Quantum variational optimization: The role of entanglement and problem hardness

Pablo Díez-Valle,[*] Diego Porras, and Juan José García-Ripoll

Instituto de Física Fundamental IFF-CSIC, Calle Serrano 113b, Madrid 28006, Spain

Quantum variational optimization has been posed as an alternative to solve optimization problems faster and at a larger scale than what classical methods allow. In this paper we study systematically the role of entanglement, the structure of the variational quantum circuit, and the structure of the optimization problem, in the success and efficiency of these algorithms. For this purpose, our study focuses on the variational quantum eigensolver (VQE) algorithm, as applied to quadratic unconstrained binary optimization (QUBO) problems on random graphs with tunable density. Our numerical results indicate an advantage in adapting the distribution of entangling gates to the problem's topology, specially for problems defined on low-dimensional graphs. Furthermore, we find evidence that applying conditional value at risk type cost functions improves the optimization, increasing the probability of overlap with the optimal solutions. However, these techniques also improve the performance of *Ansätze* based on product states (no entanglement), suggesting that a new classical optimization method based on these could outperform existing NISQ architectures in certain regimes. Finally, our study also reveals a correlation between the hardness of a problem and the Hamming distance between the ground- and first-excited state, an idea that can be used to engineer benchmarks and understand the performance bottlenecks of optimization methods.

I. INTRODUCTION

In the last years, quantum variational optimization [1, 2] has emerged as a framework that exploits the computational power of near-term intermediate scale quantum circuits [3]. In this framework, a parametrized quantum circuit is sequentially tuned to maximize or minimize a given cost function, which is computed as a function of measured single- or multiqubit observables. While it is clear that the variational approach has broad applications in the study of complex quantum systems [4, 5], such as in quantum chemistry, there is also ongoing research on its application to NP-complete and NP-hard combinatorial optimization problems [6–18], such as those represented by quadratic unconstrained binary optimization (QUBO) formulas.

In this work, we address several open questions that are open in the context of quantum discrete optimization problems. One key question is the choice of the quantum variational *Ansatz*: this is the circuit that we tune to find the optimal classical solution. This circuit should efficiently sample the Hilbert space, but it also should be as simple as possible, with the minimal number of entangling operations and quantum gates, to allow their use in NISQ architectures [19, 20]. Here, it is crucial to understand the role of entanglement and correlations, and the relation of the entanglement structure and the problem to optimize [21]. These questions also tie intimately to other aspects, such as the definition of the cost function [22, 23], and the hybrid quantum-classical algorithms which we use to optimize the variational state [24–31].

In our research we address these questions with a focus on the variational quantum eigensolver (VQE) method

[32]. We explore the performance of the VQE for a family of QUBO problems defined on random graphs with tunable density. We explore these problems by using variational *Ansätze* with an entanglement structure that either mimics the original graph, or uses a random or predefined architecture. Finally, we consider both a trivial cost function, formed by the average of the QUBO Hamiltonian, or we use the average over a fixed percentage of the lowest sampled configurations, in the spirit of the conditional Value at risk (CVaR) method [23]. All these problems are analyzed with a broad variety of classical optimization strategies (gradient descent, constrained optimization by linear approximation, simultaneous perturbation stochastic approximation, etc.).

Our first result is to confirm the importance of stochastic optimizers when working with an estimation of the cost function from measurements. Moreover, we conclude that in this realistic scenario, it is best to use the CVaR cost function [23], rather than the average energy. The former leads to wave functions that have a greater overlap probability with the actual ground state, while the latter produces binomial distributions that can fail unconditionally, where repetition does not help. Our study also shows that for problems with a structure—i.e., graphs with low or intermediate densities—it is computationally advantageous to imitate the structure of the problem with the entanglement structure. This type of advantage vanishes when facing dense problems, such as those that are found in quantum finance scenarios. Interestingly, we find that the depth of the variational circuits is only moderately relevant. The success probability saturates for relatively shallow circuits, while the evaluation cost increases with size. This means that, for some problems, repeating the same optimization with product or weakly entangled states, can be more efficient than using more entanglement. In fact, we find that a product-state *Ansatz* combined with the CVaR cost function yields good results that point to new quantum-inspired classi-

* pablo.diez@csic.es

cal optimization methods. Finally, we characterized the hardness of our problems, correlating the success probability to the Hamming distance between the ground state and the first-excited state. Intuitively, this confirms the idea that for large distances, the optimizer gets trapped in low-energy solutions that are macroscopically different from the true ground state. We expect that this will be a useful indication or technique to engineer hard problems, as well as a hint that we need to engineer quantum algorithms that enhance the probability of macroscopic tunneling between solutions.

This paper is structured as follows: In Sec. II we describe the combinatorial optimization problems that we focus on in this work, QUBO problems, and its straightforward mapping as a Hamiltonian in order to be solved by a variational hybrid method. In Sec. III we introduce the details of the resolution methods. Later, in Sec. IV we present our numerical results obtained from ideal simulations with exact state calculation and finite sampling. These results highlight the improvement introduced by a CVaR style cost function (Sec. IV A), they show the impact of the structure of the variational form (Sec. IV B), they give us intuition into how the problem characteristics can affect the performance of the algorithm (Sec. IV C), and they compare the efficiency of *Ansätze* based on entangled states with that based on product states (Sec. IV D). Finally, we conclude by summarizing the main ideas of the paper in Sec. V.

II. QUADRATIC UNCONSTRAINED BINARY OPTIMIZATION PROBLEMS ON RANDOM GRAPHS

Combinatorial optimization is one of the most common areas studied in NISQ quantum computing and quantum annealing. A broad spectrum of combinatorial optimization problems can be represented as QUBO problems, with a cost function E that involves products of two Boolean variables:

$$E(\vec{x}) = \sum_{i,j} x_i Q_{ij} x_j \; ; \quad x_i \in \{0,1\}. \tag{1}$$

Here $\vec{x}$ is an N-vector of binary variables, and Q is an N-by-N square symmetric matrix of coefficients. QUBO problems have been extensively studied in the literature [33, 34]; for instance, solving optimization problems on graphs, resources allocation problems, clustering problems, satisfiability, sequencing and ordering problems, facility locations problems, and various forms of assignment problems.

Thanks to the matrix structure, QUBO problems can be regarded as the representation of an undirected graph with N vertices, connected by undirected edges $i \leftrightarrow j$ that have associated weights $Q_{ij} = Q_{ji}$. A classical solution to the QUBO problem corresponds to labeling each vertex with 0 or 1, so as to maximize or minimize the cost function E [see Fig. 1(b)]. In this work we classify

QUBO problems according to the size of the corresponding graph, given by the number of vertices N, and the *density of the graph,* defined as the ratio

$$D = \frac{2E}{N(N-1)}, \tag{2}$$

between the number of edges E and the maximum number of potential connections $N(N-1)/2$.

A QUBO problem can always be mapped to an Ising spin model. Thereby the optimization of the cost function E, Eq. (1), becomes a search of the minimum-energy state of an Ising Hamiltonian for an N-qubit system where each qubit represents one vertex of the graph.The transformation from a QUBO to an Ising formulation is implemented through the change of variables $x_i \rightarrow \frac{1}{2}(1 + \sigma_i^z)$, where σ_i^z is the Pauli Z matrix acting on the i-th qubit. We can express the previous function as

$$f(\sigma_i^z) = -\frac{1}{2}H + \frac{1}{4}\sum_{i,j} Q_{ij},$$

$$\text{with } H = \sum_{i<j} J_{ij}\sigma_i^z\sigma_j^z + \sum_i h_i\sigma_i^z, \tag{3}$$

using the coupling matrix $J_{ij} = -Q_{ij}$ and the magnetic fields $h_i = -\sum_j Q_{ij}$. In this paper we only consider frustrated systems, where the matrices J and Q combine both positive and negative coefficients, and the optimization is potentially harder.

III. METHODOLOGY

A. Hybrid quantum-classical algorithms

Hybrid quantum-classical algorithms are those that combine the use of quantum and classical computers to solve a computational problem. In quantum variational algorithms [2, 35], the quantum computer is used to prepare and characterize a complex wave function, defined in terms of parametrized quantum operations

$$|\Psi(\vec{\theta})\rangle = U_L(\theta_L)...U_1(\theta_1)|\Psi_0\rangle. \tag{4}$$

The real values θ_i determine how this state is to be constructed in a quantum computer by composing unitary operations $U_i(\theta_i)$. The parameters of this wave function are tuned according to some classical algorithms, until the wave function maximizes or minimizes a measure of "fitness" or cost function. In this work the fitness function will be a QUBO problem, and we will consider two hybrid algorithms: the original VQE and its modification based on CVaR measures.

1. Variational quantum eigensolver algorithm.

The variational quantum eigensolver, or VQE [32], is a hybrid quantum-classical algorithm that seeks the min-

imal eigenvalue $\lambda_{\min}$ and corresponding eigenstate of a problem Hamiltonian H, such as the QUBO models described above (3). The algorithm relies on a parametrized family of wave functions (4). The variational principle states that the average energy of this state is strictly larger or equal to the energy of the ground state we seek:

$$\lambda_{\min} \leq E(\vec{\theta}) = \langle \Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle \,. \qquad (5)$$

Thus, if our parametrization (4) is dense in the Hilbert space, we can obtain a very good approximation to the ground state by solving the problem

$$\vec{\theta}_{\mathrm{opt}} = \operatorname{argmin}_{\vec{\theta}} E(\vec{\theta}). \qquad (6)$$

VQE solves this problem by iteratively preparing the wave function $|\Psi(\vec{\theta}^{(k)})\rangle$ and measuring its energy $E(\vec{\theta}^{(k)})$ in the quantum computer. The VQE relies on a classical optimization algorithm—gradient descent, SPSA, Adam, etc—to build a progressively optimized sequence of parameters $\{\vec{\theta}^{(k)}\}$ that classically aims to minimize our black-box functional $E(\vec{\theta})$.

VQE is particularly well suited to address QUBO problems, where the energy functional (3) is written as a sequence of mutually commuting observables. In this case, to solve a QUBO problem with N variables, the algorithm relies on a quantum computer with N qubits that are prepared in some trial state $|\Psi(\vec{\theta})\rangle$. The energy of this state is approximated by simultaneously measuring all qubits in the Pauli z basis. This produces a string of N values $\{s_i \in \{-1, 1\}\}$, which can be translated to a bit string $\{x_i\}$ and a QUBO cost function (1) that we label E_n. By averaging these values over K repetitions, we obtain a stochastic approximation

$$\bar{E}_K(\vec{\theta}) := \frac{1}{K} \sum_{k=1}^{K} E_k \to E(\vec{\theta}), \; K \to \infty \qquad (7)$$

that is iteratively optimized using the classical methods mentioned before.

2. Conditional value at risk-variational quantum eigensolver

A problem with the VQE is that it regards all outcomes of the measurement process with equal importance, even outlier states that may have an energy far away from our goal. If we call $P(E; \vec{\theta})$ the distribution of energies associated with the states we build, the VQE cost function is

$$\bar{E}_K(\vec{\theta}) \simeq \mathbb{E}[E(\vec{\theta})] = \langle \Psi(\vec{\theta})|H|\Psi(\vec{\theta})\rangle = \int_{-\infty}^{\infty} P(E; \vec{\theta})EdE. \qquad (8)$$

Recently, Barkoutsos et al. [23] proposed a variation of VQE that replaces the cost function with a new estimator that is inspired by the CVaR or the *expected shortfall*

definition applied in finance [36]. This new estimator only works with a fraction $\rho \in (0,1)$ of the bit-strings that the quantum computer produces, selecting them to be those that have the lowest energy. If we sort all sampled energies $\{E_1 \leq E_2 \leq \ldots \leq E_K\}$, and we take the fraction ρ of lowest energy outcomes,

$$\mathrm{CVaR}_\rho(E(\vec{\theta})) \approx \frac{1}{\lfloor \rho K \rfloor} \sum_{k=1}^{\lfloor \rho K \rfloor} E_k. \qquad (9)$$

In the limit of many measurements, the CVaR-VQE estimator approximates the cost function

$$\mathrm{CVaR}_\rho(E(\vec{\theta})) = \mathbb{E}[E(\vec{\theta})|E \leq F_E^{-1}(\rho)] =$$
$$= \frac{1}{\rho} \int_{-\infty}^{E_\rho(\vec{\theta})} P(E; \vec{\theta})EdE \;, \qquad (10)$$

where F_E represents the cumulative density function of E and $E_\rho(\vec{\theta})$ is computed as

$$\int_{-\infty}^{E_\rho(\vec{\theta})} P(E; \vec{\theta})dE = \rho \,. \qquad (11)$$

The intuition behind the idea is that improving the properties of the best measurement outcomes may be more efficient than improving the average of all outcomes, by concentrating the probability of the wave function in the lowest-energy states.

B. Quantum variational wave functions

A crucial ingredient in all variational algorithms is the choice of parametrized wave function or *Ansatz* (4). Typically, these *Ansätze* are physically motivated and combine local gates with entangling operations that correlate the qubits in a hardware efficient manner. More generally, the choice of gates and the topology of the entangling operations can affect the efficiency of the outcome. In this study we focus on four types of variational states: three types wave functions that are created by the action of control Z gates, and also a family of cheap product states where no correlation is required.

1. Entangled states

We start by considering a set of entangled variational wave functions as originally proposed for the VQE method [2],

$$|\Psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle \,, \qquad (12)$$

with $U(\vec{\theta}) = \prod_{l=1}^{L} \left[\prod_{n=1}^{N} e^{i\theta_{nl}\sigma_n^y} U_{\mathrm{ent}} \right] \prod_{n=1}^{N} e^{i\theta_{n0}\sigma_n^y}$.

This wave function combines L layers of entangling gates U_{ent} with real-valued rotations generated by the qubits' Pauli y matrices σ_n^y. The $N \times (L+1)$ angles of
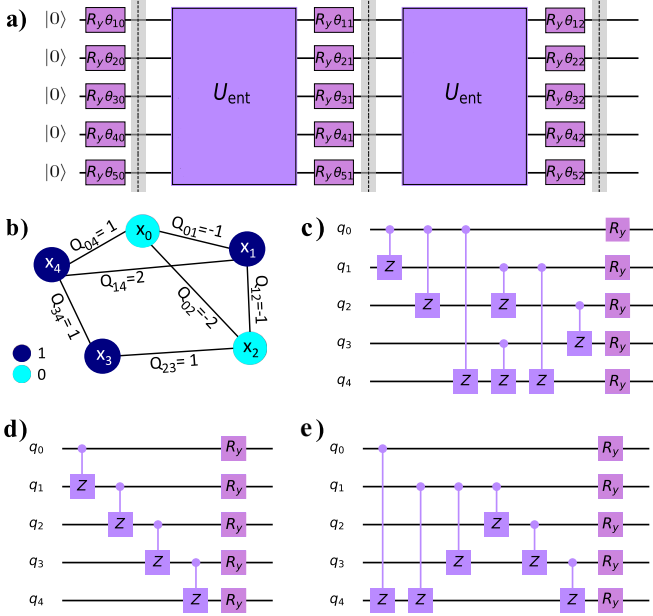
FIG. 1. (a) Variational form based on single-qubit σ^y rotations for five qubits with two layers and fifteen variational parameters. (b) Example of a particular optimized graph. (c) Entangling layer compatible with the graph structure. (d) Entangling layer with a linear entanglement. (e) Entangling layer with a random entanglement.

these rotations θ_{nl} are the variational parameters of the wave function (cf. Fig. 1). The entangling unitary, U_{ent}, determines the correlation structure of the wave function. We will test three such structures, built from two-qubit control Z gates:

(i) Linear entanglement: The entangling gates are independent of the structure of the QUBO problem. Every entangling layer is made up of two-qubit control Z gates between each qubit with its nearest-neighbor qubit in a linear quantum processor topology [see Fig. 1(d)].

$$ U_{\text{ent}} = \prod_i e^{i \frac{\pi}{4} \left(\mathbb{I} - \sigma_i^z \right) \left(\mathbb{I} - \sigma_{i+1}^z \right)} \ , \tag{13} $$

where σ^z is the Pauli Z matrix acting on the ith qubit.

(ii) Compatible entanglement: Entangling gates are chosen to reflect the structure of the QUBO problem. Every entangling layer is made up of two-qubit control Z gates between each qubit with nearest neighbors in the QUBO graph [see Fig. 1(c)].

$$ U_{\text{ent}} = \prod_{\langle i,j \rangle} e^{i \frac{\pi}{4} \left(\mathbb{I} - \sigma_i^z \right) \left(\mathbb{I} - \sigma_j^z \right)} \ , \tag{14} $$

where $\langle i,j \rangle$ denote that i and j are nearest neighbors, i.e., $Q_{ij} \neq 0$ with Q being the QUBO matrix.

(iii) Random entanglement: Every entangling layer is made up of two-qubit control Z gates between random qubits. The number of entangled qubits is given by the number of nonzero weights in the QUBO graph, i.e., it is not a fully random entanglement but it is dependent on the density of the problem [see Fig. 1(e)].

$$ U_{\text{ent}} = \prod_{i,j} e^{i \frac{\pi}{4} \left(\mathbb{I} - \sigma_i^z \right) \left(\mathbb{I} - \sigma_j^z \right)} \ , \tag{15} $$

where $i \neq j$ are E random pairs.

In our variational algorithms we typically initialize the VQE to a uniform superposition of all states in the computational basis, using Hadamard-like rotations $\theta_{n0} = \pi/4$, and making all remaining layers weak perturbations over this state, $\theta_{nl} \approx 10^{-2}$ for $l \geq 1$.

2. Product state method

In this work we also explore fully separable variational states, built from σ^y rotations of the N qubits

$$ |\Psi(\vec{\theta})\rangle = \bigotimes_{j=1}^{N} |\psi(\theta_j)\rangle \,, \tag{16} $$

with $|\psi(\theta_j)\rangle = \cos\theta_j |0\rangle + \sin\theta_j |1\rangle$.

The state (16) can be understood as a borderline case of the variational form (12) exposed in the previous section in which the number of layers L is 0. The elimination of the entangling gates makes the state hardware efficient and more accurate, but it also makes the whole variational procedure efficiently simulable in a classical way. We can regard this family of states as a new type of classical algorithms, much like the spin-vector quantum Monte Carlo methods that have been used to benchmark quantum annealers [37, 38].

C. Efficiency indicators

To study the efficiency of the algorithms we use the algorithm's runtime and the success as indicators. The algorithm's speed is quantified by the number of the cost function evaluations needed to converge to an optimal solution (global or local). Note that this is also the number of queries made to the quantum processor, the most time-consuming step in the whole process.

The algorithm's success is quantified by the probability that the final wave function $|\Psi_{\text{out}}\rangle$ has a significant overlap with the target state $|\Psi_{\text{solution}}\rangle$. This target state is previously computed by a classical method. We deem the algorithm successful $S = 1$ whenever this probability exceeds a fixed cut-off β:

$$ S \equiv \begin{cases} 0 & \text{if } |\langle\Psi_{\text{out}}|\Psi_{\text{solution}}\rangle|^2 < \beta \ , \\ 1 & \text{if } |\langle\Psi_{\text{out}}|\Psi_{\text{solution}}\rangle|^2 \geq \beta \ . \end{cases} \tag{17} $$
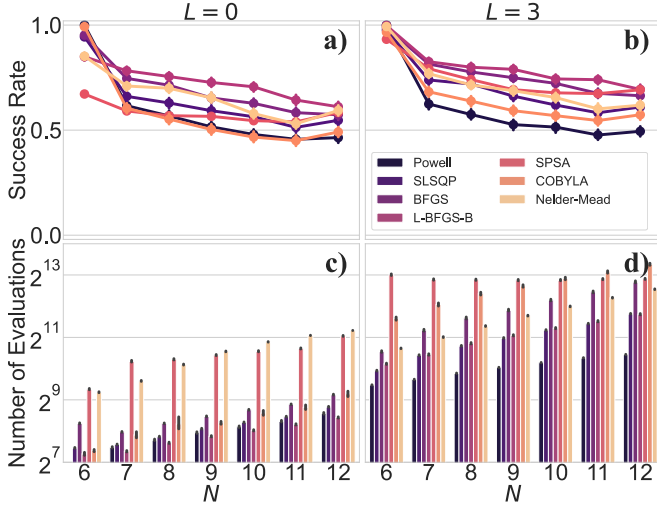
FIG. 2. Comparison of VQE performance with a variety of classical optimization methods to optimize the variational parameters using exact quantum states resulting from simulation. From left to right we increase the number of layers L of the *Ansatz*. On the x axis we plot the size of the problem: (a),(b) success rate, (c),(d) objective function evaluations needed to converge. The results show the average of 1600 instances and a 95% confidence interval.
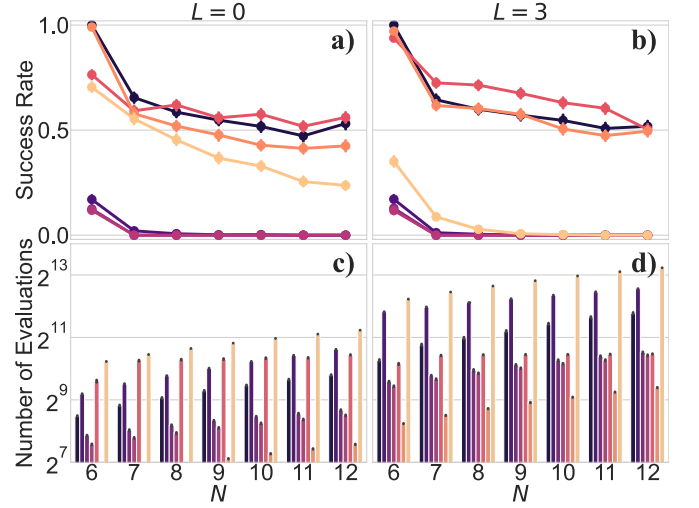

FIG. 3. Comparison of VQE performance with a variety of classical optimization methods to optimize the variational parameters using approximate quantum states computed by a finite set of measurements on the qubits resulting from simulation. From left to right we increase the number of layers L of the *Ansatz*. On the x axis we plot the size of the problem. (a),(b) success rate, (c),(d) objective function evaluations needed to converge. The results show the average of 1600 instances and a 95% confidence interval.

Let us point out that the algorithm could converge to a local minimum of the cost function quite different from $|\Psi_{solution}\rangle$, since the classical optimizers use standards convergence criteria based on the change of the cost function value in each optimization step. Therefore the convergence of the algorithm does not guarantee $S = 1$. Given a set of problems, we call the *success rate* the average of S over all such instances. The value of β is somewhat arbitrary, but once fixed it lower bounds the probability of obtaining the right solution by measuring over the final state k times

$$P(\Psi_{\text{solution}}|S = 1) \geq 1 - (1 - \beta)^k. \qquad (18)$$

In our simulation we set $\beta = 0.1$ so that the success of the algorithm implies we obtain the right state with 99.997% probability after $k = 100$ measurements of the variational state.

D. Classical benchmark of the algorithms

Actual variational quantum algorithms are expected to run on real quantum hardware. However, the goal of this work is to perform a neutral benchmark focused on the influence of variational states and cost functions in the algorithm, leaving aside experimental imperfections, such as noise or disconnected qubits. For this reason, we conduct our benchmarks by using a classical simulator of an ideal quantum computer. All simulations as well as the optimization of the variational parameters were performed on nodes with two Intel Sandybridge E5-2670

processors, each one with eight cores operating at 2.6 GHz and a cache of 20MB and 64 GB of RAM memory (i.e., 4 GB/core) running Centos 7.4.

Another important ingredient in these simulations is whether we consider a finite number of measurements— i.e., K is finite in Eqs. (7) and (9)—, or whether we study the limit of infinite measurements, where the quantum uncertainty is eliminated. In this work we begin with the second set of results, which are obtained by simulating the complete wave function and evaluating the exact expectation values $E(\vec{\theta})$ or $\text{CVaR}_\rho(E)$. These results are later on complemented by stochastic simulations that emulate a finite number of measurements and the randomness of a quantum computer.

E. Choice of classical optimizer

In our study of the variational method we have tested seven different optimizers: modified Powell's conjugate direction method [39], sequential least squares programming (SLSQP) [40], Broyden–Fletcher–Goldfarb–Shanno (BFGS) [41–44] and its limited-memory version (L-BFGS-B) [45], the simultaneous perturbation stochastic approximation (SPSA) [46], the constrained optimization by linear approximation [47] and the Nelder-Mead algorithm [48]. We use the optimization methods straightforwardly provided by Scipy 1.3.2 [49], except for SPSA, which is a homemade implementation.

Before our full study of the variational algorithms, we benchmarked these optimizers in a limited set of QUBO

problems. Figure 2 shows the success rates and number of evaluations of these methods vs the number of qubits. The results have been averaged over 1600 randomly generated problems, optimized in the limit of infinite number of measurements—i.e., wave function simulations—. Figure 3 shows similar results, but emulating 9000 measurements of a quantum computer per evaluation.

These figures show a stark contrast in the actual performance of the optimizer. Gradient-free optimizers such as SPSA, COBYLA, Powell, and Nelder-Mead perform well even when the information of the objective function is not complete and its computation presents some stochasticity. Gradient-based optimizers such as SLSQP, BFGS, and L-BFGS-B perform very well in wave function simulations, even outperforming the other methods. However, these algorithms fail when the cost function is evaluated with some uncertainty, becoming trapped in local minima due to an imperfect estimate of the optimal descent direction. Consequently, in the remainder of this work we use L-BFGS-B when working with wave function simulations, while we use SPSA in scenarios with a finite number of measurements.

IV. RESULTS AND DISCUSSION

Let us now show the main results we obtained by benchmarking the VQE and CVaR-VQE algorithms on random instances of QUBO problems. These problems were generated by creating regular and unstructured random graphs with the Python package *networkx* [50], and assigning to the edges of the graph a uniformly sampled random integer weight on the $[-10, 10]$ interval. We performed the experiments for various numbers of qubits N and graph densities D, generating $N_{ins} = 1600$ instances of each configuration. A different set of 1600 instances was used in each experiment. The final results are shown in Figs. 2-10 with 95% confidence intervals, which are obtained by resampling the output data with the bootstrap method [51, 52].

A. CVaR-VQE vs. standard VQE performance

First, we present a comparison between the performance of standard VQE and CVaR-VQE (with $\rho = 10\%$), using an *Ansatz* with three entangling layers ($L = 3$) and a product state wave function.

Final state wave function: CVaR-VQE enhances success probability.- Figure 4 shows the overlap between the exact optimized wave function and the actual solution of QUBO problems with $N = 12$ qubits. This plot illustrates a qualitatively different behavior in both algorithms. On the one hand, the standard VQE algorithm converges almost always to a classical state, which either coincides with the global minimum or is completely orthogonal to it. The only exception occurs for the product state *Ansatz* in problems with small con-
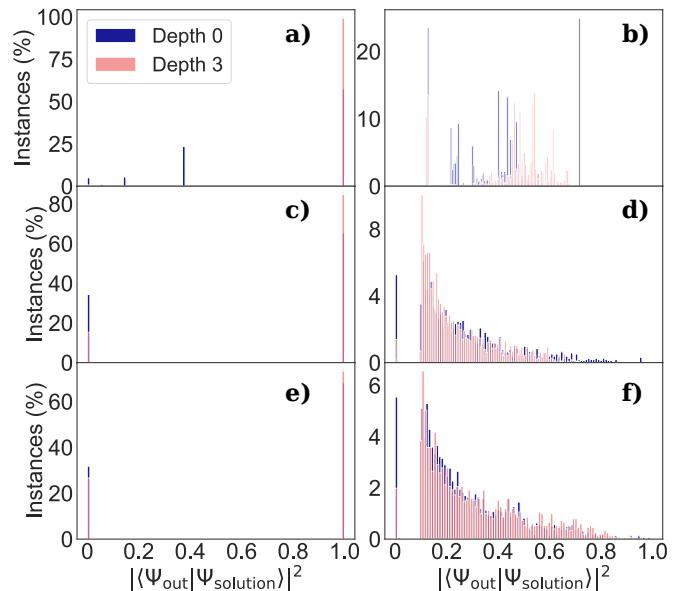

FIG. 4. Distribution of the overlap of the final state $|\Psi_{out}\rangle$ resulting from the convergence of the algorithm with the exact ground state $|\Psi_{solution}\rangle$ computed classically. We plot the percentages obtained from $N_{ins} = 1600$ instances for an *Ansatz* with $L = 0$ and $L = 3$ entangling layers, $N = 12$ qubits, and using L-BFGS-B as classical optimizer. We show on the left the results from VQE standard algorithm, and on the right the results from CVaR-VQE with $\rho = 10\%$. From top to bottom, the density is (a), (b) 0.045; (c), (d) 0.258; and (e), (f) 0.894. In this paper we consider successful trials those whose final overlap is at least 0.1.

nectivity $D \ll 1$. The CVaR-VQE method, on the other hand, converges to quantum superpositions of classical configurations. This makes the CVaR-VQE wave function have a smaller overlap with the global minimum on average, but it greatly increases the probability of recovering the optimal solution by repeated measurements. In contrast, when the VQE algorithm fails, we have to restart the optimization process again from a different initial condition. The intuition behind this result is that any local minimum of the VQE cost function (7) is likely to be a classical state, while any state with an overlap x with the ground state of (7) such that $\rho < x$ is a global minimum of the CVaR-VQE cost function (9).

Success rate and speed: CVaR-VQE is more efficient than standard VQE.- In Fig. 5 we compare the performance of standard and CVaR-VQE by looking at the success rate, defined by the criterion of Sec. III C, and the number of evaluations. We investigate QUBO problems with different sizes and densities. In Figs. 5(a) and 5(b), we show results with a product state *Ansatz* ($L = 0$), whereas Figs. 5(c) and 5(d) presents results with $L = 3$ and a linear entanglement pattern. In both figures we observe that CVaR-VQE clearly outperforms standard VQE, both in terms of success rate and speed, for any choice of problem size, density, and entanglement structure. Based on these results, from now on the
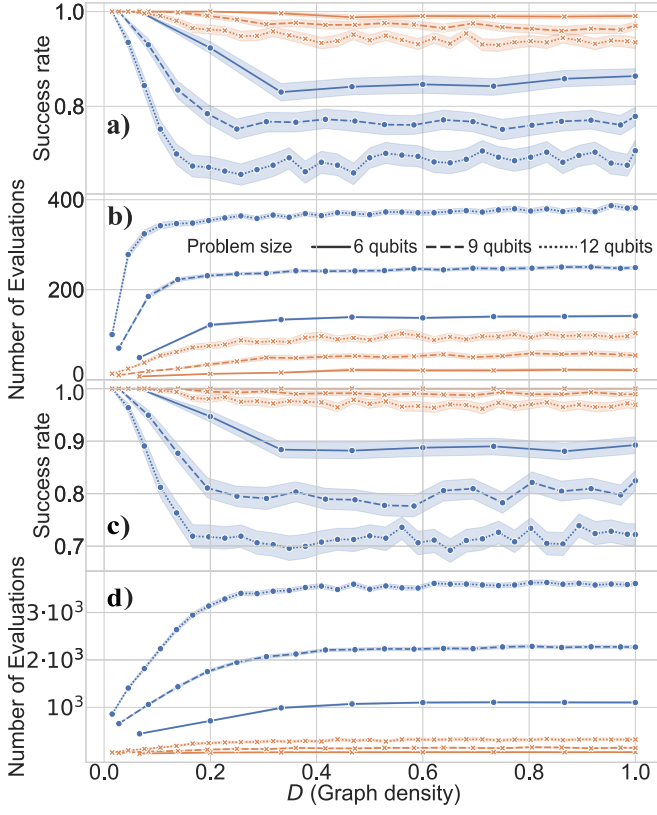
FIG. 5. Idealized performance of standard VQE (blue circles) and CVaR-VQE with $\rho = 10\%$ (orange X symbols), as a function of graph density D for a classical simulation with the full wave function ($N_{\text{shots}} \to \infty$). We compare (a), (b) a product state with (c), (d) an entangled state with $L = 3$ layers for problems with different $N = 6$, 9 and 12 qubits. We plot (a), (c) the success rate and (b), (d) the number of function evaluations vs the density of the underlying graph. The lines show the average over $N_{\text{ins}} = 1600$ problem instances, surrounded by a colored 95% confidence interval.

whole study will be carried out considering exclusively the CVaR-VQE algorithm.

B. Entangling layer structure

Let us now study the influence of correlations in the variational algorithm. Our goal is to understand whether an entanglement structure that imitates the topology of the QUBO problem produces significant advantages. In this section we also introduce the use of a finite number of measurements, exploring how stochasticity affects the different entanglement patterns.

Shallow entanglement patterns.- Figure 6 illustrates the performance of CVaR-VQE using one or no entanglement layers of the three types mentioned in Sec. III B. We computed the success rate and function evaluations over 1600 instances of QUBO problems with $N = 12$ qubits, for different values of the associated graph density D. We observe that, for small graph den-

sities, a compatible entanglement has a marginal advantage in the success rate and convergence speed. However, as the density of the QUBO graph increases this advantage disappears, and the success rate is matched or outperformed by the classical product state and the linear entanglement wave functions, at slightly lower numbers of function evaluations.

Efficiency as a function of circuit depth.- We have repeated the previous study with increasing number of layers. Figure 7 illustrates the success rate and number of evaluations, for graphs with intermediate density ($D = 0.258$, left column) and dense graphs ($D = 0.894$, right column). We observe that in both cases the convergence of the algorithm slows down with the number of layers, proportionally to the growth in the number of parameters. Despite this cost, the success rate reaches a plateau with a few layers, an effect that is especially prominent at smaller densities. In some cases, such as the use of compatible or linear entanglement, the success rate saturates at one layer [cf. Fig. 7(a)]. These are shallow enough circuits which have been simulated exactly in a classical computer. Hence, we cannot attribute the effect to vanishing gradients or barren plateaus [53, 54], but to an intrinsic limitation of the algorithm.

Influence of the number of experimental shots.- Until now, our study has focused on exact simulations of the variational wave function, computing the exact expectation values. We will now consider a realistic scenario with a finite number of experimental shots for each observable. In this case, we no longer compute the exact CVaR-VQE cost function, but the random estimator. As explained above, this also requires us to use the SPSA classical optimizer. The results are summarized in Fig.
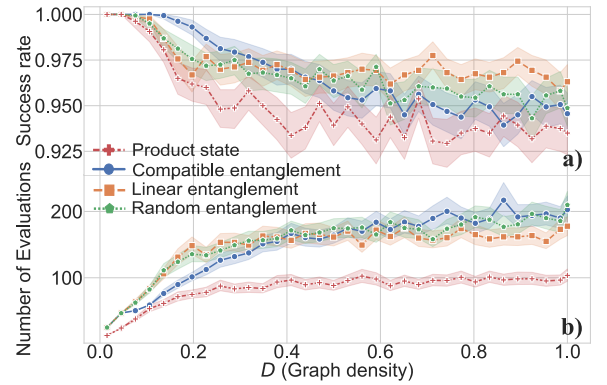

FIG. 6. Performance of CVaR-VQE ($\rho = 10\%$) for different entanglement structures and product states, as a function of the QUBO problem's graph density. We simulated problems with $N = 12$ qubits, using exact wave functions with $L = 0$ or 1 entanglement layer, and the L-BFGS-B classical optimizer. We plot the average (a) success rate and (b) number of evaluations computed over $N_{\text{ins}} = 1600$ instances, surrounded by the 95% confidence interval.

8, for simulations with $L = 1$ layers. Note how Figs. 8 (a) and 8(b) imitate earlier plots, for different numbers of measurements, while Figs. 8 (c) and 8(d) focus on a fixed density and study the variation with increasing number of shots.

We can extract several conclusions. First, we see that the compatible entanglement is the most successful *Ansatz* when a sufficient number of shots is available. The contrast with Fig. 6 must be attributed to the change in the optimizer, which also affects the number of evaluations. Second, this advantage is clearly wiped out when we consider a scenario with a finite number of measurements, where the cost function is randomly estimated. Third, we see that the slowdown in convergence produced by the increased graph density is aggravated when the algorithm has incomplete information of the wave function. Third, the previous remark does not apply to the product state *Ansatz*, as discussed in more detail in Sec. IV D.

Effect of adapting the entanglement to the problem.- Although briefly mentioned in the preceding paragraphs, let us summarize under what circumstances matching the entanglement to the problem structure offers an advantage over other *Ansätze*. With exact wave functions simulations (see Fig. 6) and when the information of the cost function is quite incomplete (see 3000 shots case in Fig. 8), compatible entanglement is marginally better than other approaches only when the problem density is low. However, the adaptation of entanglement to the network connections loses its advantageous effect when the density increases since the specific network structure becomes less noticeable. We also observe a sharper deterioration in the speed of the algorithm when compatible entanglement is applied to high-
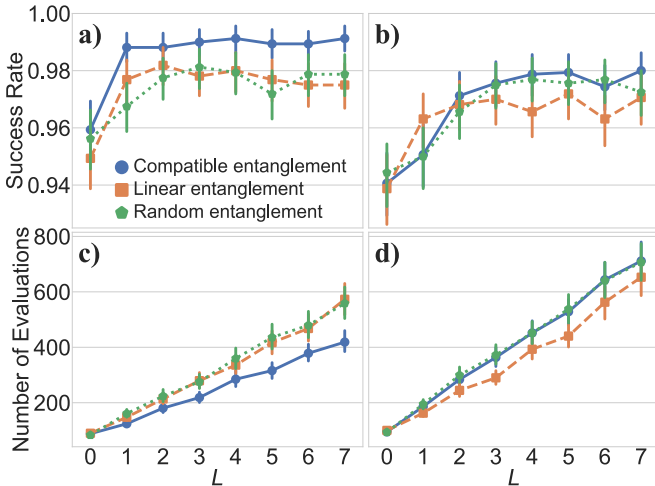

FIG. 7. Success rate and average number of cost function evaluations needed for convergence, as a function of the number of layers of the variational form. The results show the average of $N_{ins} = 1600$ instances and a 95% confidence interval. Plots (a), (c) and (b), (d) correspond to problems with graph density equal to 0.258 and 0.894, respectively.
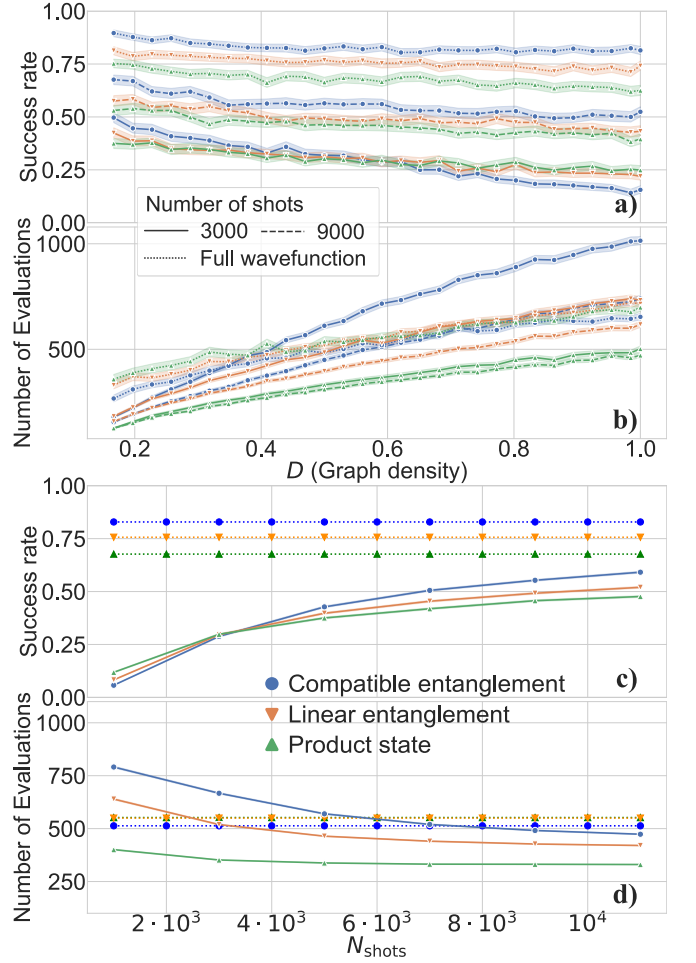

FIG. 8. Performance of CVaR-VQE ($\rho = 10\%$) in a realistic simulation with a finite number of measurements, using the SPSA classical optimizer. Calculations have been done using $N = 12$ qubits and $L = 0, 1$ entanglement layers. We plot both the average and a 95% confidence interval. Plots (a), (b) show the success rate and the speed of convergence respectively as a function of the connectivity of the problem for 3000 and 9000 measurements of the variational wave function. Plots (c), (d) illustrate the success rate and speed of convergence as we change the number of measurements, with the ideal wave function simulations ($N_{shots} \to \infty$) in dotted lines.

density problems. If we reduce the uncertainty in the estimation of the objective function raising the number of measurements to 9000 (see Fig. 8), we can see that a graph-adapted entanglement is able to slightly improve the success rate of the algorithm for problems of any density without a significant delay in the convergence.

C. Hardness analysis

At this point we have discussed the efficiency of optimizing QUBO problems according to the characteristics of the variational algorithm such as the cost function or the *Ansatz*. In this section we analyze the problem from
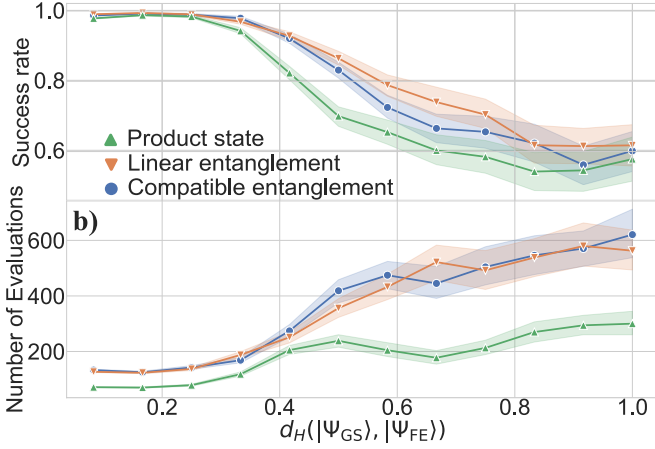
FIG. 9. Performance of CVaR-VQE ($\rho = 10\%$) as a function of the QUBO's problem Hamming distance between the optimal and the first-excited subspace. We used $N = 12$ qubits with $L = 0$ or 1 layers, computing the exact energy with the full wave function ($N_{\text{shots}} \to \infty$). The plots show (a) the average success rate and (b) the number of cost function evaluations, with a 95% confidence interval.
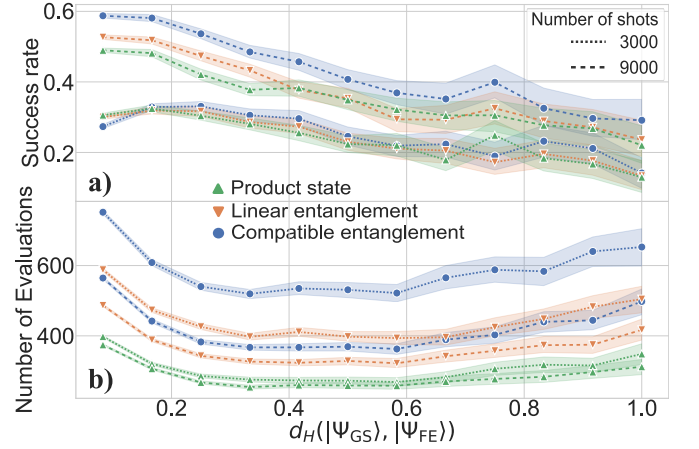

FIG. 10. Performance of CVaR-VQE ($\rho = 10\%$) as a function of the QUBO problems' Hamming distance between the ground state and lowest-excited manifold, when we consider a limited number of measurements in the estimation of the cost function. Simulations used $N = 12$ qubits and $L = 0$ and 1 entanglement layers. We plot the (a) average success rate and (b) average number of evaluations, with a 95% confidence interval.

a different point of view. We now try to get some insight into how the specific properties of each QUBO problem affect the hardness of the optimization in terms of success rate and speed of convergence. We define as a hard problem one for which the variational algorithm requires a high number of evaluations to converge, while still obtaining a low success rate.

Success rate and speed as a function of graph density.- We have already concluded from Figs- 4-6 that the density of the graph severely impacts the performance of the variational algorithm. These studies lead to some general remarks. (a) On average we found that graphs with many connected vertices—i.e., QUBO with dense matrices—are harder problems than sparse problems. (b) When the entanglement structure imitates the graph, we saw that the number of successful instances decreases with the graph connectivity (see Fig. 6). (c) However, for variational forms whose design is independent of the problem, we observed that there is a critical density in which the trend steadies around a minimum success rate, despite increasing the number of edges in the graph (see Figs. 5 and 6). Such critical density might be related to the percolation point of the network. (d) We observe equivalent trends when we consider the number of function evaluations.

Impact of low-energy state structure: a measure of hardness.- Optimization algorithms typically fail because the *Ansatz* gets trapped in a local minimum that is different from the global minimum we seek. We expect that this phenomenon will be more likely when there are states that have a value of the cost function close to the optimum, but are far from the optimal state in Hilbert space. To test this hypothesis, we have correlated the performance of the variational quantum op-

timizer to the minimum Hamming distance between the ground state space of the QUBO problem and the (possibly degenerate) manifold of its first excitations. Let us recall that the Hamming distance between two bit strings $|\psi_1\rangle = |s_1, \ldots, s_N\rangle$ and $|\psi_2\rangle = |t_1, \ldots, t_N\rangle$, is given by $d_{\text{H}}(|\psi_1\rangle, |\psi_2\rangle) = \frac{1}{N} \sum_j^N |s_j - t_j|$. We have computed the minimum such distance between the ground state and first-excited manifolds of randomly generated QUBO problem, exploring 29 values of the density and averaging over 1600 random regular graphs for each density.

The variational algorithm was simulated using the full wave function and also a finite number of measurements, as shown in Figs. 9 and 10, respectively. These results show a strong correlation between the Hamming distance from the fundamental state to the first-excited state and the difficulty of the algorithm to solve the optimization problem, both in terms of success rate and speed of convergence. This conclusion holds for all variational *Ansätze*.

It must be remarked that the correlation of the difficulty to the Hamming distance is much clearer than to the density (cf. Figs. 6 and 7). While the increase of the graph connectivity led to a maximum loss of success rate of approximately 3%, in this case the difference can be of up to 40% for full simulations, and 30% when using 9000-shot measurements. Moreover, the number of objective function evaluations also grows significantly with the Hamming distance. We thus conclude that the Hamming distance is a clearer indicator than the plain graph density to predict the hardness of an optimization problem [55, 56].

D. Product state vs. entangled *Ansätze*

So far we have been considering on equal footing the *Ansätze* with $L \geq 1$ entangling layers, and the special case in which we eliminate all entanglers and work with product states. It is important to do a careful comparison between both situations, because only the former case argues for an advantage of using the quantum computer. Indeed, using product states, with CVaR or pure VQE, can be regarded as a new classical optimization algorithm, similar to other methods proposed in the last decade [37, 38].

At a glance in Figs. 2-10, we observe a slight improvement in the success rate of VQE when using some entangling operations, as compared with the $L = 0$ product state method. This success comes with two costs. First, we need to use a quantum computer, which in itself involves a non-negligible overhead. Second, the increased number of parameters in the variational *Ansatz* also leads to a larger number of function evaluations and slower optimization. Let us analyze these metrics with greater care.

Product state-based *Ansatz* seems better for few shots.- The advantage of entangled states decreases and even disappears when we consider the statistical errors induced by a finite number of quantum measurements. This was seen in Figs. 8(a) and 10(a), and more clearly in Fig. 8(b). A decrease in the number of samples reduces the success gap, to a point where at 3000 shots the product state can outperform other methods. Moreover, as seen in Figs. 8(c) and 10(b), and especially in Fig. 8(d), as we decrease the number of measurements, the product state also saturates in the number of function evaluations, while all other *Ansätze* grow. We therefore conclude that the use of a product state variational form (and its fully classical simulation) is advantageous when we are limited in the precision with which we can estimate the cost function.

The role of entanglement as a function of graph density.- One may also wonder whether entanglement can provide an advantage in highly- or weakly-connected problems. Inspecting Figs. 6 and 8, this does not seem to be the case. At large densities with no statistical uncertainty, the success rate is only marginally better in the entangled *Ansatz* case, by 2-3%. However, in this region there is a significant performance hit due to the larger number of function evaluations as compared with the product states (see Fig. 6). If we consider the stochasticity of quantum measurements, the situation is reversed: entangled *Ansätze* achieve greater success rates, by 20% or higher, with faster convergence, regardless of the density of the optimized graph.

The role of entanglement in different hardness regimes.- We have seen that the graph density is not a good quantifier for hardness. It is therefore interesting to look at the correlation between the Hamming distance and the performance of both types of methods, as shown in Table I. From inspecting this table and Figs. 9 and

		3000 shots			9000 shots			Full wave function		
		(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
Success	Product	31%	22%	16%	46%	34%	26%	98%	69%	55%
	Entangled	31%	25%	20%	57%	40%	30%	99%	83%	61%
Speed	Product	335	274	327	316	262	298	73	206	288
	Entangled	496	538	625	463	372	461	129	390	560

TABLE I. Performance of the product-state and entangled-state *Ansätze*, for varying hardness and number of shots. The columns (A), (B), and (C) correspond to problems with Hamming distances $d_{\mathrm{H}} \leq 0.3$, $d_{\mathrm{H}} \in [0.4, 0.7]$, and $d_{\mathrm{H}} \geq 0.8$, respectively. The table displays the success rate and the average number of evaluations of the cost function.

10, we conclude that entanglement provides a moderate advantage in the success rate, ranging from 10% at easy problems, up to $4 - 5\%$ for hard problems, always in the realistic case with a limited number of measurements. However, the product-state *Ansatz* can be efficiently simulated in the limit of infinite measurements (full wave function). Therefore, in practical terms, we would achieve 25% to 41% higher success rates, with up to an order of magnitude reduction in the number of evaluations, simulating the product-state *Ansatz*.

Let us sum up the comparison between product-state and entangled *Ansätze*. We have found that entanglement yields a slight improvement only in the case of negligible measurement error. On the contrary, the product-state *Ansatz* is advantageous, if there is a significant uncertainty in the cost function evaluation (see Figs. 8 and 10, and Table I). Our analysis implies that product-state outperforms entangled *Ansätze* in practical terms, since they require fewer resources and can even be classically simulated.

V. CONCLUSIONS

In this work we have studied the optimization of QUBO problems using the original and CVaR formulations of the VQE algorithm. We have compared the performance of this type of solvers for different variational *Ansätze* and for different types of problems, analyzing the success rates, convergence speeds, types of states that are created, and the behavior of the algorithm both in idealized scenarios (unlimited number of measurements) and limited resources.

Our work corroborates the advantage of using a CVaR cost function over conventional averages. We have also verified that adapting the structure entangling operations to the topology of the problem is marginally advantageous, in line with recent experimental results [21]. We also found that in a scenario with limited numbers of measurements, entanglement also provides a quantitative advantage over other approaches when enough shots can be performed, but this advantage quickly saturates with the depth of the *Ansatz*. Our study also finds a correlation between the practical hardness of a problem and

a metric that characterizes the structure of low-energy excitations.

One important takeaway is that CVaR metrics with product state *Ansatze* provide a classical optimization capable of outperforming the variational quantum eigensolver. This method can afford very large numbers of effective measurements, reaching success rates and performance metrics that exceed those that we see for conventional hybrid quantum-classical methods. This result highlights the need to investigate further the existing algorithms for quantum optimization, including other alternatives such as QAOA [1]. It also illustrates the possibilities that quantum inspired alternatives can bring to the field of classical optimization in the near term.

[1] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), arXiv:1411.4028 [quant-ph].

[2] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).

[3] J. Preskill, Quantum **2**, 79 (2018).

[4] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, arXiv e-prints , arXiv:2101.08448 (2021), arXiv:2101.08448 [quant-ph].

[5] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and et al., Nature Reviews Physics **3**, 625–644 (2021).

[6] M. Willsch, D. Willsch, F. Jin, H. De Raedt, and K. Michielsen, Quantum Information Processing **19**, 10.1007/s11128-020-02692-8 (2020).

[7] G. Nannicini, Phys. Rev. E **99**, 013304 (2019).

[8] D. Headley, T. Müller, A. Martin, E. Solano, M. Sanz, and F. K. Wilhelm, Approximating the quantum approximate optimisation algorithm (2020), arXiv:2002.12215 [quant-ph].

[9] J. Choi, S. Oh, and J. Kim, Appl. Sci. **10**, 10.3390/app10207116 (2020).

[10] Z. Hamid Saleem, International Journal of Quantum Information **18**, 10.1142/S0219749920500112 (2020).

[11] Utkarsh, B. K. Behera, and P. K. Panigrahi, Solving vehicle routing problem using quantum approximate optimization algorithm (2020), arXiv:2002.01351 [quant-ph].

[12] A. Bengtsson, P. Vikstål, C. Warren, M. Svensson, X. Gu, A. F. Kockum, P. Krantz, C. Križan, D. Shiri, I.-M. Svensson, G. Tancredi, G. Johansson, P. Delsing, G. Ferrini, and J. Bylander, Phys. Rev. Applied **14**, 034010 (2020).

[13] P. Vikstål, M. Grönkvist, M. Svensson, M. Andersson, G. Johansson, and G. Ferrini, Phys. Rev. Applied **14**, 034009 (2020).

[14] S. Mugel, E. Lizaso, and R. Orus, Use cases of quantum optimization for finance (2020), arXiv:2010.01312 [q-fin.GN].

[15] S. Fernández-Lorenzo, D. Porras, and J. J. García-Ripoll, Quantum Science and Technology **6**, 034010 (2021).

[16] E. Osaba, E. Villar-Rodriguez, I. Oregi, and A. M.-F. de Leceta, Focusing on the hybrid quantum computing – tabu search algorithm: new results on the asymmetric salesman problem (2021), arXiv:2102.05919 [quant-ph].

[17] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Hybrid quantum-classical algorithms for approximate graph coloring (2020), arXiv:2011.13420 [quant-ph].

[18] R. Herrman, J. Ostrowski, and T. e. a. Humble, Quantum Inf Process **20**, 10.1007/s11128-021-03001-7 (2021).

[19] T. Haug, K. Bharti, and M. Kim, PRX Quantum **2**, 10.1103/prxquantum.2.040309 (2021).

[20] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus (2021), arXiv:2101.02138 [quant-ph].

[21] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo, and et al., Nature Physics 10.1038/s41567-020-01105-y (2021).

[22] L. Li, M. Fan, M. Coram, P. Riley, and S. Leichenauer, Physical Review Research **2**, 10.1103/physrevresearch.2.023074 (2020).

[23] P. K. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli, and S. Woerner, Quantum **4**, 256 (2020).

[24] X. Liu, A. Angone, R. Shaydulin, I. Safro, Y. Alexeev, and L. Cincio, Layer vqe: A variational approach for combinatorial optimization on noisy quantum computers (2021), arXiv:2102.05566 [quant-ph].

[25] S. Sim, J. R. Fontalvo, J. F. Gonthier, and A. A. Kunitsa, Quantum Science and Technology (2021).

[26] D. Wang, O. Higgott, and S. Brierley, Phys. Rev. Lett. **122**, 140504 (2019).

[27] J.-G. Liu, Y.-H. Zhang, Y. Wan, and L. Wang, Phys. Rev. Research **1**, 023025 (2019).

[28] A. Garcia-Saez and J. I. Latorre, Addressing hard classical problems with adiabatically assisted variational quantum eigensolvers (2018), arXiv:1806.02287 [quant-ph].

[29] I. G.Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov, J. Chem. Theory Comput. **14** (2018).

[30] H. Grimsley, S. Economou, E. Barnes, and et al., Nat Commun **10** (2019).

[31] S. M. Harwood, D. Trenev, S. T. Stober, P. Barkoutsos, T. P. Gujarati, and S. Mostame, Improving the variational quantum eigensolver using variational adiabatic

quantum computing (2021), arXiv:2102.02875 [quant-ph].

[32] A. Peruzzo, J. McClean, and P. Shadbolt et al, Nat Commun **5 4213** (2014).

[33] G. Kochenberger, JK.Hao, and F. et al., J Comb Optim **28**, 58 (2014).

[34] G.A.Kochenberger, F.Glover, and B. et al., OR Spectrum **26**, 237 (2004).

[35] N. M. et al, Quantum Sci. Technol **3 030503** (2018).

[36] C. Acerbi and D. Tasche, Journal of Banking and Finance **26(7)**, 1487 (2002).

[37] J. A. Smolin and G. Smith, Frontiers in Physics **2**, 10.3389/fphy.2014.00052 (2014).

[38] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, How "Quantum" is the D-Wave Machine? (2014), arXiv:1401.7087v2 [quant-ph].

[39] M. Powell, Computer Journal **7**, 155 (1964).

[40] D. Kraft, A software package for sequential quadratic programming, Report DFVLR-FR 88-28 (Deutsche Forschungs- und Versuchsanstalt fr Luft- und Raumfahrt, 1988).

[41] C. G.Broyden, J lnst Maths Appl **6**, 76 (1970).

[42] R.Fletcher, Computer journal **13**, 317 (1970).

[43] D.Goldfarb, Mathematics of Computation **24**, 23 (1970).

[44] D.F.Shanno, Mathematics of Computation **24**, 647 (1970).

[45] R. H. Byrd, P. Lu, and J. Nocedal, SIAM Journal on Scientific Computing **16**, 1190 (1995).

[46] J. Spall, IEEE Transactions on Automatic Control **37**, 332 (1992).

[47] M. Powell, Advances in Optimization and Numerical Analysis , 51 (1994).

[48] J.A.Nelder and R.Mead, The Computer Journal **7**, 308 (1965).

[49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors, Nature Methods **17**, 261 (2020).

[50] A. Hagberg, P. Swart, and D. S Chult, *Exploring network structure, dynamics, and function using NetworkX*, Tech. Rep. (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008).

[51] B. Efron, The Annals of Statistic **7**, 1 (1979).

[52] B. Efron and R. J.Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, Inc., 1993).

[53] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications **9**, 10.1038/s41467-018-07090-4 (2018).

[54] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nature Communications **12**, 10.1038/s41467-021-27045-6 (2021).

[55] E. Farhi, J. Goldstone, S. Gutmann, and D. Nagaj, International Journal of Quantum Information **06** (2008).

[56] A. Callison, N. Chancellor, F. Mintert, and V. Kendon, New Journal of Physics **21**, 123022 (2019).