$$

where $\mathbf{U}(\theta)$ is a parameterized quantum circuit (PQC) acting on an easily prepared initial state $|\Psi_0\rangle$, such as the uniform

superposition. The classical computer then performs an optimization over the parameters θ to solve

$$\min_{\theta} \mathcal{L}(\theta) \quad \text{with} \quad \mathcal{L}(\theta) := \langle \Psi(\theta) | \mathbf{C} | \Psi(\theta) \rangle. \quad (5)$$

With optimal parameters θ^* , the optimized state $|\Psi(\theta^*)\rangle$ approximates the true ground state $|\mathbf{q}^*\rangle$. This is a consequence of the variational principle, which ensures that $\langle \psi | \mathbf{C} | \psi \rangle$ is always lower bounded by the ground state energy of the system [5].

While the classical optimization methods used in VQAs are often well established, much of the research effort focuses on designing well-behaved quantum ansatzes, which define both the PQC structure and the measurement strategy for estimating the loss function, such as in Eq. (3). Two of the most widely studied VQAs are the Variational Quantum Eigensolver (VQE) [58], initially developed for quantum chemistry, and the Quantum Approximate Optimization Algorithm (QAOA) [59], a VQE variant explicitly designed for the field of combinatorial optimization. Both algorithms have been extensively explored for a variety of general optimization tasks [60–63] since then, making them and their adaptations central candidates in the search for near-term quantum advantage.

Figure 1 illustrates representative ansatz circuits for both algorithm classes, which are also used in the experiments presented in Sec. IV B. For VQE, a typical example is a hardware-efficient ansatz, consisting of alternating layers of single-qubit rotations and entangling operations, followed by a global measurement. QAOA, in contrast, follows a problem-inspired approach, implementing a trotterized version of adiabatic quantum optimization [46]. It consists of alternating applications of a shift operator $\mathbf{U}_P(\gamma) = e^{-i\gamma \mathbf{H}_P}$ based on the problem Hamiltonian \mathbf{H}_P , and a mixing operator $\mathbf{U}_M(\beta) = e^{-i\beta \mathbf{H}_M}$, where the default choice is a sum of Pauli- \mathbf{X} matrices on each qubit, i.e., $\mathbf{H}_M = \sum_{i=1}^n \mathbf{X}_i$. The resulting QAOA ansatz is composed of p layers, parameterized by $2p$ angles, with the unitary evolution

$$\mathbf{U}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \mathbf{U}_M(\beta_p) \mathbf{U}_P(\gamma_p) \cdots \mathbf{U}_M(\beta_1) \mathbf{U}_P(\gamma_1). \quad (6)$$

The unitary is applied to an initial state—typically the uniform superposition—and the objective is evaluated as in Eq. (5).

A recently proposed alternative to these well-established algorithms is the Block Encoding Quantum Optimizer (BENQO) [64,65], which introduces a novel VQA architecture while demonstrating promising results on small-scale problems. Designed explicitly for solving Ising problems, BENQO employs block encoding [66], a technique that embeds the cost operator $\mathbf{C} \in \mathbb{R}^{2^n \times 2^n}$ (2) into a larger unitary matrix at the expense of one additional qubit:

$$\mathbf{U} := \begin{bmatrix} \sin(\hat{\mathbf{C}}) & \cos(\hat{\mathbf{C}}) \\ \cos(\hat{\mathbf{C}}) & -\sin(\hat{\mathbf{C}}) \end{bmatrix} \quad \text{with} \quad \hat{\mathbf{C}} := \mathbf{C}/K, \quad (7)$$

where K is a scaling factor that ensures a monotonic relationship between $\langle \mathbf{U} \rangle$ and $\langle \mathbf{C} \rangle$. Then $\langle \mathbf{U} \rangle$ can be used interchangeably with $\langle \mathbf{C} \rangle$ in the optimization context. Unlike conventional VQAs, which rely on global measurements to estimate expectation values (cf. Fig. 1), BENQO circumvents this step by using the Hadamard test [67]. This method

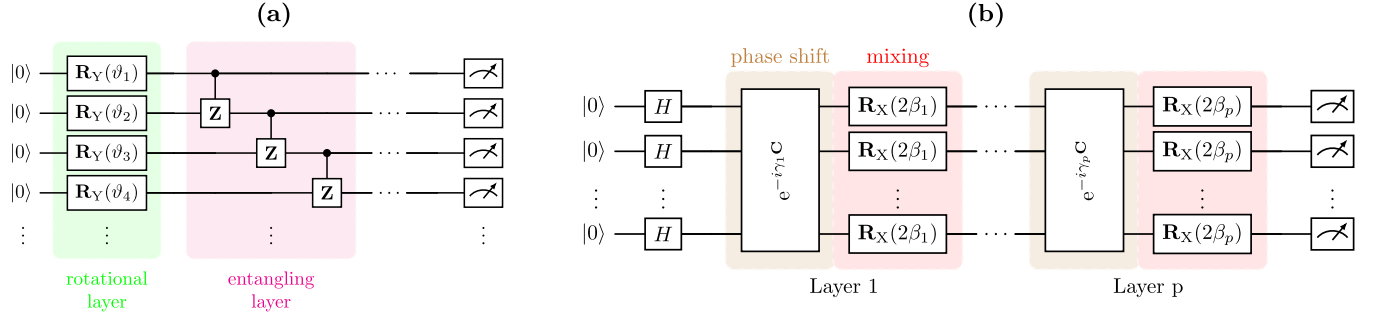


FIG. 1. Representative VQA ansatz circuits. (a) Two-local VQE ansatz. After initializing all qubits to $|0\rangle^{\otimes n}$, alternating layers of single-qubit rotations $R_Y(\vartheta)$ and linear entangling operations via controlled-Z gates are applied. A global measurement follows to evaluate the circuit's output. "Two-local" refers to the ansatz structure, where each entangling gate acts only on pairs of qubits. (b) Conventional QAOA ansatz. After initializing all qubits to the equal superposition state $|+\rangle$ using Hadamard gates, p alternating layers of phase shift and mixing operators are applied as described in Eq. (6), again, followed by a global measurement.

projects the eigenvalues of C onto an ancillary qubit, allowing for their determination via single-qubit measurements.

Figure 2 illustrates BENQO's ansatz circuit. The algorithm first applies a rotational layer $|\Psi(\theta)\rangle = \bigotimes_{i=1}^n R_Y(\theta_i)|0\rangle$, parameterized by $\theta \in \mathbb{R}^n$, bringing the system into an initial state including an ancillary qubit:

$$|0\rangle_a \otimes |\Psi_{\text{in}}\rangle, \quad \text{with } |\Psi_{\text{in}}\rangle = |0\rangle_c \otimes |\Psi(\theta)\rangle. \quad (8)$$

The Hadamard test then projects the system onto the eigenspace of U using the projectors $P_{\pm} = \frac{I \pm U}{2}$. The resulting output state is

$$|\Psi_{\text{out}}\rangle = |0\rangle_a \otimes (P_+|\Psi_{\text{in}}\rangle) + |1\rangle_a \otimes (P_-|\Psi_{\text{in}}\rangle). \quad (9)$$

Measuring the ancillary qubit allows BENQO to extract the expectation value of U via the probabilities

$$p_0 = \|P_+|\Psi_{\text{in}}\rangle\|^2, \quad p_1 = \|P_-|\Psi_{\text{in}}\rangle\|^2. \quad (10)$$

From this, the cost function is obtained as $\langle U \rangle \equiv p_0 - p_1$, leading to

$$\langle C \rangle = K \arcsin(p_0 - p_1). \quad (11)$$

This unique measurement strategy distinguishes BENQO from conventional VQAs, providing an alternative route for quantum optimization. To showcase that the benchmarking and scaling analysis is broadly applicable beyond standard

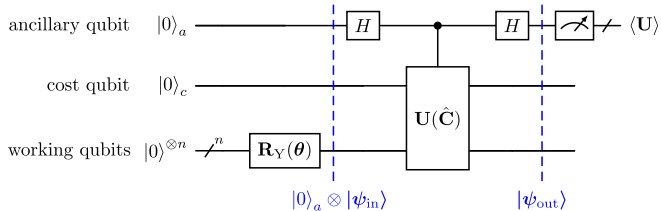


FIG. 2. BENQO's ansatz circuit. The quantum system begins in the initial state $|0\rangle_a \otimes |\Psi_{\text{in}}\rangle$ (8), including an ancillary qubit. The Hadamard test then projects the system onto the eigenspace of U via the projectors P_{\pm} , producing the output state $|\Psi_{\text{out}}\rangle$ (9). Measuring the ancillary qubit yields probabilities p_0 and p_1 , from which the expectation value of U and ultimately C can be extracted as in Eq. (11). This enables direct measurement of the cost function using a single qubit.

VQA methods, BENQO is included in the comparative loss function studies in Sec. IV B.

IV. EXPERIMENTAL SETUP

To ensure a generalizable scaling analysis, we account for the fact that the three main components of VQAs—problem encoding, the quantum subroutine (for loss evaluation), and the classical subroutine (for optimization)—each contributes differently to overall performance. To isolate the impact of classical optimization, the central component of VQAs, we fix the other two components under representative conditions, ensuring that our analysis directly reflects the capabilities of parameter optimizers rather than external constraints. Figure 3 provides a schematic overview of the experimental setup:

(1) *Problem encoding*: Random QUBO instances are used with the additional stipulation that the diagonal of the QUBO matrices contains only negative values. QUBO can be regarded as the standard formulation for combinatorial quantum optimization in the VQA context and provides a realistic benchmark (cf. Sec. IV A).

(2) *Quantum subroutine*: The loss function is derived from BENQO, as it exhibits a particularly well-behaved loss landscape compared to other established methods (cf. Sec. IV B).

(3) *Noise modeling*: Instead of relying on hardware-specific noise models, we introduce effective Gaussian noise directly to the loss function, ensuring a highly generalizable and systematic approach to model stochastic errors (cf. Sec. IV C) regardless of their origin.

(4) *Classical subroutine*: To eliminate biases arising from specific optimizer choices, we include a full suite of state-of-the-art methods in our experiments (cf. Sec. IV D).

(5) *Statistical evaluation*: By assigning a success index to each optimization run and repeating experiments 100 times, we derive statistical measures of problem solvability, enabling a systematic scaling analysis (cf. Sec. IV E).

This structured approach ensures that our findings are not dependent on one specific problem, optimizer, or noise type, allowing us to draw broader conclusions about the scalability of variational quantum optimization. The following sections provide the technical details behind each of these considerations.

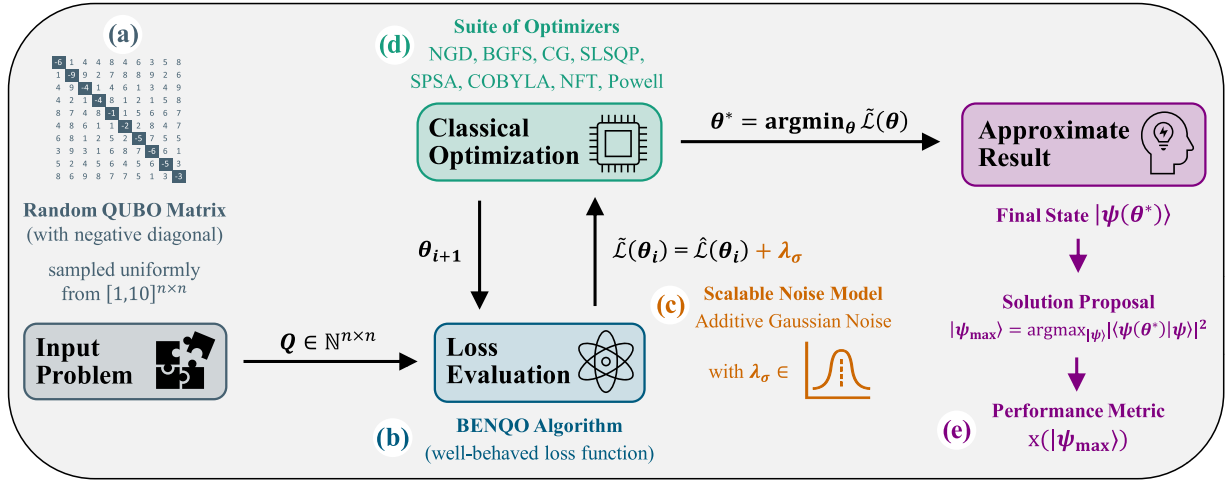


FIG. 3. Schematic overview of the experimental setup. (a) A random QUBO problem is sampled as input to the VQA pipeline (Sec. IV A). (b) The BENQO algorithm is simulated to obtain exact loss values for the given problem, chosen for its particularly well-behaved loss landscape (Sec. IV B). (c) The exact loss values are perturbed by additive Gaussian noise before being passed to a classical optimizer (Sec. IV C). (d) A suite of well-established classical optimization methods (Sec. IV D) attempts to find the optimal parameters minimizing the noisy loss function. (e) The results are then evaluated based on a predefined success criterion (Sec. IV E).

A. Random problem instances

Rather than focusing on a particular graph problem and mapping it to its Ising equivalent [57], this study employs randomly generated problem instances for the experiments. Each instance is uniquely seeded, with all entries of the $n \times n$ QUBO matrix Q sampled from a discrete uniform distribution over the interval $[1, 10]$:

$$Q_{ij} \sim \begin{cases} +\mathcal{U}(1, 10) & \text{if } i \neq j \\ -\mathcal{U}(1, 10) & \text{if } i = j \end{cases}, \quad \forall i, j \in \{1, \dots, n\}. \quad (12)$$

Since many practically relevant problems involve constraints that translate into quadratic penalty terms [54], the resulting QUBO matrices often feature negative diagonal elements (cf. Appendix B). To better reflect this structure, we also impose negative terms in Q (12). Setting the sampling range to start at 1 instead of 0 further ensures that the matrices represent fully connected graph problems, including self-connections. Lastly, the exact magnitude or precision of the weights is not critical, as the resulting loss values are normalized across all experiments.

B. Well-behaved quantum loss functions

Trainability in VQA architectures is generally hindered by two key challenges: the barren plateau (BP) phenomenon, characterized by an exponential concentration of the loss function and its gradients [10], and other unfavorable loss landscape structures which complicate optimization [13, 14]. While empirical studies on the latter are computationally demanding due to the high dimensionality of the parameter space [68], BPs can be detected more efficiently by analyzing the sample variance of the quantum loss function [9, 24].

Consider the exact loss function $\mathcal{L}(\theta)$, parameterized by $\theta \in \mathbb{R}^n$. To ensure consistent quantitative comparisons of the loss variance across different system sizes, this function must

be normalized, as suggested in [24]:

$$\hat{\mathcal{L}}(\theta) = \frac{\mathcal{L}(\theta)}{|\mathcal{L}(\theta)|_{\max}} \quad \text{with} \quad |\mathcal{L}(\theta)|_{\max} = \max_{\theta \in \mathbb{R}^n} |\mathcal{L}(\theta)|. \quad (13)$$

This normalization guarantees that the loss landscape is bounded within $[-1, 1]$, independent of the problem size n . The sample variance can thus be calculated as

$$\text{Var}[\{\hat{\mathcal{L}}_1, \dots, \hat{\mathcal{L}}_N\}] = \frac{1}{N-1} \sum_{i=1}^N (\hat{\mathcal{L}}_i - \bar{\hat{\mathcal{L}}})^2, \quad (14)$$

using a set of normalized values $\{\hat{\mathcal{L}}_i = \hat{\mathcal{L}}(\theta_i)\}_{i=1, \dots, N}$ sampled uniformly across the parameter space.

In Fig. 4 we illustrate the scaling of variances across three different VQA ansatzes: a conventional QAOA implementation from QISKIT, a hardware-efficient two-local VQE ansatz, consisting of one rotational layer and one linear entangling layer, and the BENQO algorithm (cf. Sec. III B). Each ansatz is parameterized such that the number of parameters matches the problem size n [69], ensuring a fair and unbiased comparison. Specifically, for the QAOA ansatz this amounts to $p = n/2$ layers, and therefore requires n to be even.

Based on a characterization of their dynamical Lie algebra (DLA) structures, we know neither the Two Local ansatz used here nor the problem-dependent BENQO ansatz exhibits exponentially scaling DLA dimensions [70, 71]. Conversely, QAOA when applied to the Ising cost function defined in Eq. (2) is known to have an exponentially scaling DLA dimension and therefore exhibits a barren plateau [24, 70]. While various mitigation strategies, such as smart parameter initialization methods (discussed in detail in Sec. VII D), can substantially improve the trainability of such ansatzes, this study explicitly concentrates on ansatzes that inherently avoid barren plateaus. Among the two remaining ones tested, BENQO consistently maintains the highest absolute variance in its loss values with increasing system size, suggesting the best parameter sensitivity. Therefore, BENQO is selected for

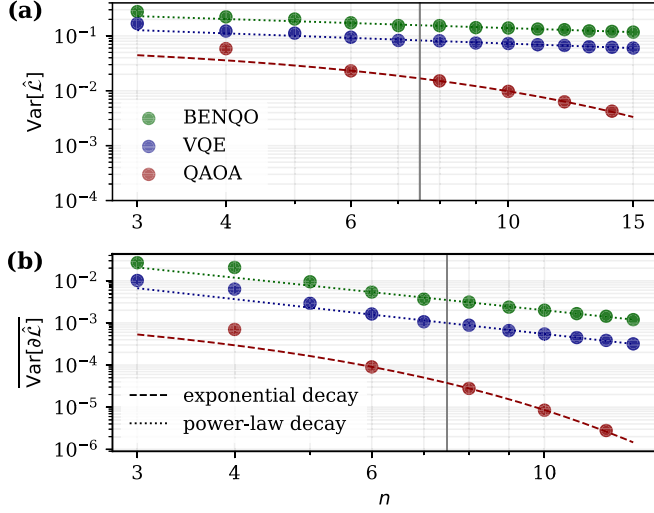


FIG. 4. Variance analysis of BENQO, VQE, and QAOA. (a) Sample variance of the normalized loss values $\hat{\mathcal{L}}(\theta)$ computed across 10000 parameter sets θ , uniformly sampled within $[-2\pi, 2\pi]^n$, for increasing system sizes n . (b) Mean sample variance of the n partial derivatives $\partial_\theta \hat{\mathcal{L}}(\theta)$, calculated as Eq. (17), and evaluated at the same 10000 parameter sets. Both plots include fitted theoretical curves representing exponential and power-law decay, with fitting performed only for $n \geq 8$ (indicated by the vertical line) to reduce the influence of small-scale effects.

the subsequent experiments as a representative scenario for a favorable case for classical optimization.

C. Scalable Gaussian noise model

In our experiments, effective noise is introduced as a post-processing step before the evaluated loss values are passed to the classical optimizers. This ensures that the original, simulated quantum output remains unchanged (cf. Fig. 3). Since parameter optimizers treat the quantum expectation value as black-box function, they respond only to the total error magnitude rather than its specific source, it is reasonable to approximate the cumulative effect of various noise sources using additive Gaussian noise, as motivated in Appendix C (cf. [39]). The noise model is applied as follows:

$$\tilde{\mathcal{L}}_\sigma(\theta) = \hat{\mathcal{L}}(\theta) + \lambda_\sigma \quad \text{with} \quad \lambda_\sigma \sim \mathcal{N}(0, \sigma). \quad (15)$$

Here $\hat{\mathcal{L}}(\theta)$ is the normalized exact loss value (13), and λ_σ is a random variable drawn from a normal distribution with mean 0 and standard deviation σ .

Although this method does not explicitly capture hardware-level error propagation, as explored in other studies [17,31,32], it may offer several advantages: First, it is highly scalable, allowing systematic evaluations across different problem sizes without the computational overhead of full quantum noise simulations. Second, it provides a generalizable and broadly applicable framework by collapsing all potential noise sources into a single tunable noise parameter σ , independent of the chosen quantum ansatz. As a result, other studies can map their findings onto a comparable σ value by computing the standard deviation of their total error,

enabling direct comparison regardless of the specific hardware or noise model used.

D. State-of-the-art optimizers

In VQAs, the principal optimization routine is handled by classical optimizers, which can be broadly classified into gradient-based and gradient-free methods. Gradient-based approaches are often preferred for their quick convergence in practice, but calculating derivatives can be computationally expensive or even infeasible in some cases. Consequently, gradient-free methods are widely used when gradients are unreliable, impractical, or costly to compute.

This study includes both categories of optimizers, ensuring a diverse representation of optimization strategies. A selection of well-established methods from QISKIT [72] and SciPy [73] as chosen (cf. Table I), along with a normalized gradient descent method (NGD) [74,75], proposed as such by Kuete Meli *et al.* [64], which has demonstrated promising performance in prior studies [64,65,75]. For the Qiskit optimizers, one should refer to the documentation and references listed in Table I.

NGD uses a normalized gradient as the search direction and an exponentially decaying step size for its update rule, ensuring stable convergence:

$$\theta_{k+1} \leftarrow \theta_k - \sqrt{\frac{\pi}{2}} n \exp\left(-\frac{4k^2}{k_{\max}^2}\right) \frac{\nabla_\theta \mathcal{L}(\theta_k)}{\|\nabla_\theta \mathcal{L}(\theta_k)\|_2}. \quad (16)$$

Here the only hyperparameter is the total number of optimization steps, k_{\max} , set to $k_{\max} = 20$, as recommended in previous studies [64].

Note that for optimizers requiring exact gradients $\nabla_\theta \mathcal{L}(\theta) = (\partial_{\theta_1} \mathcal{L}(\theta), \dots, \partial_{\theta_n} \mathcal{L}(\theta))^T$, the parameter-shift rule [84,85] was employed:

$$\frac{\partial}{\partial \theta_i} \mathcal{L}(\theta) = \frac{\mathcal{L}(\theta_{i,+}) - \mathcal{L}(\theta_{i,-})}{2}, \quad (17)$$

where $\theta_{i,\pm} := (\dots, \theta_i \pm \frac{\pi}{2}, \dots)$. This exact calculation method requires $2n$ function evaluations per gradient computation, and a total of $(2n + n)k_{\max}$ evaluations for the full optimization.

To ensure a fair comparison among all optimizers, we fix their available computational resources, using default settings from QISKIT. The initial parameters for each optimization run are sampled from a multivariate standard normal distribution centered at $\mathbf{0} \in \mathbb{R}^n$, a strategy shown to be effective in previous research [64,65,86]. Each experiment consists of a single optimization run with a fixed optimizer (from Table I), a specific noise level σ , and a random QUBO problem of size n . To account for statistical variation, we repeat each experiment $N = 100$ times with different problem instances.

E. Measures of solvability

The solution of a variational optimization task is defined as the n -dimensional parameter vector

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta), \quad (18)$$

which (approximately) minimizes the quantum loss function $\mathcal{L}(\theta)$. A widely used metric to evaluate the quality of this

TABLE I. Overview of tested optimizers. A representative set of both gradient-based and gradient-free parameter optimizers was selected based on a prescreening process to ensure the inclusion of methods that demonstrated promising performance and efficiency. Their respective class names in QISKIT are shown in the last column.

	Name	Description	Reference	In QISKIT
Gradient-based	NGD	Normalized gradient descent method	Kuete Meli <i>et al.</i> [64]	Custom
	BFGS	Broyden-Fletcher-Goldfarb-Shannon algorithm	Byrd <i>et al.</i> [76]	L_BFGS_B
	CG	Conjugate gradient method	Fletcher [77]	CG
	SLSQP	Sequential least squares programming method	Kraft [78]	SLSQP
Gradient-free	SPSA	Simultaneous perturbation stochastic approximation	Spall [79,80]	SPSA
	COBYLA	Constrained optimization by linear approximation	Powell [81,82]	COBYLA
	NFT	Nakanishi-Fuji-Todo algorithm	Nakanishi <i>et al.</i> [30]	NFT
	Powell	Conjugate direction method	Powell [83]	POWELL

solution is the approximation ratio (AR),

$$Q_{\text{AR}}(|\psi^*\rangle) = \frac{\langle \psi^* | \mathbf{C} | \psi^* \rangle - C_{\text{max}}}{C_{\text{min}} - C_{\text{max}}}, \quad Q_{\text{AR}} \in [0, 1], \quad (19)$$

which quantifies how close the solution state $|\psi^*\rangle = |\psi(\theta^*)\rangle$ is to the optimal state $|q^*\rangle$, based on the energy of the underlying quantum system. Here \mathbf{C} is the Hamiltonian cost operator (2), $C_{\text{min}} = \langle q^* | \mathbf{C} | q^* \rangle$ represents the minimum energy, and C_{max} is the maximum, according to Eq. (3):

$$C_{\text{max}} = \sum_{i=1}^n |\mathcal{C}_{ii}| + \sum_{i < j} |\mathcal{C}_{ij}|. \quad (20)$$

For classical optimization problems, the optimal state $|q^*\rangle$ corresponds to a computational basis state and can be expressed as a binary string. Thus, instead of directly evaluating $|\psi^*\rangle$ using the AR, the most probable basis state of the quantum output,

$$|\psi_{\text{max}}\rangle = \arg \max_{\psi_i} |\langle \psi^* | \psi_i \rangle|^2, \quad (21)$$

should be considered the solution candidate of the VQA. This choice is motivated by the practical perspective that real-world users aim to directly sample a high-quality solution, rather than undertake the impractical effort of searching for one within a delocalized quantum output distribution. Note that in a real experiment, $|\psi_{\text{max}}\rangle$ would be chosen based on the basis state with the highest measured shot count. However, since we are working with simulated results, we can use the exact definition.

To provide a more representative performance indicator, we use the approximation index x_t [64,65], which assesses whether the loss value of $|\psi_{\text{max}}\rangle$ is within a specified threshold $t \in [0, 1]$ of the global minimum:

$$x_t(|\psi_{\text{max}}\rangle) = \begin{cases} 1, & \text{if } Q_{\text{AR}}(|\psi_{\text{max}}\rangle) \geq t \\ 0, & \text{otherwise.} \end{cases} \quad (22)$$

For example, $x_1 = 1$, if and only if the most probable basis state $|\psi_{\text{max}}\rangle$ is the exact optimal solution $|q^*\rangle$.

Using this index, the probability of successfully finding an optimal (or near-optimal) solution—referred to as the “solvability”—is estimated as

$$\hat{p}_t = \frac{1}{N} \sum_{i=0}^N x_t(|\psi_{\text{max}}\rangle_i), \quad \hat{p}_t \in [0, 1], \quad (23)$$

where N is the total number of optimization runs. This probability reflects the fraction of runs among all experiments that yielded a solution within the specified threshold t . In the results presented in Sec. V, we denote the probability of finding the exact optimal solutions ($t = 1$) as \hat{p}_{opt} , while near-optimal solutions within a 5% and 1% margin of the optimal are monitored as $\hat{p}_{95\%}$ and $\hat{p}_{99\%}$, respectively.

The standard error of the sample success probability is given by

$$\sigma(\hat{p}_t) = \sqrt{\frac{\hat{p}_t(1 - \hat{p}_t)}{N}}, \quad (24)$$

which follows from the normal approximation of a binomial distribution and is valid for sufficiently large N .

V. RESULTS OF SCALING ANALYSIS

The experiments described above are conducted for system sizes $n \in [3, 10]$ and Gaussian noise levels $\sigma \in (10^{-3}, 10^1)$. Figure 5 presents the combined results as two-dimensional grids, showing the measured success probabilities \hat{p}_{opt} , $\hat{p}_{99\%}$, and $\hat{p}_{95\%}$ as contours. Note that the optimizers BFGS, CG, and SLSQP are excluded due to their poor performance across all tested noise settings, and their results are therefore omitted from further analysis.

The visualizations in Fig. 5 reveal both optimizer-specific differences and common trends across all methods. In particular, a well-defined transition emerges between regions where optimization is successful and those where the success probability drops rapidly to near zero. For exact optimality (\hat{p}_{opt}), this separation appears as a line, while for $\hat{p}_{95\%}$, the boundary takes on a different shape. The following sections provide a detailed quantitative analysis of these observations.

A. Solvability under additive Gaussian noise

To analyze the impact of increasing noise levels on solvability, the first scaling analysis is conducted at fixed system size. Figure 6 shows the experimental data for an exemplary system size of $n = 6$. On a logarithmic scale in the noise level σ , the curves resemble sigmoidal functions with horizontal asymptotes for $\ln(\sigma) \rightarrow \pm\infty$. As a typical sigmoidal function, the hyperbolic tangent defined as $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ is

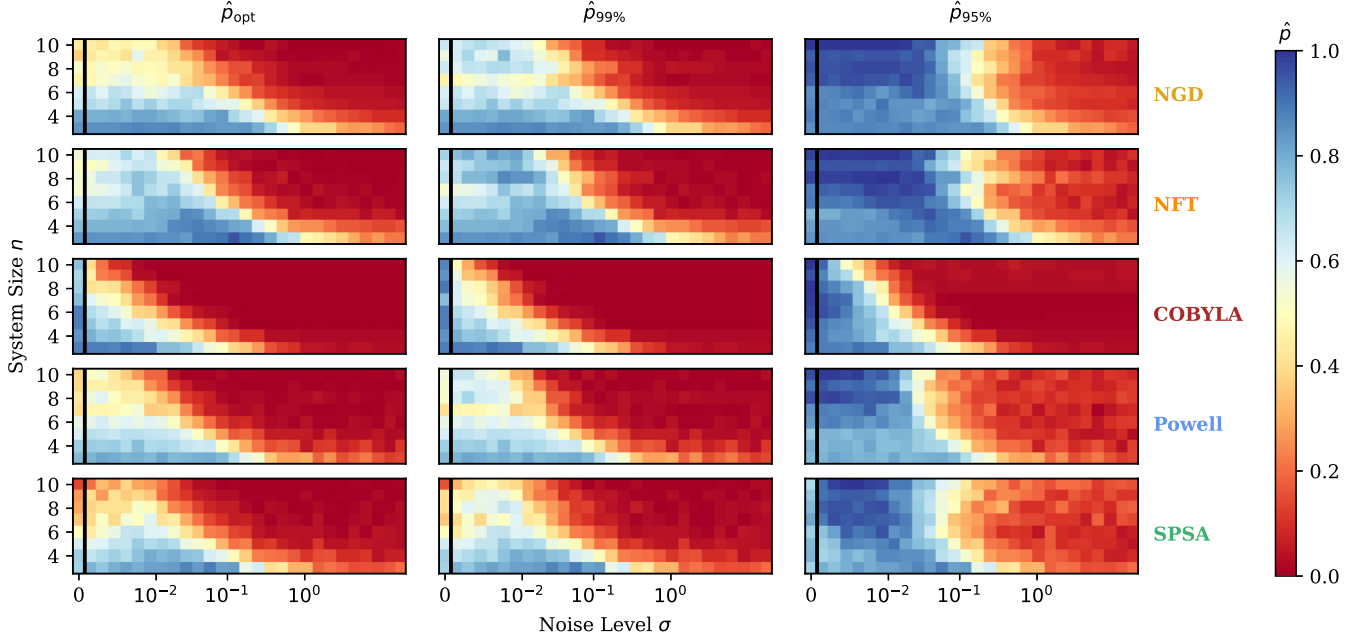


FIG. 5. Success probability across noise levels and system sizes. Contour plots show the proportions \hat{p}_{opt} , $\hat{p}_{99\%}$, and $\hat{p}_{95\%}$ for varying noise levels σ (x axis, logarithmic scale) and system sizes n (y axis) of the tested optimizers. The leftmost column in each grid represents the no-noise results, which are displayed outside the logarithmic scale for completeness.

used to fit the measured success probability curves. For easier interpretation, the following form is used:

$$p_{\text{fit}}(\sigma) = \frac{p_u - p_l}{2} \tanh[-b \ln(\sigma) + c] + \frac{p_u + p_l}{2}. \quad (25)$$

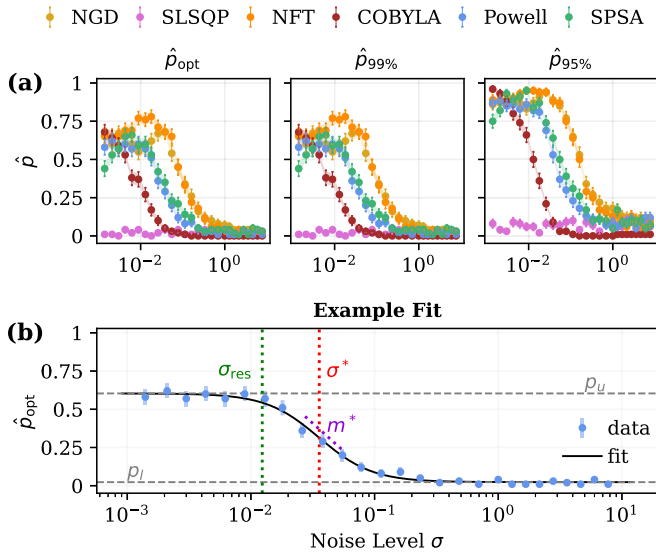


FIG. 6. (a) Scaling of success probabilities with increasing noise level. The proportions of optimal (\hat{p}_{opt}) and near-optimal ($\hat{p}_{99\%}$, $\hat{p}_{95\%}$) solutions for problems of size $n = 60$ are plotted against the noise levels σ (logarithmic scale) for a suite of classical optimizers. (b) An exemplary tanh fit of \hat{p}_{opt} for the Powell optimizer is shown, with the parameters of noise resilience, σ^* (27), m^* (28), and σ_{res} (29), indicated.

Here p_u and p_l represent the upper and lower asymptotes of performance, respectively, and are defined as

$$\lim_{\sigma \rightarrow +\infty} p_{\text{fit}}(\sigma) = p_l, \quad \lim_{\sigma \rightarrow 0} p_{\text{fit}}(\sigma) = p_u. \quad (26)$$

The parameters b and c quantify the gradient and location of the performance decrease on the logarithmic scale, and offer insights into the noise resilience of the tested solver. It is noteworthy, however, that the curves of the NFT and SPSA optimizer deviate significantly from a purely sigmoidal shape. Instead, both optimizers exhibit a clear peak in performance at a nonzero noise level before their success probabilities decline. This suggests that a small level of noise can enhance their ability to find optimal solutions by helping to escape saddle points and local minima, a phenomenon also reported in the literature [87–90].

To characterize the scaling behavior of the solvability measure \hat{p} , several metrics were derived to describe the noise resilience of classical optimizers based on the fit parameters:

(1) The *point of steepest descent*, σ^* , indicates the location of the performance drop and computes to

$$\sigma^* = \exp\left(\frac{c}{b}\right). \quad (27)$$

(2) The *slope at steepest descent*, m^* , quantifies the absolute performance decrease at $\sigma = \sigma^*$ and corresponds to the derivative of $p_{\text{fit}}(\sigma)$ at that point:

$$m^* = \frac{b(p_l - p_u)}{2} \exp\left(-\frac{c}{b}\right). \quad (28)$$

(3) The *point of resilience*, σ_{res} , serves as another indicator of noise tolerance, defined as the noise level at which

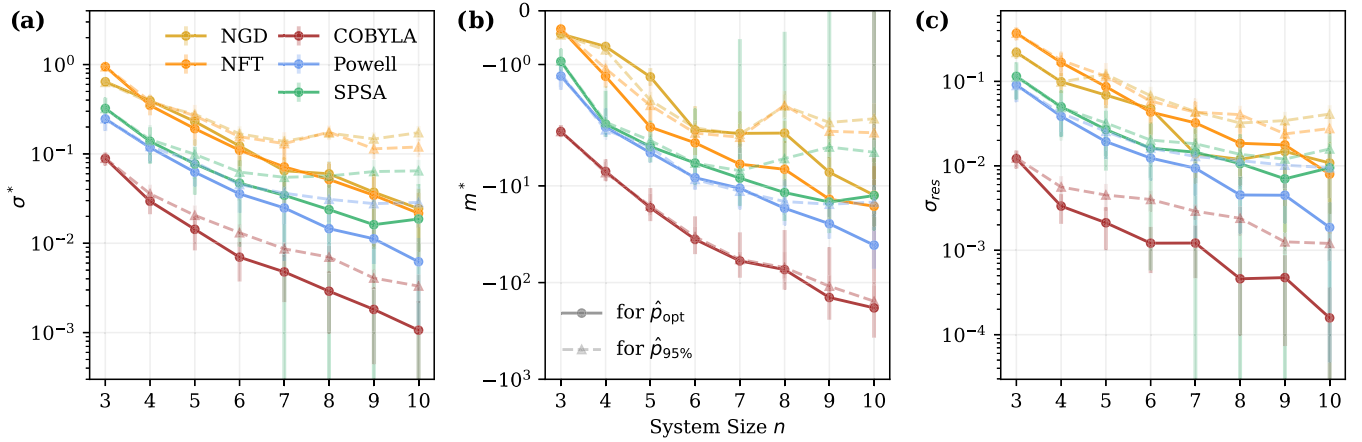


FIG. 7. Noise resilience parameters as a function of system size. (a) The point of steepest descent σ^* , (b) slope at steepest descent m^* , and (c) point of resilience σ_{res} are shown for increasing system sizes n , derived from the tanh fits of $\hat{p}_{\text{opt}}(\sigma)$ (solid line) and $\hat{p}_{95\%}(\sigma)$ (dashed line), for exact and approximate solvability of various optimizers, respectively. All plots use logarithmic y axes for better visualization.

the optimizer's performance drops to 90% of its no noise level p_u :

$$\sigma_{\text{res}} = p_{\text{fit}}^{-1}(0.9 p_u). \quad (29)$$

Here the 90% threshold is an arbitrary choice made for consistency across experiments.

These metrics offer valuable insights into the onset and rate of performance loss experienced by classical optimizers as noise levels increase. Their behavior across different problem sizes is systematically analyzed in the following section.

B. Solvability depending on system size

To investigate the transitions dividing solvable and unsolvable regions in Fig. 5, the noise resilience parameters derived in the previous section are used. The point of steepest descent σ^* (27), characterizes the critical threshold beyond which optimization fails. Figure 7 illustrates the behavior of this point, alongside its slope m^* (28), and the point of resilience σ_{res} (29) across different system sizes. Together, these parameters quantify the practical limitations of each optimizer, providing a clear ranking of their ability to handle noise: NGD and NFT perform best, followed by SPSA, Powell, and finally COBYLA. These findings align well with related work [30,33,35,36]. For a more detailed ranking and further discussion, refer to Appendix D.

Beyond ranking the robustness of the optimizers, we examine the overall scaling behavior of noise resilience metrics, which appear to follow consistent trends across different optimizers. Due to the strong similarity in results between \hat{p}_{opt} and $\hat{p}_{99\%}$, Fig. 7 focuses on exact solvability (\hat{p}_{opt}) and $\hat{p}_{95\%}$ for approximate solvability.

The point of steepest descent σ^* and the point of resilience σ_{res} should, in principle, exhibit similar scaling behavior, as both correspond to specific positions along the same tanh curve. All three metrics, including the slope of the descent's tangent line m^* , show a notable decline with increasing system size. While this overall trend is consistent across both exact and approximate solvability, the lower success threshold $\hat{p}_{95\%}$ intuitively exhibits a slightly softer decay compared to \hat{p}_{opt} .

However, since the practical usefulness of the approximate solvability metric is subject to limitations (see discussion in Sec. VII C and Appendix E), the subsequent analysis focuses on the scaling behavior of exact solvability. We emphasize, however, that the analysis was conducted for both thresholds, and full data are reported for each (see Appendix F). The overall qualitative conclusions remain unchanged, with only minor quantitative differences in the experimental curves.

Given the limited experimental data—constrained by the computational cost of simulating larger system sizes—any conclusions about the exact scaling behaviors or asymptotic trends require cautious analysis. To better understand the nature of the observable decay trends in Fig. 7, we therefore fit three different decay functions to the boundary line $\sigma^*(n)$:

$$\text{exponential : } f_{\text{exp}}(n, k, \gamma) = k \exp(-\gamma n), \quad (30)$$

$$\text{power-law : } f_{\text{pl}}(n, k, \gamma) = k n^{-\gamma}, \quad (31)$$

$$\text{logarithmic : } f_{\text{ln}}(n, k, \gamma) = k \ln(n)^{-\gamma}. \quad (32)$$

The results of these fits are illustrated in Fig. 8 for \hat{p}_{opt} . Visually, the power-law fit appears to align best with the majority of optimizer curves. However, the fit residuals do not necessarily support this hypothesis. They are given by

$$e(n) = \sigma^*(n) - f(n, \hat{k}, \hat{\gamma}), \quad (33)$$

where $e(n)$ represents the total difference between the measured $\sigma^*(n)$ values and the fitted decay function $f(n)$. Notably, none of the goodness-of-fit metrics were conclusive (cf. Appendix F), indicating that the precise scaling behavior of noise resilience may vary across different optimizers. A larger data set with a broader range of optimizers would be required to confirm this hypothesis.

Despite uncertainties regarding the exact form of decay, one clear conclusion emerges: noise resilience declines rapidly across all tested methods. This decline is intuitively attributed to the corresponding shrinkage of loss function variance [cf. Eq. (14)]. However, even after normalizing the noise level σ to account for this effect (cf. Sec. VI), this decline persists. This suggests that loss variance decay cannot fully

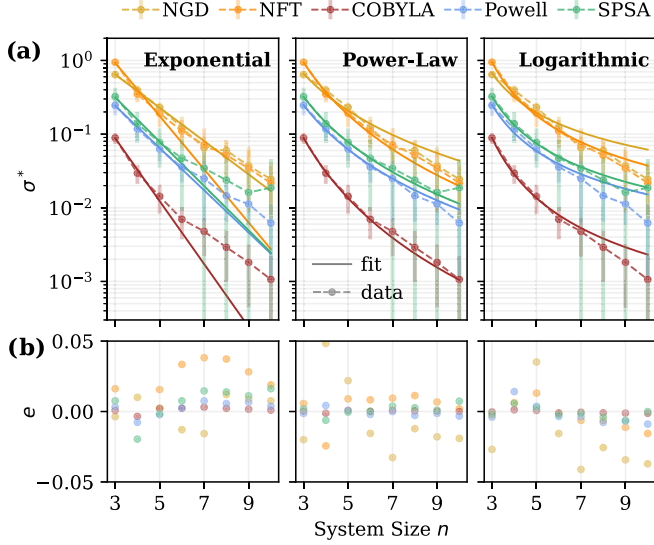


FIG. 8. Comparison of decay function fits. (a) Three theoretical decay functions [cf. Eqs. (30)–(32)] are fitted to the experimental data, modeling the decline of noise resilience σ^* with increasing system size n of the classical optimizers. (b) The final fit residuals e (33) show no significant differences among the three fit models, suggesting comparable accuracy in capturing the observed decay behavior.

explain the observed drop in noise resilience among classical optimizers.

Several other factors must therefore contribute to this behavior: For instance, Fontana *et al.* [91] demonstrate that quantum noise can disrupt parameter symmetries, eliminating degeneracies in the loss landscape and complicating optimization. Additionally, Wang *et al.* [16] identify the emergence of barren plateaus caused by noise, which severely impairs the trainability of VQAs. Other factors, such as the growing number of local minima in variational loss landscapes as system size grows [68] and the issue of overparameterization [92], may also play a role. More broadly, Leymann and Barzen [15] provide a summary of the challenges associated with the trainability of VQAs, beyond the issue of barren plateaus.

For reproducibility, the final parameters, k^* and γ^* , of all fitted theory curves [cf. Eqs. (30)–(32)] are provided in Appendix F.

VI. PRACTICAL IMPLICATIONS

In recent years, numerous significant studies have highlighted the practical limitations of variational quantum algorithms [13,14,16–19,22,93]. However, the implications of these findings are often left to the reader or deferred to future research. In this section we aim to bridge this gap by not only discussing the key insights from our results but also by explicitly demonstrating their practical relevance and broader impact.

Our analysis demonstrates that the noise resilience of VQA optimization decreases rapidly as the system size increases. This behavior was quantified in terms of the standard deviation of the total error on the loss function σ^* , beyond which the optimizer performance dropped to less than 50%

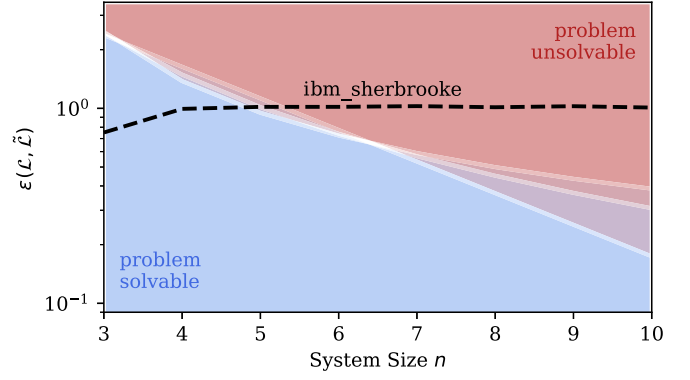


FIG. 9. Exemplary hardware error and problem solvability regions. The three possible separation lines represent the three fitted functions (cf. Fig. 8) of the critical noise threshold σ^* for the most noise-resilient optimizer, NGD, converted into the RAE metric ϵ^* (35) based on measurements from Sec. IV B. Above these lines, the probability of finding the optimal solution becomes negligible or effectively zero. The dashed line shows the empirically measured hardware error obtained using the `ibm_sherbrooke`¹ backend model.

of its noise-free level. Among the tested optimizers, the NGD (normalized gradient descent) algorithm [64] showed the best noise resilience.

To generalize these findings in this section, we use an error measure that is independent of the shrinking variance of the tested loss function: the relative absolute error (RAE), defined as

$$\epsilon(\mathcal{L}, \tilde{\mathcal{L}}) = \frac{\sum_{i=1}^N |\tilde{\mathcal{L}}(\theta_i) - \mathcal{L}(\theta_i)|}{\sum_{i=1}^N |\mathcal{L}(\theta_i) - \bar{\mathcal{L}}|}, \quad (34)$$

where $\tilde{\mathcal{L}}$ denotes the noisy loss values, and $\bar{\mathcal{L}}$ is the sample mean. The RAE normalizes the mean absolute error by the mean absolute deviation of the loss values, making it a relative measure independent of the chosen normalization [cf. Eq. (13)]. Assuming an approximately normal distribution of the errors, the noise threshold, $\sigma^*(n)$, characterized in this work, can be translated into the RAE metric as

$$\epsilon^*(n) = \frac{\sigma^*(n)}{\sqrt{\text{Var}(n)}}, \quad (35)$$

where $\text{Var}(n)$ is the sample variance of the loss function, as depicted in Fig. 4.

A. Hardware development

Despite significant advancements in quantum hardware development, current devices still face persistent limitations [94], with no demonstration of practical quantum advantage achieved to date. Mapping the RAE of loss functions when evaluated on current devices to the noise thresholds identified in this work provides valuable insights into these challenges.

¹Device specifications, based on the Eagle processor family: 7.095×10^{-3} (ECR error), 2.184×10^{-4} (SX error), 1.270×10^{-3} (readout error), 280.82 μs (t_1), and 204.19 μs (t_2).

Figure 9 illustrates where the error produced by a representative IBM hardware noise model lies relative to the resilience thresholds of the most resilience optimizer, NGD (cf. Fig. 8 and Appendix F). The intersection of the two curves indicates that problems with more than six nodes may already be unsolvable on current hardware, according to our metric of solvability [cf. Eq. (23)].

While other qubit technologies or newer hardware may exhibit lower noise levels for the algorithm under consideration, the implications of the near-exponential scaling of precision requirements are clear: drastic reductions in hardware error rates are required to achieve practical applicability. This critical relationship between system size and noise levels has also been highlighted in previous studies [17,19].

According to IBM's Development & Innovation Roadmap [95], the Starling processor, set for release in 2029, is expected to be error corrected with 200 logical qubits. Similarly, Quantinuum's roadmap projects a nearly fault-tolerant quantum computer with 1000 logical qubits by 2030 [96]. Comparable goals have been outlined by companies such as IonQ, IQM, and Google. This raises another critical question: will fully fault-tolerant hardware be able to address the scaling challenges and noise constraints highlighted in this work? In the following section, we seek to address this question.

B. The curse of finite sampling

In the case of fault-tolerant hardware, the only persistent source of stochastic noise is produced by finite measurement sampling. Independent of whether VQAs will be used with fault-tolerant systems, this can be considered the ideal setting for these algorithms. This section, therefore, deals with the question of how the noise thresholds for problem solvability identified in Sec. V translate to the required number of measurement shots and what implications their scaling behavior has for achieving a practical quantum speed-up. In the following, we consolidate all previous findings, adopting the most optimistic scenarios.

Based on the results of Sec. VB, the decline in an optimizer's noise resilience as the system size n increases, denoted by $\sigma^*(n)$, may follow an exponential, logarithmic, or power-law trend (cf. Fig. 8). To convert these resilience thresholds of the most noise-resilient optimizer, NGD, into the general measure of the RAE (34), they need to be normalized by the variance of the corresponding loss function [cf. Eq. (35)], as obtained for BENQO. The most optimistic scaling for the loss variance, which also provided a good fit, followed a power-law decay [97]. Using the fit parameters from Fig. 4, the maximum RAE $\epsilon^*(n)$ that an optimizer can tolerate for a given system size n computes to

$$\epsilon^*(n) \approx \begin{cases} 8.5 n^{0.4} e^{-0.5n} & \text{for exponential} \\ 22.0 n^{-1.8} & \text{for power-law} \\ 2.5 n^{0.4} \ln(n)^{-3.2} & \text{for logarithmic} \end{cases}, \quad (36)$$

depending on the realized scaling law. These results, while based on numerical fits with inherent uncertainties, strongly suggest that the decline in resilience persists regardless of how the empirical data are fitted [98].

To map this to measurement shot requirements, we first analyze the error introduced by finite sampling in evaluating a

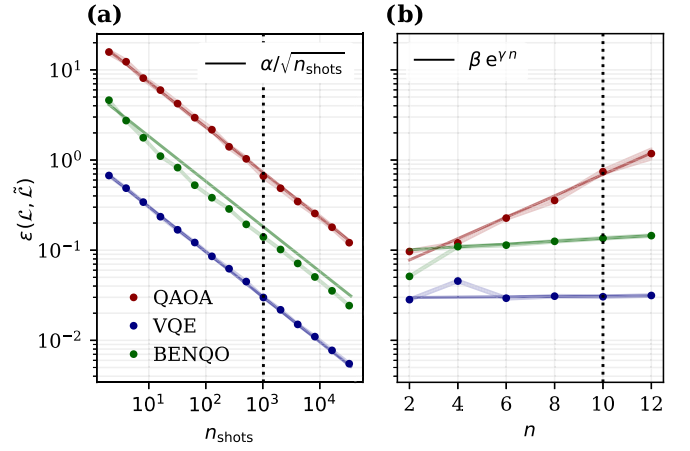


FIG. 10. Relative absolute error ϵ induced by finite sampling of the loss function. (a) The RAE (34) is plotted as a function of measurement shots n_{shots} (logarithmic scale) for a fixed system size $n = 10$, along with corresponding $1/\sqrt{n_{\text{shots}}}$ -fit functions. (b) The RAE is shown for increasing system sizes n at a fixed $n_{\text{shots}} = 1024$ shots, a common default in experimental computations, with exponential fit functions applied for $n \geq 6$ to reduce small-scale effects. Vertical dotted lines indicate the region of overlap between both plots. All error values are computed from 1000 loss evaluations per n , with parameters uniformly sampled from $[-2\pi, 2\pi]^n$.

loss function, denoted as ϵ_{FS} . Figure 10 illustrates the behavior of the RAE (34), computed for three VQAs (cf. Sec. IV B), including BENQO, as a function of the number of shots n_{shots} (at a fixed system size) and the system size n (at a fixed number of shots). As expected from a multinomial probability distribution, the error decreases proportionally to $1/\sqrt{n_{\text{shots}}}$ when n is fixed. Conversely, when n_{shots} is kept constant, the error grows exponentially, reflecting the increasing range of loss values alongside a simultaneously shrinking variance as the system size increases. For an analytical derivation of this scaling behavior, see Appendix C (cf. [39]).

Assuming that this general scaling behavior is consistent across all sampling rates and system sizes, we derive an approximate function for the RAE due to finite sampling for the examined loss function. Using the measured error value at $(n = 10, n_{\text{shots}} = 1024)$, where both experimental curves align, and applying the fitted parameters for BENQO from Fig. 10, the finite sampling error for system size n and n_{shots} measurement shots can be expressed as

$$\epsilon_{\text{FS}}(n, n_{\text{shots}}) \approx 4.0 \frac{e^{0.04n}}{\sqrt{n_{\text{shots}}}}. \quad (37)$$

For a random QUBO instance to be solvable with a VQA—specifically BENQO—the solvability condition

$$\epsilon_{\text{FS}}(n, n_{\text{shots}}) < \epsilon^*(n) \quad (38)$$

must hold for all n . Substituting Eq. (36) and Eq. (37), we derive that the required number of measurement shots grows exponentially with system size, for the scaling assumed in Eq. (37). Specifically,

$$n_{\text{shots}}(n) > \left(\frac{4.0 e^{0.04n}}{\epsilon^*(n)} \right)^2 \quad (39)$$

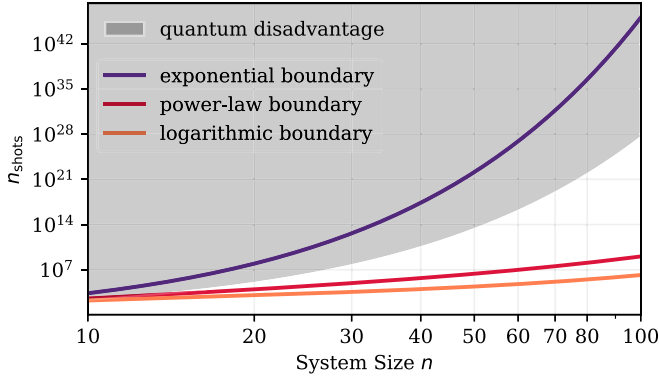


FIG. 11. Required shots n_{shots} for solving a random QUBO problem of size n . The minimum number of shots required by the most noise resilient optimizer, NGD, to reach three specified noise thresholds is shown as solid lines. For any value above these theoretical curves, solving a QUBO problem of size n using the BENQO algorithm becomes feasible. The shaded region represents the regime where the number of required quantum circuit executions surpasses the number of classical brute-force trials, rendering the quantum algorithm less efficient than the worst classical solver.

is always of order $O(k^n)$ for some $k > 1$, regardless of the specific scaling law in Eq. (36). This aligns with previous studies reporting exponential resource requirements in variational quantum optimization [18,99,100].

Furthermore, Scriva *et al.* [18] emphasized that the quantum computational cost, quantified as $n_{\text{shots}} \cdot n_{\text{calls}}$, must strictly remain below 2^n to avoid quantum disadvantage, i.e., the point where classical brute-force algorithms become more efficient than quantum computing. Note that this is the case even without accounting for quantum gate execution times.

In our experiments, the most noise-resilient optimizer, NGD, requires only 20 iterations regardless of n , which each consist of one quantum call to evaluate the loss function and $2n$ additional calls to evaluate the gradient [cf. Eq. (17)]. For this optimizer, the maximum number of allowed shots before encountering quantum disadvantage is given by

$$n_{\text{shots}}(n) < \frac{2^n}{20(1 + 2n)}. \quad (40)$$

This implies that for problem sizes below nine, achieving a quantum speed-up is impossible, as brute-force solutions are always more efficient. Figure 11 graphically summarizes these findings for system sizes $n \geq 10$.

The figure clearly illustrates that if the boundary line identified in Sec. VB were to follow an exponential decay, the number of required measurement shots would already grow faster than the number of brute-force trials needed to solve the problem classically. This strongly suggests that quantum advantage for optimization tasks under these conditions is fundamentally unattainable. In the alternative scenarios of power-law and logarithmic scaling, a small window of potential quantum advantage remains for problem sizes $n \geq 20$ (logarithmic boundary) and $n \geq 25$ (power-law boundary), where the shot requirement still falls below the critical threshold of quantum disadvantage. However, these potential advantages are only theoretical, as they are based

purely on resource scaling and do not yet translate into practical runtime improvements.

Therefore, we extend this analysis to the total runtime t required for the optimization process. Since this time varies significantly across hardware platforms, Scriva *et al.* [18] proposed a general lower bound for the runtime:

$$t_{\text{run}} < n_{\text{shots}} \cdot n_{\text{calls}} \cdot D \cdot t_{\text{gate}}. \quad (41)$$

Here n_{calls} denotes the number of quantum loss function evaluations, each measured with n_{shots} . The circuit depth of the used ansatz is denoted by D , and t_{gate} is the hardware-specific time required to execute a single gate. Notably, n_{calls} does not necessarily correspond to the number of optimization iterations, as many optimizers require additional function evaluations per step, particularly for gradient estimation [e.g., $2n$ calls using the parameter-shift rule (17)].

For problem sizes of practical relevance, $n \geq O(100)$, even under the most optimistic assumptions—constant circuit depth $D = O(1)$ and gate execution times of approximately 100 ns—the lower bound on the time to solution would already exceed one hour. Given the fast scaling of this time and the fact that this comparison is only made against the worst classical algorithms (while state-of-the-art classical solvers can handle problem instances orders of magnitude beyond 100 [101–103] in less time), any potential quantum speed-up appears highly impractical.

These findings further highlight the substantial gap that quantum devices still have to bridge to outperform classical solvers and achieve real world impact in optimization. They also cast significant doubt on the practical feasibility of achieving meaningful quantum advantage in this domain using variational quantum algorithms.

VII. VALIDATION AND DISCUSSION

A common concern in empirical studies like ours is to what degree the observed results are influenced by specific choices in the experimental setup (cf. Fig. 3). Factors such as problem formulation, quantum ansatz, noise model, optimizer selection, and performance metrics may, in principle, affect the conclusions drawn. One could, therefore, argue that the critical behavior identified in this work merely arises from unlucky experimental choices rather than inherent scalability challenges of hybrid variational quantum optimization. While each of our decisions was carefully justified to favor well-behaved scenarios and ensure a fair and optimistic evaluation, this section provides additional validation and discussion about the reliability and generality of our findings.

A. Independent confirmation of scaling trends

A study by Scriva *et al.* [18] examined shot noise effects using a VQE ansatz on an Ising chain problem and found that the success probability for a fixed system size follows the same sigmoidal pattern as observed in our experiments. Their results also demonstrated that the optimal number of function calls required to reach a fixed success probability scales exponentially with system size for certain parameter initializations, reinforcing our findings in Sec. VIB.

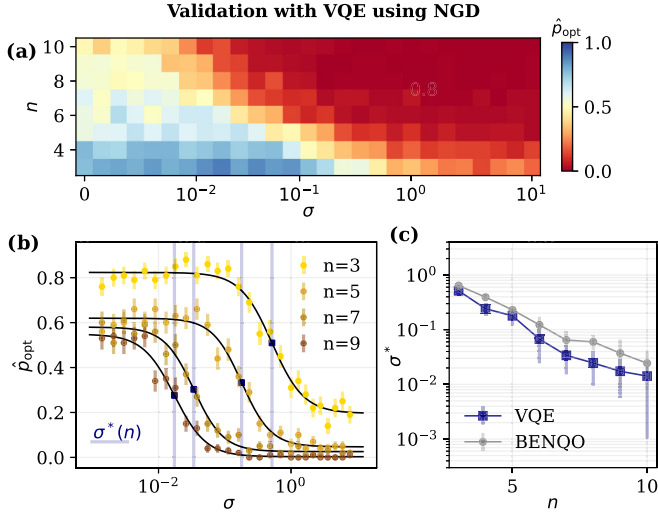


FIG. 12. Validation of NGD solvability scaling with VQE. (a) Contour plot of the empirical NGD solvability experiments (cf. Fig. 5), this time using VQE as the underlying quantum loss function. (b) Scaling of the success probability \hat{p}_{opt} across different noise levels for fixed system sizes n , fitted with sigmoidal theory curves as in Fig. 6. Blue vertical lines indicate the points of steepest descent σ^* (27), obtained from tanh fits [cf. Eq. (25)]. (c) The main noise resilience parameter σ^* is plotted against increasing system sizes and directly compared to the noise threshold obtained for BENQO (cf. Sec. VB). The curves exhibit a close agreement.

Similarly, Sung *et al.* [31] analyzed a QAOA ansatz applied to three distinct optimization problems, including the MaxCut problem on three regular graphs. Unlike our study, they employed global and alternative classical optimizers and modeled noise at the gate level of the quantum ansatz. Regardless of these differences in setup, they also observed a sigmoidal drop in success probability as noise increased and an unfavorable scaling of the required time to solution for fixed precision levels.

Despite the differences in quantum algorithms, noise models, classical optimizer selection, and problem types, both studies independently observed the same key trends as those presented in this work.

B. Validation with a VQE ansatz

We repeat our experiments to further validate our conclusions using another well-behaved quantum loss function: the two-local shallow VQE ansatz described in Sec. IV B. The results for the NGD optimizer are summarized in Fig. 12, closely matching those obtained using BENQOs. This agreement is not only qualitative but also quantitative. Normalizing the noise resilience threshold $\sigma^*(n)$ by the variance of the respective loss functions (cf. Fig. 4)—thus mapping both results onto the RAE metric (34)—would likely reveal an even stronger correspondence between the two ansatzes.

To directly validate the corresponding resource scaling reported in Sec. VIB, we conduct additional experiments with the VQE algorithm, this time introducing noise solely through finite sampling rather than the additive Gaussian noise model. The shallowness of the VQE ansatz enables simulations with larger system sizes than in previous experiments, extending

the range to $n \in [3, 13]$. The following results focus on the COBYLA solver, as it demonstrated the highest resource efficiency in terms of quantum function calls and mean runtime (cf. Appendix D).

Figure 13 presents the results of this analysis. As expected, we observe a sigmoidal drop in success probability for fixed system sizes n as the number of measurement shots n_{shots} decreases. By fitting the data using Eq. (25), this time with n_{shots} as the variable, we extract the steepest descent positions n_{shots}^* [cf. Eq. (27)] and analyze their scaling behavior with increasing system sizes. A power-law function provides the best visual agreement and is therefore used to extrapolate the required shot numbers to practically relevant system sizes. This allows for a direct comparison with Fig. 11. As before, the boundary of quantum disadvantage is derived by dividing the number of brute-force trials 2^n by the number of function calls [cf. Eq. (40)].

These validation results align more closely with the power-law scaling scenario discussed in Sec. VIB and a power-law scaling of the finite sampling error ϵ_{FS} (37) in n , meaning that there is a theoretical window of quantum advantage—though only when compared to the most inefficient classical solution method. Interestingly, our direct shot requirement analysis suggests slightly lower resource demands than initially estimated. However, under the same optimistic assumptions, i.e., a constant circuit depth $D = 1$ and gate execution times of approximately 100 ns, the lower bound for time to solution (41) would still exceed one hour for problems of size $n \geq 243$. While this is slightly above the $n = 100$ estimate from Sec. VIB, the difference remains minor given that these calculations assume best-case scenarios with only finite sampling errors—an assumption that is valid only in fully error-corrected quantum devices.

C. Reconsidering the success metric

Another point of discussion of our study is the choice of success metric. In this study we measured the sample success probability of finding a solution within a specific relative threshold t of the optimal energy [cf. Eq. (23)], with thresholds set at $t = 100\%$ (exact optimality), 99%, and 95% (approximate solvability). However, most of the scalability issues discussed are derived from the probability of finding the optimal solution \hat{p}_{opt} . One could argue that the primary goal of quantum optimization, particularly for NP-hard problems, is not necessarily to find the absolute global optimum but rather to generate high-quality approximations that outperform classical solutions.

Our experimental results indicate that the scaling behavior of $\hat{p}_{99\%}$ and $\hat{p}_{95\%}$ differs slightly from that of the exact solution probability (cf. Fig. 7). In particular, the further the relative energy threshold t deviates from 100%, the more its scaling diverges from the solvability scaling discussed earlier. In some cases—especially in the absence of noise—the near-optimal success probability $\hat{p}_{95\%}$ even increases with system size (cf. Fig. 5), suggesting that as the system grows, optimizers tend to converge to local minima that lie within a close vicinity of the global optimum.

This trend can be explained by examining the distribution of loss values across all possible solution states

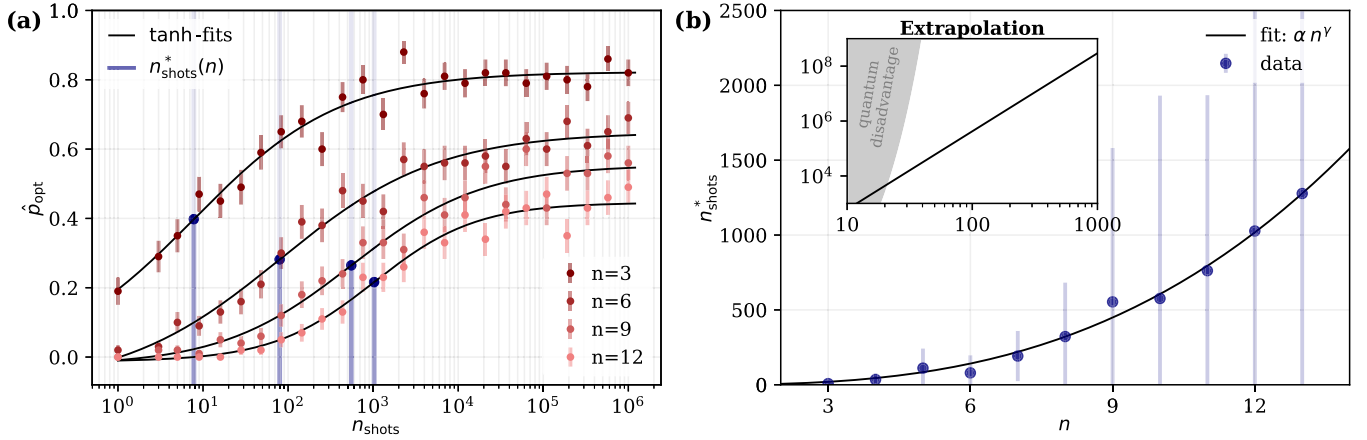


FIG. 13. Validation of measurement shot scaling with VQE. (a) Success probability \hat{p}_{opt} as a function of the used numbers of measurement shots n_{shots} in simulation, for fixed system sizes n . Blue vertical lines mark the positions of steepest descent n_{shots}^* , obtained from tanh fits [cf. Eqs. (25) and (27)]. (b) Required number of measurement shots for system solvability as a function of system size. A power-law fit of the data enables extrapolation to larger system sizes, as shown in the inset, with the region of quantum disadvantage shaded in gray (cf. Fig. 11).

(cf. Appendix E). In our QUBO formulation, feasible solutions are enforced via quadratic penalty terms (cf. Appendix B), which can—depending on the chosen penalty factor—significantly expand the loss value range in one direction, often leading to $|C_{\text{max}}| \gg |C_{\text{min}}|$. Due to the chosen normalization, the maximum loss value remains 1, while the majority of loss values concentrate (possibly exponentially) near zero (cf. Sec. IV B), leading to an asymmetry in the loss value distribution. As the system size increases, this effect intensifies, effectively pushing most local minima closer to the global optimum. Consequently, the approximation ratio in Eq. (22) approaches 1, shifting the threshold for identifying feasible solutions to much higher energy levels, sometimes even beyond the 99% threshold used in this study [36,104]. This explains why approximate solutions remain relatively accessible, even when finding the exact optimum becomes increasingly difficult.

However, this also raises doubts about the practical utility of near-optimal solutions using standard metrics as in Eq. (22), particularly if they fail to satisfy feasibility constraints. One approach to mitigate this issue is to explore alternative problem encodings that yield a higher proportion of feasible solutions [105]. Another possibility is to redefine the chosen success probability metric [cf. Eq. (23)]. For this, rather than relying on the approximation ratio as defined in Eq. (19), a more informative metric could be a normalized approximation ratio that considers the energy of a totally mixed state as the upper bound C_{max} . This adjustment would provide a more meaningful indicator of high-quality solutions, rather than merely achieving proximity to the optimal value under a potentially misleading scale.

D. On ansatz choice and smart initializations

In this study, we deliberately focused on two variational ansatzes that avoid barren plateaus, maintaining nonvanishing gradient and loss values to ensure more favorable trainability properties on average (cf. Sec. IV B). However, their absence of barren plateaus is ultimately linked to a nonexponentially scaling dimension of the underlying dynamical Lie algebra

[25,26], which in turn implies efficient classical simulability [11,12]. As such, these ansatzes may not offer a viable path toward achieving quantum advantage within VQAs.

To address this, one might intentionally consider ansatz structures that do exhibit BPs, such as QAOA or generic hardware-efficient ansatzes (HEAs) [24], in order to retain the possibility of quantum advantage in the loss computation. In these cases, the key challenge becomes overcoming the exponential suppression of loss and gradient magnitudes. A leading approach to mitigating this issue is the use of smart parameter initialization strategies, which have shown promising performance in both QAOA [18,106–110] and HEA circuits [86,111–114]. An important direction for future work is to investigate how the presence of Gaussian uncertainty interacts with such initialization strategies, and whether the observed impact of stochastic noise differs when applied to barren-plateau-prone ansatzes initialized with trainability-enhancing methods.

VIII. CONCLUSIONS AND OUTLOOK

The potential of variational quantum algorithms for optimization fundamentally depends on their ability to handle stochastic noise, which arises both from current hardware imperfections and finite sampling, the latter remaining even in fault-tolerant quantum computers. While numerous theoretical and empirical studies have highlighted how hardware noise can impair or even eliminate quantum advantage in VQAs (cf. Sec. II), there has been less focus on the impact of isolated stochastic uncertainty on system solvability. This work provides a rigorous numerical investigation into how classical optimizers—the main computational engine of VQAs—respond to additive Gaussian noise as the size of quantum loss functions scales up. Our findings, validated through extensive testing, empirically demonstrated that stochastic noise alone may impose fundamental limitations on the scalability of variational quantum optimization to practically relevant problem sizes, even when using fault-tolerant hardware.

We find, that for fixed system sizes, the probability of finding high-quality solutions—including those within 1% and 5% of the optimal energy—decreases in a sigmoidal pattern with respect to the logarithmic noise level. This reveals a sharp performance drop at a critical noise threshold σ^* , which varies significantly across the different optimizers. Using tanh fits, we extracted these thresholds for system sizes $n \in [3, 10]$, suggesting a common scaling trend across all optimizers: their noise resilience declines rapidly, potentially exponentially, as system size increases. Notably, this decline is steeper than what can be explained by a simultaneously shrinking loss variance, confirming that larger systems inherently present a more challenging loss landscape.

While it is well known that current hardware error rates far exceed what is required for classically relevant problem sizes, our results highlight another possible limitation imposed by finite sampling errors. Extrapolating the observed decline in noise resilience showed that, to achieve the loss precision necessary for problem solvability, the required number of measurement shots scales prohibitively bad with system size. This leaves minimal room before exceeding the computational resources of classical brute-force sampling, which is highly impractical. These findings suggest that achieving meaningful quantum advantage with VQAs for large-scale optimization problems may face significant challenges.

We therefore advocate for a shift in research focus, away from introducing yet new heuristic hybrid algorithms with idealized testing, and toward strategies that directly address the challenges demonstrated in this and prior work. If quantum utility [115] in optimization remains the goal, future research should prioritize alternative algorithmic frameworks (and possibly different problem domains) that can provably circumvent these barriers.

Apart from the methodological suggestions made in Secs. VIIC and VIID, one promising avenue is improved problem encoding, which aims to reduce the number of required qubits and thereby mitigate some of the scaling issues discussed here [116–120]. However, while these methods can offer potential efficiency gains, they often lack both theoretical performance guarantees and large-scale empirical validation in noise-exposed settings. Moreover, given that the scalability challenges demonstrated here may be inherent in the hybrid nature of VQAs, which rely on classical optimization, it seems to be more promising to explore fully quantum alternatives outside the variational paradigm. Recent works [120–123] have begun such investigations, and a similar shift away from hybrid frameworks can be seen in the quantum machine learning community [124]. Additionally, this study has not yet considered alternative computational models, such as Quantum Boson Sampling [125–127] or Quantum Annealing [128,129], which may warrant further investigation.

Beyond algorithmic strategies, a reevaluation of the types of problems targeted for near-term quantum advantage may be necessary. Rather than tackling intrinsically classical problems such as Ising spin glasses, growing attention is being given to quantum-native optimization problems, such as the simulation of quantum mechanical systems [130,131], where exponential speed-ups are both theoretically proven and more intuitive. For recent advances in quantum simulation, see Refs. [132,133].

Ultimately, our work provides a general and robust framework for rigorously investigating the scalability and limitations of not only VQAs but also other quantum algorithms in the presence of stochastic errors. We encourage researchers to apply and extend our methodology to more promising alternative setups—whether by testing different quantum ansatzes, optimizer setups, problem instances, or hardware platforms. If future studies uncover scenarios where problem solvability scales more favorably with system size and noise levels, this would be an exciting breakthrough. However, if the scaling behavior observed here proves to be universal and inherent to the classical optimization problem inherent in hybrid architectures like VQAs, it would strongly suggest that quantum advantage in hybrid variational quantum optimization is not realistic. In fact, our results indicate that the challenges facing VQAs—in both the NISQ and fault-tolerant era—may be more fundamental than what can be resolved through hardware improvements, better classical optimizers, or ansatz tuning alone.

ACKNOWLEDGMENTS

This work is supported by the Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag through the project QuaST–Quantum-enabling Services and Tools for Industrial Application. The authors would like to express their gratitude to Frank Pollmann for providing valuable advice on this work and carefully reviewing this manuscript. We also thank Marita Oliv, Amine Ben, and Nicola Franco for insightful discussions and valuable feedback, which have contributed to the refinement of this study.

DATA AVAILABILITY

Key fitting parameters for reproducing final models are available within the article. Complete experimental datasets corresponding to the findings of this study are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: MAPPING QUBO TO ISING PROBLEMS

To translate the general QUBO form (1)

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n Q_{ij} x_i x_j \quad (\text{A1})$$

into the Ising form of Eq. (2), one can apply the mapping $x_i = \frac{z_i + 1}{2}$ between the binary variables $x_i \in \{0, 1\}$ and the spin variables $z_i \in \{-1, +1\}$ as follows:

$$\begin{aligned} g(z) &= \sum_{i=1}^n \sum_{j=1}^n Q_{ij} \left(\frac{z_i + 1}{2} \right) \left(\frac{z_j + 1}{2} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{Q_{ij}}{4} z_i z_j + \frac{Q_{ij}}{4} z_i + \frac{Q_{ij}}{4} z_j + \underbrace{\frac{Q_{ij}}{4}}_{\text{const.}} \right) \end{aligned}$$

$$\sim \sum_{i \neq j} \frac{Q_{ij}}{4} z_i z_j + \sum_{i=1}^n \left(\sum_{j=1}^n \frac{Q_{ij}}{4} + \sum_{j=1}^n \frac{Q_{ji}}{4} \right) z_i. \quad (\text{A2})$$

In the final step, all constant offsets were disregarded as they do not affect the minimization of the loss function.

Assuming that \mathbf{Q} is a symmetric matrix, and by replacing the spin variables z with Pauli- \mathbf{Z} operators, the final Ising operator depending on the QUBO matrix elements becomes

$$\mathbf{C}(\mathbf{Q}) = \sum_{i < j} \frac{Q_{ij}}{2} \mathbf{Z}_i \mathbf{Z}_j + \sum_{i=1}^n \left(\sum_{j=1}^n \frac{Q_{ij}}{2} \right) \mathbf{Z}_i. \quad (\text{A3})$$

Generally, this reformulation means that the Ising weight matrix $\mathbf{C} = \{C_{ij}\}$ [cf. Eq. (2)] can be retrieved from the QUBO matrix $\mathbf{Q} = \{Q_{ij}\}$ as follows:

Off-diagonal elements: $C_{ij} = Q_{ij}/2$,

Diagonal elements: $C_{ii} = \sum_{j=1}^n Q_{ij}/2$.

A similar mapping between the QUBO and Maximum Cut (MaxCut) problem was derived in Ref. [134].

Lastly, note that this transformation is not unique. An equivalent formulation can be obtained by applying the mapping $x_i = \frac{1-z_i}{2}$, which transforms the binary variables $x \in \{0, 1\}$ into the spin variables $z \in \{+1, -1\}$. This leads to an Ising form with a negative sign for the diagonal terms. In this formulation the ground state configuration would have zeroes where the original formulation had ones, and vice versa. Thus, this mapping would result in an inverted ground state compared to the initial transformation.

APPENDIX B: QUADRATIC PENALTY TERMS IN PERMUTATION PROBLEMS

In graph-partitioning problems, such as the well-known MaxCut problem, the number of nodes in the problem graph directly corresponds to the number of binary variables in the problem formulation. However, permutation problems, such as the highly relevant Traveling Salesperson Problem (TSP), are more complex as they require assigning an order instead of a partition to the nodes. This can be achieved using one-hot-encoding, where permutations are represented by n^2 binary decision variables x_i^α , indicating whether node i is assigned to a position α in the permutation. In the case of the TSP, i would represent a specific city, and α denotes the time step at which the city is visited along the route.

This binary encoding is therefore subject to two permutational constraints:

Each position α must hold exactly one node:

$$\sum_i x_i^\alpha = 1 \quad \forall \alpha = 1, \dots, n.$$

Each node i must occur exactly once:

$$\sum_\alpha x_i^\alpha = 1 \quad \forall i = 1, \dots, n.$$

Glover *et al.* [54] proposed incorporating these constraints into a QUBO loss function like Eq. (1) using quadratic penalty terms with a penalty factor $P > 1$, resulting in the new QUBO formulation:

$$C(\mathbf{x}) = C_0(\mathbf{x}) + P \left[\sum_{i=1}^n \left(1 - \sum_{\alpha=1}^n x_i^\alpha \right)^2 + \sum_{\alpha=1}^n \left(1 - \sum_{i=1}^n x_i^\alpha \right)^2 \right]. \quad (\text{B1})$$

Here $C_0(\mathbf{x})$ represents the actual unconstrained loss function of the problem, defined by the QUBO matrix \mathbf{Q}_0 , while the latter terms impose the corresponding permutational constraints. In the TSP, $C_0(\mathbf{x})$ would be a function returning the total path length.

To ensure that feasible solutions are enforced after minimizing Eq. (B1), P must be larger than a problem-specific threshold dependent on C_0 . To obtain the QUBO matrix representation from the loss function in Eq. (B1), we translate its terms into the matrix elements $Q_{ij}^{\alpha\beta}$, which encode the interaction between variables $x_{i\alpha}$ and $x_{j\beta}$:

$$Q_{ij}^{\alpha\beta} = \{\mathbf{Q}_0\}_{ij}^{\alpha\beta} + P[\delta_{ij}(1 - \delta_{\alpha\beta}) + \delta_{\alpha\beta}(1 - \delta_{ij})] - 4P\delta_{ij}\delta_{\alpha\beta}. \quad (\text{B2})$$

From Eq. (B2), it is immediately apparent that the penalties terms affect the diagonal and off-diagonal elements of \mathbf{Q} differently:

(1) The off-diagonal penalty terms [second term of (B2)] introduce additional contributions of order P to the original problem matrix \mathbf{Q}_0 .

(2) The diagonal penalty terms [third term of (B2)] shift the diagonal elements in \mathbf{Q}_0 downward by values of order $4P$. Since P is chosen to be significantly larger than each element of \mathbf{Q}_0 , all diagonal elements of \mathbf{Q} become negative. This confirms that adding equality constraints as quadratic penalty terms transforms the problem's QUBO matrix into one with negative diagonal entries, justifying the formulation used in Sec. IV A.

Lastly, note that certain permutational graph problems can be reduced to $(n-1)^2$ binary variables by fixing the starting point of each permutation. This is done by setting $x_1^1 = 1$, $x_i^\alpha = 0$ for all $\alpha \neq 1$, and $x_i^i = 0$ for all $i \neq 1$, which eliminates the cyclic permutation symmetry inherent in the problem, effectively reducing the search space by a factor n .

APPENDIX C: DISTRIBUTION OF FINITE SAMPLING ERRORS

As quantum computing advances toward fully error-corrected systems, the only remaining source of stochastic noise will be that arising from the finite sampling of circuit measurements. In the VQA paradigm (cf. Sec. III B), the number of measurement shots n_{shots} directly determines the precision of the estimated expectation value $\mathbb{E}[C]$ relative to the true loss value $\langle C \rangle$.

For a single-qubit measurement with n_{shots} shots, the count of “0” outcomes (k_0) follows a binomial distribution. According to the central limit theorem [135], as $n_{\text{shots}} \rightarrow \infty$, this distribution approaches a normal distribution with mean $\mu(k_0) = n_{\text{shots}} p_0$ and standard deviation $\sigma(k_0) = \sqrt{n_{\text{shots}} p_0(1-p_0)}$, where p_0 is the probability of measuring “0.” The corresponding standard error of the measured proportion $\hat{p}_0 = \frac{k_0}{n_{\text{shots}}}$ is given by

$$\sigma(\hat{p}_0) = \sqrt{\frac{p_0(1-p_0)}{n_{\text{shots}}}}. \quad (\text{C1})$$

In standard VQA measurement schemes, such as with VQE and QAOA, the cost operator (2) is used as the measurement

observable. Thus, for a parametrized quantum state $|\psi(\theta)\rangle$, the expectation value is

$$\langle\psi(\theta)|C|\psi(\theta)\rangle = \sum_{j=1}^{2^n} p_j \langle q_j|C|q_j\rangle. \quad (C2)$$

Here $p_j = |\langle\psi(\theta)|q_j\rangle|^2$ is the probability of measuring the system in the basis state $|q_j\rangle$, which can only be estimated from measurement samples on a quantum device. The measurement counts of basis states follow a multinomial distribution, such that each estimated proportion \hat{p}_j carries the standard error given in Eq. (C1).

By applying the linear approximation of error propagation and using the fact that variances of multiple independent variables add in quadrature, the statistical uncertainty of the estimated expectation value can be approximated by

$$\begin{aligned} \sigma(\mathbb{E}[C]) &\approx \sqrt{\sum_{j=1}^{2^n} \left(\frac{d\mathbb{E}[C]}{d\hat{p}_j} \Big|_{\hat{p}_j=p_j} \sigma(\hat{p}_j) \right)^2} \\ &= \sqrt{\sum_{j=1}^{2^n} (\langle q_j|C|q_j\rangle)^2 \sigma(\hat{p}_j)^2}. \end{aligned} \quad (C3)$$

Here the dependence on the system size n is embedded in the expectation value of the cost operator, as given by Eq. (3), while the dependence on n_{shots} is embedded in the standard error described by Eq. (C1). This behavior is empirically demonstrated in Fig. 10, confirming the scaling of statistical uncertainty as a function of system size and sampling rates.

In Sec. IV C we argue that these finite sampling errors can be reasonably modeled by adding Gaussian noise to the exact values of the loss function, assuming that this is a good-enough characterization of how statistical errors are distributed across the parameter space during loss evaluation. To illustrate this, Fig. 14 presents the distribution of total normalized errors $\delta\mathcal{L}$ when sampling the loss at a randomly chosen point $\theta \in [-2\pi, 2\pi]^n$ for a system of size $n = 6$ with $n_{\text{shots}} = 1024$ shots. These distributions are computed across the three VQAs discussed in Sec. IV B and compared to a normal distribution with a mean of zero and the same standard deviation as the empirical data. The close visual agreement between the histogram and the theoretical Gaussian curve validates the central limit theorem, which predicts that for a sufficiently large number of shots, the binomial (or multinomial) distribution from finite sampling converges to a normal distribution.

Figure 14 also investigates the relationship between the total error magnitude $s[\delta\mathcal{L}]$ and the loss value at each evaluation point. This analysis helps determine whether finite sampling noise behaved in an additive or multiplicative manner. To quantify this, linear fits of the sample standard deviations performed, yielding values for the slope m and y intercept t of the lines: If $m \ll t$, the noise is predominantly additive, whereas if $m \gg t$, the noise exhibits a multiplicative behavior.

For the BENQO algorithm, the y-intercept t is significantly larger than the slope m , meaning that the error is almost entirely additive. For the other two algorithms, the error displays both additive and multiplicative components, with a larger multiplicative contribution. Their increased mul-

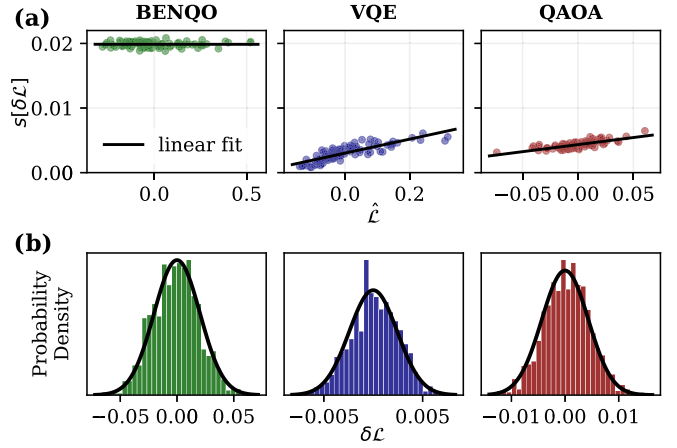


FIG. 14. Distribution of finite sampling errors across loss landscape. (a) Standard deviation of the total errors from 1000 loss evaluations with 1024 measurement shots, at 100 randomly selected points across the loss landscape of an $n = 6$ QUBO problem instance (cf. Sec. IV A). The relationship between the normalized loss value and the error magnitude is analyzed for three different VQAs, with linear fits indicating whether the noise behaves additively or multiplicatively. (b) Probability density histograms of 1000 normalized errors at a single representative point, compared to the theoretical normal distribution with zero mean and the same standard deviation, validating the central limit theorem.

tiplicative component arises from the strong dependence of the statistical error (C3) on $\langle C \rangle$. BENQO, however, almost entirely loses this dependency due to its implicit measurement strategy [64].

Despite these differences, additive errors tend to distort the loss landscape more severely than multiplicative ones [39]. Therefore, when modeling total finite sampling-induced errors, a worst-case scenario can always be represented by simply adding Gaussian noise to the loss function. This justifies the noise model used in our methodology (cf. Sec IV C).

APPENDIX D: RANKING OF TESTED OPTIMIZERS

Inspired by Singh *et al.* [35], we present a performance ranking of the tested classical optimizers in Fig. 15. This comparative ranking examines the trade-off between solver efficiency and noise resilience, where efficiency was measured by the mean runtimes across all experiments, directly correlating with the number of loss function calls. A clear trend emerges: optimizers with greater noise resilience tend to require more function evaluations to converge, forming an inherent trade-off between these two critical performance measures. This suggests that more noise-tolerant optimizers compensate for the lost precision by requiring additional function calls, thereby sacrificing efficiency to maintain stability in noisy settings.

Certain optimizers—CG, SLSQP, and BFGS—were excluded from this ranking due to their poor performance across all tested noise levels. Their failure likely stems from their reliance on precise gradient calculations, which become very unreliable in the presence of noise. This supports previous findings [33,35], indicating that gradient-based op-

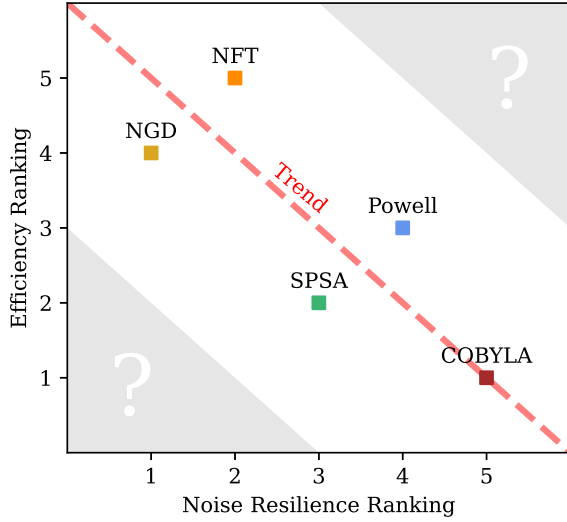


FIG. 15. Performance ranking of classical optimizers. The relationship between each optimizer's noise resilience and its efficiency ranking, based on measured mean runtimes in the same noisy experiments, is shown. The dashed diagonal line suggests a potential trend: optimizers requiring more function evaluations tend to exhibit higher noise resilience.

timizers, while efficient in ideal conditions, struggle in noisy settings, whereas gradient-free strategies prove significantly more robust.

Lastly, note that while this ranking provides a comparative assessment, it is important to acknowledge its limitations. The approach, inspired by Singh *et al.* [35], assigns discrete ranks rather than directly plotting continuous performance metrics. As a result, it does not fully capture quantitative performance differences between optimizers. A more precise analysis would involve directly plotting the number of function calls n_{calls} against the noise resilience threshold σ^* (27), allowing for a more detailed examination of the performance trade-offs.

APPENDIX E: OPTIMALITY DISTRIBUTION OVER CLASSICAL SOLUTION SPACE

This section examines the distribution of all possible solutions within the loss landscape, assessing the significance of solutions falling within the 99%, and 95% thresholds. Figure 16 depicts the results of this investigation. The data for

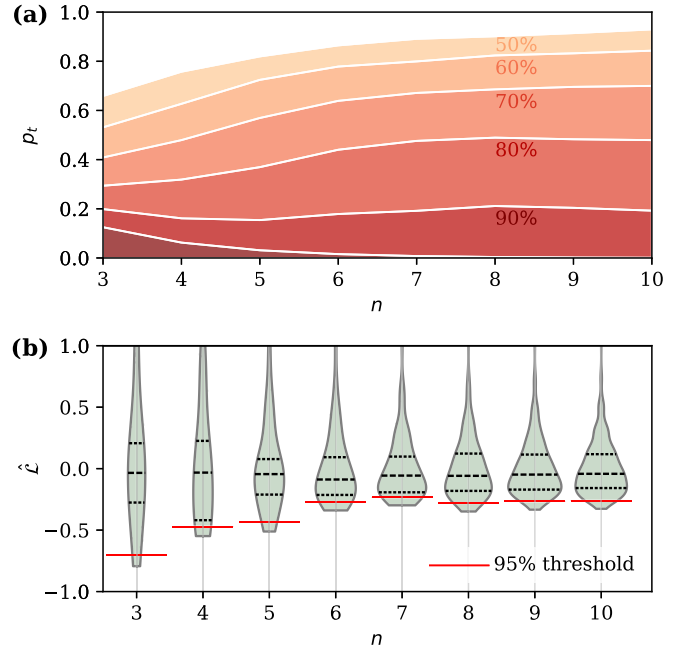


FIG. 16. Distribution of solution optimality. (a) The percentage of solutions p_t that achieve a loss value within a relative threshold t of the optimal loss value, across varying system sizes n , is depicted as contours. (b) Violin plots illustrate the distribution of all 2^n possible classical solution states within the normalized loss landscape \hat{L} . The loss value corresponding to the 95% optimality threshold is marked for each n , providing insight into the spread and concentration of near-optimal solutions.

these plots are derived from the same 100 problem instances used in our optimizer studies.

The upper panel of Fig. 16 presents the total percentage of solutions p_t that achieve a loss value within different relative thresholds t of the optimal loss [cf. Eq. (23)] across all 2^n possible classical solution states, i.e., bit strings. Note that this corresponds to the success probabilities of random guessing. Interestingly, this percentage appears to converge for system sizes beyond seven or eight. More importantly, nearly 20% of all possible solution states still yield loss value within 90% of the optimum, indicating that the solution space is highly structured rather than uniformly random.

The lower panel of Fig. 16 shows the corresponding normalized loss distribution of all solution state, which also seems to converge to a fixed shape, Especially, as system

TABLE II. MSE of decay function fits. The MSE values for each optimizer's noise resilience threshold $\sigma^*(n)$ w.r.t. both \hat{p}_{opt} (upper) and $\hat{p}_{95\%}$ (lower) are computed for three decay models, exponential (exp), power-law (pl), and logarithmic (log), following Eqs. (30)–(32). The lowest MSE for each optimizer is marked for better comparability.

\hat{p}_{opt}	NGD	NFT	COBYLA	Powell	SPSA
exp	1.0×10^{-4}	1.1×10^{-3}	5.0×10^{-6}	2.8×10^{-5}	1.6×10^{-4}
pl	6.7×10^{-4}	1.3×10^{-4}	4.1×10^{-7}	5.6×10^{-6}	1.5×10^{-5}
log	1.6×10^{-3}	8.4×10^{-5}	8.8×10^{-7}	5.5×10^{-5}	1.3×10^{-5}
$\hat{p}_{95\%}$	NGD	NFT	COBYLA	Powell	SPSA
exp	3.5×10^{-3}	6.4×10^{-3}	1.4×10^{-5}	2.0×10^{-4}	1.1×10^{-3}
pl	1.4×10^{-3}	2.6×10^{-3}	1.8×10^{-6}	4.0×10^{-5}	4.7×10^{-4}
log	8.3×10^{-4}	1.1×10^{-3}	7.4×10^{-7}	1.9×10^{-5}	2.3×10^{-4}

TABLE III. Final parameters of decay function fits. The optimized parameters k^* and γ^* for the exponential (exp), power-law (pl), and logarithmic (log) decay functions [cf. Eqs. (30)–(32)] are listed for the measured noise resilience thresholds $\sigma^*(n)$ (27) and $\epsilon^*(n)$ (35). Results are provided for all tested optimizers and both solvability measures \hat{p}_{opt} (upper part) based on Fig. 8 and $\hat{p}_{95\%}$ (lower part).

For \hat{p}_{opt}		NGD		NFT		COBYLA		Powell		SPSA	
		k^*	γ^*	k^*	γ^*	k^*	γ^*	k^*	γ^*	k^*	γ^*
exp	σ^*	3.1(2)	0.52(2)	11(3)	0.83(7)	1.7(3)	0.98(6)	1.8(2)	0.66(3)	2.5(5)	0.69(6)
	ϵ^*	7.3(7)	0.37(3)	18(3)	0.58(5)	2.8(4)	0.73(5)	3.5(2)	0.46(2)	4.3(7)	0.46(4)
pl	σ^*	8(2)	2.3(2)	32(3)	3.21(8)	5.1(4)	3.68(6)	4.9(3)	2.71(5)	6.8(5)	2.78(6)
	ϵ^*	16(4)	1.7(2)	49(2)	2.44(4)	7.7(3)	2.91(3)	9(1)	2.1(1)	10.6(7)	2.03(5)
log	σ^*	0.91(08)	3.2(3)	1.43(02)	4.38(8)	0.142(3)	4.9(1)	0.36(2)	3.8(2)	0.47(1)	3.85(8)
	ϵ^*	3.1(4)	2.5(4)	4.7(1)	3.4(1)	0.46(1)	4.0(1)	1.20(8)	3.0(2)	1.53(5)	2.9(1)
for $\hat{p}_{95\%}$		NGD		NFT		COBYLA		Powell		SPSA	
		k^*	γ^*	k^*	γ^*	k^*	γ^*	k^*	γ^*	k^*	γ^*
exp	σ^*	1.4(3)	0.30(6)	5(2)	0.6(1)	0.8(2)	0.73(7)	1.1(2)	0.52(6)	1.2(4)	0.5(1)
	ϵ^*	2.9(6)	0.13(4)	7(2)	0.28(6)	1.3(2)	0.48(4)	2.0(3)	0.30(4)	1.7(5)	0.19(6)
pl	σ^*	3.1(7)	1.5(2)	12(3)	2.3(2)	2.1(2)	2.89(9)	2.8(4)	2.2(1)	2.8(9)	2.0(2)
	ϵ^*	5(1)	0.7(1)	15(4)	1.4(2)	3.2(2)	2.12(5)	4.4(5)	1.48(9)	3(1)	1.1(2)
log	σ^*	0.77(5)	2.2(2)	1.27(7)	3.3(2)	0.130(2)	3.99(7)	0.34(1)	3.2(1)	0.41(3)	2.9(2)
	ϵ^*	2.5(2)	1.2(2)	4.0(3)	2.2(2)	0.43(1)	3.06(9)	1.09(4)	2.2(1)	1.3(1)	1.7(3)

size grows, this distribution exhibits a left-skewed shape, with a higher concentration of solutions near the minimum energy. This deviation from a normal distribution explains why solutions above high t values are more common than one might expect from a purely random search. This tendency of solutions to cluster closer to the global minimum than to the global maximum directly influences the behavior of $\hat{p}_{99\%}$, and $\hat{p}_{95\%}$ in our experiments.

APPENDIX F: FIT ANALYSIS AND PARAMETERS FOR OBSERVED NOISE RESILIENCE DECAY

Figure 8 presents the fitted decay functions [cf. Eqs. (30)–(32)] for the declining noise resilience threshold $\sigma^*(n)$ with respect to \hat{p}_{opt} across all tested optimizers. To assess the

quality of these fits, we compute the mean squared error (MSE):

$$d_{\text{mse}} = \frac{1}{8} \sum_{n=3}^{10} [\sigma^*(n) - f(n)]^2. \quad (\text{F1})$$

The resulting MSE values, shown in Table II, serve as a measure of the goodness-of-fit for each optimizer and decay model. Unfortunately, these values remain inconclusive, as no single decay model consistently provides the best fit across all optimizers, with the lowest MSE dependent on the optimizer. Consequently, these results do not allow for a definitive statement on the precise nature of the decay of $\sigma^*(n)$ with increasing system size. Finally, for reproducibility, Table III provides the final parameters of all fitting experiments shown in Fig. 8.

- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, and D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [2] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, *Appl. Phys. Rev.* **6**, 021314 (2019).
- [3] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, and Y. Hu *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
- [4] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [5] 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).
- [6] 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).
- [7] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, and T. Menke *et al.*, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
- [8] A. Abbas, A. Ambainis, B. Augustino, A. Bäertschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, and B. G. Elmegreen *et al.*, Challenges and opportunities in quantum optimization, *Nat. Rev. Phys.* **6**, 718 (2024).
- [9] 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).

- [10] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, A review of barren plateaus in variational quantum computing, *Nat. Rev. Phys.* **7**, 174 (2025).
- [11] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing, [arXiv:2312.09121](https://arxiv.org/abs/2312.09121).
- [12] M. L. Goh, M. Larocca, L. Cincio, M. Cerezo, and F. Sauvage, Lie-algebraic classical simulations for quantum computing, [arXiv:2308.01432](https://arxiv.org/abs/2308.01432).
- [13] L. Bittel and M. Kliesch, Training variational quantum algorithms is NP-hard, *Phys. Rev. Lett.* **127**, 120502 (2021).
- [14] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, *Nat. Commun.* **13**, 7760 (2022).
- [15] F. Leymann and J. Barzen, The bitter truth about gate-based quantum algorithms in the NISQ era, *Quantum Sci. Technol.* **5**, 044007 (2020).
- [16] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, *Nat. Commun.* **12**, 6961 (2021).
- [17] G. González-García, R. Trivedi, and J. I. Cirac, Error propagation in NISQ devices for solving classical optimization problems, *PRX Quantum* **3**, 040326 (2022).
- [18] G. Scriva, N. Astrakhantsev, S. Pilati, and G. Mazzola, Challenges of variational quantum optimization with measurement shot noise, *Phys. Rev. A* **109**, 032408 (2024).
- [19] D. S. França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, *Nat. Phys.* **17**, 1221 (2021).
- [20] Y. Shao, F. Wei, S. Cheng, and Z. Liu, Simulating noisy variational quantum algorithms: A polynomial approach, *Phys. Rev. Lett.* **133**, 120603 (2024).
- [21] T. Schuster, C. Yin, X. Gao, and N. Y. Yao, A polynomial-time classical algorithm for noisy quantum circuits, [arXiv:2407.12768](https://arxiv.org/abs/2407.12768).
- [22] V. Akshay, H. Philathong, M. E. S. Morales, and J. D. Biamonte, Reachability deficits in quantum approximate optimization, *Phys. Rev. Lett.* **124**, 090504 (2020).
- [23] 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).
- [24] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, *Quantum* **6**, 824 (2022).
- [25] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, Characterizing barren plateaus in quantum ansätze with the adjoint representation, *Nat. Commun.* **15**, 17171 (2024).
- [26] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits, *Nat. Commun.* **15**, 17172 (2024).
- [27] S. Bravyi, D. Gosset, and R. Movassagh, Classical algorithms for quantum mean values, *Nat. Phys.* **17**, 337 (2021).
- [28] A. Deshpande, P. Niroula, O. Shtanko, A. V. Gorshkov, B. Fefferman, and M. J. Gullans, Tight bounds on the convergence of noisy random circuits to the uniform distribution, *PRX Quantum* **3**, 040329 (2022).
- [29] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, Limitations of variational quantum algorithms: A quantum optimal transport approach, *PRX Quantum* **4**, 010309 (2023).
- [30] K. M. Nakanishi, K. Fujii, and S. Todo, Sequential minimal optimization for quantum-classical hybrid algorithms, *Phys. Rev. Res.* **2**, 043158 (2020).
- [31] K. J. Sung, J. Yao, M. P. Harrigan, N. C. Rubin, Z. Jiang, L. Lin, R. Babbush, and J. R. McClean, Using models to improve optimizers for variational quantum algorithms, *Quantum Sci. Technol.* **5**, 044008 (2020).
- [32] W. Lavrijsen, A. Tudor, J. Muller, C. Iancu, and W. de Jong, Classical optimizers for noisy intermediate-scale quantum devices, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Denver, CO, 2020), pp. 267–277.
- [33] A. Pellow-Jarman, I. Sinayskiy, A. Pillay, and F. Petruccione, A comparison of various classical optimizers for a variational quantum linear solver, *Quantum Info. Proc.* **20**, 202 (2021).
- [34] V. P. Soloviev, P. Larrañaga, and C. Bielza, Quantum parametric circuit optimization with estimation of distribution algorithms, in *Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO'22* (ACM, Boston, Massachusetts, 2022), pp. 2247–2250.
- [35] H. Singh, S. Majumder, and S. Mishra, Benchmarking of different optimizers in the variational quantum algorithms for applications in quantum chemistry, *J. Chem. Phys.* **159**, 044117 (2023).
- [36] L. Palackal, B. Poggel, M. Wulff, H. Ehm, J. M. Lorenz, and C. B. Mendl, Quantum-assisted solution paths for the capacitated vehicle routing problem, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Bellevue, WA, 2023), pp. 648–658.
- [37] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, and T. E. O'Brien, Performance comparison of optimization methods on variational quantum algorithms, *Phys. Rev. A* **107**, 032407 (2023).
- [38] A. Pellow-Jarman, S. McFarthing, I. Sinayskiy, D. K. Park, A. Pillay, and F. Petruccione, The effect of classical optimizers and ansatz depth on QAOA performance in noisy devices, *Sci. Rep.* **14**, 16011 (2024).
- [39] A. Bärligea, Scalability of variational quantum optimization in the presence of noise, Master's thesis, Technical University of Munich, Garching, Germany, 2024.
- [40] V. Kungurtsev, G. Korpas, J. Marecek, and E. Y. Zhu, Iteration complexity of variational quantum algorithms, *Quantum* **8**, 1495 (2024).
- [41] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms* (Springer, Berlin, 2018).
- [42] R. M. Karp, Reducibility among combinatorial problems, *Complexity of Computer Computations*, edited by R. E. Miller and J. W. Thatcher, The IBM Research Symposia Series (Springer, New York, 1972), pp. 85–103.
- [43] G. Ausiello, A. Marchetti-Spaccamela, P. Crescenzi, G. Gambosi, M. Protasi, and V. Kann, *Complexity and Approximation* (Springer, Berlin, 1999).
- [44] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth*

- Annual ACM Symposium on Theory of Computing, STOC '96* (Association for Computing Machinery, New York, 1996), pp. 212–219.
- [45] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2016).
- [46] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, Quantum computation by adiabatic evolution, [arXiv:quant-ph/0001106](#).
- [47] A. Bochkarev, R. Heese, S. Jäger, P. Schiewe, and A. Schöbel, Quantum computing for discrete optimization: A highlight of three technologies, [arXiv:2409.01373](#).
- [48] S. Heng, D. Kim, T. Kim, and Y. Han, How to solve combinatorial optimization problems using real quantum machines: A recent survey, *IEEE Access* **10**, 120106 (2022).
- [49] E. Zahedinejad and A. Zaribafian, Combinatorial optimization on gate model quantum computers: A survey, [arXiv:1708.05294](#).
- [50] K. Sankar, A. Scherer, S. Kako, S. Reifenstein, N. Ghadermarzy, W. B. Krayenhoff, Y. Inui, E. Ng, T. Onodera, P. Ronagh, and Y. Yamamoto, A benchmarking study of quantum algorithms for combinatorial optimization, *npj Quantum Inf.* **10**, 64 (2024).
- [51] M. Dupont, B. Sundar, B. Evert, D. E. B. Neira, Z. Peng, S. Jeffrey, and M. J. Hodson, Benchmarking quantum optimization for the maximum-cut problem on a superconducting quantum computer, *Phys. Rev. Appl.* **23**, 014045 (2025).
- [52] Y. Fu and P. W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation, *J. Phys. A: Math. Gen.* **19**, 1605 (1986).
- [53] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, and Y. Wang, The unconstrained binary quadratic programming problem: A survey, *J. Comb. Optim.* **28**, 58 (2014).
- [54] F. Glover, G. Kochenberger, R. Hennig, and Y. Du, Quantum bridge analytics I: A tutorial on formulating and using QUBO models, *Ann. Oper. Res.* **314**, 141 (2022).
- [55] F. Glover, G. Kochenberger, and Y. Du, Applications and computational advances for solving the QUBO model, in *The Quadratic Unconstrained Binary Optimization Problem: Theory, Algorithms, and Applications*, edited by A. P. Punnen (Springer International Publishing, Cham, 2022), pp. 39–56.
- [56] F. Barahona, On the computational complexity of Ising spin glass models, *J. Phys. A: Math. Gen.* **15**, 3241 (1982).
- [57] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5 (2014).
- [58] 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).
- [59] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](#).
- [60] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, Quantum optimization using variational algorithms on near-term quantum devices, *Quantum Sci. Technol.* **3**, 030503 (2018).
- [61] G. Nannicini, Performance of hybrid quantum-classical variational heuristics for combinatorial optimization, *Phys. Rev. E* **99**, 013304 (2019).
- [62] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The variational quantum eigensolver: A review of methods and best practices, *Phys. Rep.* **986**, 1 (2022).
- [63] K. Blekos, D. Brand, A. Ceschini, C.-H. Chou, R.-H. Li, K. Pandya, and A. Summer, A review on quantum approximate optimization algorithm and its variants, *Phys. Rep.* **1068**, 1 (2024).
- [64] N. Kuete Meli, F. Mannel, and J. Lellmann, A universal quantum algorithm for weighted maximum cut and Ising problems, *Quantum Info. Proc.* **22**, 279 (2023).
- [65] A. Bärligea, B. Poggel, and J. M. Lorenz, Solving combinatorial optimization problems with a block encoding quantum optimizer, [arXiv:2404.14054](#).
- [66] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand unification of quantum algorithms, *PRX Quantum* **2**, 040203 (2021).
- [67] D. Aharonov, V. Jones, and Z. Landau, A polynomial quantum algorithm for approximating the Jones polynomial, *Algorithmica* **55**, 395 (2009).
- [68] J. Rajakumar, J. Golden, A. Bärtschi, and S. Eidenbenz, Trainability barriers in low-depth QAOA landscapes, in *Proceedings of the 21st ACM International Conference on Computing Frontiers, CF'24* (ACM, Ischia, Italy, 2024), pp. 199–206.
- [69] This is a conservative assumption on parameter scaling, as larger problem instances typically require deeper circuits with more parameters, leading to higher-dimensional loss landscapes and potentially better optimization conditions [136]. Fixing this relationship therefore ensures a fair scalability comparison without biases from varying parameter growth.
- [70] R. Wiersema, E. Kökcü, A. F. Kemper, and B. N. Bakalov, Classification of dynamical Lie algebras of 2-local spin systems on linear, circular and fully connected topologies, *npj Quantum Inf.* **10**, 110 (2024).
- [71] E. Kökcü, R. Wiersema, A. F. Kemper, and B. N. Bakalov, Classification of dynamical Lie algebras generated by spin interactions on undirected graphs, [arXiv:2409.19797](#).
- [72] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, Quantum computing with Qiskit, [arXiv:2405.08810](#).
- [73] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, and J. Bright *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).
- [74] R. Murray, B. Swenson, and S. Kar, Revisiting normalized gradient descent: Fast evasion of saddle points, *IEEE Trans. Autom. Control* **64**, 4818 (2019).
- [75] Y. Suzuki, H. Yano, R. Raymond, and N. Yamamoto, Normalized gradient descent for variational quantum algorithms, in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Broomfield, CO, 2021), pp. 1–9.
- [76] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.* **16**, 1190 (1995).
- [77] R. Fletcher, Function minimization by conjugate gradients, *Comput. J.* **7**, 149 (1964).

- [78] D. Kraft, A software package for sequential quadratic programming (1988), https://degenerateconic.com/uploads/2018/03/DFVLR_FB_88_28.pdf.
- [79] J. C. Spall, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE Trans. Autom. Control* **37**, 332 (1992).
- [80] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, *Johns Hopkins APL Tech. Dig.* **19**, 482 (1998).
- [81] M. J. D. Powell, Direct search algorithms for optimization calculations, *Acta Numer.* **7**, 287 (1998).
- [82] M. J. D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in Optimization and Numerical Analysis* (Springer Netherlands, Amsterdam, 1994), pp. 51–67.
- [83] M. J. D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Comput. J.* **7**, 155 (1964).
- [84] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [85] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [86] K. Zhang, L. Liu, M.-H. Hsieh, and D. Tao, Escaping from the barren plateau via Gaussian initializations in deep variational quantum circuits, in *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, New Orleans, LA, 2022), Vol. 35, pp. 18612–18627.
- [87] J. Branke and C. Schmidt, Selection in the presence of noise, in *Genetic and Evolutionary Computation—GECCO 2003* (Springer, Berlin, 2003), pp. 766–777.
- [88] S. Stich and H. Harshvardhan, Escaping local minima with stochastic noise, in *OPT2021: 13th Annual Workshop on Optimization for Machine Learning* (NeurIPS, San Diego, CA, 2021), pp. 18612–18627.
- [89] J. Liu, F. Wilde, A. A. Mele, L. Jiang, and J. Eisert, Stochastic noise can be helpful for variational quantum algorithms, *Phys. Rev. A* **111**, 052441 (2025).
- [90] E. Kaminishi, T. Mori, M. Sugawara, and N. Yamamoto, Impact of measurement noise on escaping saddles in variational quantum algorithms, [arXiv:2406.09780](https://arxiv.org/abs/2406.09780).
- [91] E. Fontana, M. Cerezo, A. Arrasmith, I. Rungger, and P. J. Coles, Non-trivial symmetries in quantum landscapes and their resilience to quantum noise, *Quantum* **6**, 804 (2022).
- [92] J. Kim and Y. Oz, Quantum energy landscape and circuit optimization, *Phys. Rev. A* **106**, 052424 (2022).
- [93] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, Evaluating the noise resilience of variational quantum algorithms, *Phys. Rev. A* **104**, 022403 (2021).
- [94] O. Ezratty, Where are we heading with NISQ? [arXiv:2305.09518](https://arxiv.org/abs/2305.09518).
- [95] I. Quantum, Development & innovation roadmap (2024), https://www.ibm.com/quantum/assets/IBM_Quantum_Development_&_Innovation_Roadmap.pdf.
- [96] Quantinuum, Quantinuum Unveils Accelerated Roadmap to Achieve Universal, Fully Fault-Tolerant Quantum Computing by 2030 (2024), <https://www.quantinuum.com/press-releases/quantinuum-unveils-accelerated-roadmap-to-achieve-universal-fault-tolerant-quantum-computing-by-2030>.
- [97] Using an exponential fit for the loss variance—consistent with a barren plateau assumption—would imply an increase in the noise resilience for system sizes $n \geq 61$ in the power-law case (31), and $n \geq 27$ for logarithmic decay (32). While this would not change the overall implications of Sec. VI B, this scenario is unlikely.
- [98] If σ^* were to decay logarithmically, ϵ^* would eventually start increasing again due to the faster power-law decay of the loss variance. However, since this effect only becomes relevant for $n > e^8 \approx 2981$, it remains negligible within the scope of our findings.
- [99] G. Mazzola and G. Carleo, Exponential challenges in unbiasing quantum monte carlo algorithms with quantum computers, [arXiv:2205.09203](https://arxiv.org/abs/2205.09203).
- [100] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, Can error mitigation improve trainability of noisy variational quantum algorithms? *Quantum* **8**, 1287 (2024).
- [101] D. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, Concorde TSP solver (2003), <https://www.math.uwaterloo.ca/tsp/concorde/index.html>.
- [102] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*, Princeton Series in Applied Mathematics (Princeton University Press, Princeton, 2007).
- [103] J. Pawłowski, J. Tuziński, P. Tarasiuk, A. Przybysz, R. Adamski, K. Hendzel, L. Paweła, and B. Gardas, VeloxQ: A fast and efficient QUBO solver, [arXiv:2501.19221](https://arxiv.org/abs/2501.19221).
- [104] O. Salehi, A. Glos, and J. A. Miszczak, Unconstrained binary models of the travelling salesman problem variants for quantum optimization, *Quantum Info. Proc.* **21**, 67 (2022).
- [105] M. Schnaus, L. Palackal, B. Poggel, X. Runge, H. Ehm, J. M. Lorenz, and C. B. Mendl, Efficient encodings of the travelling salesperson problem for variational quantum algorithms, in *2024 IEEE International Conference on Quantum Software (QSW)* (IEEE, Shenzhen, China, 2024), pp. 81–87.
- [106] D. J. Egger, J. Mareček, and S. Woerner, Warm-starting quantum optimization, *Quantum* **5**, 479 (2021).
- [107] S. H. Sack and M. Serbyn, Quantum annealing initialization of the quantum approximate optimization algorithm, *Quantum* **5**, 491 (2021).
- [108] S. Boulebnane and A. Montanaro, Predicting parameters for the quantum approximate optimization algorithm for MAX-CUT from the infinite-size limit, [arXiv:2110.10685](https://arxiv.org/abs/2110.10685).
- [109] N. Jain, B. Coyle, E. Kashefi, and N. Kumar, Graph neural network initialisation of quantum approximate optimisation, *Quantum* **6**, 861 (2022).
- [110] A. N. Cáliz, J. Riu, J. Bosch, P. Torrente, J. Miralles, and A. Riera, A coherent approach to quantum-classical optimization, *Commun. Phys.* **8**, 197 (2025).
- [111] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, *Quantum* **3**, 214 (2019).
- [112] Y. Wang, B. Qi, C. Ferrie, and D. Dong, Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization, *Phys. Rev. Appl.* **22**, 054005 (2024).
- [113] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, *Quantum Sci. Technol.* **6**, 025008 (2021).

- [114] C.-Y. Park, M. Kang, and J. Huh, Hardware-efficient ansatz without barren plateaus in any depth, [arXiv:2403.04844](#).
- [115] N. Herrmann, D. Arya, M. W. Doherty, A. Mingare, J. C. Pillay, F. Preis, and S. Prestel, Quantum utility—Definition and assessment of a practical quantum advantage, in *2023 IEEE International Conference on Quantum Software (QSW)* (IEEE, Chicago, IL, 2023), pp. 162–174.
- [116] M. Sciorilli, L. Borges, T. L. Patti, D. García-Martín, G. Camilo, A. Anandkumar, and L. Aolita, Towards large-scale quantum optimization solvers with few qubits, *Nat. Commun.* **16**, 476 (2025).
- [117] P. Bermejo and R. Orús, Variational quantum non-orthogonal optimization, *Sci. Rep.* **13**, 9840 (2023).
- [118] F. B. Maciejewski, B. G. Bach, M. Dupont, P. A. Lott, B. Sundar, D. E. B. Neira, I. Safro, and D. Venturelli, A multilevel approach for solving large-scale QUBO problems with noisy hybrid quantum approximate optimization, [arXiv:2408.07793](#).
- [119] J. B. Larsen, M. D. Grace, A. D. Baczewski, and A. B. Magann, Feedback-based quantum algorithms for ground state preparation, *Phys. Rev. Res.* **6**, 033336 (2024).
- [120] J. Jiang, P. Huang, Z. Wu, X. Sun, Z. Guo, W. Huang, L. Zhang, Y. Zhou, J. Zhang, W. Guo, X. Linpeng, S. Liu, W. Ren, Z. Tao, J. Chu, J. Niu, Y. Zhong, and D. Yu, Improving quantum optimization to achieve quadratic time complexity, [arXiv:2501.13469](#).
- [121] T. Bennett, L. Noakes, and J. Wang, Non-variational quantum combinatorial optimisation, [arXiv:2404.03167](#).
- [122] A. B. Catli, S. Simon, and N. Wiebe, Exponentially better bounds for quantum optimization via dynamical simulation, [arXiv:2502.04285](#).
- [123] S. P. Jordan, N. Shutty, M. Wootters, A. Zalcman, A. Schmidhuber, R. King, S. V. Isakov, T. Khattar, and R. Babbush, Optimization by decoded quantum interferometry, [arXiv:2408.08292](#).
- [124] P.-W. Huang and P. Rebentrost, Post-variational quantum neural networks, [arXiv:2307.10560](#) [quant-ph].
- [125] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian boson sampling, *Phys. Rev. Lett.* **119**, 170501 (2017).
- [126] M. Słysz, K. Kurowski, and G. Waligóra, Solving combinatorial optimization problems on a photonic quantum computer, [arXiv:2409.13781](#).
- [127] B. Go, C. Oh, and H. Jeong, Quantum computational advantage of noisy boson sampling with partially distinguishable photons, [arXiv:2501.13433](#).
- [128] 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).
- [129] S. Yarkoni, E. Raponi, T. Bäck, and S. Schmitt, Quantum annealing for industry applications: Introduction and review, *Rep. Prog. Phys.* **85**, 104001 (2022).
- [130] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
- [131] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [132] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, *Nature (London)* **607**, 667 (2022).
- [133] B. Fauseweh, Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges, *Nat. Commun.* **15**, 2123 (2024).
- [134] F. Barahona, M. Jünger, and G. Reinelt, Experiments in quadratic 0-1 programming, *Math. Program.* **44**, 127 (1989).
- [135] H. Fischer, *A History of the Central Limit Theorem: From Classical to Modern Probability Theory* (Springer, New York, 2011).
- [136] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, *Nat. Comput. Sci.* **3**, 542 (2023).