Energy landscapes for the quantum approximate optimization algorithm

Choy Boy and David J. Wales

Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 3 January 2024; accepted 8 May 2024; published 3 June 2024)

Variational quantum algorithms (VQAs) have demonstrated considerable potential in solving NP-hard combinatorial problems in the contemporary noisy intermediate-scale quantum (NISQ) era. The quantum approximate optimization algorithm (QAOA) is one such algorithm, used in solving the maximum cut (Max-Cut) problem for a given graph by successive implementation of L quantum circuit layers within a corresponding Trotterized ansatz. The challenge of exploring the cost function of VQAs arising from an exponential proliferation of local minima with increasing circuit depth has been well documented. However, fewer studies have investigated the impact of circuit depth on QAOA performance in finding the correct Max-Cut solution. Here we employ basin-hopping global optimization methods to navigate the energy landscapes for QAOA ansätze for various graphs, and analyze QAOA performance in finding the correct Max-Cut solution. The structure of the solution space is also investigated using discrete path sampling to build databases of local minima and the transition states that connect them, providing insightful visualizations using disconnectivity graphs. We find that the corresponding landscapes generally have a single funnel organization, which makes it relatively straightforward to locate low-lying minima with good Max-Cut solution probabilities. In some cases below the adiabatic limit the second lowest local minimum may even yield a higher solution probability than the global minimum. This important observation has motivated us to develop broader metrics in evaluating QAOA performance, based on collections of minima obtained from basin-hopping global optimization. Hence we establish expectation thresholds in elucidating useful solution probabilities from local minima, an approach that may provide significant gains in elucidating reasonable solution probabilities from local minima.

DOI: 10.1103/PhysRevA.109.062602

I. INTRODUCTION

The initial setback of implementing practical quantum algorithms utilizing the quantum phase estimation (QPE) architecture onto current-day noisy intermediate-scale quantum (NISQ) devices [1–3], which typically possess short decoherence times [4] and significant quantum noise [5], has prompted the rapid development of variational quantum algorithms (VQAs) with shorter quantum circuit depths [6–8]. VQAs typically operate within a hybrid classical-quantum optimization framework [9], where an initial quantum state is evolved by a parameterized circuit ansatz on a quantum device. After the final evolved wave function is measured, a classical optimizer evaluates the cost function from the measurement and subsequently suggests new parameters that are fed back into the parameterized quantum circuit. This interface between classical and quantum computers iterates until a suitable convergence criterion is attained. VQAs have proved to be surprisingly robust in tackling various sources of noise attributed to NISQ devices, such as decoherence [10] and depolarization [11]. These properties are ascribed to the innate variational nature acting as a parametric lever, which can be flexibly adjusted even under noisy environments [12]. Coupled with the recent advances in error-mitigation methods [13–15], it is anticipated that VQAs will enable the realization of practical quantum advantage before the advent of fault-tolerant devices equipped with error-correction protocols [16].

The quantum approximate optimization algorithm (QAOA) is a VQA that monotonically improves in performance as the number of quantum circuit layers in the ansatz increases [17]. Variational quantum eigensolver (VQE) algorithms typically utilize the global minimization of the cost landscape directly as the solution of interest [18]. In contrast, QAOA uses the resulting final wave function to obtain approximate solutions based on the states in the computational basis with the highest frequency when measurement after circuit evolution is carried out. Hence QAOA is a particularly attractive algorithm for solving combinatorial optimization problems, such as Max-Cut [19], with promising applications in portfolio optimization [20] and chemistry [21,22].

The expansion of VQAs has also motivated the study of their potential shortcomings in solving practical large-scale problems from a software perspective. In particular, for various VQA ansätze there is an exponential growth in the barren plateau problem as the number of qubits and circuit layers required to encode a given problem increases [23–25]. Recently, more insight has also been gained into the challenges of exploring the cost landscapes of VQAs that arise from a proliferation in the number of local minima and other stationary points as the complexity of the problem increases [26,27]. However, further analysis of the organization of the global cost landscapes of VQAs, and how this structure

impacts the quality of the solutions obtained, is needed, especially for QAOA [28,29].

Here we seek to address these gaps in understanding using the well-established theory and associated computational methodology of molecular energy landscapes [30,31]. We characterize the cost optimization landscapes of QAOA for various weighted and unweighted graphs in solving the Max-Cut problem using basin-hopping global optimization methods [32–34] to locate global minima, and discrete path sampling [35,36] to create connected databases of minima and the transition states that connect them. Recently, energy landscape techniques have demonstrated considerable utility for quantum computing in the analysis of hardware-efficient ansätze for the VQE algorithm [37], and optimization of electronic wave functions in a combined discrete space of operators and continuous parameter amplitudes [38]. Our results for QAOA show that the solution landscapes below the adiabatic limit generally possess single-funneled structures associated with self-organizing systems where locating the global minimum is relatively straightforward [30,39,40]. Furthermore, we find that local minima sufficiently close in energy to the global minimum may also exhibit good solution probabilities for the Max-Cut problem. In some instances, the second lowest minimum has a higher solution probability than the global minimum, highlighting the importance of studying the VQA solution landscape globally. This observation leads us to introduce metrics that take into account the distribution of minima in evaluating the performance and robustness of QAOA. We also utilize the convex hull of the solution space in estimating expectation cutoffs for the location of local minima with reasonable solution probabilities. We hope that these techniques can advance the feasibility of implementing QAOA for problems with numerous local minima in noisy environments.

II. METHODOLOGY

Given an undirected graph $G = (V, E)$, with weights w_{ij} assigned to edges $e_{ij} \in E$ for connected vertices $i, j \in V$, the Max-Cut problem seeks to partition V into two distinct sets such that the sum of weights between the two sets is maximized. If $w_{ij} = 1$ for all e_{ij}, then G is said to be an unweighted graph; otherwise G is a weighted graph. It follows that the Max-Cut problem can be mapped to a two-spin Ising-type cost Hamiltonian $\hat{H}_C$ corresponding to N implementable qubits:

$$\hat{H}_C = \frac{1}{2}\sum_{e_{ij}\in E} w_{ij}(Z_i \otimes Z_j), \tag{1}$$

where Z is the Pauli matrix Z, and the states $|s\rangle = \{|\alpha\rangle, |\beta\rangle\}^{\otimes N}$ encode the desired solution strings to the Max-Cut problem, with $|\alpha\rangle = |0\rangle$ if and only if $|\beta\rangle = |1\rangle$, and vice versa. Thus, the aim of QAOA is to approximate the ground-state energy or the lowest eigenvalue of $\hat{H}_C$ via a suitable ansatz with unitary operator $\hat{U}(\boldsymbol{\theta})$ to evolve an initial state $|\psi_0\rangle$, and subsequently use the final evolved state $|\Psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\psi_0\rangle$ to approximate $|s\rangle$. This objective can be achieved on a quantum device by performing a certain number of shots per experiment and measuring all qubits in the computational basis, taking the state possessing the greatest number of shots to best approximate $|s\rangle$ for that experiment. We seek to simulate this procedure classically by considering the probability of measuring the state $|s\rangle$ in the computational basis, $p(|s\rangle)$:

$$p(|s\rangle) = |\langle s|\Psi(\boldsymbol{\theta})\rangle|^2. \tag{2}$$

The objective function to be minimized by the classical computer is the expectation of $\hat{H}_C$, $\langle \hat{H}_C\rangle$:

$$\langle \hat{H}_C\rangle = E(\boldsymbol{\theta}) = \langle\psi_0|\hat{U}^\dagger(\boldsymbol{\theta})\hat{H}_C\hat{U}(\boldsymbol{\theta})|\psi_0\rangle. \tag{3}$$

The QAOA ansatz with parameters $\boldsymbol{\theta} = \{\boldsymbol{\gamma}, \boldsymbol{\delta}\}$ can be assembled as a Trotterized variational schedule, comprising a cost circuit layer with unitary operator $\hat{U}_C(\gamma)$, followed by a mixer circuit layer with unitary operator $\hat{U}_M(\delta)$ up to a circuit depth L:

$$|\Psi(\boldsymbol{\gamma}, \boldsymbol{\delta})\rangle = \prod_{l=1}^{L}\hat{U}_M(\delta_l)\hat{U}_C(\gamma_l)|\psi_0\rangle, \tag{4}$$

where $|\psi_0\rangle = |+\rangle^{\otimes N}$ is the state encoding for all possible partitions of V with equal probability. The cost layer encoding $\hat{H}_C$ can be compiled as a sequence of two-qubit parameterized R_{zz} quantum gates for qubits q_i and q_j, with γ scaled based on the weights of e_{ij}:

$$\begin{aligned}\hat{U}_C(\gamma) &= e^{-i\gamma\hat{H}_C}\\ &= \prod\nolimits_{e_{ij}\in E} R_{zz}(-w_{ij}\gamma).\end{aligned} \tag{5}$$

The mixer layer performs a time evolution of the mixer Hamiltonian $\hat{H}_M = -\sum_{i=1}^{N} X_i$, where X is the Pauli matrix X, of which $\hat{H}_M$ anticommutes with $\hat{H}_C$ and has $|\psi_0\rangle$ as an eigenvector. The mixer layer can be realized as a parallelization of single-qubit parameterized R_x quantum gates:

$$\begin{aligned}\hat{U}_M(\delta) &= e^{-i\delta\hat{H}_M}\\ &= \bigotimes_{i=1}^{N} R_x(2\delta).\end{aligned} \tag{6}$$

It has been shown that QAOA conforms to the adiabatic theorem; i.e., for $L \to \infty$ the final evolved state $|\Psi(\boldsymbol{\theta})\rangle$ converges exactly to the ground state of $\hat{H}_C$ and thus gives the optimal $p(|s\rangle)$ [17]. In practice, such an implementation is unfeasible in the NISQ regime, hence we are interested in considering L_{ad} for a given system, defined as the minimum number of circuit layers required to reach the adiabatic limit, assuming that L_{ad} can be attained. As we will demonstrate in our analysis of the energy landscapes of QAOA, it is also important to distinguish L_{ad} from $L_{\min}$, where $L_{ad} \geqslant L_{\min}$. Here $L_{\min}$ is the minimum number of layers needed to achieve the maximum $p(|s\rangle)$ in the corresponding global minimum. Local minima with lower $p(|s\rangle)$ may be present due to underparametrization of the circuit ansatz, and hence a less thorough exploration of states in the Hilbert space may be sufficient to obtain a useful solution. We hypothesize that the exponential increase in the number of local minima is attributable to circuit ansätze with $1 \leqslant L \leqslant L_{\min}$ layers. The behavior of local minima may vary for $L_{\min} < L \leqslant L_{ad}$ layers if $L_{\min} < L_{ad}$, and we observe for various graphs that the number of local minima may increase first before decreasing to the adiabatic limit, or instead decrease monotonically.

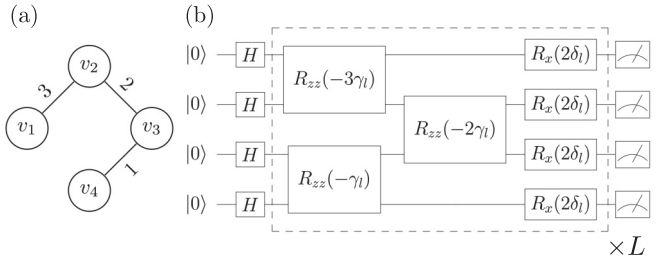

FIG. 1. (a) (left) Weighted graph G_1 encoded with four qubits. (b) (right) Corresponding QAOA ansatz for G_1

For each graph considered and L, we generate an initial set of minima via basin-hopping global optimization [32–34] using the GMIN program [41]. The analytic gradients of the parameterized rotation gates were calculated via the parameter-shift rule [42] (see Appendix A for more details):

$$\frac{\partial E(\boldsymbol{\theta})}{\partial \theta_i} = \frac{1}{2}\left[E\left(\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_i\right) - E\left(\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_i\right)\right], \tag{7}$$

where $\mathbf{e}_i$ is the unit vector of θ_i. Local minimization for the basin-hopping steps employed a limited-memory Broyden [43], Fletcher [44], Goldfarb [45], Shanno [46] (L-BFGS) procedure [47,48] equipped with the Metropolis criterion for accepting or rejecting steps [49]. The resulting minima were then employed as the starting points for construction of a kinetic transition network [50–52]. Discrete path sampling [35,36] (DPS) was used via connection attempts for selected minima pairs with final states $|\Psi(\boldsymbol{\theta}_\mu)\rangle$ and $|\Psi(\boldsymbol{\theta}_\nu)\rangle$. The doubly-nudged [53,54] elastic band [55–58] approach was used to locate candidates for accurate transition state refinement by hybrid eigenvector following [59–61]. The missing connection algorithm [62] was used to select pairs of minima to fill in the gaps in incomplete pathways via Dijkstra's shortest path algorithm [63] combined with a distance metric based on the state overlap between local minima, $S_{\mu\nu}$:

$$S_{\mu\nu} = 1 - |\langle\Psi(\boldsymbol{\theta}_\mu)|\Psi(\boldsymbol{\theta}_\nu)\rangle|. \tag{8}$$

Any new minima are added to the database along with the transition states and connection information. The resulting cost landscapes can be visualized using disconnectivity graphs, where the local minima are segregated into disjoint sets for regular thresholds in the energy [39,64]. In these graphs, the bottom of a branch corresponds to the energy of a local minimum, and the branches are joined when the corresponding minima can interconvert via a pathway below the given threshold. Visualization of the energy landscape can be further enhanced by coloring each minimum with the corresponding probability of finding the correct Max-Cut solution; we find this construction especially useful in comparing the solution landscapes as L varies.

As an example, consider the weighted graph G_1 with four vertices [Fig. 1(a)], where the Max-Cut problem can be encoded as a four-qubit QAOA ansatz with varying L [Fig. 1(b)]. We note that although the ansatz cost circuit layer can be compiled in numerous ways, the arrangement in Fig. 1(b) is ideal in reducing the overall circuit depth. This ordering does not require swap gates to permute nonneighboring qubits corresponding to their respective edges, which is an important consideration when transpiling QAOA onto planar superconducting processors [65]. The cost landscape of the $L = 1$ circuit ansatz features a global and a local minimum connected by a transition state [Fig. 2(a)], and the corresponding disconnectivity graph is shown in Fig. 2(b), where the branches are colored with the probabilities of finding the state $|\alpha\beta\alpha\beta\rangle$, which corresponds to the Max-Cut solution of G_1.

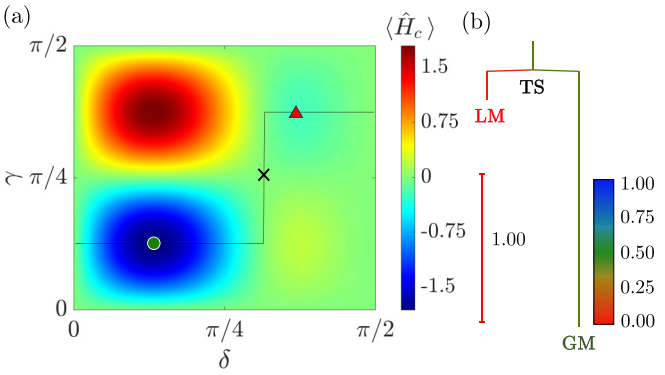

FIG. 2. Contour plot of $\langle\hat{H}_C\rangle$ for graph G_1 and its QAOA ansatz with $L = 1$ against the parameters γ and δ. The solid line depicts the minimum of each vertical slice of the contour plot, with the global minimum (GM, green circle), local minimum (LM, red triangle), and the transition state connecting them (TS, black cross) situated on the pathway that is also plotted. (b) Corresponding disconnectivity graph of the contour plot in (a), with both the GM and LM colored (grayscaled) based on their respective probabilities of obtaining the correct Max-Cut solution of $|\alpha\beta\alpha\beta\rangle$.

III. RESULTS

A. Complete unweighted graphs

We begin by examining the QAOA circuit ansätze for the complete unweighted graph series K_N from $N = 3$ to $N = 8$

TABLE I. Number of minima M (top value) and the highest correct Max-Cut probability (HCMP) (bottom value) for graphs K_3 to K_8 with varying L obtained from basin-hopping global optimization. For the case of K_8 and $L = 3$, the HCMP is not equal to the maximum value of 1, and the number of decimal places used for all HCMPs is chosen to be the same as this case for ease of comparison.

Graph	$L = 1$	$L = 2$	$L = 3$
K_3	1	1	1
	1.000000	1.000000	1.000000
K_4	1	1	1
	0.739106	1.000000	1.000000
K_5	1	4	1
	0.975990	1.000000	1.000000
K_6	1	23	324
	0.671340	0.994239	1.000000
K_7	1	37	598
	0.951350	0.999619	1.000000
K_8	1	46	3418
	0.629727	0.991483	0.999997

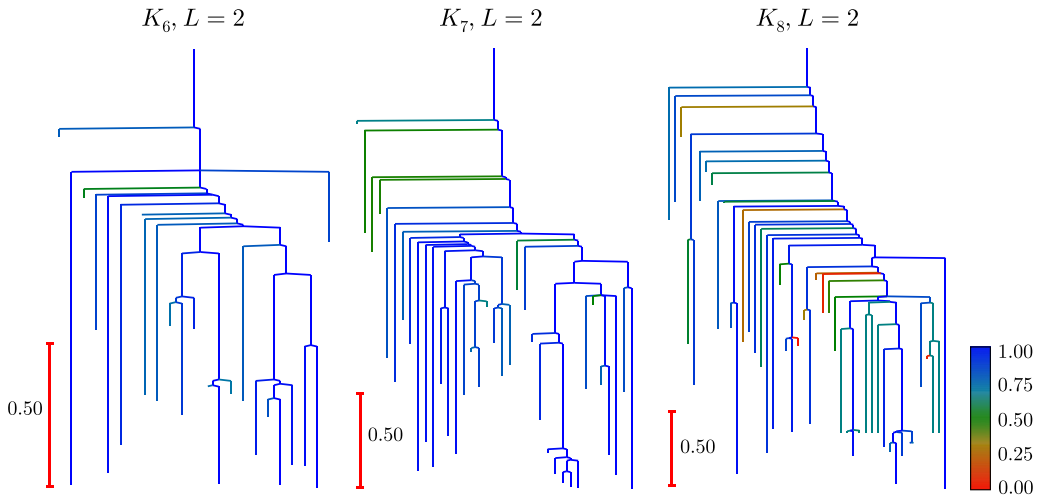

FIG. 3. Disconnectivity graphs of K_6, K_7, and K_8 for $L = 2$.

and from $L = 1$ to $L = 3$, where each vertex is connected to every other vertex with unit weight. It follows that the Max-Cut solution of K_N is the set of all possible tensor product permutations of $\lfloor N/2 \rfloor$ number of $|\alpha\rangle$ states and $\lceil N/2 \rceil$ number of $|\beta\rangle$ states: thus the total number of Max-Cut solutions for odd N is $2N!/\{\lfloor N/2 \rfloor! \lceil N/2 \rceil!\}$, and for even N it is $N!/\{2(N/2)!\}$ solutions.

Table I summarizes the number of minima M and the highest correct Max-Cut probability (HCMP) from the collections of minima obtained via basin-hopping global optimization for $L = 1$ to $L = 3$. The complete graphs with odd N generally possess higher HCMPs than their even counterparts, mainly due to their greater number of accepted Max-Cut solutions that contribute to their corresponding probabilities. We also find that although for K_6 to K_8 the expected exponential increase in M is observed as L increases, for K_5 there was a decrease in the number of minima from $L = 2$ to $L = 3$, leading to a simplification in the energy landscape from $L_{\min} = 2$ to $L_{ad} = 3$. Looking at the disconnectivity graphs of K_6 to K_8 for $L = 2$ (Fig. 3) and $L = 3$ (Fig. 4), we observe that the majority of the local minima generally possess very high correct Max-Cut probabilities, particularly those closer to the global minimum, especially as L and M increase. The well-funneled organization of the landscape also becomes more apparent as L increases, and this structure is expected to outweigh the challenges associated with solving the Max-Cut problem for higher N, particularly for K_8, where local minima with low probabilities are increasingly interspersed with other local minima corresponding to higher probabilities.

To further evaluate the performance of QAOA ansätze for various L based on the databases of minima and their respective Max-Cut probabilities and $\langle \hat{H}_C \rangle$ values, we introduce the weighted average metric F:

$$F = \frac{1}{M|\langle \hat{H}_C \rangle_{\min}|} \sum_{m=1}^{M} |\langle \hat{H}_C \rangle_{\min} - \langle \hat{H}_C \rangle_m| [1 - p_m(|s\rangle)], \quad (9)$$

where $\langle \hat{H}_C \rangle_{\min}$ is the $\langle \hat{H}_C \rangle$ value of the global minimum. This formulation of F is advantageous for two reasons. First, it distinguishes circuit ansätze with L_{ad} and $L_{\min}$ layers, since for L_{ad}, $F = 0$ because only the global minimum is present, compared to $L_{\min}$ where other local minima with lower Max-Cut probabilities are present. Second, F reflects the contribution

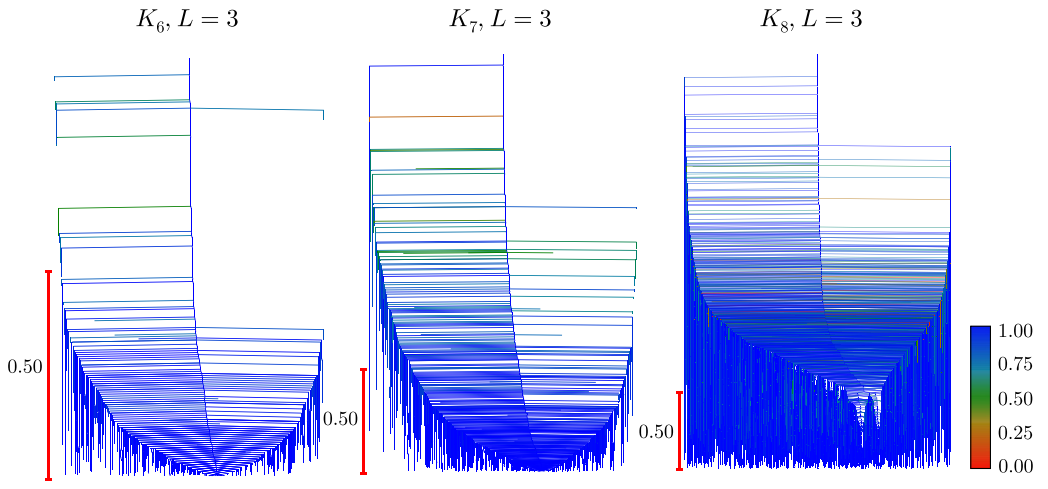

FIG. 4. Disconnectivity graphs of K_6, K_7, and K_8 for $L = 3$.

TABLE II. The weighted average metric F for graphs K_5 to K_8 of varying L.

Graph	$L=1$	$L=2$	$L=3$
K_5	0.002276	0.067219	0.000000
K_6	0.083438	0.060651	0.007907
K_7	0.005269	0.119087	0.026351
K_8	0.081604	0.190701	0.038062

of minima for circuit ansätze with $L < L_{\min}$ more accurately, since it is possible for minima with lower $\langle \hat{H}_C \rangle$ values, including the global minimum, to possess lower Max-Cut probabilities than their counterparts with higher $\langle \hat{H}_C \rangle$ values. As L increases, a decrease in the value of F can generally be interpreted as an improvement in QAOA performance, since it corresponds to an increase in the proportion of local minima with better probabilities. However, the converse situation, where F increases as L increases, may not necessarily signify a drop in QAOA performance, as the well-funneled organization of the cost landscape and the guarantee of obtaining a better Max-Cut probability, may outweigh the trade-off in obtaining a lower proportion of local minima with relatively good probabilities. Nevertheless, we propose choosing circuit ansätze with L layers that feature sufficiently low values of F when simulating QAOA on noisy quantum devices, as choosing circuits with higher L may also increase the impact of quantum gate and qubit decoherence noise.

Analyzing the F values for K_5 to K_8 (Table II), we observe that for K_5, $F = 0$ for $L_{ad} = 3$, differentiating it from $L_{\min} = 2$, as expected. Interestingly, with the exception of K_6, F appears to increase for $L = 2$ before decreasing for $L = 3$. The increase in F can mainly be attributed to the general increase in the number of local minima for $L = 2$ with lower probabilities than that of the single global minimum for $L = 1$. This trend is evident in the scatter plots of the probabilities of the correct Max-Cut solution against $\langle \hat{H}_C \rangle$ for the databases of minima for varying L (Fig. 5). The somewhat triangular convex hull of the solution space, C_s, seems to become better defined with the transition from $L = 2$ to $L = 3$. We also see a proliferation of local minima towards the apex of the global minimum, which would explain the observed subsequent decrease in F. Thus, a choice of $L = 3$ would be adequate for solving the Max-Cut problem for graphs K_5 to K_8 based on their F values.

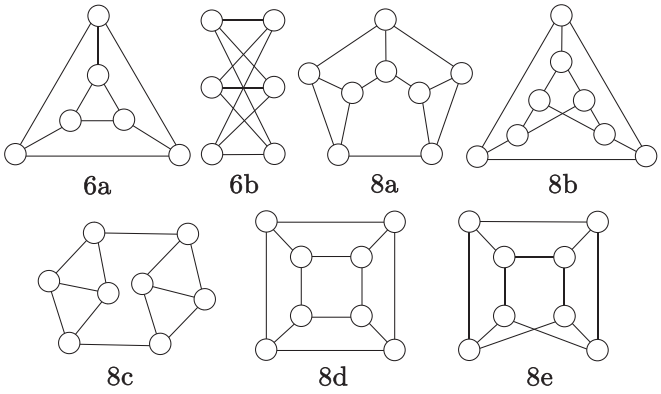

FIG. 6. 3-regular unweighted graphs 6a–8e investigated in this study.

B. 3-regular unweighted graphs

Next, we analyzed all connected 3-regular unweighted graphs with six and eight vertices [66] from $L = 1$ to $L = 4$, labeled 6a–8e, respectively (Fig. 6). In terms of the number of minima obtained from basin-hopping runs, the 3-regular graphs generally possess lower M values than the complete graphs of K_6 and K_8, particularly for graphs with eight vertices (Table III). For graph 6b, starting from $L = 1$ its QAOA ansatz gives rise to more than one minimum, which subsequently produces a more rapid increase in M compared to its counterpart 6a. This phenomenon can largely be attributed to the relatively more complex analytic expression of $\langle \hat{H}_C \rangle$ with the $L = 1$ circuit ansatz [67] for 6b. Another interesting pair of graphs is 8d and 8e, where, although their $L = 1$ analytic expressions of $\langle \hat{H}_C \rangle$ are identical and thus give rise to the same $\langle \hat{H}_C \rangle_{\min}$ value [67], we find that the trends of M and HCMPs for higher L are markedly different. Although the 3-regular series of graphs give rise to lower M for similar L, and hence somewhat simpler energy landscapes compared to their complete graph counterparts, their HCMPs are also comparatively lower, hence requiring a greater L to achieve a

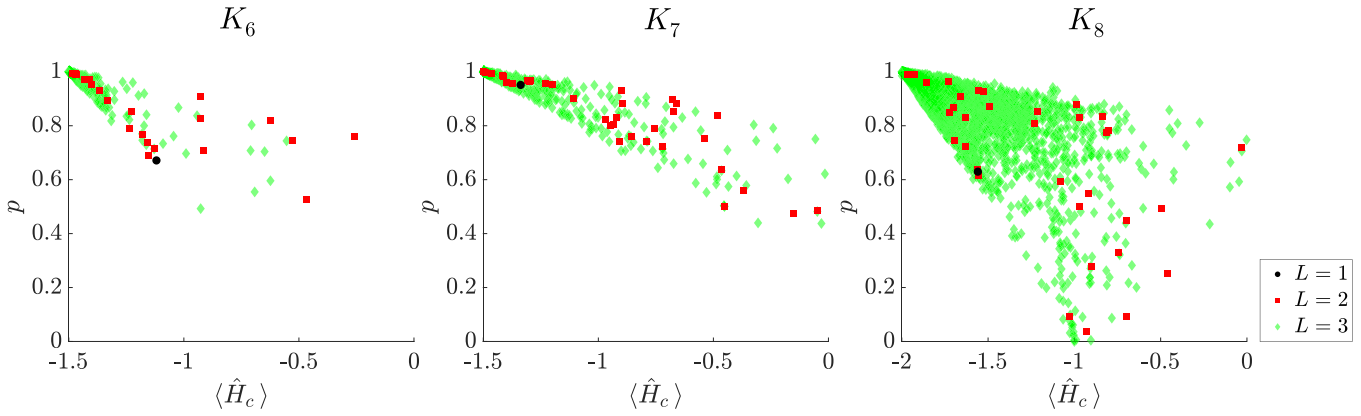

FIG. 5. Scatter plots of probabilities of the correct Max-Cut solutions against $\langle \hat{H}_C \rangle$ for individual minima of graphs K_6, K_7, and K_8 for $L = 1$ (black circle), $L = 2$ (red square), and $L = 3$ (green diamond).

TABLE III. Number of minima M (top value), HCMPs (middle value) and F (bottom value) for graphs 6a–8e of varying L obtained from basin-hopping global optimization. HCMPs with asterisks indicate that they correspond to the next highest local minimum rather than the global minimum.

Graph	$L=1$	$L=2$	$L=3$	$L=4$
	1	5	23	145
6a	0.400816	0.720917	0.933445*	0.996304
	0.254240	0.224891	0.167129	0.093993
	3	20	191	1451
6b	0.274835	0.659823	0.915348	0.978625
	0.772861	0.615867	0.510817	0.340108
	1	4	16	83
8a	0.142701	0.349371	0.616672	0.748746
	0.427166	0.363476	0.497637	0.428079
	1	4	15	97
8b	0.232056	0.420045*	0.638057	0.767138
	0.355057	0.325065	0.426828	0.380643
	1	4	18	82
8c	0.077352	0.182753	0.536762	0.701246
	0.516636	0.560946	0.519856	0.429701
	1	6	31	151
8d	0.186302	0.520680	0.871573	0.972013
	0.500506	0.360078	0.258910	0.141219
	1	4	15	110
8e	0.321737	0.574918*	0.769975	0.918878*
	0.286668	0.241691	0.302882	0.300042

sufficiently high HCMP. In some cases, the HCMPs were derived not from the global minimum, but from the next highest local minimum. (See Appendix B detailing the differences in Max-Cut probabilities and $\langle \hat{H}_C \rangle$ values between the two minima for these cases.) This phenomenon appears sporadically without much predictability, most notably for graph 8e, where the HCMP corresponds to the next highest local minimum for $L=2$ and $L=4$. Overall, these observations further underline the importance of evaluating QAOA performance based on the correct Max-Cut probabilities of individual minima independently alongside their $\langle \hat{H}_C \rangle$ values.

The distributions of minima for the 3-regular graphs with varying L differ significantly from their complete graph counterparts (Fig. 7). We observe that the convex hulls of the 3-regular graphs tend to take on a more compact shape, with greater correlation between the $p(|s\rangle)$ and $\langle \hat{H}_C \rangle$ values for the individual minima. However, for graphs 8a, 8b, and 8e there is a notable absence of minima with intermediate $p(|s\rangle)$ and $\langle \hat{H}_C \rangle$ values, particularly for higher L. This structure is also reflected in their disconnectivity graphs (refer to Appendix C for the disconnectivity graphs of 6a–8e). Another major difference of the 3-regular graphs is the much reduced energy differences between minima with low proximity to the global minimum and their connected transition states, producing more streamlined and single-funneled disconnectivity graphs than for K_6 and K_8. However, even though the energy landscapes of the 3-regular graphs appear less complex and easier to navigate than their complete graph counterparts, the local minima in their energy landscapes give rise to a larger range of $p(|s\rangle)$. Hence a greater proportion of local minima with high energies possess low Max-Cut probabilities. This trend is captured by comparing F values between the 3-regular graphs and the complete graphs K_6 and K_8, where the former graphs typically have much higher F values than the latter.

Comparing F values among the 3-regular graphs, we observe that graphs 6a and 6b follow a smooth downward trend with increasing L, while the 8-vertex graphs tend to peak at $L=2$ and $L=3$ before decreasing, with the exception of 8d, which follows a similar trend to the 6-vertex graphs. The performance of the 8-vertex graphs at $L=2$ and $L=3$ can be attributed to an increase in the number of local minima with low Max-Cut probabilities that outweigh the general improvement in HCMPs and low-lying minima, while for graph 8d this effect is reversed, with an increase in HCMPs and minima with good probabilities. At $L=4$, a greater proportion of minima with high $p(|s\rangle)$ appear for all 3-regular graphs, and we therefore recommend a minimum of $L=4$ when employing QAOA for these cases.

Another factor that supports the choice of $L=4$ comes from the construction of heuristic expectation thresholds that aim to identify minima with sufficiently high $p(|s\rangle)$ values. This analysis can be carried out by finding the intercepts of the corresponding convex hulls with a suitable probability cutoff p_{op}. For the 3-regular graphs we choose $p_{op}=0.5$ and define the difference in $\langle \hat{H}_C \rangle$ values from the global minimum to the two intercepts as the worst-case and best-case expectation cutoffs, d_1 and d_2, respectively, where $d_1 < d_2$. We observe that the expectation thresholds generally expand as L increases (Table IV), with d_1 and d_2 attaining their highest values at $L=4$. For $L \leqslant L_{\min}$ the widening and stabilizing of expectation thresholds is significant, along with the increase in M as L increases. We see that a greater number of minima that possess a wider range of $\langle \hat{H}_C \rangle$ values with a sufficiently high $p(|s\rangle)$ exist within the solution landscape for the QAOA ansatz.

C. Competing QAOA Max-Cut solutions

Finally, we explore competing Max-Cut solutions $|\alpha\beta\alpha\beta\rangle$ and $|\alpha\alpha\beta\beta\rangle$ for a series of four-vertex weighted graphs with a common variable weight x, where $x=(0,3,4,5)$ correspond to the graphs (G_2, G_3, G_4, G_5), respectively (Fig. 8). These graphs open up two modes of analysis: they allow comparison between G_2 with the more complex graphs $G_3 - G_5$, particularly with G_3, since both sets of graphs have $|s\rangle = |\alpha\beta\alpha\beta\rangle$. Comparisons between $G_3 - G_5$ can also be carried out, since G_5 possesses a different correct Max-Cut solution of $|\alpha\alpha\beta\beta\rangle$, while for G_4 both $|\alpha\beta\alpha\beta\rangle$ and $|\alpha\alpha\beta\beta\rangle$ are correct Max-Cut solutions. We will denote the alternative Max-Cut solution for a given graph as $|t\rangle$, thus for G_3, $|t\rangle = |\alpha\alpha\beta\beta\rangle$; for G_5, $|t\rangle = |\alpha\beta\alpha\beta\rangle$, and G_4 has no alternative solution.

We find that implementing QAOA for the weighted graphs $G_2 - G_5$ is more difficult than for the complete unweighted graph K_4, as their energy landscapes are much more complex due to an increase in M and a decrease in their respective HCMPs (Table V). The disconnectivity graphs of $G_2 - G_5$ exhibit similar topological features to the 3-regular unweighted

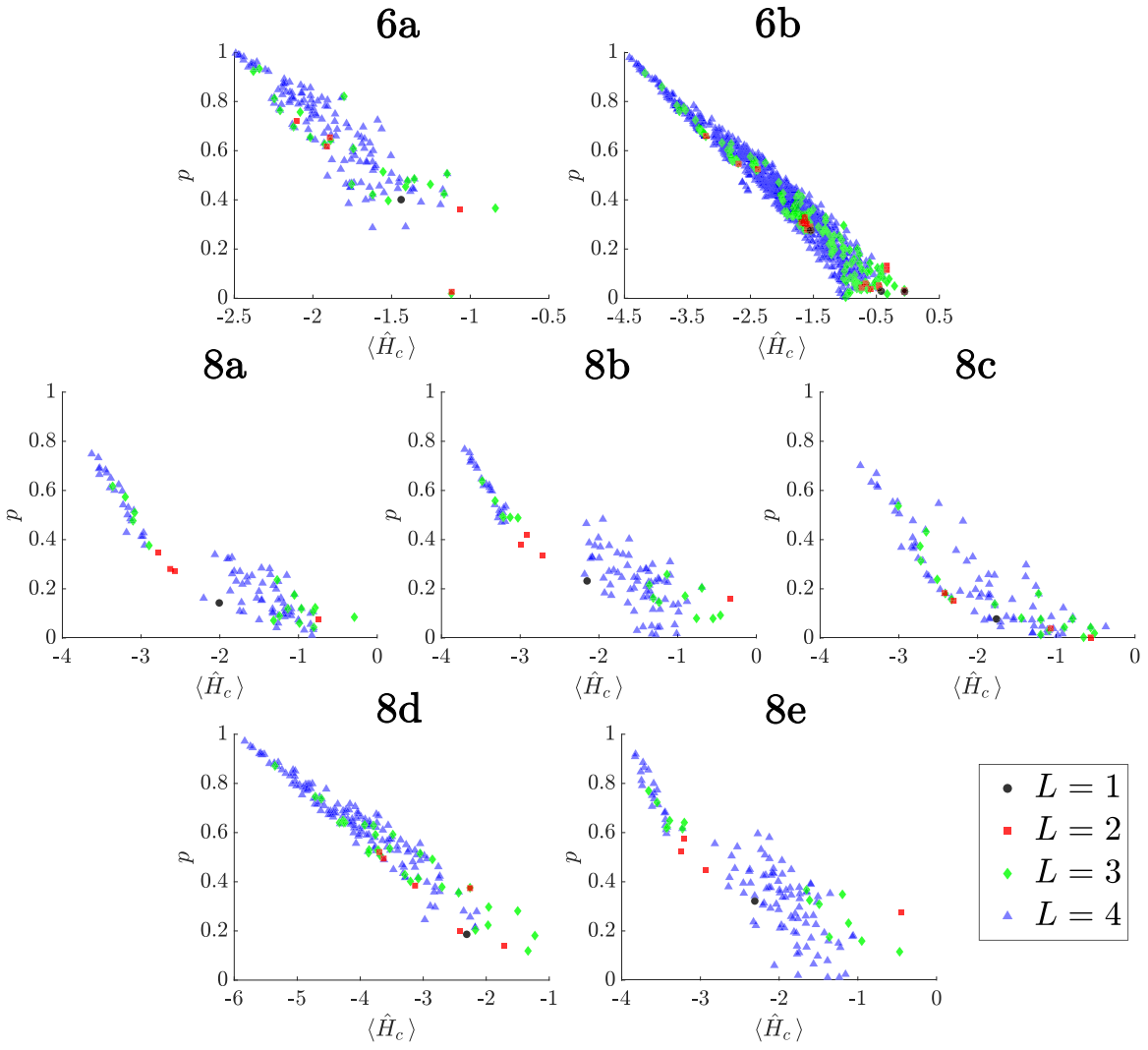

FIG. 7. Scatter plots of probabilities of the correct Max-Cut solutions against $\langle \hat{H}_C \rangle$ for connected minima of graphs 6a–8e for $L = 1$ (black circle), $L = 2$ (red square), $L = 3$ (green diamond), and $L = 4$ (blue triangle).

graphs, possessing a well-funneled organization and minima featuring a wide range of $\langle \hat{H}_C \rangle$ and Max-Cut probabilities [see Appendix D for the disconnectivity graphs of $G_2 - G_5$ colored based on $p(|s\rangle)$, and Appendix E for the disconnectivity graphs of G_3 and G_5 colored based on $p(|t\rangle)$]. Unsurprisingly, G_2 has a lower M and thus a simpler landscape than the graphs $G_3 - G_5$, although its collection of minima with modest Max-Cut probabilities produces a comparatively high F value up to $L = 3$. In the range $G_3 - G_5$, it is interesting that even though G_5 with a different Max-Cut solution has a more complex energy landscape than G_3, it yields a comparatively higher HCMP, while G_4 exhibits the best QAOA performance as it has two distinct Max-Cut solutions. The phenomenon where the HCMP arises for the next highest local minimum rather than the global minimum was also observed for $G_3 - G_5$, especially for G_5 (refer to Appendix B for differences in Max-Cut probabilities and $\langle \hat{H}_C \rangle$ values between the two minima for these cases). All four graphs feature a monotonic decrease in F as L increases. Hence, as for $K_5 - K_8$, we recommend choosing $L = 3$ in solving the Max-Cut problem for $G_2 - G_5$. Finally, it is noteworthy that for G_4 and the $L = 4$ ansatz, the number of minima greatly exceeds that of the $L_{\min} = 3$ ansatz, in contrast to the behavior observed for K_5, highlighting the general unpredictability of the energy landscape complexity after $L_{\min}$.

For graphs $G_3 - G_5$, the scatter plots of the $\langle \hat{H}_C \rangle$ values for connected minima and Max-Cut probabilities were used to construct both the convex hulls of the correct Max-Cut solution C_s and the alternative solution C_t (Fig. 9). We find that, similar to the unweighted graphs, both C_s and C_t take on more definite shapes as they become populated with more minima for increasing L. This trend allowed us to investigate d_1 and d_2 by finding the intercepts of C_s with the horizontal line p_{op}, setting $p_{op} = 0.5$ for G_3 and G_5, and $p_{op} = 0.25$ for G_4 (Table VI). Additionally, one may also use the intercepts between C_s and the left edge of C_t as expectation and threshold cutoffs to identify higher-quality minima with a high probability of finding $|s\rangle$ and a low probability of obtaining $|t\rangle$. (See Appendix F for more details.) As for the 3-regular graphs, the divergence of d_1 and d_2 with increasing L makes $L = 3$ a good choice for graphs $G_3 - G_5$.

TABLE IV. Expectation thresholds d_1 (top value) and d_2 (bottom value) for graphs 6a–8e of varying L.

Graph	$L=2$	$L=3$	$L=4$
6a	0.313189	0.546894	0.578616
	0.643070	1.254700	1.342016
6b	0.640787	1.453239	1.660319
	0.951659	2.160089	2.470976
8a	—	0.215689	0.338515
	—	0.673507	1.251349
8b	—	0.250008	0.408135
	—	0.886500	1.652028
8c	—	0.046924	0.315151
	—	0.183791	1.203264
8d	0.052613	1.580115	1.857498
	0.202901	2.457872	2.962362
8e	0.096120	0.779749	0.645478
	0.731083	1.578323	1.977320

IV. CONCLUSION

In this work we explore the solution landscapes of QAOA ansätze applied to a variety of weighted and unweighted graphs by means of the energy landscapes framework, using disconnectivity graphs to visualize their topological features. We find that the corresponding landscapes are largely funneled, suggesting that location of low-lying minima should not be particularly difficult. Under practical conditions when simulating QAOA on a quantum device, the optimization regime is thus more likely to find a minimum close to the global minimum with a good correct Max-Cut solution probability. Even under the worst-case scenario where each experiment finds a different local minimum, so long as the local minimum is sufficiently close to the global minimum, a significant proportion of the number of shots per experiment will correspond to the Max-Cut solution. This result further demonstrates the robustness of QAOA in solving the Max-Cut problem.

We have also developed a weighted average metric F to evaluate the performance of QAOA ansätze from their corresponding databases of minima. This parameter allows one to choose a suitable number of circuit layers that balances the likelihood of obtaining good solution probabilities from local minima with an adequate circuit depth that minimizes the impact from quantum noise.

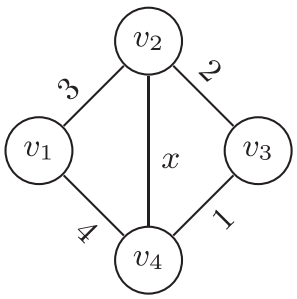

FIG. 8. Four-vertex weighted graph with a variable central weight x. The graphs (G_2, G_3, G_4, G_5) correspond to $x = (0, 3, 4, 5)$, respectively.

TABLE V. Number of minima M (top value), HCMPs (middle value) and F (bottom value) for graphs $G_2 - G_5$ with varying L obtained from basin-hopping global optimization. HCMPs with asterisks indicate that they were obtained from the next highest local minimum instead of the global minimum.

Graph	$L=1$	$L=2$	$L=3$
G_2	3	16	87
	0.477824	0.866404	0.981615
	0.594432	0.338167	0.163886
G_3	7	109	1835
	0.370522*	0.784475	0.948704*
	0.526539	0.300520	0.171897
G_4	4	36	308
	0.540426*	0.894074	1.000000
	0.461355	0.245864	0.111340
G_5	9	183	3426
	0.468114*	0.881732*	0.972620*
	0.498733	0.328245	0.174754

Finally, we have established two ways in which expectation thresholds can be established to determine the cutoff for minima with high $p(|s\rangle)$. The solution landscapes we have characterized suggest that QAOA is a good VQA candidate to demonstrate practical quantum advantage. In future work we plan to extend these results to quantum machine learning (QML) algorithms, such as variational quantum classifiers (VQCs), which minimize a given cost function to classify data [68].

The GMIN, OPTIM, and PATHSAMPLE programs are available for use under the Gnu General Public License. They can be downloaded from the Cambridge Landscape Database at [69].

APPENDIX A: BASIN-HOPPING GLOBAL OPTIMIZATION WITH GMIN

For each basin-hopping optimization run, we performed 10 000 basin-hopping steps for the unweighted graphs $K_3 - K_8$, 6a−8e, and for the weighted graphs $G_2 - G_5$, using the GMIN program for varying L. Each local minimization had a minimum root-mean squared (RMS) gradient convergence criterion of 10^{-10} a.u., where the analytic gradients of the parameterized rotation gates of the cost and mixer layers of the QAOA ansatz were evaluated with the parameter-shift rule using Eq. 7. To accept or reject basin-hopping steps we employed a Metropolis criterion with a basin-hopping temperature of 1.0 a.u. If the minimum at step j has an expectation value $\langle \hat{H}_C \rangle_j$ that is lower than the preceding iteration, i.e., $\langle \hat{H}_C \rangle_j < \langle \hat{H}_C \rangle_{j-1}$, then the corresponding angular coordinates $\boldsymbol{\theta}_j$ are accepted and used for the next step. If $\langle \hat{H}_C \rangle_j \geqslant \langle \hat{H}_C \rangle_{j-1}$, then $\boldsymbol{\theta}_j$ is accepted with a probability of $\exp(-(\langle \hat{H}_C \rangle_j - \langle \hat{H}_C \rangle_{j-1})/kT)$. Otherwise, the new minimum is rejected.

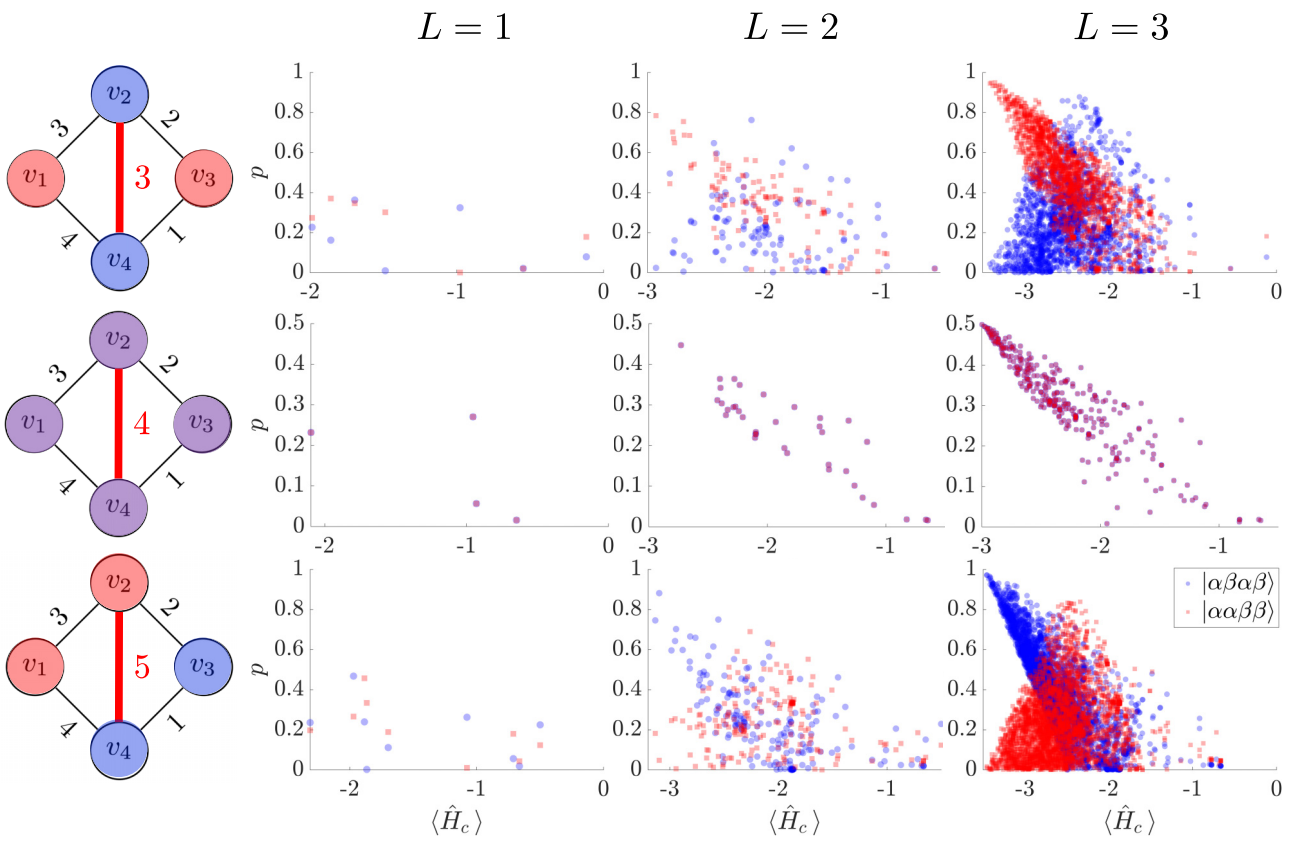

FIG. 9. Scatter plots of probabilities of competing Max-Cut solutions $|\alpha\beta\alpha\beta\rangle$ (red square) and $|\alpha\alpha\beta\beta\rangle$ (blue circle) against $\langle\hat{H}_C\rangle$ for connected minima of graphs G_3, G_4, and G_5 with varying L.

Basin-hopping moves were proposed by random perturbations of up to 1.0 rad for each angular coordinate in $\boldsymbol{\theta}_j$. At the end of each run, the collection of minima that differ by at least 1.0×10^{-9} a.u in their $\langle\hat{H}_C\rangle$ values were saved to provide a starting database for construction of the energy landscape using the OPTIM and PATHSAMPLE programs.

APPENDIX B: DIFFERENCES IN MAX-CUT PROBABILITY AND EXPECTATION VALUES FOR NONGLOBAL HCMP CASES

Table VII shows the expectation differences $\Delta\langle\hat{H}_C\rangle = \langle\hat{H}_C\rangle_{\min} - \langle\hat{H}_C\rangle_{\text{HCMP}}$ and correct Max-Cut probability differences $\Delta p(|s\rangle) = \text{HCMP} - p_{\text{GM}}(|s\rangle)$ for cases where the next highest local minimum has a Max-Cut probability greater than the global minimum. In general, cases with a large circuit depth L tend to have lower $\Delta\langle\hat{H}_C\rangle$ and $\Delta p(|s\rangle)$ as their HCMPs approach the optimal value of 1. As mentioned in the main text, this phenomenon appears to occur sporadically, as exemplified by 8e and G_3, where it does not occur for the intermediate circuit depth of $L = 3$. A notable case is G_5, which features a higher Max-Cut probability in the next highest local minimum for all sampled L. We propose to investigate this phenomenon in future work to see if it arises systematically for particular classes of connected graphs.

TABLE VI. Expectation thresholds d_1 (top value) and d_2 (bottom value) for graphs $G_3 - G_5$ of varying L.

Graph	$L = 2$	$L = 3$
G_3	0.099242	0.433945
	1.124694	2.008400
G_4	0.489057	0.536238
	1.447505	1.722258
G_5	0.217591	0.459177
	1.566966	1.896063

TABLE VII. Expectation differences $\Delta\langle\hat{H}_C\rangle$ and Max-Cut probability differences $\Delta p(|s\rangle)$ between the global and next highest local minima for nonglobal HCMP cases.

Graph	L	$\Delta\langle\hat{H}_C\rangle$	$\Delta p(\|s\rangle)$
6a	3	0.037856	0.010121
8b	2	0.075696	0.040029
8e	2	0.046450	0.052876
8e	4	0.006337	0.009846
G_3	1	0.127864	0.096306
G_3	3	0.015904	0.003739
G_4	1	1.142149	0.076922
G_5	1	0.342188	0.231787
G_5	2	0.036357	0.135824
G_5	3	0.001111	0.001329

APPENDIX C: DISCONNECTIVITY GRAPHS OF THE 3-REGULAR GRAPHS

Figures 10–16 depict the disconnectivity graphs for the 3-regular graphs 6a–8e. The 3-regular graphs generally display a wider variation and higher correlation of $p(|s\rangle)$ and $\langle \hat{H}_C \rangle$ values for individual minima as compared to the complete graphs, along with a largely well-funneled landscape organization. For graphs 8a, 8b, and 8e (Figs. 12, 13, 16) there is a notable dearth of minima with intermediate $p(|s\rangle)$ and $\langle \hat{H}_C \rangle$ values.

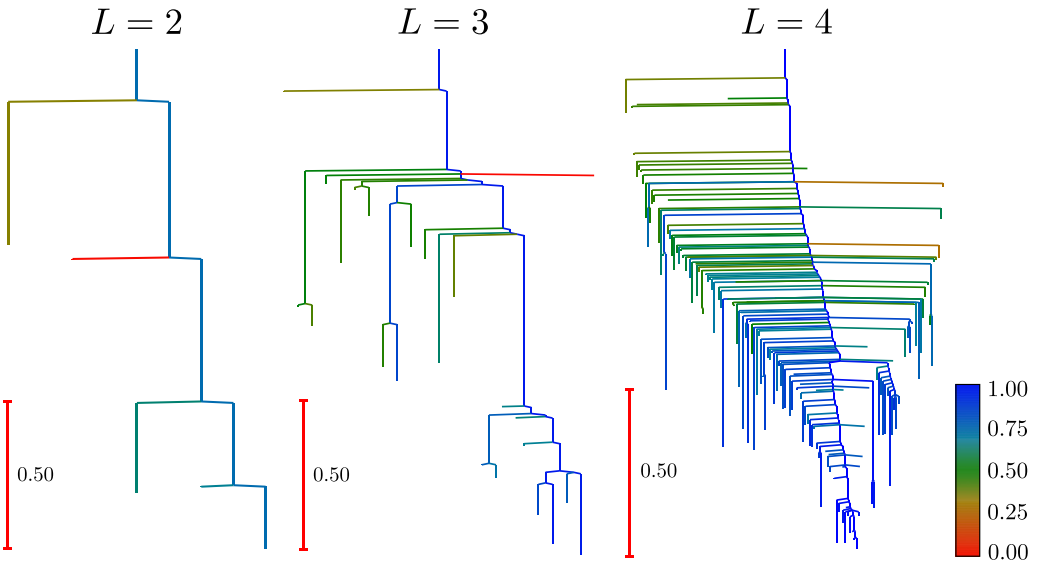

FIG. 10. Disconnectivity graphs for 6a with varying circuit depth L.

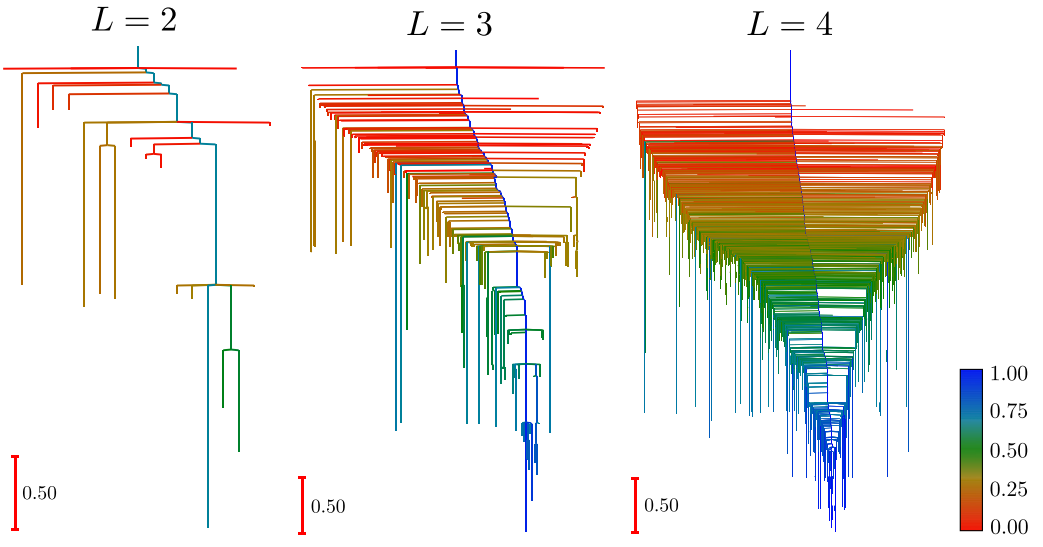

FIG. 11. Disconnectivity graphs for 6b with varying circuit depth L.

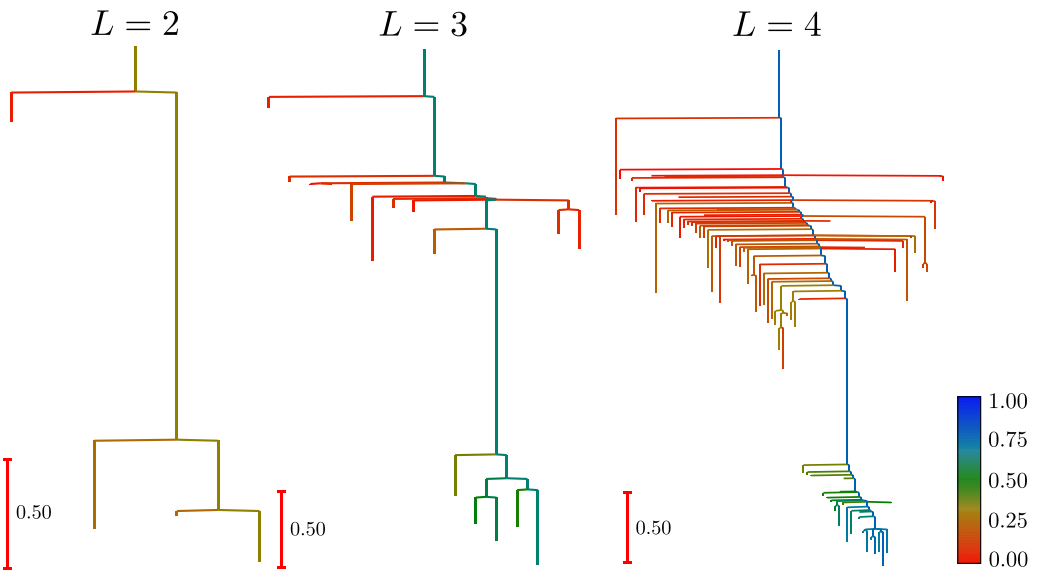

FIG. 12. Disconnectivity graphs for 8a with varying circuit depth L.

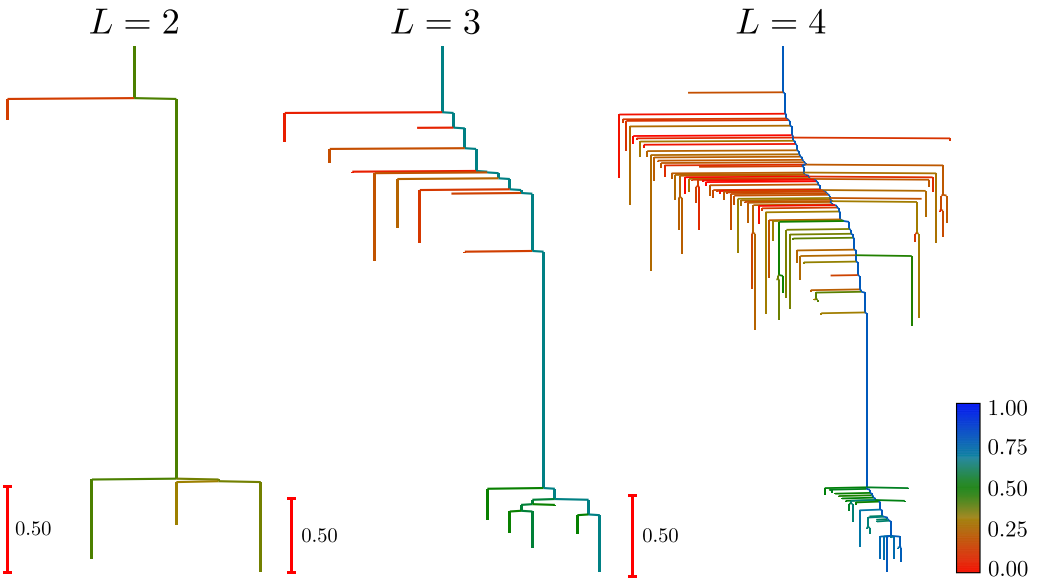

FIG. 13. Disconnectivity graphs for 8b with varying circuit depth L.

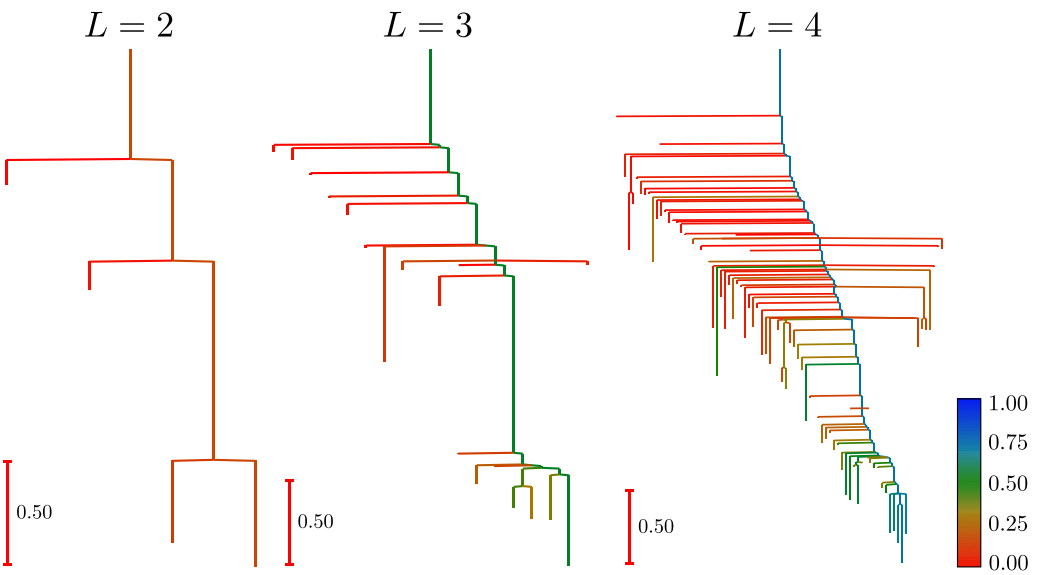

FIG. 14. Disconnectivity graphs for 8c with varying circuit depth L.

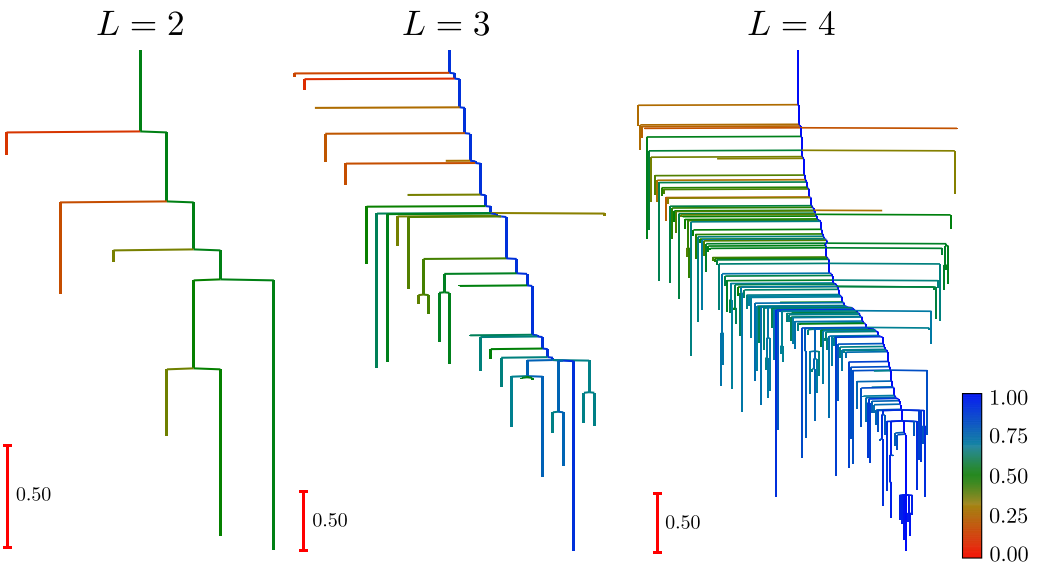

FIG. 15. Disconnectivity graphs for 8d with varying circuit depth L.

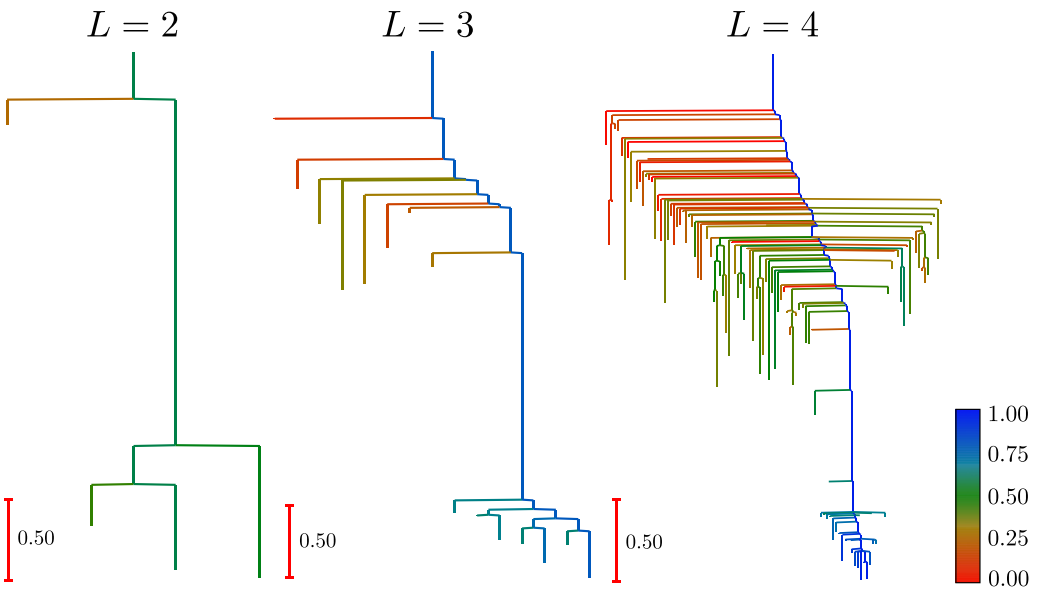

FIG. 16. Disconnectivity graphs for 8e with varying circuit depth L.

APPENDIX D: DISCONNECTIVITY GRAPHS OF THE CORRECT MAX-CUT SOLUTIONS FOR $G_2 - G_5$

Figures 17–20 depict the disconnectivity graphs of the 4-vertex graph series $G_2 - G_5$, with individual minima coloured based on their respective probabilities of finding the correct Max-Cut solution. Similar to the disconnectivity graphs of the 3-regular graphs featured in Appendix C, the disconnectivity graphs of $G_2 - G_5$ feature a broader range of $\langle \hat{H}_C \rangle$ and correct Max-Cut probabilities.

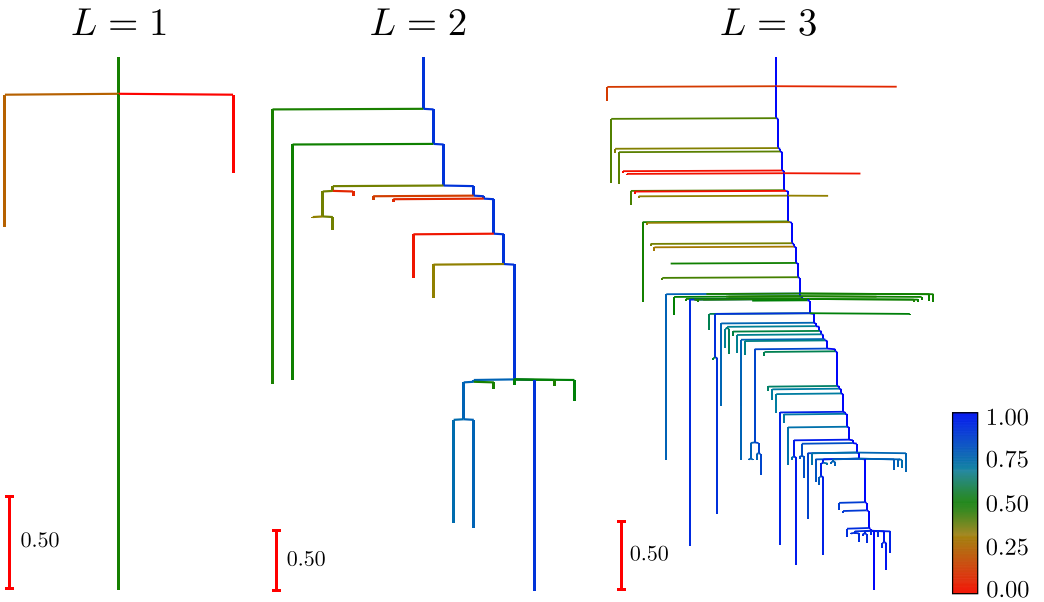

FIG. 17. Disconnectivity graphs for G_2 with varying circuit depth L. Minima are colored based on the probability of obtaining the optimal Max-Cut state of $|\alpha\beta\alpha\beta\rangle$.

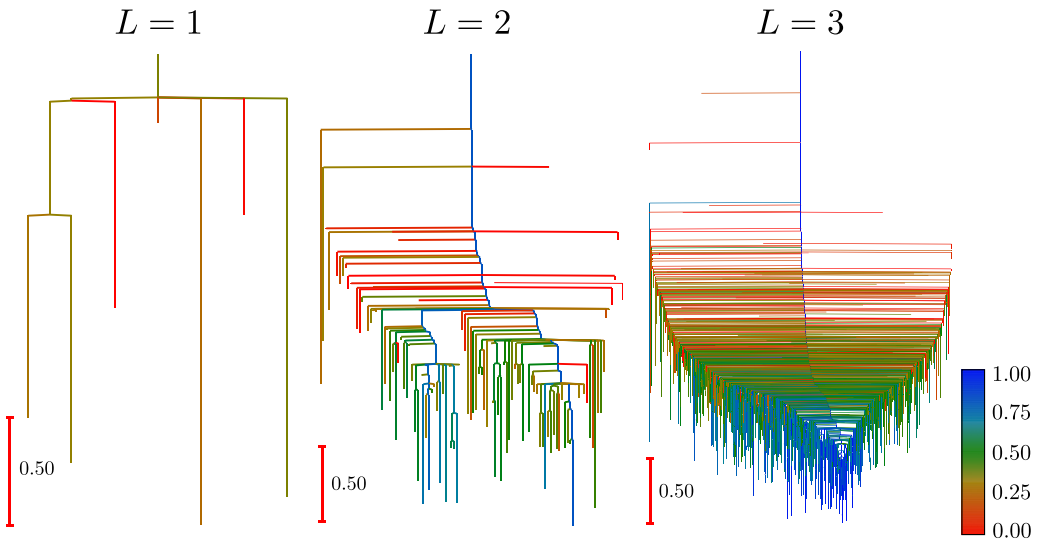

FIG. 18. Disconnectivity graphs for G_3 with varying circuit depth L. Minima are colored based on the probability of obtaining the optimal Max-Cut state of $|\alpha\beta\alpha\beta\rangle$.

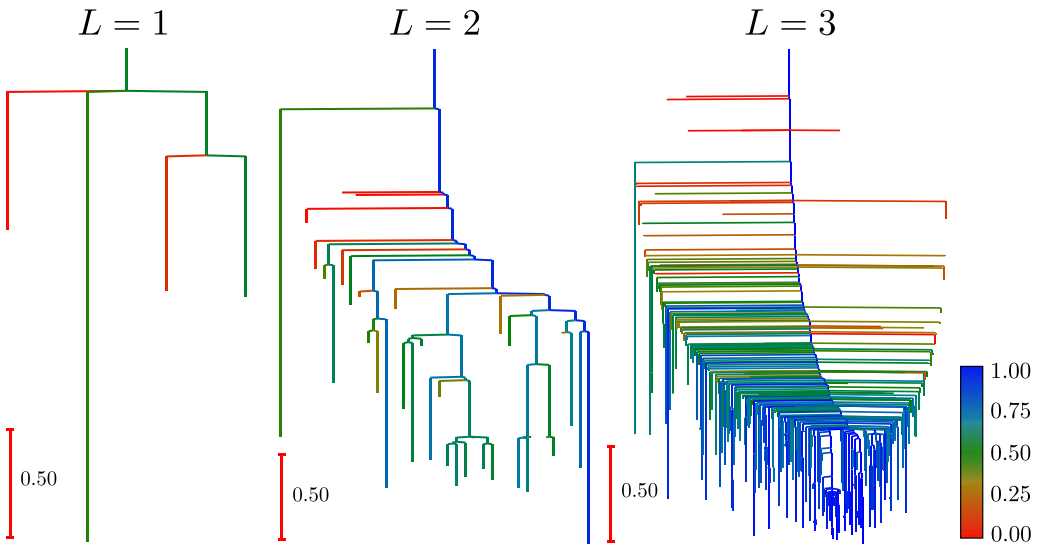

FIG. 19. Disconnectivity graphs for G_4 with varying circuit depth L. Minima are colored based on the probability of obtaining the optimal Max-Cut states of $|\alpha\beta\alpha\beta\rangle$ and $|\alpha\alpha\beta\beta\rangle$.

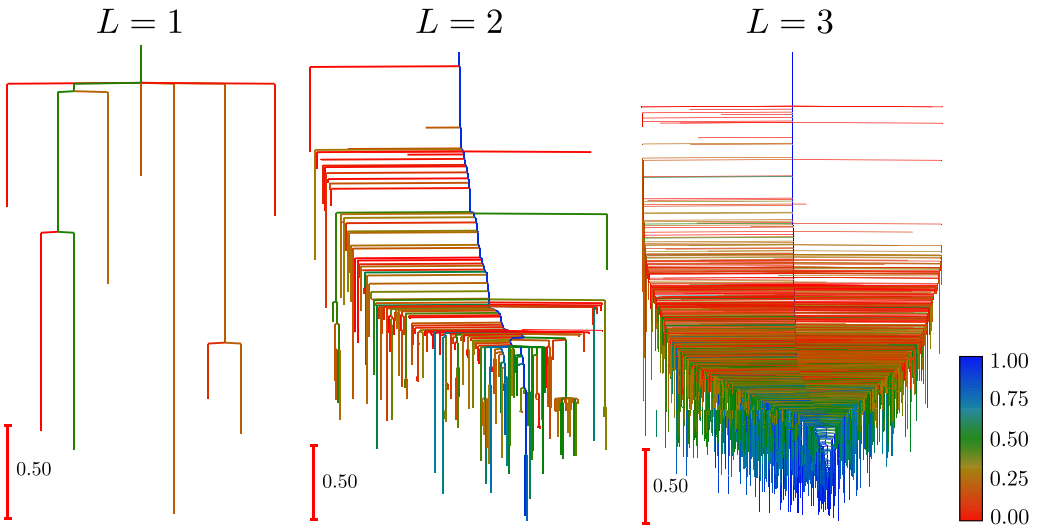

FIG. 20. Disconnectivity graphs for G_5 with varying circuit depth L. Minima are colored based on the probability of obtaining the optimal Max-Cut state of $|\alpha\alpha\beta\beta\rangle$.

APPENDIX E: DISCONNECTIVITY GRAPHS OF THE ALTERNATIVE MAX-CUT SOLUTIONS FOR G_3 AND G_5

Figures 21 and 22 depict the disconnectivity graphs of the 4-vertex graph series G_3 and G_5, with individual minima colored based on their respective probabilities of finding the alternative Max-Cut solutions of $|\alpha\alpha\beta\beta\rangle$ and $|\alpha\beta\alpha\beta\rangle$, respectively. In contrast to their respective correct Max-Cut counterparts (Fig. 18 and 20), minima with the highest alternative-solution Max-Cut probabilities possess intermediate $\langle\hat{H}_C\rangle$ values.

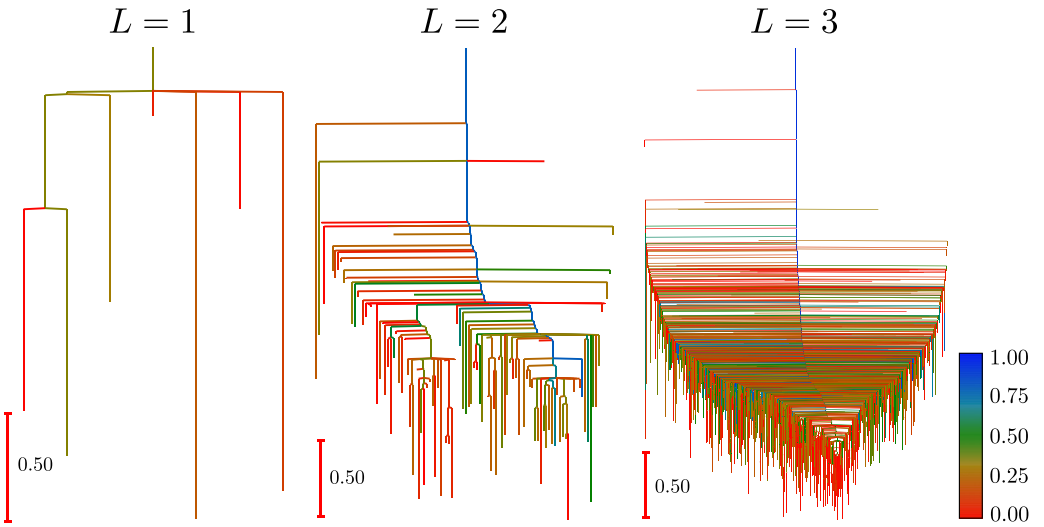

FIG. 21. Disconnectivity graphs for G_3 with varying circuit depth L. Minima are colored based on the probability of obtaining the opposing Max-Cut state of $|\alpha\alpha\beta\beta\rangle$.

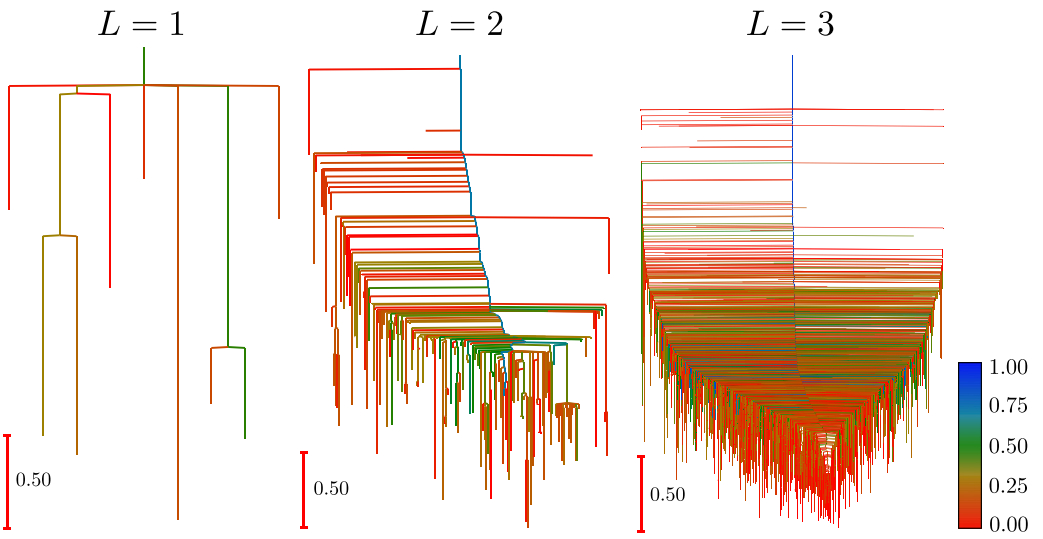

FIG. 22. Disconnectivity graphs for G_5 with varying circuit depth L. Minima are colored based on the probability of obtaining the opposing Max-Cut state of $|\alpha\beta\alpha\beta\rangle$.

APPENDIX F: EXPECTATION AND PROBABILITY THRESHOLDS FROM THE CONVEX HULL INTERCEPTS FOR G_3 AND G_5

Tables VIII and IX summarize the two intercepts $(\langle\hat{H}_C\rangle_{\min} + d_3, p_1)$ and $(\langle\hat{H}_C\rangle_{\min} + d_4, p_2)$ of C_s with the left edge of the triangular convex hull C_t for graphs G_3 and G_5, as shown in Fig. 9. This summary provides an alternative method to establish expectation cutoffs if additional information for alternative Max-Cut states is available. As for the expectation thresholds d_1 and d_2 in Table VI, d_3 and d_4 also increase with increasing L, which further justifies the choice of $L = 3$ in refining minima within C_s that possess both good correct Max-Cut probabilities and low opposing Max-Cut probabilities. The probability cutoff of $p_{op} = 0.5$ used in Table VI also appears to be situated between the probability thresholds p_1 and p_2, suggesting that this choice of p_{op} is optimal.

TABLE VIII. Expectation thresholds d_3 (top value) and d_4 (bottom value) for graphs G_3 and G_5 of varying L, where $d_3 < d_4$.

Graph	$L = 2$	$L = 3$
G_3	0.111125	0.439036
	0.546385	0.794808
G_5	0.329259	0.476747
	0.856478	0.819369

TABLE IX. Max-Cut probability thresholds p_1 (top value) and p_2 (bottom value) for graphs G_3 and G_5 of varying L, corresponding to the expectation thresholds d_3 and d_4, respectively.

Graph	$L = 2$	$L = 3$
G_3	0.465937	0.495366
	0.665681	0.799744
G_5	0.373801	0.482335
	0.680284	0.790380

[1] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding *et al.*, Scalable quantum simulation of molecular energies, Phys. Rev. X **6**, 031007 (2016).

[2] F. Leymann and J. Barzen, The bitter truth about gate–based quantum algorithms in the NISQ era, Quantum Sci. Technol. **5**, 044007 (2020).

[3] U. Skosana and M. Tame, Demonstration of Shor's factoring algorithm for $N = 21$ on IBM quantum processors, Sci. Rep. **11**, 16599 (2021).

[4] J. J. Burnett, A. Bengtsson, M. Scigliuzzo, D. Niepce, M. Kudra, P. Delsing, and J. Bylander, Decoherence benchmarking of superconducting qubits, npj Quantum Inf. **5**, 54 (2019).

[5] S. Chen, J. Cotler, H. Huang, and J. Li, The complexity of NISQ, Nat. Commun. **14**, 6001 (2023).

[6] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru–Guzik, The theory of variational hybrid quantum–classical algorithms, New J. Phys. **18**, 023023 (2016).

[7] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru–Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[8] K. Bharti, A. Cervera–Lierta, T. H. Kyaw, T. Haug, S. Alperin–Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke *et al.*, Noisy intermediate–scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[9] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[10] G. Ravi, K. N. Smith, P. Gokhale, A. Mari, N. Earnest, A. Javadi-Abhari, and F. T. Chong, VAQEM: A variational approach to quantum error mitigation, in *2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA), Seoul, Korea, Republic of (2022)*, pp. 288–303.

[11] K. Ito, W. Mizukami, and K. Fujii, Universal noise–precision relations in variational quantum algorithms, Phys. Rev. Res. **5**, 023025 (2023).

[12] E. Fontana, N. Fitzpatrick, D. M. Ramo, R. Duncan, and I. Rungger, Evaluating the noise resilience of variational quantum algorithms, Phys. Rev. A **104**, 022403 (2021).

[13] S. Halder, C. Shrikhande, and R. Maitra, Development of zero-noise extrapolated projective quantum algorithm for accurate evaluation of molecular energetics in noisy quantum devices, J. Chem. Phys. **159**, 114115 (2023).

[14] A. Weidinger, G. B. Mbeng, and W. Lechner, Error mitigation for quantum approximate optimization, Phys. Rev. A **108**, 032408 (2023).

[15] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, The future of quantum computing with superconducting qubits, J. Appl. Phys. **132**, 160902 (2022).

[16] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[17] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm arXiv:1411.4028.

[18] Y. Zhang, L. Cincio, C. F. A. Negre, P. Czarnik, P. J. Colesa, P. M. Anisimov, S. M. Mniszewski, S. Tretiak, and P. A. Dub, Variational quantum eigensolver with reduced circuit complexity, npj Quantum Inf. **8**, 96 (2022).

[19] J. Wurtz and P. J. Love, Counterdiabaticity and the quantum approximate optimization algorithm, Quantum **6**, 635 (2022).

[20] G. Buonaiuto, F. Gargiulo, G. D. Pietro, M. Esposito, and M. Pota, Best practices for portfolio optimization by quantum computing, experimented on real quantum devices, Sci. Rep. **13**, 19434 (2023).

[21] V. Kremenetski, T. Hogg, S. Hadfield, S. J. Cotton, and N. M. Tubman, Quantum alternating operator ansatz (QAOA) phase diagrams and applications for quantum chemistry arXiv:2108.13056.

[22] S. Hadfield, T. Hogg, and E. G. Rieffel, Analytical framework for quantum alternating operator ansätze, Quantum Sci. Technol. **8**, 015017 (2023).

[23] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced Barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[24] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[25] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent Barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[26] D. Wierichs, C. Gogolin, and M. Kastoryano, Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer, Phys. Rev. Res. **2**, 043246 (2020).

[27] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nat. Commun. **13**, 7760 (2022).

[28] S. H. Sack, R. A. Medina, R. Kueng, and M. Serbyn, Recursive greedy initialization of the quantum approximate optimization algorithm with guaranteed improvement, Phys. Rev. A **107**, 062404 (2023).

[29] G. E. Crooks, Performance of the quantum approximate optimization algorithm on the maximum cut problem arXiv:1811.08419.

[30] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).

[31] D. J. Wales, Exploring energy landscapes, Annu. Rev. Phys. Chem. **69**, 401 (2018).

[32] Z. Li and H. A. Scheraga, Monte Carlo-minimization approach to the multiple-minima problem in protein folding, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).

[33] D. J. Wales and J. P. K. Doye, Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms, J. Phys. Chem. A **101**, 5111 (1997).

[34] D. J. Wales and H. A. Scheraga, Global optimization of clusters, crystals and biomolecules, Science **285**, 1368 (1999).

[35] D. J. Wales, Discrete path sampling, Mol. Phys. **100**, 3285 (2002).

[36] D. J. Wales, Some further applications of discrete path sampling to cluster isomerization, Mol. Phys. **102**, 891 (2004).

[37] B. Choy and D. J. Wales, Molecular energy landscapes of hardware–efficient ansätze in quantum computing, J. Chem. Theory Comput. **19**, 1197 (2023).

[38] H. G. A. Burton, D. Marti–Dafcik, D. P. Tew, and D. J. Wales, Exact electronic states with shallow quantum circuits from global optimisation, npj Quantum Inf. **9**, 75 (2023).

[39] D. J. Wales, M. A. Miller, and T. R. Walsh, Archetypal energy landscapes, Nature (London) **394**, 758 (1998).

[40] T. V. Bogdan and D. J. Wales, New results for phase transitions from catastrophe theory, J. Chem. Phys. **120**, 11090 (2004).

[41] D. J. Wales, GMIN: A program for finding global minima and calculating thermodynamic properties from basin-sampling, https://www-wales.ch.cam.ac.uk/GMIN/ (2023).

[42] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).

[43] C. G. Broyden, The convergence of a class of double-rank minimization algorithms 1. General considerations, J. Inst. Math. Appl. **6**, 76 (1970).

[44] R. Fletcher, A new approach to variable metric algorithms, Comput. J. **13**, 317 (1970).

[45] D. Goldfarb, A family of variable metric updates derived by variational means, Math. Comp. **24**, 23 (1970).

[46] D. F. Shanno, Conditioning of quasi-Newton methods for function minimization, Math. Comp. **24**, 647 (1970).

[47] D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Program. **45**, 503 (1989).

[48] J. Nocedal, Updating quasi-Newton matrices with limited storage, Math. Comp. **35**, 773 (1980).

[49] M. L. Paleico and J. Behler, A flexible and adaptive grid algorithm for global optimization utilizing basin hopping Monte Carlo, J. Chem. Phys. **152**, 094109 (2020).

[50] F. Noé and S. Fischer, Transition networks for modeling the kinetics of conformational change in macromolecules, Curr. Opin. Struct. Biol. **18**, 154 (2008).
[51] D. Prada-Gracia, J. Gómez-Gardenes, P. Echenique, and F. Falo, Exploring the free energy landscape: From dynamics to networks and back, PLoS Comput. Biol. **5**, e1000415 (2009).
[52] D. J. Wales, Energy landscapes: Some new horizons, Curr. Opin. Struct. Biol. **20**, 3 (2010).
[53] S. A. Trygubenko and D. J. Wales, A doubly nudged elastic band method for finding transition states, J. Chem. Phys. **120**, 2082 (2004).
[54] D. Sheppard, R. Terrell, and G. Henkelman, Optimization methods for finding minimum energy paths, J. Chem. Phys. **128**, 134106 (2008).
[55] G. Mills, H. Jónsson, and G. K. Schenter, Reversible work transition state theory: Application to dissociative adsorption of hydrogen, Surf. Sci. **324**, 305 (1995).
[56] H. Jónsson, G. Mills, and K. W. Jacobsen, Nudged elastic band method for finding minimum energy paths of transitions, in *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, Singapore, 1998), pp. 385–404.
[57] G. Henkelman, B. P. Uberuaga, and H. Jónsson, A climbing image nudged elastic band method for finding saddle points and minimum energy paths, J. Chem. Phys. **113**, 9901 (2000).
[58] G. Henkelman and H. Jónsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, J. Chem. Phys. **113**, 9978 (2000).
[59] L. J. Munro and D. J. Wales, Defect migration in crystalline silicon, Phys. Rev. B **59**, 3969 (1999).
[60] G. Henkelman and H. Jónsson, A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives, J. Chem. Phys. **111**, 7010 (1999).
[61] Y. Kumeda, L. J. Munro, and D. J. Wales, Transition states and rearrangement mechanisms from hybrid eigenvector-following and density functional theory: Application to C_{10} H_{10} and defect migration in crystalline silicon, Chem. Phys. Lett. **341**, 185 (2001).
[62] J. M. Carr, S. A. Trygubenko, and D. J. Wales, Finding pathways between distant local minima, J. Chem. Phys. **122**, 234903 (2005).
[63] E. W. Dijkstra, A note on two problems in connexion with graphs, Numer. Math. **1**, 269 (1959).
[64] O. M. Becker and M. Karplus, The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics, J. Chem. Phys. **106**, 1495 (1997).
[65] M. P. Harrigan, K. J. Sung, M. Neeley, K. J. Satzinger, F. Arute, K. Arya, J. Atalaya, J. C. Bardin, R. Barends, S. Boixo *et al.*, Quantum approximate optimization of non–planar graph problems on a planar superconducting processor, Nat. Phys. **17**, 332 (2021).
[66] G. G. Guerreschi and A. Y. Matsuura, QAOA for max–cut requires hundreds of qubits for quantum speed–up, Sci. Rep. **9**, 6903 (2019).
[67] M. Fernández–Pendás, E. F. Combarro, S. Vallecorsa, J. Ranilla, and I. F. Rúa, A study of the performance of classical minimizers in the Quantum Approximate Optimization Algorithm, J. Comput. Appl. Math. **404**, 113388 (2022).
[68] J. Jäger and R. V. Krems, Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines, Nat. Commun. **14**, 576 (2023).
[69] www-wales.ch.cam.ac.uk.