



Warm start of variational quantum algorithms for quadratic unconstrained binary optimization problems

Yahui Chai^{1*}, Karl Jansen^{1,2}, Stefan Kühn¹, Tim Schwägerl^{3,1} and Tobias Stollenwerk⁴

*Correspondence:

yahui.chai@desy.de

¹Deutsches Elektronen-Synchrotron
DESY, Platanenallee 6, 15738
Zeuthen, Germany

Full list of author information is
available at the end of the article

Abstract

Variational Quantum Eigensolver (VQE) is widely used in near-term hardware. However, their performances remain limited by the poor trainability and are dependent on random parameter initialization. In this work, we propose a warm start method inspired by imaginary time evolution, allowing for determining initial parameters that prioritize lower energy states in a resource-efficient way. Using classical simulations, we demonstrate that this warm start method significantly improves the success rate and reduces the number of iterations required for the convergence of VQE. The numerical results also indicate that the warm start approach effectively mitigates statistical errors arising from a finite number of measurements, and to a certain extent alleviates the effect of barren plateaus.

Keywords: Warm start; Variational quantum algorithm; Combinatorial optimization

1 Introduction

The Variational Quantum Eigensolver (VQE) is a classical–quantum hybrid algorithm that uses parametrized quantum circuits as ansätze [1, 2]. Because these circuits can be flexible and relatively shallow, VQE is well suited for noisy intermediate-scale quantum (NISQ) devices. The method has been applied to a wide range of problems, including eigenvalue calculations in physics and chemistry [3, 4] and various combinatorial optimization tasks [5–9]. Despite this broad applicability, two major challenges limit its practical usefulness: statistical errors arising from finite measurement shots [10] and the presence of barren plateaus in the optimization landscape [11]. Both issues can prevent the optimization from converging to the global minimum. Consequently, designing suitable ansätze and choosing effective initial parameters are important for improving VQE performance.

In this work, we investigate the quadratic unconstrained binary optimization (QUBO) problem [12] and develop a warm start approach. By approximating the action of imaginary time evolution step by step, this warm start is able to provide an initialization prioritizing lower energy states and offering a more effective starting point than the typical universal superposition or random states. To effectively obtain the warm start parameters, we introduce a *structure-inspired ansatz* (SIA). While the SIA alone does not outperform

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

the hardware-efficient ansatz (HEA) [13] with finite shots, it gains significant improvement combined with the warm start, highlighting the importance of the initialization for the VQE. Our numerical simulation results indicate that VQE with the warm start is able to result in a higher success rate in finding the optimal solution with fewer iterations. Besides, the warm start shows a potential to mitigate the problems of statistical error and the barren plateaus.

The paper is organized as follows. Section 2 outlines the QUBO problem and the simulation setup. Subsequently, we introduce a warm start approach in Sect. 3.2, as well as low-cost simplification thereof in Sect. 3.3, which allows for determining the initial set of parameters classically. The improvement in performance of these methods is demonstrated using numerical simulation. Section 4 shows the potential of the warm start to mitigate statistical errors and to alleviate the effect of barren plateaus. Finally, Sect. 5 concludes with a summary and prospects for future research.

2 QUBO and CVaR-VQE

In this paper we focus on the CVaR-VQE, a variant of the VQE proposed for combinatorial optimization problems [14], and apply it to QUBO problems, both of which we review below. The QUBO problem is defined by a vector of N binary variables $\mathbf{x} = \{x_0, x_1, \dots, x_{N-1}\}$ and the symmetric coefficient matrix $Q \in \mathbb{R}^{N \times N}$ as well as a cost function [12]

$$f(\mathbf{x}) = \mathbf{x}^T Q \mathbf{x} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x_i \cdot Q_{ij} \cdot x_j. \quad (1)$$

The objective is to find the assignment of binary variables which minimizes the cost function for a given Q . A wide range of combinatorial optimization problems can be expressed in this mathematical framework, among them the optimal flight to gate assignment at an airport, particle tracking, job scheduling, and many others [5–8]. Every QUBO problem can be associated with a graph by interpreting the matrix Q as the adjacency matrix of a weighted, undirected graph $G(V, E)$. V corresponds the set of vertices of size $|V| = N$, and E represents the set of edges between vertex pairs (i, j) for which $Q_{ij} = Q_{ji} \neq 0$.

QUBO problems can be straightforwardly represented as Hamiltonians on N qubits which are polynomials of Pauli-Z operators [15]. In particular, the corresponding Hamiltonian for $f(\mathbf{x})$ can be obtained by substituting the binary variables x_i with the operator $(\mathbb{1} - \sigma_i^z)/2$, where $\mathbb{1}$ denotes the identity operator and σ_i^z represents the Pauli-Z operator acting on the i -th qubit. Using this substitution one finds for the Hamiltonian up to a constant

$$H = \sum_{i \in V} h_i \cdot \sigma_i^z + \sum_{(i,j) \in E} J_{ij} \cdot \sigma_i^z \sigma_j^z, \quad (2)$$

where the coefficients h_i, J_{ij} are related to the original QUBO problem's Q_{ij} . Note that the above Hamiltonian only contains Pauli-Z operators, which renders it diagonal in the computational basis, and its eigenstates are the computational basis states. This is a general property of boolean functions, as they can all be represented as diagonal matrices in the computational basis [15]. For the purpose of this study, we consider random QUBOs and sample the coefficients h_i, J_{ij} uniformly in the range $[-1, 1]$ with four-digit precision.

For combinatorial optimization problems, the conditional value at risk (CVaR) [16] has been proposed as a cost function for the VQE, and it has been shown to significantly enhance VQE's performance [5, 14, 17]. Specifically, taking S measurements in the computational basis and arranging the corresponding eigenvalues associated with the basis states sampled in ascending order, $\{E_1 \leq E_2 \leq \dots \leq E_S\}$, the CVaR corresponds to the average of the fraction α of the lowest energies

$$\text{CVaR}_\alpha = \frac{1}{\lceil \alpha S \rceil} \sum_{k=1}^{\lceil \alpha S \rceil} E_k. \quad (3)$$

Note that in the limit $\alpha = 1$, the CVaR_α is nothing but the usual sample mean for estimating the energy with S measurements used in regular VQE. In the opposite limit, $\alpha \rightarrow 0$, CVaR_α does nothing but selecting the measurement which resulted in the lowest energy. For $0 < \alpha < 1$, the CVaR_α cost function can be understood as trying to push more weight in the low-energy tail of the measurements, thereby enhancing computational basis states corresponding low-energy solutions in the wave function.

In this work we focus on $\text{CVaR}_{0.01}$ unless stated otherwise. This choice balances two considerations: emphasizing the low-energy sector while maintaining sufficient sampling statistics for a stable estimate. Since our main simulations use $S = 10,000$ measurement shots per iteration, setting $\alpha = 0.01$ ensures that the lowest 1% of outcomes still contain 100 samples, which is enough to yield a reliable average. Such a small value of α therefore emphasizes the relevant low-energy sector while maintaining robustness against statistical fluctuations.

For the rest of the paper, we use the VQE with CVaR_α as a cost function as a test bed for the ansatz and the warm start procedures we develop in the following sections. To have a consistent set of test cases, we create a 100 random QUBO instances corresponding to complete graphs for each problem size N , where N takes even numbers ranging from 12 to 24, that we use for all our numerical experiments. Throughout a VQE run, we monitor the fidelity of the final wave function with the exact solution, i.e., the probability of the optimal solutions in the quantum state, with the optimal solution obtained by brute force. A problem solution for a particular instance is deemed as successful if the maximum fidelity throughout all iterations is larger than 1%. As a metric to judge the performance of the VQE, we use the success rate among the problem ensemble, that is, the fraction of instances that were successful.

3 Warm start inspired by imaginary time evolution

In this section, we first explain how the decomposition of the Imaginary time evolution (ITE) operator motivates the structure of the ansatz in Sect. 3.1. The actual warm start procedure to determine good initial parameters for this ansatz will be introduced separately in Sect. 3.2 and Sect. 3.3.

3.1 Structure-inspired ansatz

Given a Hamiltonian H , a low-energy state can be obtained by evolving a given initial state $|\psi_0\rangle$ in imaginary time. In particular, the state tends to be the ground state, $|E_0\rangle$, in the limit

$$|E_0\rangle = \lim_{\tau \rightarrow \infty} \frac{\exp(-\tau H) |\psi_0\rangle}{\sqrt{\langle \psi_0 | \exp(-2\tau H) | \psi_0 \rangle}}, \quad (4)$$

provided the initial state $|\psi_0\rangle$ has nonvanishing overlap with $|E_0\rangle$. Note that the ITE operator, $\exp(-\tau H)$, is nonunitary and does not preserve the norm, thus we explicitly include the norm factor in the expression above. Although ITE provides theoretical guarantees of convergence to the global minimum [18], it cannot be implemented directly on quantum hardware because of its nonunitary nature, motivating the development of several indirect realizations. Variational approaches approximate the ITE dynamics using parametrized circuits [19–22], which typically require nontrivial resources for accurate parameter estimation. In addition, classical–quantum hybrid Monte Carlo schemes have been proposed, where imaginary-time observables are reconstructed through importance sampling of real-time evolution [23, 24]. These methods avoid variational optimization and rely on repeated short-time quantum circuit executions.

For combinatorial optimization problems, where the eigenstates are given by computational basis states, the target is to find the single optimal bitstring, so it is not necessary to execute the ITE exactly to get the same resulting state. Instead, it suffices to increase the probability of the optimal (or near-optimal) bitstrings, and the distribution of higher-energy states is largely irrelevant in this case. Thus, we propose to approximate the ITE in a resource-efficient way, provide a good initial state that can guide VQE's optimization. The CVaR-VQE applied afterward will focus on refining the low-energy subspace. In this subsection, we introduce the ansatz inspired by the structure of the ITE operator, which is essential for deriving the warm start parameters that mimic the ITE process. The detailed procedure for determining these parameters will be presented in the following subsections.

For the ITE process, a suitable choice of initial state is the uniform superposition of all computational basis states, $|\psi_0\rangle = |+\rangle^{\otimes N}$, where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Moreover, the terms of the Hamiltonian in Eq. (2) commute with each other, thus the corresponding imaginary time evolution can be decomposed exactly as

$$e^{-\tau H} = \prod_{(i,j) \in E} e^{-\tau J_{ij} \sigma_i^z \sigma_j^z} \times \prod_{i \in V} e^{-\tau h_i \sigma_i^z}. \quad (5)$$

Utilizing the above decomposition, we can understand the effect of the imaginary time evolution on this initial state step-by-step. Considering the second term, i.e. the single-body terms, we see that

$$e^{-\tau h_i \sigma_i^z} |+\rangle = \frac{1}{\sqrt{2}} (e^{-\tau h_i} |0\rangle + e^{\tau h_i} |1\rangle). \quad (6)$$

Up to normalization, the resulting state on the right hand side can also be generated by applying the unitary rotation gate $R_y(\theta_i^*) = \exp(-i\theta_i^* \sigma_i^y/2)$ to $|+\rangle$, with appropriately chosen angle θ_i^* . Hence, up to normalization, we have

$$|\psi_1\rangle = \prod_{i \in V} R_y(\theta_i^*) |+\rangle^{\otimes N} \sim \prod_{i \in V} e^{-\tau h_i \sigma_i^z} |+\rangle^{\otimes N}. \quad (7)$$

For the two-body terms of the imaginary time evolution operator, we can proceed similarly. Starting from $|\psi_1\rangle$, the effect of the different $e^{-\tau J_{ij} \sigma_i^z \sigma_j^z}$ in Eq. (5) can again be described by a unitary U_{ij} up to normalization

$$|\psi_{k+1}\rangle = U_{ij} |\psi_k\rangle \sim e^{-\tau J_{ij} \sigma_i^z \sigma_j^z} |\psi_k\rangle, \quad k \geq 1. \quad (8)$$

The operator U_{ij} might be very non-local to get the same resulting state as ITE. For simplicity, we can only consider the operators acts on the same qubits to approximate the effect of ITE. Since the ITE operator $e^{-\tau J_{ij} \sigma_i^z \sigma_j^z}$ is a purely real matrix without complex phase, we can consider the following two-qubit operator to approximate it:

$$U_{ij}(\theta_{ij}) = e^{-i(\theta_{ij,1} \cdot \sigma_i^z \sigma_j^y + \theta_{ij,0} \cdot \sigma_i^y \sigma_j^z)/2}, \quad (9)$$

Although there are some other real unitaries like $e^{-i\theta \cdot \sigma_i^x \sigma_j^y}$, $e^{-i\theta \cdot \sigma_i^x \sigma_j^z}$ can be considered, which does not have an effect in this case as shown in Appendix A.

Assuming that the optimal parameters $\theta_{ij,0}^*$, $\theta_{ij,1}^*$ to approximate the action of ITE on the current state $|\psi_k\rangle$ are available, we apply the operator $U_{ij}(\theta_{ij}^*)$ to mimic the effect of ITE on the qubit pair (i, j) . We then proceed to the next relevant qubit pair and repeat the same procedure. By iterating this process over all qubit pairs with non-zero couplings J_{ij} in the Hamiltonian, we obtain the final state:

$$|\psi_{(V,E)}(\theta^*)\rangle := \prod_{(i,j) \in E} U_{ij}(\theta_{ij}^*) \times \prod_{i \in V} R_y(\theta_i^*) |+\rangle^{\otimes N}, \quad (10)$$

which aims to mimic the full ITE process and yields a higher probability for low-energy solutions compared to the uniform superposition. Since the resulting ansatz reflects the structure of the problem graph, we refer to it as the *structure-inspired ansatz* (SIA). We emphasize that the novelty lies not in the ansatz structure itself, but in how the optimal parameters θ^* are obtained to enable a warm start, this procedure is detailed in the next subsection.

3.2 Warm start by measurement

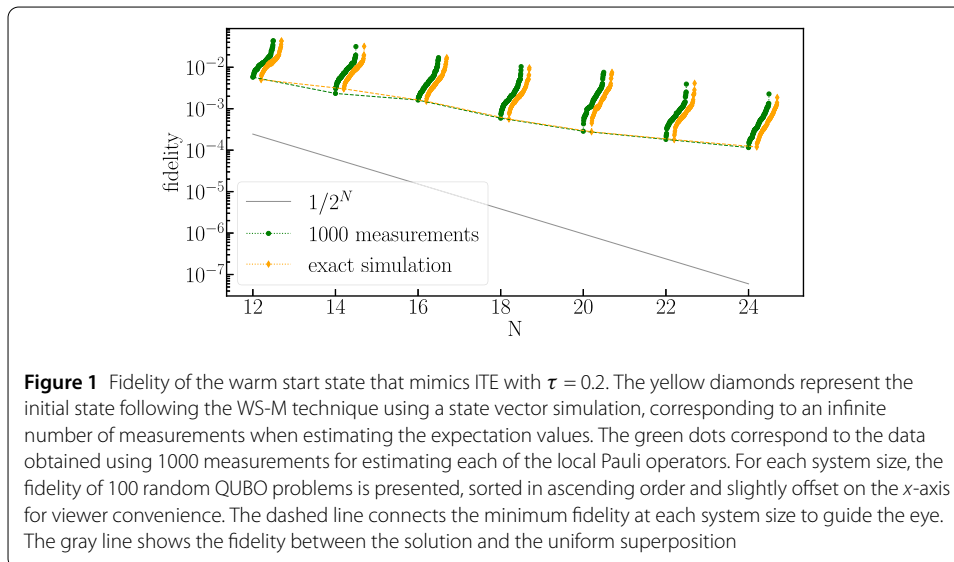
To approximate ITE step by step, we determine parameters that mimic each local ITE factor. For single qubit term, the optimal parameter can be got analytically: $\theta_i^* = 2 \arctan(-\exp(-2\tau h_i)) + \pi/2$. In contrast, the two-body operator is more complicated and depends on the state $|\psi_k\rangle$ obtained from the previous step. Specifically, the optimal parameters of $U_{ij}(\theta_{ij})$ in Eq. (8) can be determined by maximizing the overlap between the exact ITE-evolved state and the variationally transformed state

$$\begin{aligned} f_{\tau,k}(\theta_{ij,0}, \theta_{ij,1}) \\ = \langle \psi_{k-1} | e^{-\tau J_{ij} \sigma_i^z \sigma_j^z} \cdot e^{-i(\theta_{ij,1} \cdot \sigma_i^z \sigma_j^y + \theta_{ij,0} \cdot \sigma_i^y \sigma_j^z)/2} | \psi_{k-1} \rangle, \end{aligned} \quad (11)$$

The above formula corresponds to the expectation value of a two-qubit operator, which can be expanded as a sum of Pauli strings $\hat{P} \in \{I, \sigma^x, \sigma^y, \sigma^z\}^{\otimes 2}$, and the expectation $\langle \psi_{k-1} | \hat{P} | \psi_{k-1} \rangle$ of the individual Paul strings can be measured efficiently. As detailed in Appendix A, after obtaining the expectation value of those Pauli operators, the overlap $f_{\tau,k}$ for a fixed time τ only depends on the parameters $\theta_{ij,0}$, $\theta_{ij,1}$. The optimal parameters to maximize this overlap function can be easily found classically by solving

$$\theta_{ij,0}^*, \theta_{ij,1}^* = \arg \max f_{\tau,k}(\theta_{ij,0}, \theta_{ij,1}). \quad (12)$$

We then apply the unitary gate U_{ij} with these optimal parameters, and the state is updated by $|\psi_{k+1}\rangle = U_{ij}(\theta_{ij,0}^*, \theta_{ij,1}^*) |\psi_k\rangle$. This procedure is repeated for every two-body term until

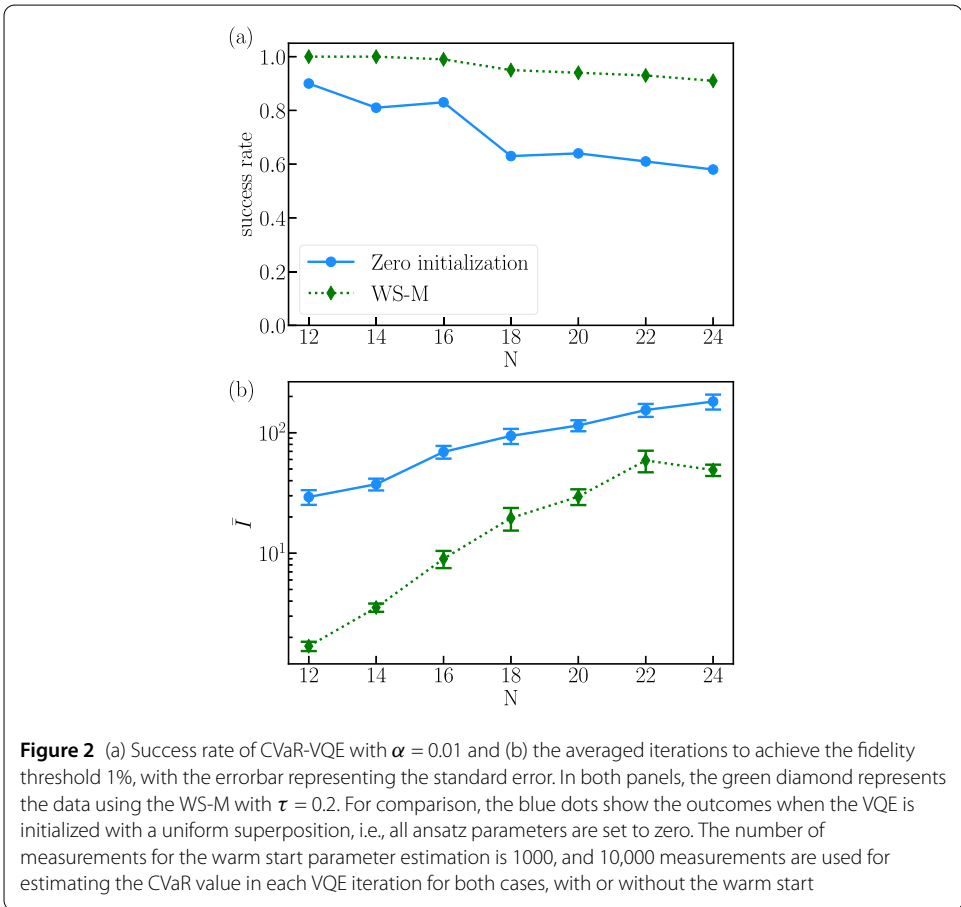


all corresponding gates in Eq. (10) have been assigned their optimal parameters. For the rest of the paper, we dub the method discussed above *warm start by measurement* (WS-M) to distinguish it from the approximate warm start technique proposed in the next section.

Note that this warm start will only provide an approximation to the ITE. Since we have chosen a specific ansatz structure for the two-qubit gates, it can not produce the effect of the ITE operator up to arbitrarily large τ . For intermediate values of τ , we expect that the WS-M approach can provide a good initial state for the VQE.

Figure 1 demonstrates the efficiency of our warm start approach, by showing the fidelity of the state with initial parameters according to the WS-M procedure compared to the uniform superposition. Across all random QUBO instances we study, the warm start consistently achieves a significantly higher fidelity than the uniform superposition. For clearer visualization, we connect the minimum fidelity at each system size with an orange dashed line in Fig. 1. Compared to the baseline given by the uniform superposition, these minimum fidelities decrease much more slowly with system size and exceed the baseline by several orders of magnitude across the entire range studied. Besides, using a modest budget of 1000 measurements to estimate the expectation value of the local Pauli operators in Eq. (11), the simulated data taking into account a finite number of measurements is close to the exact state vector simulation. Remarkably, this consistency does not decay with the number of qubits, which indicates the necessary number of measurements does almost not increase with the number of the qubits for the entire range of problem sizes we consider. This is likely because the WS-M only requires the estimation of expectations of two-qubit operators.

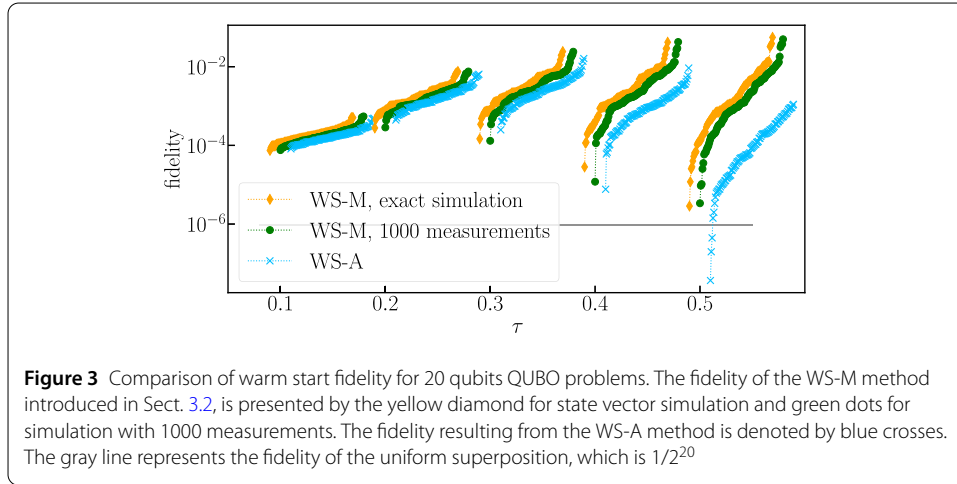
To show that the WS-M not only increases the fidelity of the initial state with the optimal solution, but generally also leads to a better VQE performance, we also compare the performance of the VQE using the SIA ansatz with and without warm start. For the case without warm start, all parameters in the circuit Eq. (10) are set to zero initially, which has a comparable performance as the hardware-efficient ansatz with the same initialization as shown in Appendix B. The classical minimization is performed with constrained optimization by linear approximation (COBYLA) [25], with the maximum number of it-



erations set to $50 \times N$, to provide a benchmark with a number of cost function evaluations realistic on current quantum hardware.

As shown in Fig. 2(a), warm-started VQE has a much higher success rate compared to starting VQE with the uniform superposition state. The separation in success rate between the cases grows slightly as the problem size increases. In addition, for the successful instances, we also record the minimal number of iterations to achieve the fidelity threshold 1%. Figure 2(b) shows the average number of iterations, \bar{I} , among the successful instances to reach the desired fidelity threshold. As the figure reveals, \bar{I} is noticeably smaller when using the warm start initial state, indicating a faster convergence rate of the VQE with warm start parameters towards the optimal solution. Specifically, for 24 qubits, \bar{I} of VQE without warm start is around 180, whereas with warm start, it reduces to about 50, thus achieving the fidelity threshold up to 3.6 times more rapidly. While our system sizes are still relatively small, these results strongly suggest that the warm start enhances VQE convergence and robustness. We plan to explore its scaling properties in follow-up work among larger system sizes.

For completeness, we also examined whether an appropriate warm start initialization could improve the performance of the HEA ansatz, as discussed in Appendix C. When we use the solution of the relaxed QUBO problem [26] as the warm start, the initial state is slightly better than the uniform superposition, but this does not lead to improved VQE optimization outcomes for HEA. This contrast further highlights the importance of the SIA ansatz together with the physically motivated warm start strategy employed in our work.



3.3 Warm start by approximation

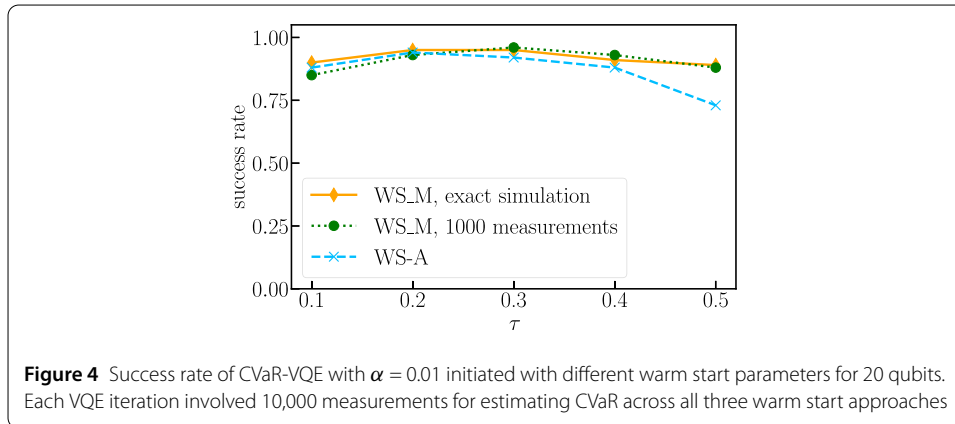
The WS-M approach allows for improving the performance of the VQE using $CVaR_\alpha$, as we have demonstrated in the last section. However, this comes at the expense of having to perform additional measurements. Here we demonstrate an approximate way for warm starting the VQE. As shown in Appendix A, the expectation value of a Pauli operator \hat{P} with respect to the intermediate state $|\psi_{k-1}\rangle$ can be approximated as:

$$\langle \psi_{k-1} | \hat{P} | \psi_{k-1} \rangle = N^{\otimes} \langle + | \hat{P} | + \rangle^{\otimes N} + \mathcal{O}(\tau^2). \tag{13}$$

Hence, provided τ is small, this allows for approximating the warm start parameters by evaluating the expectation values of Pauli operators \hat{P} in the state $|+\rangle^{\otimes N}$. These can be obtained analytically, and evaluate either to 0 or ± 1 , so that an approximation of the warm start parameters can be obtained without running the quantum circuit. Hence, we dub this technique *warm start by approximation* (WS-A). As Fig. 3 shows, for values of $\tau \leq 0.3$, the WS-A yields fidelities with the exact solution comparable to those of the WS-M discussed previously. Only for larger values of τ , the fidelity of the state initialized with WS-A decreases, which can be attributed to the increasing error term $\mathcal{O}(\tau^2)$ in estimating the expectation value of Pauli operators using state $|+\rangle^{\otimes N}$. Further comparisons between these two warm start approaches, in particular, for a larger number of ansatz layers are presented in Appendix E, revealing that the WS-M approach has a more stable performance using a larger number of layers.

In Fig. 4, we show the success rate of the VQE initialized with the warm start parameters derived from different approaches for 20 qubits. This outcome exhibits a pattern consistent with that observed in Fig. 3: the success rate for the WS-M approach with 1000 measurements is on par with that of infinite shots, and the approximation approach has a comparable performance when τ is small, but performs worse when τ increasing. Subsequent sections will focus on $\tau = 0.2$, where both warm start methods show promising performance with a single layer. For stable performance across different layers, we will use the WS-M method to generate good initial parameters for VQE.

As our results demonstrate, for a single layer there is essentially no difference in performance between the different warm start approaches, and for the following section, we will simply focus on the approach WS-M.



4 Effects of statistical errors and avoidance to barren plateaus

To elucidate the factors underlying the effectiveness of the warm start strategy, we first analyze the impact of statistical errors arising from a finite number of measurements. We then examine the cost concentration of the SIA ansatz and confirm that the ansatz itself exhibits barren plateaus [27] when initialized with random parameters. Importantly, the warm start procedure does not eliminate barren plateaus in general. Instead, it provides an initialization with gradients that do not decay with the system size, as shown in Sect. 4.2. This advantageous starting point enables the optimization to avoid flat regions of the landscape during the early stages of VQE and leads to an efficient minimization of the cost function, as illustrated in Sect. 4.2.

4.1 Statistical errors due to a finite number of measurements

To study the effect of a finite number of measurements, we utilize the QUBO instances with 18 qubits as a test bed. For these instances, the statevector simulation of the VQE shows nearly perfect success rates using the SIA ansatz: 99% for all initial parameters set to zero, 100% for warm start. Thus, this setting serves as a testbed for analyzing how statistical errors influence success rates in each methodology.

Figure 5(a) illustrates how the success rate varies with the number of measurements used in every iteration, clearly indicating that the VQE with a warm start is more robust against the finite shots. Using the warm start, one consistently observes success rates close to 100% as soon as the number of measurements is larger than 70,000. Even for as little as 2000 measurements, one can achieve a high success rate of 90%. In contrast, the VQE without a warm start achieves only maximum success rates of approximately 70%, even for 150,000 measurements. To obtain further insight into the mechanism behind the robustness of the warm start, we calculate the standard error of the expectation value for the cost function $CVaR_\alpha$ with coefficient α and a number of measurements S . The standard error is given by

$$\sigma = \sqrt{\frac{\sum_{k=1}^{\lceil \alpha \cdot S \rceil} (E_k - CVaR_\alpha(E))^2}{\lceil \alpha \cdot S \rceil (\lceil \alpha \cdot S \rceil - 1)}}, \tag{14}$$

where $\alpha = 0.01$ in this study and the obtained from each of the measurements are sorted in ascending order: $E_1 \leq E_2 \leq \dots \leq E_S$. A larger standard error indicates greater uncertainty in $CVaR_{0.01}$ when estimating the cost function with finite number of measurements. This

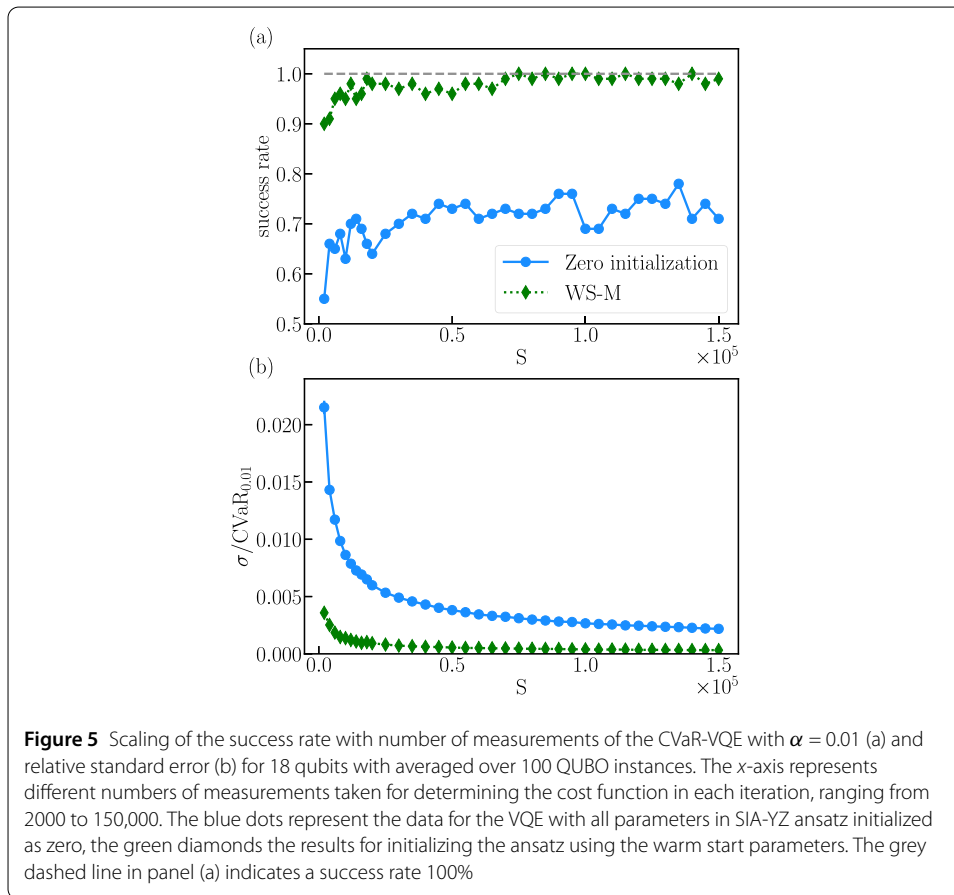


Figure 5 Scaling of the success rate with number of measurements of the CVaR-VQE with $\alpha = 0.01$ (a) and relative standard error (b) for 18 qubits with averaged over 100 QUBO instances. The x-axis represents different numbers of measurements taken for determining the cost function in each iteration, ranging from 2000 to 150,000. The blue dots represent the data for the VQE with all parameters in SIA-YZ ansatz initialized as zero, the green diamonds the results for initializing the ansatz using the warm start parameters. The grey dashed line in panel (a) indicates a success rate 100%

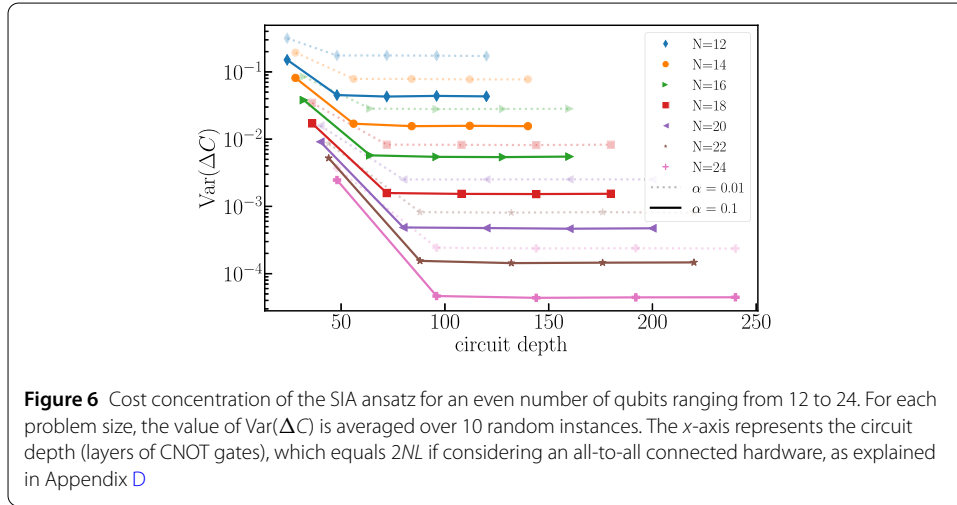
might mislead the classical optimizer when selecting a new set of parameters to minimize the cost function value. The relative standard error $\sigma / \text{CVaR}_{0.01}$ of the 10th iteration in VQE is plotted in Fig. 5(b), demonstrating that the standard error is significantly reduced when utilizing a warm start. This might explain the improved performance of the warm start with finite measurements. In turn, the reduced standard error is as a result of the warm start procedure mimicking the ITE initially pushing more weights into the low-energy states, resulting in a more concentrated lowest- α tail of the measured eigenvalue distribution.

4.2 Avoidance to barren plateaus

The problem of barren plateaus is one of the most difficult problems for the VQE, as it can prevent the trainability of ansätze [11]. Thus, it is pertinent to investigate if the SIA ansatz and warm start can mitigate this problem. To this end, we examine the scaling of the cost concentration with the number of qubits for the SIA ansatz, which was proven to be exponentially decaying if the cost landscape has the problem of barren plateaus [27].

To study to what extent the SIA ansatz is affected by barren plateaus with increasing circuit depth, we consider the ansatz with a number of layers $L \geq 1$ which is given by

$$\prod_{l=1}^L \left(\prod_{(i,j) \in E} U_{ij}(\theta_{ij}^l) \times \prod_{i \in V} R_y(\theta_i^l) \right) |+\rangle^{\otimes N}. \quad (15)$$



Furthermore, in this section, besides $\alpha = 0.01$, we also consider the $\alpha = 0.1$, because some instances for $N = 12, 14, 16$ might gain fidelity larger than 0.01 in the warm start, where the global minimal is already achieved for $\alpha = 0.01$, resulting in zero gradient. This could cause an increase in gradient as the number of qubits increases, so we also consider $\alpha = 0.1$ to verify the scaling of the gradients with the problem size.

Firstly, we explore the variance of the cost difference with increasing number of ansatz layers. For each random QUBO instance and each number of layers L , we sample 2000 random parameter vectors, $\boldsymbol{\theta} = \{\theta_i^l, \theta_{jk}^l | i \in V, (j, k) \in E, 1 \leq l \leq L\}$, and evaluate the variance of the cost difference as in Eq. (13) in Ref. [27]

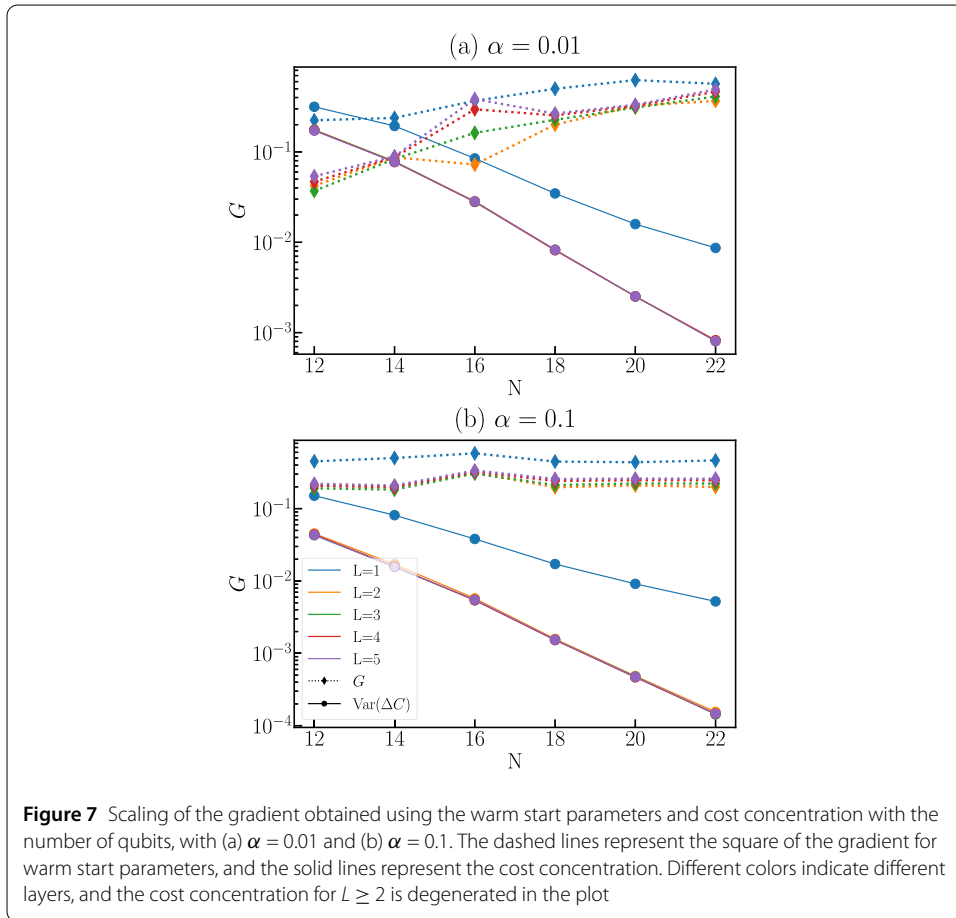
$$\text{Var}(\Delta C) = \text{Var}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta}) - \mathbb{E}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta})]]. \quad (16)$$

As throughout the rest of the paper, the cost function $C(\boldsymbol{\theta})$ is given by the CVaR_{α} in Eq. (3).

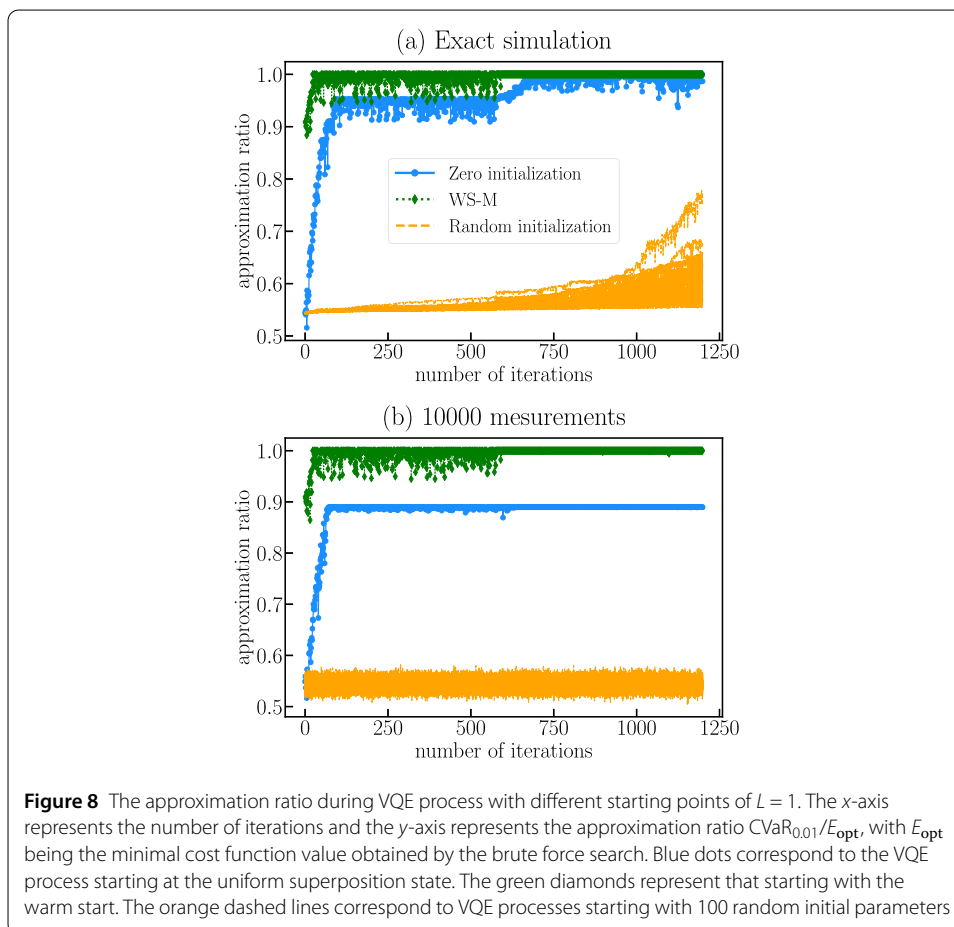
The numerical simulation result is shown in Fig. 6. As the depth of a single layer of our ansatz circuit depends on the number of qubits N , we show $\text{Var}(\Delta C)$ as a function of circuit depth for better comparability of results for different problem sizes. Looking at Fig. 6, we initially observe an exponential decay, before reaching a plateau at $L \geq 2$. This indicates the SIA ansatz will form a unitary 2-design upon reaching sufficient depth [11]. In addition, we observe that the value of the plateau reached for a fixed problem size decays exponentially with the number of qubits N . Our results indicate that the SIA generally shows barren plateaus as L increases.

Furthermore, we also study the gradient observed using the warm start. Note that in the case of $l \geq 2$, the parameters of single-qubit gates in the WS-M approach should also be determined by maximizing the overlap function similar to Eq. (11), because the state they are acting on is no longer a product state anymore. To facilitate a direct comparison with the cost concentration, which correlates with the gradients' variance, we explore the squared gradient values normalized by the dimension of the parameter space, $\dim(\boldsymbol{\theta})$, for a given instance

$$G = \frac{1}{\dim(\boldsymbol{\theta})} \sum_{i=1}^{\dim(\boldsymbol{\theta})} (\partial_i C(\boldsymbol{\theta}))^2. \quad (17)$$



The data for G are shown in Fig. 7 for various problem sizes. As the figure reveals, the gradient obtained using the warm start parameters is not decaying with an increasing number of qubits. In particular, for the smaller value of $\alpha = 0.01$ we observe an increase in gradient magnitude. This can be primarily attributed to the high fidelity of the warm start for smaller system sizes. As we have shown in Fig. 1, for $N = 12$ most instances have a fidelity exceeding 0.01 using the warm start parameters. Thus, for $\alpha = 0.01$ the cost function is essentially already optimal, and the gradient will be zero. As the number of qubits increases, the fidelity of the warm start eventually decreases below α , leading to a nonvanishing gradient and, thus, to the observed increase in gradient magnitude, as can be seen in Fig. 7(a). For the data with $\alpha = 0.1$, shown in Fig. 7(b), we observe essentially a constant value of the gradient throughout the entire range of system sizes we study. This is a result of the fact that in this case the warm start does not reach fidelities larger than α (see Fig. 1), and the cost function for smaller N is not close to the optimal value. In both case, $\alpha = 0.01, 0.1$, increasing the number of layers in the ansatz from 1 to 2 will result in a better warm start, resulting in a smaller value of the gradient. However, increasing the number of layers further beyond 2 does not necessarily enhance fidelity and consequently does not reduce the value of gradient further, explaining why the data for $L \geq 2$ is approximately the same for large N . In summary, initializing the ansatz using the warm start method does not lead to exponentially vanishing gradients, while naively increasing the number of layers will not improve the performance systematically.



The warm start provides a set of initial parameters with a large gradient, ensuring a trainable first step in the VQE. However, there are still questions about the trainability of subsequent iteration steps, and whether the warm start will lead the optimization to global or local minima. The high success rate and fast convergence to the optimal solution shown in Fig. 2 indicates that the warm start parameter should benefit these problems. To demonstrate this in a more straightforward way, we choose a random instance with $N = 24$ qubits and a single ansatz layer to examine the optimization process of the VQE with different parameter initialization strategies, aiming to confirm that the warm start provides a starting point that is easier to be optimized to the optimal solution. For simplicity, we only present the approximation ratio during the VQE of $\alpha = 0.01$, the conclusions remain consistent for $\alpha = 0.1$. As shown in Fig. 8, a random choice of initial parameters essentially leads to an untrainable ansatz, especially in the case of using a finite number of measurements. Even with an exact state vector simulation (see Fig. 8 (a)), corresponding to an infinite number of measurements, most instances show barely any improvement of the cost function during the training process. The few instances that do show some improvement converge rather slow and stay far away from the minimum cost function value over the course of 1250 iterations. As soon as we use a finite number of measurements $S = 10,000$ (cf. Fig. 8 (b)), we observe no substantial improvement of the cost function throughout the training process for randomly chosen initial parameters. This observation aligns with our findings on barren plateaus for the SIA ansatz discussed previously.

Initializing with all parameters being zero, corresponding to the uniform superposition as initial state, will result in a large gradient in the beginning, which enables the optimizer to quickly find a direction where the cost function decreases. However, even with an infinite number of measurements it takes a long time to get close to the global minimum and the optimization process revolves around a local minimum for a long time, as Fig. 8(a) demonstrates. Using a finite number of measurements, Fig. 8(b) illustrates that this initialization tends to get stuck in local minima, explaining the low success rate observed previously in Fig. 9(b). This issue can be attributed to the large statistical errors as discussed in Sect. 4.1.

In contrast, initializing parameters using the warm start method results in an already good approximation ratio initially, and the optimization procedure reliably converges to the optimal solution, in scenarios with both infinite and finite shots, as the results in Fig. 9 reveals. This indicates that our warm start is able to provide a good starting point in the cost function landscape that is close to the narrow gorges in which the cost function is concentrated and can help to overcome trainability issues.

5 Summary and outlook

In this work, we propose a resource-efficient warm-start method that mimics the imaginary time evolution of a uniform superposition over all computational basis states. Performing classical simulations with up to 24 qubits we numerically benchmarked the performance of the warm start for the VQE using the conditional value at risk as a cost function. Our data demonstrate that the warm start procedure allows for obtaining a good performance, even with a modest number of measurements per iteration in the VQE. The boost in performance through the warm start method can be primarily attributed to the fact that the initial state generated has a dominant component of low-energy states, resulting in a reduction of statistical errors when estimating expectation values with a finite number of measurements. Moreover, we examined how the warm start parameters avoid barren plateau regions. While we observe that the ansatz in general suffers from exponential cost concentration, our numerical results indicate that the warm start provides a point in parameter space sufficiently close to one of the narrow gorges where the cost function does not look flat.

It is worth mentioning that in our warm start approach, we determined the optimal parameter of each gate in the ansatz individually while keeping the others fixed to mimic the ITE. While this is the most resource-efficient approach in terms of overhead require for the warm start, optimizing multiple parameters at the same time might allow for faithfully emulating the effect of imaginary time evolution up to longer time-scales. This could provide an initial state that has even larger overlap with the solution as our current technique. Evaluating the trade-off between the cost of obtaining good warm start parameters and the resulting performance enhancement in the subsequent VQE will be an interesting aspect for further studies. Beyond VQE, the proposed warm-start strategy may be applicable to other quantum algorithms. For example, it could serve as an initial state for Grover search, increasing the initial overlap with the target solution and reducing the number of required iterations. Exploring such extensions is another promising avenue for further investigation.

Moreover, in our simulations considered an ideal quantum device performing a finite number of measurement. For the future it would be interesting to realize the simulation

on current noisy quantum hardware and to investigate if the warm start shows a similar performance improvement in the presence of noise and in combination with state of the art error mitigation methods [28–32]. In addition, it is an interesting question different kinds of hardware noise affect the warm start procedure and how these could possibly be mitigated.

Appendix A: Details of the warm start procedures

For the WS-M approach, we have to maximize the expectation value in Eq. (11) by choosing appropriate variational parameters for the two-qubit gate acting on qubits i and j . Utilizing the identities $e^{-a\hat{P}} = \cosh(a) - \sinh(a)\hat{P}$ and $e^{-ia\hat{P}} = \cos(a) - i\sin(a)\hat{P}$, for a real coefficient a and \hat{P} denoting one of the Pauli operators, $\hat{P} \in \{I, \sigma^x, \sigma^y, \sigma^z\}^{\otimes 2}$, we can rewrite the overlap f_k as follows

$$\begin{aligned}
 & f_{\tau,k}(\theta_{ij,0}, \theta_{ij,1}) \\
 &= \langle \psi_{k-1} | e^{-\tau J_{ij} \sigma_i^z \sigma_j^z} \cdot e^{-i(\theta_{ij,1} \sigma_i^z \sigma_j^y + \theta_{ij,0} \sigma_i^y \sigma_j^z)/2} | \psi_{k-1} \rangle \\
 &= \cos\left(\frac{\theta_{ij,1}}{2}\right) \cos\left(\frac{\theta_{ij,0}}{2}\right) \left(\cosh(\tilde{\tau}_{ij}) - \sinh(\tilde{\tau}_{ij}) \cdot \langle \sigma_i^z \sigma_j^z \rangle \right) \\
 &\quad - i \cos\left(\frac{\theta_{ij,1}}{2}\right) \sin\left(\frac{\theta_{ij,0}}{2}\right) \left(\cosh(\tilde{\tau}_{ij}) \langle \sigma_i^y \sigma_j^z \rangle + i \sinh(\tilde{\tau}_{ij}) \cdot \langle \sigma_i^x \rangle \right) \\
 &\quad - i \sin\left(\frac{\theta_{ij,1}}{2}\right) \cos\left(\frac{\theta_{ij,0}}{2}\right) \left(\cosh(\tilde{\tau}_{ij}) \langle \sigma_i^z \sigma_j^y \rangle + i \sinh(\tilde{\tau}_{ij}) \cdot \langle \sigma_j^x \rangle \right) \\
 &\quad - \sin\left(\frac{\theta_{ij,1}}{2}\right) \sin\left(\frac{\theta_{ij,0}}{2}\right) \left(\cosh(\tilde{\tau}_{ij}) \langle \sigma_i^x \sigma_j^x \rangle + \sinh(\tilde{\tau}_{ij}) \cdot \langle \sigma_i^y \sigma_j^y \rangle \right).
 \end{aligned} \tag{A.1}$$

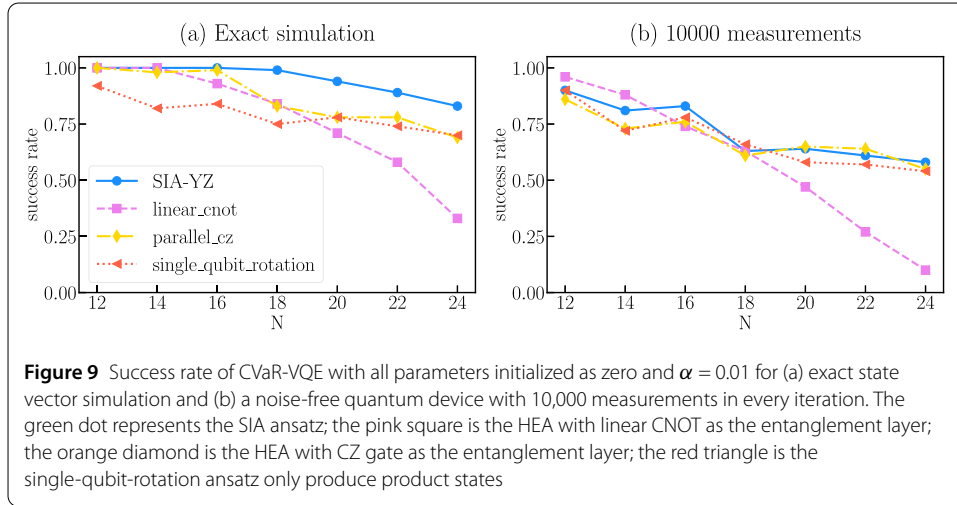
In the above expression, we have used the short-hand notations $\tilde{\tau}_{ij} = \tau J_{ij}$ and $\langle \hat{P} \rangle = \langle \psi_{k-1} | \hat{P} | \psi_{k-1} \rangle$. Notice that $\langle \sigma_i^z \sigma_j^y \rangle = 0 = \langle \sigma_i^y \sigma_j^z \rangle$, because these two operators are imaginary and hermitian, resulting in a vanishing expectation in a real state.

A similar calculation can be performed for the operators $\sigma_i^x \sigma_j^y, \sigma_i^y \sigma_j^x$ in the ansatz, the corresponding overlap will be given by

$$\begin{aligned}
 & f'_{\tau,k}(\theta_{ij,0}, \theta_{ij,1}) \\
 &= \langle \psi_{k-1} | e^{-\tau J_{ij} \sigma_i^z \sigma_j^z} \cdot e^{-i(\theta_{ij,1} \sigma_i^x \sigma_j^y + \theta_{ij,0} \sigma_i^y \sigma_j^x)/2} | \psi_{k-1} \rangle \\
 &= e^{-\tilde{\tau}_{ij}} \cdot \left(\cos\left(\frac{\theta_{ij,1}}{2}\right) \cos\left(\frac{\theta_{ij,0}}{2}\right) - \sin\left(\frac{\theta_{ij,1}}{2}\right) \sin\left(\frac{\theta_{ij,0}}{2}\right) \cdot \langle \sigma_i^z \sigma_j^z \rangle \right).
 \end{aligned} \tag{A.2}$$

This expression is maximized when $\theta_{ij,0} = 0 = \theta_{ij,1}$, which explains why we did not include the operator $\sigma_i^x \sigma_j^y, \sigma_i^y \sigma_j^x$ in our ansatz.

Considering the expectation of the Pauli operators in Eq. (A.1), with operator $\hat{P} = \sigma_i^x, \sigma_j^x, \sigma_i^x \sigma_j^x, \sigma_i^y \sigma_j^y, \sigma_i^z \sigma_j^z$, the expectation $\langle \hat{P} \rangle$ can be expanded by using the definition of



$|\psi_{k-1}\rangle$ from Eq. (8)

$$\begin{aligned}
 \langle \hat{P} \rangle &= \langle \psi_{k-1} | \hat{P} | \psi_{k-1} \rangle \\
 &= \langle \psi_{k-2} | e^{i(\theta_{pq,1} \cdot \sigma_p^z \sigma_q^y + \theta_{pq,0} \cdot \sigma_p^y \sigma_q^z)/2} \cdot \hat{P} \cdot e^{-i(\theta_{pq,1} \cdot \sigma_p^z \sigma_q^y + \theta_{pq,0} \cdot \sigma_p^y \sigma_q^z)/2} | \psi_{k-2} \rangle \\
 &= \langle \psi_{k-2} | \hat{P} | \psi_{k-2} \rangle + \langle \psi_{k-2} | i \left[\sigma_p^z \sigma_q^y + \sigma_p^y \sigma_q^z, P \right] | \psi_{k-2} \rangle \cdot \mathcal{O}(\tau) + \mathcal{O}(\tau^2), \\
 &= N^{\otimes} \langle + | \hat{P} | + \rangle^{\otimes N} + \left(\sum_{i \in V} N^{\otimes} \langle + | i [\sigma_i^y, P] | + \rangle^{\otimes N} \right. \\
 &\quad \left. + \sum_{pq \in E'} N^{\otimes} \langle + | i \left[\sigma_p^z \sigma_q^y + \sigma_p^y \sigma_q^z, P \right] | + \rangle^{\otimes N} \right) \cdot \mathcal{O}(\tau) + \mathcal{O}(\tau^2) \\
 &= N^{\otimes} \langle + | \hat{P} | + \rangle^{\otimes N} + \mathcal{O}(\tau^2).
 \end{aligned} \tag{A.3}$$

In the second line of the above equation, qubits indexed by p, q refer to the gate applied at step $k-1$. The Baker-Campbell-Hausdorff formula, $e^{\hat{A}} \cdot \hat{P} \cdot e^{-\hat{A}} = \hat{P} + [\hat{A}, \hat{P}] + \frac{1}{2} [\hat{A}, [\hat{A}, \hat{P}]] \dots$, was used for the expansion on the third line. Besides, it is assumed that the parameters $\theta_{pq,0}$ and $\theta_{pq,1}$ are of the same order as τ . Finally, all gates existing in the circuit before step k are expanded in a similar procedure. E' represents the pairs of two-qubit gates that were previously applied in the circuit, with $(p, q) \in E'$ being different from (i, j) , such that the commutator $[\sigma_p^z \sigma_q^y + \sigma_p^y \sigma_q^z, P]$ will be zero or contain σ^z or σ^y operator, resulting in a zero expectation in state $|+\rangle^{\otimes N}$, similar to the terms $[\sigma_i^y, P]$. Therefore, the expectation $\langle \hat{P} \rangle$ can be estimated in the product state with the error $\mathcal{O}(\tau^2)$.

Appendix B: Numerical performance benchmarks of the SIA ansatz

In this appendix, we benchmark the performance of the SIA ansatz compared with the HEA ansatz [13], to explore if the SIA ansatz itself can lead to a better performance. Using the same instances as in the main text, Fig. 9, shows the success rate of the SIA introduced in the main text compared to various ansätze that haven been widely used in the literature. In particular, we compare the SIA to HEAs with linear CNOT entanglement, parallel CZ

entanglement and a single-qubit rotation ansatz that produces only product states (see Appendix D, Figs. 12(a) - 12(c) for details). For a fair comparison, all the ansätze were applied to the same initial state, the uniform superposition state $|+\rangle^{\otimes N}$, with the same number of parameters N^2 and all parameters set to zero initially, as detailed in Appendix D. Note that we have not included the warm start for SIA yet. Focusing on the results obtained from the exact simulation in Fig. 9(a), corresponding to a perfect noise-free quantum computer with an infinite number of measurements, the SIA ansatz shows a slightly higher success rate than the other ansätze, in particular for large problem sizes. The HEA ansätze with linear CNOT entangling layers performs only better than the single-qubit rotation ansatz without any entanglement for the problems with less than 18 qubits. From there on they are worse (HEA with linear CNOT entangling layers) or on par (HEA with parallel CZ entangling layers) with the single-qubit rotation ansatz.

Incorporating the fact that actual quantum devices only allow for a finite number of measurements changes the picture significantly, as Fig. 9(b) reveals. Despite using a relatively large number of 10,000 measurements per iteration, the statistical noise leads to a significant drop in performance of the SIA ansätze. In particular, they lose their advantage over the single-qubit rotation ansatz, and except for the HEA with linear CNOT entangling layers, all variational ansätze show very similar performance for the entire range of problem sizes we study. Our results are in agreement with the conclusion of Ref. [17], which found that entanglement gates adapted to the problem structure are beneficial for optimization problems while the advantage will decrease in the case of a finite number of measurements. The poor performance of HEA and SIA at large problem sizes with a finite number of measurements might be the result of statistical errors as well as barren plateaus, which is discussed in Sect. 4. Based on this observation, we conclude that the SIA ansatz alone can not provide practical performance improvement, highlighting the importance of finding a good initial parameter to enhance VQE.

Appendix C: Additional tests for HEA with warm start initialization

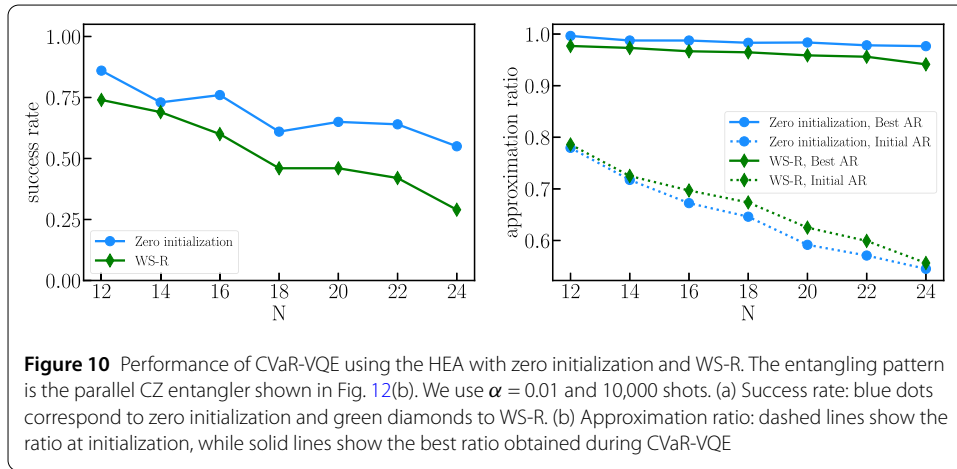
For completeness, we also examined whether a warm start procedure can improve the performance of the HEA. Since HEA does not incorporate any problem structure, identifying an efficient warm start is challenging. Nevertheless, following the approach used in Ref. [26], we constructed a warm start based on the continuous relaxation of the QUBO problem. For each instance, we solve

$$\min_{x \in [0,1]^N} x^T Q x, \quad (\text{C.1})$$

where the binary variables $x_i \in \{0, 1\}$ are relaxed to $x_i \in [0, 1]$. The optimal relaxed solution, denoted by c_i^* , can be efficiently obtained via exact diagonalization of the $N \times N$ matrix Q . We then initialize each qubit in the state $|\phi_i\rangle = R_y(\theta_i)|0\rangle$ with

$$\theta_i = 2 \arcsin\left(\sqrt{c_i^*}\right),$$

such that the probability of measuring $|1\rangle$ equals c_i^* . (As all ansätze begin with a Hadamard gate in our implementation, this is equivalent to applying $R_y(\theta_i - \pi/2)$ after a Hadamard on

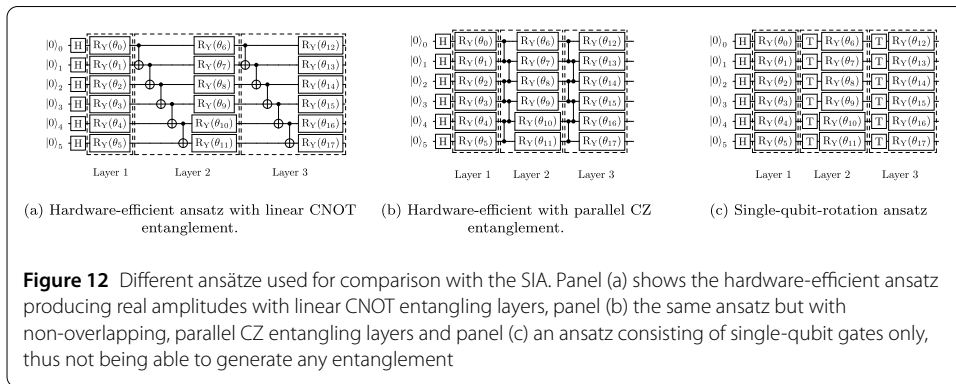
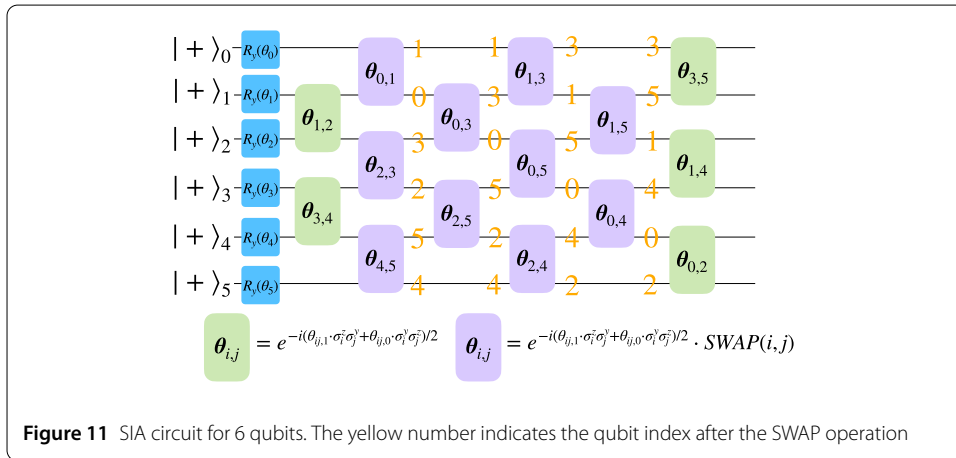


$|0\rangle$.) To avoid extreme cases $c_i^* = 0$ or 1 , we follow Ref. [26] and introduce a regularization parameter $\varepsilon = 0.25$, resulting in the initialization rule

$$\theta_i = \begin{cases} 2 \arcsin(\sqrt{c_i^*}), & c_i^* \in [\varepsilon, 1 - \varepsilon], \\ 2 \arcsin(\sqrt{\varepsilon}), & c_i^* \leq \varepsilon, \\ 2 \arcsin(\sqrt{1 - \varepsilon}), & c_i^* \geq 1 - \varepsilon. \end{cases} \quad (C.2)$$

We use these parameters to initialize the first-layer parameters of the HEA with parallel CZ entanglement (Fig. 12(b)), while all remaining parameters are set to zero. We refer to this method as *warm start by relaxation* (WS-R). Using WS-R as the initialization for the HEA, we performed CVaR-VQE simulations with $\alpha = 0.01$ and 10,000 shots. As shown in Fig. 10(a), the success rate does not improve compared to zero initialization. To confirm that WS-R indeed provides a better starting state, Fig. 10(b) compares the approximation ratio at initialization and the best approximation ratio obtained during optimization. Although WS-R yields a higher initial approximation ratio, it does not improve the subsequent VQE convergence. This may be because the HEA does not reflect the problem structure and thus cannot effectively exploit the warm start state. A similar conclusion was drawn in Ref. [26]: as shown in Fig. 4(b) of that work, the warm start is beneficial only when the mixer of QAOA is adapted to the warm start state; otherwise, using the warm start state with the standard mixer does not improve performance.

To confirm that this behavior is not specific to the parallel CZ entanglement, we also tested HEA with linear CNOT and single-qubit rotation. In both cases, we observed the same qualitative behavior: WS-R improves the initial approximation ratio but does not enhance the overall optimization performance. Since these additional results do not provide further qualitative insights beyond Fig. 10, we omit the corresponding plots from the main text. The findings support our conclusion that a warm start strategy is effective only when the ansatz is aligned with the problem or the initialization state, as is the case for the SIA, but not for problem-agnostic ansatzes such as the HEA.



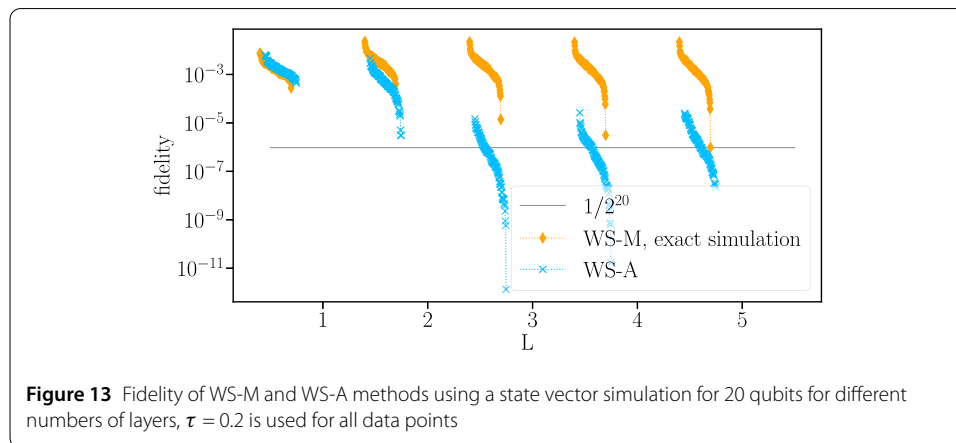
Appendix D: Circuits for different ansätze

The circuit for the SIA-YZ ansatz includes two-qubit gates for all possible qubit pairs. In general, current hardware platforms only offer limited qubit connectivity, and we need additional SWAP gates to realize all-to-all connectivity. Following the strategy in Ref. [33], we take 6 qubits as an example and show the resulting circuit assuming linear qubit connectivity in Fig. 11. Full connectivity is reached after $N - 2$ SWAP layers, and the SWAP gate is compressed with the operator $U_{ij}(\theta_{ij})^{YZ}$ in Eq. (9), thus they can be implemented with 3 CNOT gates (purple box in Fig. 11). Additionally, the first and final layers are only gates $U_{ij}(\theta_{ij})^{YZ} \in SO(4)$ that can be decomposed by using 2 CNOT gates [34] (green box in Fig. 11). Consequently, the overall structure requires $3N - 2$ CNOT layers and $\mathcal{O}(1.5N^2)$ CNOT gates. Alternatively, if the quantum device has all-to-all connectivity, no additional SWAP gates are necessary and the circuit only consists of the gates $U_{ij}(\theta_{ij})^{YZ}$ and single-qubit rotation gates. In this case the circuit depth is reduced to $2N$ CNOT layers and $\mathcal{O}(N^2)$ CNOT gates.

In addition to the SIA, we consider three alternative ansätze often used in the literature for comparison, which are illustrated in Fig. 12. For a fair comparison, all ansätze have the same number of parameters, utilizing N^2 parameters equivalent to those of a single-layer SIA, thereby necessitating N layers for the ansatz depicted in Fig. 12. Moreover, Table 1 shows the number of two-qubit gates required for each ansatz, maintaining the same parameter count as the 1-layer SIA.

Table 1 Number of CNOT gates for different ansätze with N^2 parameters, with respect to linear or all-to-all qubit connectivity of the hardware platform

Qubit connectivity	Ansatz			
	SIA	HEA with linear CNOT	HEA with parallel CZ	Single-qubit rotations
linear	$1.5N^2 - 2.5N + 1$	$(N - 1)^2$	$(N - 1)^2$	0
all-to-all	$N(N - 1)$	$(N - 1)^2$	$(N - 1)^2$	0



Appendix E: Performance of the warm start approach for more than a single layer

This appendix shows results for the performance of the warm start across different number of layers, $L = 1, 2, 3, 4, 5$. Figure 13 shows the data for 20 qubits as an example and compares the fidelity between the WS-M and WS-A. While both approaches yield comparable results for a single layer, the fidelity achieved through the WS-A decreases with an increasing number of layers. Conversely, the WS-M method maintains a relatively consistent performance for all values of L we consider. Notably, the fidelity improves slightly when going from a single layer to $L = 2$, but there is no significant improvement when further adding additional layers. This observation suggests the necessity to incorporate more non-local operators if one wants to enhance the fidelity of the initial state with the exact solution.

Author contributions

Y.C. performed the numerical simulations. All authors contributed to the development of the ideas and to writing the manuscript. All authors reviewed the manuscript.

Funding information

Open Access funding enabled and organized by Projekt DEAL. This work is funded by the European Union’s Horizon Europe Framework Programme (HORIZON) under the ERA Chair scheme with grant agreement no. 101087126 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 429529648 - TRR 306 QuCoLiMa (“Quantum Cooperativity of Light and Matter”). This work is supported with funds from the Ministry of Science, Research and Culture of the State of Brandenburg within the Centre for Quantum Technologies and Applications (CQTA). The funding bodies had no role in the design of the study, data collection and analysis, or writing of the manuscript.

Data availability

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

Declarations

Ethics declarations

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Deutsches Elektronen-Synchrotron DESY, Platanenallee 6, 15738 Zeuthen, Germany. ²Computation-Based Science and Technology Research Center, The Cyprus Institute, 20 Kavafi Street, 2121 Nicosia, Cyprus. ³Institut für Physik, Humboldt-Universität zu Berlin, Newtonstr. 15, 12489 Berlin, Germany. ⁴Institute for Quantum Computing Analytics (PGI-12) Jülich Research Center, Wilhelm-Johnen-Str., 52428 Jülich, Germany.

Received: 7 September 2025 Accepted: 3 December 2025 Published online: 12 December 2025

References

1. Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love PJ, Aspuru-Guzik A, O'Brien JL. A variational eigenvalue solver on a photonic quantum processor. *Nat Commun.* 2014;5:1–7.
2. Cerezo M, Arrasmith A, Babbush R, Benjamin SC, Endo S, Fujii K, McClean JR, Mitarai K, Yuan X, Cincio L, et al. Variational quantum algorithms. *Nat Rev Phys.* 2021;3:625–44.
3. Di Meglio A, Jansen K, Tavernelli I, Alexandrou C, Arunachalam S, Bauer CW, Borrás K, Carrazza S, Crippa A, Croft V, de Putter R, Delgado A, Dunjko V, Egger DJ, Fernandez-Combarro E, Fuchs E, Funcke L, Gonzalez-Cuadra D, Grossi M, Halimeh JC, Holmes Z, Kühn S, Lacroix D, Lewis R, Lucchesi D, Martinez ML, Meloni F, Mezzacapo A, Montangero S, Nagano L, Radescu V, Ortega ER, Roggero A, Schuhmacher J, Seixas J, Silvi P, Spentzouris P, Tacchino F, Temme K, Terashi K, Tura J, Tuysuz C, Vallecorsa S, Wiese U-J, Yoo S, Zhang J. Quantum computing for high-energy physics: state of the art and challenges. Summary of the qc4hep working group. 2023. [arXiv:2307.03236](https://arxiv.org/abs/2307.03236) [quant-ph].
4. Tilly J, Chen H, Cao S, Picozzi D, Setia K, Li Y, Grant E, Wossnig L, Rungger I, Booth GH, Tennyson J. The variational quantum eigensolver: a review of methods and best practices. *Phys Rep.* 2022;986:1–128.
5. Chai Y, Funcke L, Hartung T, Jansen K, Kühn S, Stornati P, Stollenwerk T. Optimal flight-gate assignment on a digital quantum computer. *Phys Rev Appl.* 2023;20:064025.
6. Chai Y, Epifanovsky E, Jansen K, Kaushik A, Kühn S. Simulating the flight gate assignment problem on a trapped ion quantum computer. 2023. [arXiv:2309.09686](https://arxiv.org/abs/2309.09686) [quant-ph].
7. Schwägerl T, Issever C, Jansen K, Khoo TJ, Kühn S, Tüysüz C, Weber H. Particle track reconstruction with noisy intermediate-scale quantum computers. 2023. [arXiv:2303.13249](https://arxiv.org/abs/2303.13249) [quant-ph].
8. Amaro D, Rosenkranz M, Fitzpatrick N, Hirano K, Fiorentini M. A case study of variational quantum algorithms for a job shop scheduling problem. *EPJ Quantum Technol.* 2022;9. <https://doi.org/10.1140/epjqt/s40507-022-00123-4>.
9. Nannicini G. Performance of hybrid quantum-classical variational heuristics for combinatorial optimization. *Phys Rev E.* 2019;99:013304.
10. Scriver G, Astrakhantsev N, Pilati S, Mazzola G. Challenges of variational quantum optimization with measurement shot noise. *Phys Rev A.* 2024;109. <https://doi.org/10.1103/physreva.109.032408>.
11. McClean JR, Boixo S, Smelyanskiy VN, Babbush R, Neven H. Barren plateaus in quantum neural network training landscapes. *Nat Commun.* 2018;9. <https://doi.org/10.1038/s41467-018-07090-4>.
12. Kochenberger G, Hao J-K, Glover F, Lewis M, Lü Z, Wang H, Wang Y. The unconstrained binary quadratic programming problem: a survey. *J Comb Optim.* 2014;28:58–81.
13. Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow JM, Gambetta JM. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature.* 2017;549:242–6.
14. Barkoutsos PKI, Nannicini G, Robert A, Tavernelli I, Woerner S. Improving variational quantum optimization using cvar. *Quantum.* 2020;4:256.
15. Hadfield S. On the representation of Boolean and real functions as Hamiltonians for quantum computing. *ACM Trans Quantum Comput.* 2021;2:18:1–18:21.
16. Acerbi C, Tasche D. On the coherence of expected shortfall. *J Bank Finance.* 2002;26:1487–503.
17. Diez-Valle P, Porras D, García-Ripoll JJ. Quantum variational optimization: the role of entanglement and problem hardness. *Phys Rev A.* 2021;104:062426.
18. Hartung T, Jansen K. Convergence and efficiency proof of quantum imaginary time evolution for bounded order systems. 2025. [arXiv:2506.03014](https://arxiv.org/abs/2506.03014) [quant-ph].
19. Motta M, Sun C, Tan ATK, O'Rourke MJ, Ye E, Minnich AJ, Brandão FGSL, Chan GK-L. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat Phys.* 2019;16:205–10.
20. McArdle S, Jones T, Endo S, Li Y, Benjamin SC, Yuan X. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quantum Inf.* 2019;5.
21. Sun S-N, Motta M, Tazhigulov RN, Tan ATK, Chan GK-L, Minnich AJ. Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution. *PRX Quantum.* 2021;2:010317.
22. Yuan X, Endo S, Zhao Q, Li Y, Benjamin SC. Theory of variational quantum simulation. *Quantum.* 2019;3:191.
23. Huo M, Li Y. Error-resilient Monte Carlo quantum simulation of imaginary time. *Quantum.* 2023;7:916.
24. Han M, Park H, Choi S. Quantum Zeno Monte Carlo for computing observables. *npj Quantum Inf.* 2025;11:46. [arXiv:2403.02763](https://arxiv.org/abs/2403.02763) [quant-ph].
25. Powell MJ. A direct search optimization method that models the objective and constraint functions by linear interpolation. In: *Advances in optimization and numerical analysis*. Berlin: Springer; 1994. p. 51–67.
26. Egger DJ, Mareček J, Woerner S. Warm-starting quantum optimization. *Quantum.* 2021;5:479.
27. Arrasmith A, Holmes Z, Cerezo M, Coles PJ. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Sci Technol.* 2022;7:045015.
28. Endo S, Benjamin SC, Li Y. Practical quantum error mitigation for near-future applications. *Phys Rev X.* 2018;8:031027.
29. Cai Z, Xu X, Benjamin SC. Mitigating coherent noise using Pauli conjugation. *npj Quantum Inf.* 2020;6:17.
30. Funcke L, Hartung T, Jansen K, Kühn S, Stornati P, Wang X. Measurement error mitigation in quantum computers through classical bit-flip correction. *Phys Rev A.* 2022;105:062404.
31. van den Berg E, Minev ZK, Temme K. Model-free readout-error mitigation for quantum expectation values. *Phys Rev A.* 2022;105:032620.

32. Viola L, Lloyd S. Dynamical suppression of decoherence in two-state quantum systems. *Phys Rev A*. 1998;58:2733–44.
33. Weidenfeller J, Valor LC, Gacon J, Tornow C, Bello L, Woerner S, Egger DJ. Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware. *Quantum*. 2022;6:870.
34. Vatan F, Williams C. Optimal quantum circuits for general two-qubit gates. *Phys Rev A*. 2004;69. <https://doi.org/10.1103/physreva.69.032315>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.