Solving correlation clustering with QAOA and a Rydberg qudit system: a full-stack approach

Jordi R. Weggemans,[1,2] Alexander Urech,[3,2] Alexander Rausch,[4] Robert Spreeuw,[3,2] Richard Boucherie,[5] Florian Schreck,[3,2] Kareljan Schoutens,[6,2] Jiří Minář,[6,2] and Florian Speelman[7,2]

[1]*CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands*
[2]*QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands*
[3]*Van der Waals-Zeeman Institute, Institute of Physics, University of Amsterdam,*
Science Park 904, 1098 XH Amsterdam, the Netherlands
[4]*Robert Bosch GmbH, Corporate Research, Robert-Bosch-Campus 1, 71272 Renningen, Germany*
[5]*Stochastic Operations Research, Department of Applied Mathematics,*
University of Twente, 7500 AE, Enschede, The Netherlands.
[6]*Institute for Theoretical Physics, University of Amsterdam,*
Science Park 904, 1098 XH Amsterdam, the Netherlands
[7]*Informatics Institute, University of Amsterdam,*
Science Park 904, 1098 XH Amsterdam, the Netherlands
(Dated: July 13, 2021)

We study the correlation clustering problem using the quantum approximate optimization algorithm (QAOA) and qudits, which constitute a natural platform for such non-binary problems. Specifically, we consider a neutral atom quantum computer and propose a full stack approach for correlation clustering, including Hamiltonian formulation of the algorithm, analysis of its performance, identification of a suitable level structure for ^{87}Sr and specific gate design. We show the qudit implementation is superior to the qubit encoding as quantified by the gate count. For single layer QAOA, we also prove (conjecture) a lower bound of 0.6367 (0.6699) for the approximation ratio on 3-regular graphs. Our numerical studies evaluate the algorithm's performance by considering complete and Erdős-Rényi graphs of up to 7 vertices and clusters. We find that in all cases the QAOA surpasses the Swamy bound 0.7666 for the approximation ratio for QAOA depths $p \geq 2$. Finally, by analysing the effect of errors when solving complete graphs we find that their inclusion severely limits the algorithm's performance.

I. INTRODUCTION

The Quantum Approximate Optimization Algorithm (QAOA) is a promising attempt at trying to find a quantum advantage when using near-term Noisy Intermediate-Scale Quantum (NISQ) devices [1]. The current body of literature points into mixed directions as far as the utility of QAOA is concerned: whilst it has provable advantages such as recovering near optimal query complexity in Grover's search [2], exhibiting universality [3, 4] and the possibility for quantum supremacy [5], there are also known limitations in the low depth regime [6–9]. However, current analytical tool for analysing the performance of QAOA have only been able to investigate very specific problem instances, predominantly at low depth [10–14]. A general analytical approach remains to be found, which is why a large portion of the literature resorts to numerical simulation.

Earlier work on QAOA studied problems such as MAXCUT [1, 14, 15] and MAX E3LIN2 [16], where the translation procedure entails translating constraints on a binary string into a cost Hamiltonian. In the current work we focus on the *correlation clustering* problem. First introduced by Bansal et al. [17], reductions to correlation clustering are common in machine-learning contexts, with applications ranging from data analysis [18] to image recognition and pose estimation [19]. Since it is a computationally intensive task to find these clusterings,

with approximations being APX-hard [20], it is natural to ask whether quantum algorithms can do better.

For much of the earlier work, the QAOA blueprint entailed working with qubits: the +1/-1 eigenstates of the Pauli Z-operator are chosen in direct correspondence with the variables of the underlying combinatorial optimization problem to compute the cost function, and the mixer Hamiltonian is chosen as an independent Pauli X applied to each qubit. For instance, the MAXCUT problem divides a given graph into two sets, which makes the two-qubit ZZ-operator suitable for encoding the objective function. Our situation is quite different as correlation clustering naturally encompasses many more different clusters. While it is always possible in principle to encode higher-dimensional information into multiple qubits, this does significantly complicate the interactions required.

In this paper we investigate a different route, namely encoding the cluster choice into a qudit. Qudits can be realized in a number of physical platforms including photons [21–29], ions [30–32], molecules [33–35], superconducting circuits [36–38], nuclear magnetic resonance platforms [39, 40] or NV centres [41, 42] and can offer a more resource efficient approach to quantum computing as compared to qubits [43]. Importantly, for the present problem of correlation clustering, they offer a support with the Hilbert space which is *native* to the studied problem, namely the different qudit states can encode

the labelling of nodes into their respective cluster.

In this respect, quantum computers based on trapped neutral atoms interacting via highly excited Rydberg states are of particular interest [44, 45]. These platforms underwent striking developments in recent years with advances such as the creation of arbitrary quantum processor geometries [46–48] containing hundreds of individual atoms [49, 50] or the efficient implementation of quantum gates [51, 52]. This in principle allows for efficient implementation of QAOA algorithms [53, 54] and led to intense efforts, both academic and industrial [55–58], in neutral atom based quantum computing [59, 60].

Even though QAOA was invented as a NISQ-suitable algorithm, it can be far from trivial to connect the abstract Hamiltonian formulation to a physical system. Problems include how to relate the intended Hamiltonians to manipulations on the system, how to manage the spatial structure and non-uniformity of the pairwise interactions of a system in a lattice, and careful analysis of the types of errors that might occur in these operations.

It has also become progressively clear that an efficient quantum algorithm has to be designed in a way which implicitly takes into account the relevant hardware constraints, a so-called full-stack approach [61]. In this work we describe precisely such a full-stack solution starting from a combinatorial optimization problem (the correlation clustering problem), through a QAOA Hamiltonian formulation, all the way to describing how to control and analyse a Rydberg qudit system. This entails:

- We encode the correlation clustering problem into the QAOA paradigm, specifically tailored for qudit quantum systems. We give several improvements to the vanilla QAOA (see Sec. II for the definition), guided by simulations, including experiments with various meta-optimization strategies.

- Next, we go from the abstract Hamiltonian formulation to the operations available for an actual abstract Rydberg system. This entails showing how to drive the Rydberg system in a way that corresponds to the cost and mixer Hamiltonians, where we also have to take care of the spatial aspect of the interactions. Here, we focus on the example of fermionic strontium ^{87}Sr, but a similar derivation would apply to any related system.

- We analyse the algorithm performance in the presence of noise corresponding to the 'random Pauli' method and link this error model to actual errors in Rydberg qudit systems.

- As an extra theoretical result, we apply the techniques of Wurtz and Love [13] to the case of correlation clustering on 3-regular graphs, showing that for a slightly modified QAOA at $p = 1$ the approximation ratio is at least 0.6367. From numerical results obtained through local optimization, we conjecture that a tighter bound of this approximation ratio would be 0.6699.

Related work.

Independently of the current work, other studies have considered applying QAOA to multi-cut versions of MAX-k-CUT [62], and on MAX k-VERTEX COVER [63, 64], which are problems with related properties. A key difference between correlation clustering and the other studied problems is the explicit presence of positive / negative edges in correlation clustering, and the idea that for correlation clustering the number of clusters is not determined yet as part of the input. Additionally, we optimize the creation of the QAOA formulation, ending up with a native qudit implementation, as well as taking a full stack approach: we study all the steps from the problem towards the implementation on a realistic near-term quantum device.

The paper is organized as follows. In Sec. II we briefly recap the QAOA. In Sec. III we introduce the problem of correlation clustering and its implementation as QAOA. We then describe various strategies to improve the algorithm's performance in Sec. IV, which we study in Sec. V. Next, in Sec. VI we discuss the experimental building blocks of the qudit Rydberg quantum computer. We proceed with the processor design and the associated gate count and comparison to qubits in Sec. VII. Finally, we discuss how the errors affect the algorithm in Sec. VIII, and we conclude and discuss open questions in Sec. IX

II. RECAP OF QAOA

In this section we briefly review the quantum approximate optimization algorithm (QAOA) [1]. Consider some combinatorial optimization problem with objective function $C : x \to \mathbb{R}$ acting on n-bit strings $x \in \{0,1\}^n$, domain $\mathcal{D} \subseteq \{0,1\}^n$, and objective

$$\max_{x \in \mathcal{D}} C(x). \tag{1}$$

In maximization, an approximate optimization algorithm aims to find a string x' that achieves a desired approximation ratio α, i.e.

$$\frac{C(x')}{C^*} \geq \alpha, \tag{2}$$

where $C^* = \max_{x \in \mathcal{D}} C(x)$. In QAOA, such combinatorial optimization problems are encoded into a cost Hamiltonian H_C, a mixing Hamiltonian H_M and some initial quantum state $|\psi_0\rangle$. The cost Hamiltonian is diagonal in the computational basis by design, and represents C if its eigenvalues satisfy

$$H_C|x\rangle = C(x)|x\rangle \text{ for all } x \in \{0,1\}^n. \tag{3}$$

The mixing Hamiltonian H_M depends on $\mathcal{D}$ and its structure [65], and is in the unconstrained case (i.e. when $\mathcal{D} = \{0,1\}^n$) usually taken to be the transverse field

Hamiltonian $H_M = \sum_j X_j$. Constraints (i.e. when $\mathcal{D} \subset \{0,1\}^n$) can be incorporated directly into the mixing Hamiltonian or are added as a penalty function in the cost Hamiltonian. The initial quantum state $|\psi_0\rangle$ is usually taken as the uniform superposition over all possible states in the domain. QAOA$_p$, parametrized in $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{p-1}), \beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})$, refers to a level-$p$ QAOA circuit that applies p steps of alternating time evolutions of the cost and mixing Hamiltonians on the initial state. At step k, the unitaries of the time evolutions are given by

$$U_C(\gamma_k) = e^{-i\gamma_k H_C}, \tag{4}$$

$$U_M(\beta_k) = e^{-i\beta_k H_M}. \tag{5}$$

So the final state $|\gamma, \beta\rangle$ of QAOA$_p$ is given by

$$|\gamma, \beta\rangle = \prod_{k=0}^{p-1} U_M(\beta_k) U_C(\gamma_k) |\psi_0\rangle. \tag{6}$$

The expectation value $F_p(\gamma, \beta)$ of the cost Hamiltonian for state $|\gamma, \beta\rangle$ is given by

$$F_p(\gamma, \beta) = \langle \gamma, \beta | H_C | \gamma, \beta \rangle, \tag{7}$$

and can be statistically estimated by taking samples of $|\gamma, \beta\rangle$. The achieved approximation ratio (in expectation) of QAOA$_p$ is then

$$\alpha = \frac{F_p(\gamma, \beta)}{C^*}. \tag{8}$$

The parameter combinations of γ, β are usually found through a classical optimization procedure that uses (7) as a black-box function to be maximized. The QAOA framework as has been described so far, with randomized initial points and without any other improvement strategies, will be referred to as the *vanilla QAOA* in the rest of this work.

III. CORRELATION CLUSTERING

Generally, the objective of clustering problems is to group elements into a family of subsets, named clusters, such that the elements within a cluster are more similar to one another than elements in different clusters. In case of the correlation clustering problem, we would like to cluster without specifying the number of clusters in advance based only on pairwise relations. The problem was introduced by Bansal et al. [17] to the theoretical computer science community and has applications amongst others in social psychology, statistical mechanics and biological networks.

Instances of correlation clustering problems are commonly represented as a graph problem, where the nodes are the elements to be grouped in clusters and edge weights represent similarities between these elements.

Correlation clustering is then formally defined in the following way: let $G(V, E)$ be an undirected graph, where V, E, denotes the sets of nodes and edges, respectively. Let N be the total amount of nodes, i.e. $N = |V|$. Every edge $(u, v) \in E$ is labelled either '+' or '–', depending on whether the elements are similar or dissimilar, respectively. This is the unweighted variant of correlation clustering. Additionally, one can also consider the weighted variant: edges (u, v) carry weights $w_{(u,v)} \in \mathbb{R}^+$ describing an additional measurement of similarity or dissimilarity. There are two complementary problems to correlation clustering. MAXAGREE aims to maximize the number of agreements, defined as the number of '+' edges inside clusters plus the number of '–' edges across clusters. In MINDISAGREE one wants to minimize the number of disagreements: the number of '+' edges across different clusters plus the number of '–' edges inside clusters. The decision versions of MAXAGREE and MINDISAGREE are identical and known to be NP-complete [17]. However, they differ in the approximation setting. MAXAGREE on general graphs is APX-hard: to be precise, it has been shown that for every $\epsilon > 0$, it is NP-hard to approximate both the weighted [20] and unweighted [66] versions of MAXAGREE within a factor of $79/80 + \epsilon$, and the best classical algorithm has approximation ratio $\alpha = 0.7666$ [67] via semi-definite programming (SDP) with rounding techniques. If the graph is complete and unweighted the problem becomes significantly easier although still NP-hard. For complete graphs, Bansal et al. provided a polynomial time approximation scheme (PTAS) [17]. MINDISAGREE is APX-hard on *both* complete [17] and general graphs [20], and the best approximation algorithm achieves only an approximation logarithmic in the input size [68].

Other algorithmic methods that have been proposed to solve correlation clustering include integer linear programming (ILP) and heuristics: examples are greedy methods [69–71], local search methods [72, 73], and large move making algorithms [74]. In practice, the ILP-approach can solve to about 200 nodes due to a scaling of $O(N^3)$ in the amount of constraints in its formulation. The SDP relaxation method has better scaling in its constraints ($O(N^2)$), with the best SDP solvers handling up to about several thousands of nodes [75]. Heuristics have been shown to be able to handle problems consisting over 100k nodes [74]. In this work, we will focus on unweighted correlation clustering in complete and Erdős–Rényi graphs.

A specific real-world application of the correlation clustering problem can be found in the sub-task of distinguishing between persons within multi-person pose estimation. For example, decomposing different persons is necessary in human-robot collaborations where humans fulfil tasks collaboratively with robots and a robot has to differentiate between the position of different human collaborators. In the approach presented in [19], a Convolutional Neural Network (CNN) computes candidates for the location of different body parts of different per-
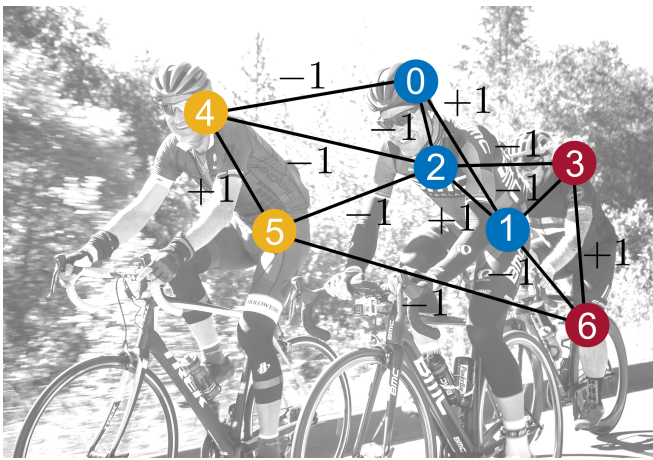
FIG. 1. Overlay of a correlation clustering instance on a real-world picture to illustrate the occurrence of correlation clustering within the problem of decomposing persons in images. The graph contains $N = 7$ nodes with weights $w_{(u,v)} = \pm 1$ while the colours indicate an optimal clustering with objective function value of 10 in the MAXAGREE setting. Note how in general the graph does not need to have a perfect (unfrustrated) clustering, i.e. where all allocated clusters match the weights, as is in this example not the case for nodes '0' and '2'.

sons within a given picture and also pairwise terms how these candidates relate to each other. Leaving out the specific labelling of body parts of each person, the detected body part candidates from the CNN can be represented as nodes and the computed pairwise terms as edge weights of a graph and thus solving the correlation clustering problem of this graph corresponds to a decomposition of the persons in a picture (cf. illustration in Fig. 1). From industrial practice we know that the objective function within the correlation clustering problem can be very sensitive to the assignments of parameters and sometimes even a sufficiently good enough local minimum cannot be found within reasonable time by using classical algorithms.

A. Hamiltonian formulation for qudit systems

Let $G = (V, E)$ be a graph with $N = |V|$ nodes that specifies the input to some correlation clustering problem instance. Every edge $(u, v) \in E$ has a weight $w_{(u,v)} \in \{-1, +1\}$, representing the '+' and '−' relationships in unweighted correlation clustering. We assume that we have access to a qudit quantum system of N qudits consisting of d levels, such that every node is described by a single qudit state $|c_u\rangle$ meaning that node u is put in (a superposition of) cluster(s) c. We write $[d]$ for the set $\{0, \ldots, d-1\}$. We define a two-body interaction

$V_{(u,v)}^d$ that acts on a two qudit sub-space according to

$$V_{(u,v)}^d = \sum_{i \neq j \in [d]} |i_u\rangle|j_v\rangle\langle i_u|\langle j_v| - \sum_{i \in [d]} |i_u\rangle|i_v\rangle\langle i_u|\langle i_v|, \quad (9)$$

which is a $d^2 \times d^2$ matrix with eigenvalues of -1 (+1) for nodes that are put in the same (different) clusters. Our full cost Hamiltonian is obtained by summing over all edges taking the product of the edge weight and (9),

$$H_C = \sum_{(u,v)\in E} w_{(u,v)} V_{(u,v)}^d. \quad (10)$$

We note that this encoding is similar to the binary encoding for MAX-k-CUT recently proposed in Refs. [62, 76]. A Hamiltonian that can mix over the single qudit sub-space was given in the work by Hadfield et al. [65], where the following single-qudit mixing Hamiltonian is proposed

$$h_M(r) = \sum_{i=1}^{r} \left((\Sigma^x)^i + (\Sigma^{x\dagger})^i\right), \quad (11)$$

where $r \in \{1, \ldots, d-1\}$, a parameter determining the connectivity of the mixer and Σ^x is the generalized Pauli X-operator, given by

$$\Sigma^x = \begin{pmatrix} 0 & & & 1 \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{pmatrix}. \quad (12)$$

One observes that for $r = 1$, the single-qudit mixer is equal to

$$h_M(r = 1) = \begin{pmatrix} 0 & 1 & & & 1 \\ 1 & & \ddots & & \\ & \ddots & & \ddots & \\ & & \ddots & & 1 \\ 1 & & & 1 & 0 \end{pmatrix} \quad (13)$$

such that every level is connected to its nearest neighbours, including periodic boundary conditions. The full mixing Hamiltonian is then

$$H_M = \sum_{u \in V} h_M. \quad (14)$$

We can pick any value of $r \in \{1, \ldots, d-1\}$, where the special cases at the boundary are called the single-qudit ring mixer for $r = 1$ and the fully-connected mixer for $r = d - 1$. We take the superposition of all qudit computational basis states as our initial state, i.e.

$$|\psi_0\rangle = \frac{1}{\sqrt{d^N}} \sum_{z \in [d]^N} |z\rangle. \quad (15)$$

Note how the cost Hamiltonian formulation (10) is not equivalent to correlation clustering in the MAXAGREE setting: instead of counting just the agreements we count the number of agreements *minus* the disagreements. However, since every clustering of two nodes (for which an edge exists), needs to be in either agreement or disagreement with the corresponding weight, the sum of the agreements and disagreements is equal to the number of edges in the unweighted setting. Therefore, for a correlation clustering problem with optimum value C^* in the MAXAGREE setting the approximation ratio of this QAOA formulation is equal to

$$\frac{F_p(\gamma, \beta) + |E|}{2C^*}. \tag{16}$$

For the numerical simulations in the algorithmic sections of this work, we implement the initial state by generalized Hadamard operations and assume that the cost and mixing unitaries are elementary operations native to our system. We adopt the Cirq framework [77], since it supports qudit systems, with custom gate operations to match our established formulation.

IV. IMPROVEMENT STRATEGIES

We will be interested in performance at low-depth and hence we only consider $r = 1$ in Eq.(11) for the simulations in sections IV and V. Through numerical evaluation of the vanilla QAOA, we found that the following strategies considerably improved the QAOA's performance for our problem:

Choice of the classical optimizer There is no such thing as a one-size-fits-all classical optimizer that performs well for all QAOA problems: performance varies amongst different problems and can greatly differ for different hyper-parameter settings [78]. We decided to compare different classical optimizers, found in the scikit-quant [78] and SciPy Python packages [79], using their off-the-shelf hyper-parameter settings. In a small study, the best performance was obtained by using BOBYQA. The results of the full comparison can be found in Appendix A.

Restarts. Local optimizers can greatly benefit from restarts (i.e. re-running the same algorithmic procedure with the same parameter settings) since they are sensitive to the quality of initial points. But even with fixed initial points, due to stochastic elements in the optimizers' procedures as well as being introduced from sampling (or gate errors), the algorithms can be made more robust by incorporating restarts. The work of Shaydulin et al. shows how multi-start methods can improve QAOA [80]. In our numerical simulations we set the number of restarts to 5.
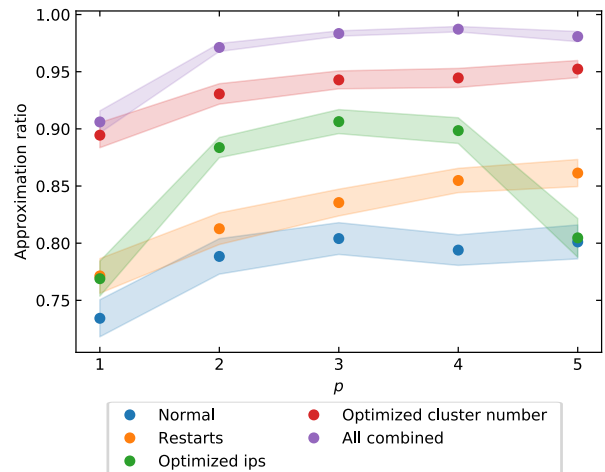

FIG. 2. Approximation ratios for different improvements added to the vanilla QAOA algorithm applied to the $N = 4$ complete graph data set. The blue plot (label 'Normal') represents the performance without any of the improvement strategies as suggested in the text. In all cases COBYLA was used as the classical optimizer. The filled circles data points indicate the average value over all 50 instances and the shaded area represents the error in the mean.

Optimised initial points. Numerical and analytical works have indicated that QAOA's parameters γ, β are concentrated in parameter space for instances that belong to 'similar classes' [81–83]. In practice, this means that we can select a single instance from a class, work very hard to find optimal parameters through variational optimization or from analytical arguments, and store these values to be used as initial points when solving other instances. For every considered N, d, we use the all-negative-weights instance to obtain the parameter values to be used as initial points. Numerical results of a study at $p = 1$ can be found in Appendix B.

Looping over clusters. Initial experiments showed that the vanilla QAOA performed well on instances for which the optimal solution required many clusters – this means that the optimal cluster number was close to the available number of clusters d – but not so when the number of clusters was low. Since our Hamiltonian formulation allows for d to be varied, we can iterate the algorithm over all possible maximum number of clusters, i.e. $d \in \{1, \ldots, N\}$, and keep the best result obtained over all iterations.

Fig. 2 shows the numerical results of the performance on a data set of 50 correlation clustering instances on complete graphs with weights randomly picked out of $\{-1, +1\}$. In the creation of the data set, we swept the probability for giving an edge weight '+1' from 0 to 1 uniformly in order to have a good representation of all possible weight configurations. We observe that looping over cluster numbers has the largest contribution to the improvement, followed by

the optimized initial points, which work particularly well for intermediate values of $2 \leq p \leq 4$. This could be due to the fact that for $p = 1$ the optimizer is able to find good points even with bad initial points, whilst for $p = 5$ the distance from the initial point and the optimal point might be larger in parameter space, which results in the initial point becoming effectively random again. The restarts are expected to have less of an impact when used in conjunction with good initial points. Examples of other variants and strategies to QAOA that one could consider, but were not used in this work, are RQAOA [76], ADAPT-QAOA [84] and parameter initialization heuristics [53].

V. PERFORMANCE

We consider correlation clustering data sets consisting of complete graphs as well as Erdős–Rényi graphs: these are random graphs with a fixed amount of nodes N but for which the edges are created according to some edge creation probability. In the creation of our Erdős–Rényi data sets we use an edge creation probability of 0.5, and additionally have the criteria that every instance needs to have at least one edge. For both graph types we again sweep the edge weight probability of giving an edge weight '+1' from 0 to 1 to create 50 random instances.

Fig. 3 shows the average values, standard deviation and total range (defined as the worst and best performance over all instances in a data set of graphs of fixed size N) of the achieved approximation ratios on instances of our data sets as functions of N and QAOA depth p, achieved by the QAOA that adopts all strategies as listed in section IV. We observe that the worst performance is slightly better for complete graphs compared to the more general Erdős–Rényi graphs, which is to be expected as the complete graph problem instances have more structure and are also more easily solved classically [17]. The average performance is comparable amongst different graph types. For $p = 1$ we observe that the worst performance on instances from our bench-marking data sets becomes comparable to the Swamy bound of 0.7666 for $N \geq 5$, but for $p = 2$ we have that the algorithm performs better than this bound on all instances in the data set

In order to investigate instance-dependence of the performance, Fig. 4 shows a scatter plot of the performance on individual instances in the complete graph data set of $N = 7$ for different values of p. We find that the algorithm at low p has the most difficulty with instances that require a low number of clusters (except the singleton cluster case, which is trivial for $d = 1$), and as p increases the most difficult instances seem to have optimal cluster numbers in the middle between the singleton and all-different clusters.

Next, Fig. 5 shows the average and worst approximation ratios a function of N giving an indication of the scalability. We observe that the decrease of the worst approximation ratio in N seems to slow down consider-
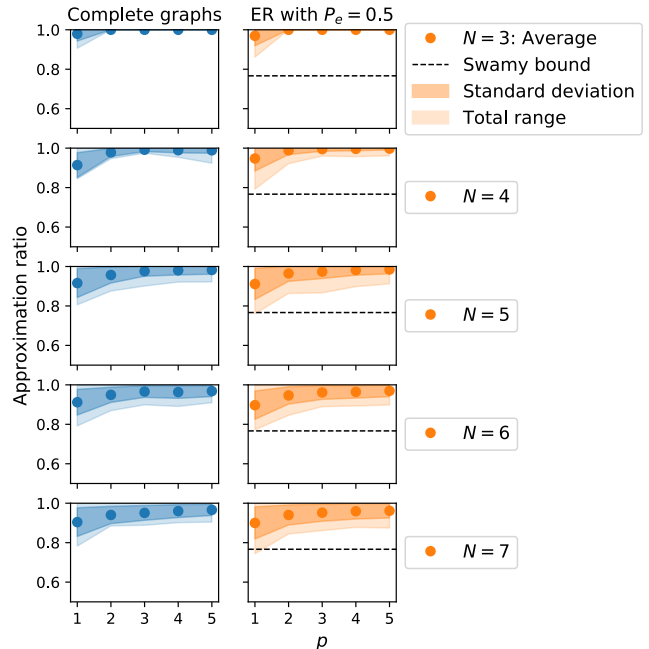

FIG. 3. Performance of QAOA on the data sets of complete graphs (blue) and Erdős–Rényi graphs (orange) with edge creation probabilities $P_e = 0.5$ as a function of the number of nodes N and p. The Swamy bound is added only for the Erdős–Rényi graphs as a PTAS exists for complete graphs. To improve readability, artificial continuity is added between the discrete data for standard deviation (dark shaded), total range (light shaded) and the Swamy bound (dashed black line).
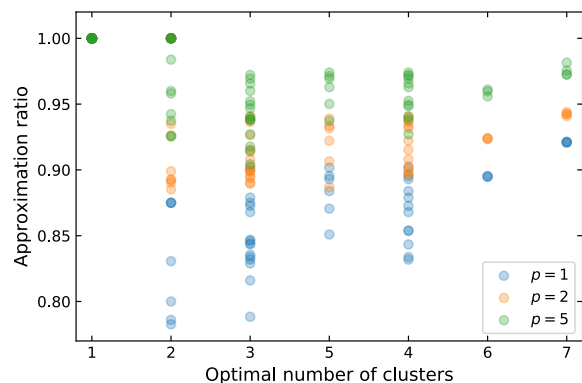

FIG. 4. Scatter plot of the achieved approximation ratios on the $N = 7$ complete graph data set as a function of the optimal number of clusters, which was obtained through brute force search. In the case when multiple different cluster numbers were optimal, the data points are plotted for all of these values.
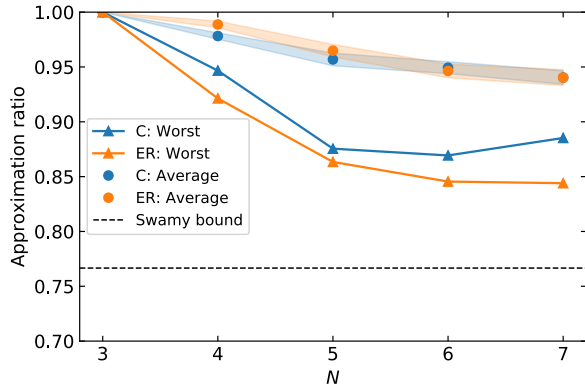
FIG. 5. Average (filled circles) and worst (triangles) approximation ratios as a function of N for $p = 2$ for the complete graph (labelled 'C') and Erdős–Rényi (labelled 'ER') data sets. The dashed line is again the Swamy bound. The shaded area represents the error in the mean.

ably for $N \geq 5$ (for complete graphs it even improves). It should be noted though that it is difficult to make too definitive statements here: we were limited to studying graphs up to $N = 7$ and have no guarantee that we used optimal parameters. However, we can show that the approximation ratio in the limit of large N can in fact become independent of N. We propose (conjecture) that in this limit QAOA$_1$, looping over $d = 1, 2, 3, 4$, has a performance guarantee of 0.6367 (0.6699) on 3-regular graphs. The full derivation of this bound can be found in Appendix C, and follows a similar method as the one proposed by Wurtz and Love [13]. However, it is not yet clear how this bound changes as a function of the graph's degree. For MAXCUT [85] and Maximum Independent Set (MIS) [86] it has already been shown that the approximation ratio at $p = 1$ decreases as the degree of the graph increases. Overall, our results indicate that QAOA, albeit with some added heuristic strategies, might achieve a competitive approximation ratios with respect to the best classical approximation algorithm in solving correlation clustering problems of low-degree graphs at low circuit depth.

VI. EXPERIMENTAL BUILDING BLOCKS

In this section we describe the proposed implementation of the qudits. While most current experimental efforts focus on the use of qubits, neutral atoms are a natural platform for qudits and we specifically focus on the example of fermionic strontium, ^{87}Sr [88–90]. The reason is that it possesses a nuclear spin $I = 9/2$ that is decoupled from the electronic spin, which features $d_{max} = 2I + 1 = 10$ hyperfine states in the ground state manifold, which are insensitive to electric and magnetic field fluctuations. Moreover, one can make use of the

long-lived excited states from the ^3P_J manifold, which has been exploited in a recent experiment to create a Bell state with fidelity reaching 99% [52]. The analysis presented below can be adapted to the analogous isotope ^{173}Yb of fermionic ytterbium [91–93], where however the maximum available $d_{max} = 6$ in the ground state manifold is smaller compared to $d_{max} = 10$ of ^{87}Sr. In the following, we refer to single and two qudit gates as 1-gates and 2-gates, respectively.

Qudit manifold. The relevant level scheme of ^{87}Sr is sketched in Fig. 6a. As stated above, the ground state manifold, which we denote with a slight abuse of notation as $|\mathbb{S}\rangle \equiv |^1S_0, F = \frac{9}{2}\rangle$, consists of $d = 2F + 1 = 10$ m_F-states, $m_F \in \{-9/2, \ldots, 9/2\}$. We also denote $|\mathbb{P}\rangle \equiv |^3P_2, F = \frac{11}{2}\rangle$ the excited state manifold, which we will use to implement the 1- and 2-gates [94] (here we choose the $^3P_2$ manifold in particular due to its long lifetime, which allows for a resonant excitation to the Rydberg state, cf. below). The optical tweezers providing the trapping potentials are typically realized with light of wavelength $\lambda_{tweezer}$ that is red-detuned from the dominant $|\mathbb{S}\rangle - |^1P_1\rangle$ trapping transition (not shown in the Fig. 6a). The choice of the $\mathbb{P}$-manifold is motivated by the fact that, unlike the other possible choices such as $|^3P_2, F = \frac{7}{2}\rangle$ or $|^3P_2, F = \frac{9}{2}\rangle$, it possesses a so-called *magic* wavelength $\lambda_{tweezer} \approx 900$ nm, for which the transition frequencies $|\mathbb{S}, m_F\rangle - |\mathbb{P}, m_{F'}\rangle$ are approximately independent of the intensity of the tweezer light for all $m_F, m_{F'}$, which ensures a position independent addressing frequency of the individual m_F states [95–97]. The actual addressing relies on the Zeeman splitting of the $\mathbb{P}$-manifold and has been experimentally demonstrated using ^{173}Yb [92]. Applying moderate values of the magnetic field B results in linear Zeeman splitting with an energy shift between the adjacent $m_{F'}$ states of $\mu_B g B/h$, where μ_B and h is the Bohr magneton and the Planck constant respectively and g is the Landé g-factor. For the $\mathbb{P}$-manifold, $g \approx 0.36$ and $\mu_B g/h \approx 0.5$ MHz/G [98]. This provides a splitting of ≈ 50 MHz between adjacent $m_{F'}$ states for a magnetic field amplitude of 100 G, allowing for both resonant and off-resonant addressing as we now discuss.

Rydberg states. The 2-gates are implemented via the Rydberg blockade provided by the interaction energy V, which for the density-density interaction between a pair of atoms separated by a distance R typically corresponds to a Van der Waals type $V(R) = C_6/R^6$, where C_6 is so-called "C_6" (or Van der Waals) coefficient [44, 45, 99, 100].

The atom can be excited from the $\mathbb{S}$-manifold through a two-photon transition via the $\mathbb{P}$-manifold to a Rydberg state $|\mathbb{R}y\rangle = |n^3S_1\rangle$ (red and green arrows in Fig 6a,b). In principle one could also use a direct one-photon transition to the $|n^1P_1\rangle$ Rydberg state (dashed purple line in Fig. 6a). However, due to current technology limita-
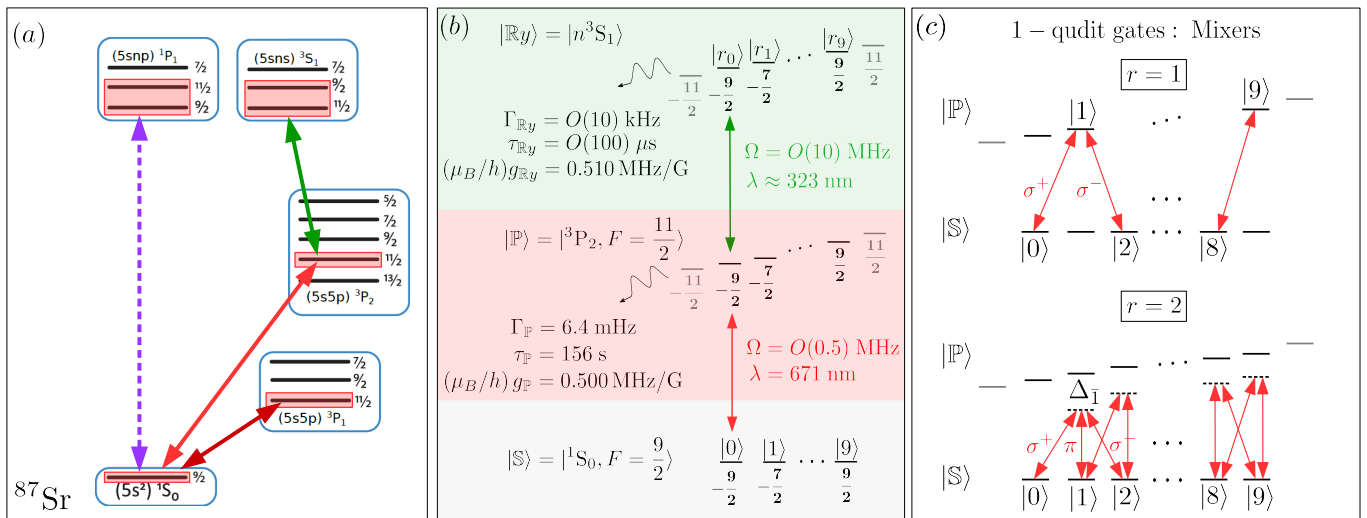
FIG. 6. **(a)** Relevant level scheme of ^{87}Sr. Proposed qudit states are realized by the ground state manifold $|\mathbb{S}\rangle = \left|^1S_0, F = 9/2\right\rangle$. The two-qudit gates are realized by exciting the states from $|\mathbb{S}\rangle$ to a Rydberg manifold $|\mathbb{R}y\rangle = \left|n^3S_1\right\rangle$ through an intermediate state from the $|\mathbb{P}\rangle = \left|^3P_2, F = 11/2\right\rangle$ manifold (red and green arrows). The dashed violet arrow indicates an alternative possibility to excite the ground state to a Rydberg manifold $\left|n^1P_1\right\rangle$, using a single photon transition. The dark red arrow indicates the transition to the $\left|^3P_1, F = 11/2\right\rangle$ manifold used for measurement of the quantum state. **(b)** Parameters of the manifolds $|\mathbb{S}\rangle, |\mathbb{P}\rangle, |\mathbb{R}y\rangle$ relevant for the qudit operations: transition wavelengths λ, typical Rabi frequencies Ω, decay rates Γ and the associated lifetimes τ from the excited $|\mathbb{P}\rangle, |\mathbb{R}y\rangle$ manifolds and the Landé g-factor quantifying the Zeeman splitting of the magnetic sublevels. The values of $\Gamma_\mathbb{P}, \tau_\mathbb{P}$ are taken from [87]. The $\mathbb{P}$-manifold states are used to realize the single qudit gates such as the mixing unitaries, which are shown in **(c)** for $r = 1$ and $r = 2$, see text for details.

tions, such as lack of lasers of sufficient power (and the associated optical elements) for the $|\mathbb{S}\rangle - \left|n^1P_1\right\rangle$ transition wavelength of $\sim 220\,$nm, we solely focus on the $|\mathbb{S}\rangle - |\mathbb{P}\rangle - |\mathbb{R}y\rangle$ scheme, cf. Fig. 6b. This allows for m_F specific addressability of the Rydberg states as well. The two-photon transition is typically operated off-resonantly but given the extremely long lifetime of the $\mathbb{P}$-manifold ($\tau_\mathbb{P} = 156\,s$ in the absence of the tweezer light) and to further reduce the timescales needed for operation we consider a resonant two-step process: the chosen m_F state is transferred as $|\mathbb{S}, m_F\rangle \rightarrow |\mathbb{P}, m_{F'}\rangle \rightarrow |\mathbb{R}y\rangle$ using two consecutive π-pulses or using stimulated rapid adiabatic passage. The properties of the relevant manifolds are summarized in Fig. 6b.

In the following, we refer to $|\mathbb{S}\rangle$ as the qudit manifold. We note that owing to the long lifetime of the $|\mathbb{P}\rangle$-manifold, one could in principle choose it as the basis for the qudit states [101] and the subsequent analysis can be readily adapted to this alternative choice.

Having identified the suitable level structure we now proceed with the design of the 1- and 2-qudit gates.

A. 1-gates

We consider the implementation of a qudit 1-gate by coupling the qudit level $|\ell\rangle$ to level $|\ell'\rangle$ by means of laser fields of Rabi frequency $\Omega_{\ell,\ell'} \equiv \Omega c_{\ell,\ell'}$, where $\Omega \in \mathbb{R}_+$ is the Rabi frequency amplitude and $c_{\ell,\ell'}$ is a (dimen-

sionless) complex number characterizing the individual couplings.

1. Implementation of hardware mixers

Let us first discuss the implementation of the mixing unitary $U_M(\beta_k) = \exp[-i\beta_k H_M]$, cf. Eqs. (5) and (14). Specifically, we will consider two specific cases, namely $r = 1$ and $r = 2$, and we further comment on $r > 2$.

$r = 1$. We propose to implement the $r = 1$ case as shown in Fig. 6c. Starting with all qudit levels in the $\mathbb{S}$-manifold, we apply the following sequence of pulses:

1. Apply $\left\lfloor \frac{d}{2} \right\rfloor$ π-polarized π-pulses on the levels $\{1, 3, \ldots, l\}$, $l = 2\left\lfloor \frac{d}{2} \right\rfloor - 1$, on the $\mathbb{S} - \mathbb{P}$ transition, which brings them from $\mathbb{S}$ to $\mathbb{P}$.

2. Apply the Rabi frequencies $\Omega_{\ell,\ell+1}$ connecting nearest-neighbour qudit levels for time τ_k.

3. Repeat the step 1. to bring the $\left\lfloor \frac{d}{2} \right\rfloor$ levels from $\mathbb{P}$ back to $\mathbb{S}$.

In Fig. 6c, the choice of couplings is depicted by the red

arrows and the coupling Hamiltonian reads

$$\Omega h_\Omega = \Omega \begin{pmatrix} 0 & c_{0,1} & & \\ c_{0,1}^* & & \ddots & \\ & \ddots & & c_{d-2,d-1} \\ & & c_{d-2,d-1}^* & 0 \end{pmatrix}. \qquad (17)$$

When the individual Rabi frequencies $\Omega_{\ell,\ell'}$ are adjusted such that $c_{\ell,\ell'=\ell+1} = 1 \ \forall \ell$, $h_\Omega \to h_M$ and the sequence 1.-3. corresponds to the mixing unitary U_M, Eq. (5), with the mixing parameter $\beta_k = \Omega \tau_k$ [102].

One important remark is in order: the mixing Hamiltonian (13) implements a "periodic boundary condition" in that it couples the levels 0 and $d-1$ (and similarly for higher r). Such coupling is typically not native to a physical qudit, which instead corresponds to "open boundary condition" as is apparent from (17). We discuss the difference between using (13) or (17) below, cf. Sec. VI A 2 and Fig. 7.

$r = 2$. For $r = 2$ we consider an *off-resonant* coupling exploiting a detuning Δ from the $\mathbb{P}$-states as shown in Fig. 6c. This allows to extend the structure of the coupling Hamiltonian Eq. (17) to include also a second diagonal. The coupling Hamiltonian (17) becomes

$$\tilde{\Omega} h_\Omega = \tilde{\Omega} \begin{pmatrix} 0 & \tilde{c}_{0,1} & \tilde{c}_{0,2} & & \\ \tilde{c}_{0,1}^* & & \ddots & \ddots & \\ \tilde{c}_{0,2}^* & \ddots & & \ddots & \tilde{c}_{d-3,d-1} \\ & \ddots & \ddots & & \tilde{c}_{d-2,d-1} \\ & & \tilde{c}_{d-3,d-1}^* & \tilde{c}_{d-2,d-1}^* & 0 \end{pmatrix}. \qquad (18)$$

Here the tilde denotes the effective Rabi frequency which in the far-detuned limit is given by $\tilde{\Omega}_{\ell,\ell'} = \Omega_{\ell,\bar{\ell}} \Omega_{\bar{\ell},\ell'} / \Delta_{\bar{\ell}}$, where $|\bar{\ell}\rangle$ is the detuned state from the $\mathbb{P}$-manifold to which $|\ell\rangle$ and $|\ell'\rangle$ are coupled. As for $r = 1$, here the Rabi frequencies $\tilde{\Omega}_{\ell,\ell'}$ have to be adjusted such that $\tilde{c}_{\ell,\ell+1} = \tilde{c}_{\ell,\ell+2} = 1 \ \forall \ell$.

$r > 2$. Going beyond $r = 2$ becomes non-trivial due to the connectivity of the coupling Hamiltonian (18), here limited to next-to-nearest qudit levels. A general strategy, exploiting the decomposition of an arbitrary single qudit unitary into a sequence of parallelized Givens rotations under the finite connectivity constraint and invoking a greedy optimization algorithm has been described in [103]. As in this work we are concerned only with $r \leq 2$ we do not analyse this situation further.

2. Performance of hardware mixers

To investigate how the hardware-specific mixers given by Eqs. (17) and (18) compare to the mixer (13) we
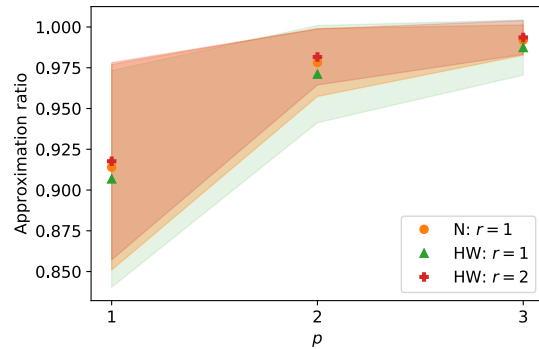

FIG. 7. Results for the different mixing Hamiltonians given by Equations (13),(17) and (18), labelled as 'N: $r = 1$', 'HW: $r = 1$' and 'HW: $r = 2$', respectively. A dot is the average approximation ratio over all 50 instances and the shaded area represents one standard deviation. The results are obtained using a complete graph with $N = |V| = 4$ data set in the same way as described in Sec. V.

run simulations similar to the ones performed in Sec. V: we consider the $N = 4$ complete graph data set as the influence on performance due to the modified boundary terms of the mixers is most profound here: for $N = 3$ (i.e. $d = 3$), Eq. (18) becomes (13) and as N (d) increases, the two boundary states $|0\rangle, |d-1\rangle$ that have different mixing constitute an increasingly small fraction over all states. All non-zero elements in (17) and (18) are set to unity, and we again generate initial points starting from the all-negative-weights graph. The numerical results in Fig. 7 show that the performance is very similar amongst the different mixers – the largest observed percentage performance difference for a single instance is about 4%. From the data we can conclude that the hardware mixer with $r = 2$ performs better on average than the normal mixer with $r = 1$, which outperforms the hardware mixer with $r = 1$. Overall, we can conclude from this data that replacing the standard mixer (13) with the hardware-specific mixers (17) and (18) is not expected to result in a significant decrease in performance.

3. Unitaries beyond mixers

Next, we specify single qudit unitaries beyond mixers, which we will exploit in the construction of the 2-gates in Sec. VI B below. For a two-level system with levels $|\ell\rangle, |\ell'\rangle$ driven on resonance with Rabi frequency $\Omega_{\ell,\ell'}$, the unitary evolution operator expressed in the $\{|\ell\rangle, |\ell'\rangle\}$ basis reads (cf. Appendix D)

$$U_{\ell,\ell'}^{2-\text{level}}(\theta, \varphi) = \begin{pmatrix} \cos\frac{\theta}{2} & -ie^{i\varphi}\sin\frac{\theta}{2} \\ -ie^{-i\varphi}\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}, \qquad (19)$$

where $\theta = 2t|\Omega_{\ell,\ell'}|$, t is the duration of the laser pulse and $\Omega_{\ell,\ell'} = |\Omega_{\ell,\ell'}|e^{i\varphi}$.

Similarly, we can write down a general unitary evolution operator for a three-level system (so-called Λ-scheme). As we will be particularly interested in performing controlled operations through driving the qudit states to a Rydberg level, we shall consider a Λ-scheme where levels $|\ell\rangle, |\ell'\rangle$ are coupled to a common Rydberg state $|r\rangle$. We will further require that at the end of the operation, there is no population in the Rydberg state. In this case the unitary for *resonant* driving and expressed in the $\{|\ell\rangle, |\ell'\rangle, |r\rangle\}$ basis reads (cf. Appendix D)

$$U_{\ell,\ell'}^{3-\text{level}}(\theta,\varphi) = -\begin{pmatrix} \cos\frac{\theta}{2} & e^{i\varphi}\sin\frac{\theta}{2} & 0 \\ e^{-i\varphi}\sin\frac{\theta}{2} & -\cos\frac{\theta}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (20)$$

where

$$\cos\frac{\theta}{2} = \frac{|\Omega_0|^2 - |\Omega_1|^2}{\Omega^2}, \quad (21a)$$

$$\sin\frac{\theta}{2} e^{i\varphi} = \frac{2\Omega_0\Omega_1^*}{\Omega^2} \quad (21b)$$

$$\Omega = \sqrt{|\Omega_0|^2 + |\Omega_1|^2} \quad (21c)$$

and $\Omega_\alpha = |\Omega_\alpha|e^{i\phi_\alpha}$, $\alpha = 0,1$ such that $\varphi = \phi_0 - \phi_1$. Here, we have denoted $\Omega_0 = \Omega_{\ell,r}, \Omega_1 = \Omega_{\ell',r}$ for simplicity. The unitary takes the form (20) for pulses of duration $t = \pi(2m-1)/\Omega$, $m \in \mathbb{N}$. In what follows we refer to the unitaries (19) and (20) simply as U and we shall specify which one is considered where appropriate. For future convenience we denote the usual Pauli X as

$$X_{\ell,\ell'} = U_{\ell,\ell'}(\pi,0) \quad (22)$$

and the phase gate (defined up to a global phase)

$$P_{\ell,\ell'}(\varphi) = U_{\ell,\ell'}(\pi,\varphi_2)U_{\ell,\ell'}(\pi,\varphi_1) = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}, \quad (23)$$

where $\varphi = \varphi_2 - \varphi_1$.

B. 2-gates

CP gate. We proceed with the construction of the cost unitary U_C, Eq. (4). Noting that the the cost Hamiltonian (10) is given by a sum of commuting operators acting on the graph edges, we consider an action of the cost unitary on a single pair of qudits. It corresponds to a controlled-phase gate of the form

$$CP(\gamma)|\ell\rangle|\bar{\ell}\rangle = [\delta_{\ell,\bar{\ell}}e^{-i\gamma} + (1 - \delta_{\ell,\bar{\ell}})]|\ell\rangle|\bar{\ell}\rangle. \quad (24)$$

It is defined up to a global phase and $\delta_{\ell,\bar{\ell}}$ is the Kronecker delta.

The key element in the construction of the CP gate Eq. (24) is a controlled-phase unitary $\mathcal{U}$ acting on a single level $|\ell\rangle$ from the qudit manifold by coupling it to the corresponding Rydberg state $|r_\ell\rangle$, namely

$$\quad (25)$$

where the first and second qudit is the target and control respectively and P is given in Eq. (23). Here, we have introduced the notation $\mathcal{U}^{(q_t|q_c)}$, where q_t, q_c label the target and control qudits, respectively. We can then construct the controlled phase gate Eq. (24) either in a manifestly symmetric way

$$CP(\gamma) = \prod_{\ell=0}^{d-1} \mathcal{U}_\ell^{(1|2)}\mathcal{U}_\ell^{(2|1)}, \quad (26)$$

or alternatively as

$$CP(\gamma) = \left[\prod_{\ell=0}^{d-2} CX_{\ell,\ell+1|\neg\ell}^{(1|2)}\right]^\dagger P_{\text{aux},0}^{(1)}(\gamma) \left[\prod_{\ell=0}^{d-2} CX_{\ell,\ell+1|\neg\ell}^{(1|2)}\right], \quad (27)$$

cf. Appendix D for derivation. The gate (27) has the advantage of reducing the cost of the gate compared to (26) in terms of the number of hardware operations (laser pulses), cf. Eq. (29) and Sec. VII. Here, $P_{\text{aux},0}^{(1)}$ is the phase gate (23) applied to the first (target) qudit and driving the level $|0\rangle$ through an auxiliary state, which here is *not* a Rydberg state. Similarly, we have introduced the controlled-X gates $CX_{\ell_t,\ell_t'|\neg\ell_c}^{(q_t|q_c)}$ where q_c denotes the control, q_t the target, ℓ_c the control qudit level and ℓ_t, ℓ_t' the pair of target levels being acted upon. Importantly, here the target levels ℓ_t, ℓ_t' are swapped when the control qubit is *not* in the state ℓ_c which is highlighted using the negation sign in the subscript, $\neg\ell_c$. We note that for a typical hardware implementation, the choice of ℓ_t, ℓ_t' is not arbitrary but upper bounded. In the present case we shall assume $|\ell_t - \ell_t'| \le 2$. The controlled-X are implemented in the standard fashion as

$$\quad (28)$$

where the X-gate acting on the target is driven through a Rydberg level (a Λ-scheme) as described by the Eq. (20). We also remark that the CP gate Eq. (27) is invariant under the exchange of the control and target qudits, i.e. $(1|2) \to (2|1)$.

In Sec. VII we will be concerned with assessing the cost of the algorithm as counted by the number of elementary 2-gate operations. To this end we introduce the notation [G] to denote the cost of a gate G as counted by the number of the CX gates (28) or its equivalents in the decomposition of G. The corresponding cost of the CP gates is thus [104]

$$\text{Eq. (26)}: [CP] = 2d \quad (29a)$$

$$\text{Eq. (27)}: [CP] = 2d - 2 \quad (29b)$$

SWAP gate. In order to perform the CP gate on a pair of

distant graph vertices, it is in general necessary to bring them together by means of a swap gate SWAP_d. Several possible implementations of the SWAP gate for qudits have been proposed [105, 106]. Here we shall consider the implementation of Ref. [105], which parallels the qubit SWAP construction consisting of three consecutive CX gates: $\mathrm{SWAP}_2 = \mathrm{CX}^{(q|\bar{q})}\mathrm{CX}^{(\bar{q}|q)}\mathrm{CX}^{(q|\bar{q})}$, where we have dropped the level indices taking $|1\rangle$ to be the control level for the qubit as customary. The qudit version of the SWAP is given by [105]

$$\mathrm{SWAP}_d = \mathrm{CX}_d^{(q|\bar{q})}\mathrm{CX}_d^{(\bar{q}|q)}\mathrm{CX}_d^{(q|\bar{q})}, \qquad (30)$$

which is the direct generalization of the SWAP for qubits. The qudit controlled-X CX_d is defined as

$$\mathrm{CX}_d|\ell\rangle|\bar{\ell}\rangle = |\ell\rangle|-\ell-\bar{\ell}\rangle \qquad (31)$$

and mod d is understood in evaluating $-\ell-\bar{\ell}$ in the last expression. It can be implemented as

$$\mathrm{CX}_d^{(q|\bar{q})} = \mathrm{QFT}_d^{(q)}\mathrm{CZ}_d^{(q|\bar{q})}\mathrm{QFT}_d^{(q)}, \qquad (32)$$

where

$$\mathrm{QFT}_d|\ell\rangle = \frac{1}{\sqrt{d}}\sum_{\ell'=0}^{d-1}\mathrm{e}^{i\frac{2\pi}{d}\ell\ell'}|\ell'\rangle \qquad (33)$$

is the quantum Fourier transform in the single qudit space (i.e. a unitary operation with matrix elements $(\mathrm{QFT}_d)_{\ell',\ell} = \exp[i(2\pi/d)\ell\ell']/\sqrt{d}$) and CZ_d is defined as

$$\mathrm{CZ}_d|\ell\rangle|\bar{\ell}\rangle = \mathrm{e}^{i\frac{2\pi}{d}\ell\bar{\ell}}|\ell\rangle|\bar{\ell}\rangle. \qquad (34)$$

We note that the QFT_d gate is a 1-gate and can thus be synthesized using methods of [103], similarly to the mixing unitary for $r > 2$, cf. also [107] for implementation of QFT_d in the context of multilevel atoms. To quantify the cost of the SWAP gate, we thus focus on the CZ_d gate (34), which can be implemented as

$$\mathrm{CZ}_d^{(q|\bar{q})} = \prod_{\bar{\ell}=1}^{d-1} X_{\bar{\ell},r_{\bar{\ell}}}^{(\bar{q})}\left[\prod_{\ell=1}^{d-1}\mathrm{P}_{\ell,r_\ell}^{(q)}\left(\frac{2\pi}{d}\ell\bar{\ell}\right)\right]X_{\bar{\ell},r_{\bar{\ell}}}^{(\bar{q})}, \qquad (35)$$

which contains $(d-1)^2$ applications of the phase gate P. Importantly, we note that the product in the brackets can be executed *in parallel* by simultaneous application of the laser pulses connecting each level $|\ell\rangle$, $\ell = 1,\ldots,d-1$, of the target to its respective Rydberg level $|r_\ell\rangle$. This is precisely an example of the parallelization offered by the qudit hardware. We thus get, in conjunction with Eq. (30), for the total cost of the qudit SWAP

$$[\mathrm{SWAP}_d] = 3(d-1). \qquad (36)$$

C. Initialisation and measurement

Initialization. The initial state (15) can be prepared by initializing all atoms in the $|0\rangle \equiv |\mathbb{S}, m_F = -9/2\rangle$ state

by standard means of optical pumping and then applying a sequence of $d-1$ unitaries (19) such that

$$|+^d\rangle = \left[\prod_{\ell=1}^{d-1}U_{\ell-1,\ell}^\dagger\left(\theta_\ell, \varphi = \frac{\pi}{2}\right)\right]^\dagger|0\rangle, \qquad (37)$$

where $\cos(\theta_\ell/2) = 1/\sqrt{d-(\ell-1)}$. Here the unitary can be implemented either via resonant or off-resonant Raman coupling described in Sec. VI A for the range $r = 1$ and $r = 2$ of the mixing unitaries respectively.

Measurement. In order to projectively measure the quantum state after each iteration of the QAOA we consider imaging the atoms using resonance fluorescence by collecting the light scattered from the strongly driven $|0\rangle - |^3\mathrm{P}_1, F' = 11/2, m_{F'} = -11/2\rangle$ transition, cf. Fig 6a. This has the advantage of simultaneously cooling the atoms in the $|0\rangle$ state while imaging them [108], cf. also [109–111] for related techniques. Exploiting the state-specific detection of individual m_F states, first only the population in the $|0\rangle$ state ($m_F = -9/2$) is being detected. In the case of negative outcome (no population in the $|0\rangle$ state), the population from the adjacent m_F state, i.e. from $|1\rangle$ ($m_F = -7/2$), is transferred to $|0\rangle$ by using optical pumping, or coherently via stimulated Raman adiabatic passage and the $|0\rangle$ is imaged again. This process is repeated until the positive detection of some qudit level $|\ell\rangle$. This allows for imaging of all the qudit states within the expected lifetime in optical tweezers of $\gtrsim 10$ seconds (as there is no active cooling of the $|1\rangle, \ldots, |d-1\rangle$ states during the imaging). This time is limited mainly by off-resonant scattering of the tweezer light and also by background gas collisions, where the latter can be further reduced by increasing the quality of the vacuum.

VII. PROCESSOR DESIGN AND GATE COUNT

Here we seek to evaluate the cost of the algorithm, cf. Sec. II and Sec. III, as quantified by the gate count. As discussed in Appendix F, the dominant errors stem mainly from the 2-gates and we thus focus on the 2-gate count. To this end we consider the total count $\mathcal{C}_{\mathrm{tot}}$ of *qubit* primitive 2-gates and specifically we will use the qubit controlled-X [CX] as our cost unit (This choice of counting will be useful when comparing the qudit vs. qubit encodings in Sec. VII A). $\mathcal{C}_{\mathrm{tot}}$ is determined by *(i)* the topology of the graph encoding the clustering problem, *(ii)* the topology of the quantum processor and, for qubits, *(iii)* the encoding scheme, which we discuss in Sec. VII A.

To proceed, let us first comment on 2-gates beyond nearest neighbours. The neutral atom and ion based systems possess a native *long-range* interaction, which typically decays as a power law $1/R^\alpha$ with distance ($\alpha = 6$ for a pair of Rydberg atoms interacting through a Van

der Waals potential). Such potential gives rise to clusters of higher connectivity on the processor, which can lead to an improvement in performance over processors with only nearest-neighbour interactions [112]. The related scaling properties of the quantum gates for trapped ensembles rather than single neutral atoms have been analysed theoretically in [113] and such ensembles occupying hundreds of sites have been realized recently experimentally [114]. Furthermore, claims have been made that up to 50 atoms can be connected without the need of a SWAP operation [57]. In the here considered implementation using Rydberg blockade and Van der Waals interaction, the price to pay for the higher connectivity is the longer duration of the gate. Taking the basic building block Eq. (25) of the CP gate as an example, the gate duration is mainly given by the duration of the P-gate applied to the target qudit. This is because while the X-gates acting on the control qudit can be performed in principle arbitrarily fast limited only by the available Rabi frequency, the Rabi frequency Ω used to realize the P-gate has to satisfy the blockade constraint $V(R)/\Omega \gg 1$ for a given atom distance R. Since $V(R) \propto 1/R^6$, for the same quality of blockade (same $V(R)/\Omega$) the gate time thus scales as $\propto 1/\Omega \propto R^6$. For this reason and for the sake of clarity, in the following we consider only nearest-neighbour 2-gates and leave the algorithm hardware optimization exploiting longer-range connectivity for future work.

Another remark is that it is desirable to *parallelize* the gate operations, cf. e.g. [51] for recent realization with Rydberg atoms, to reduce the absolute time of the algorithm run. This in principle allows one to reduce the effect of noise such as the background gas collisions or off-resonant scattering from the optical traps, cf. Sec. VIII. Such parallelization however does not change the 2-gate count and we don't elaborate on it further.

		Processor geometry	Gates for binary encoding
1D	qudit		$\widetilde{\mathrm{CP}}_{\mathfrak{q}=\log_2 d} =$
	binary		
2D	qudit		$\widetilde{\mathrm{CP}}_{\mathfrak{q}\neq\log_2 d} =$
	$\mathfrak{q}=2$		
	$\mathfrak{q}=3$		$\mathrm{C}^2(U) =$ $=$
	$\mathfrak{q}=4$		$\mathrm{C}^3(U) =$ $=$

TABLE I. *Left column:* Considered processor geometries in 1D (upper two rows) and 2D (lower four rows) for both the qudits and the binary encoding. In 2D the shaded blue and red areas highlight the effective qudit (e-dit) in the binary encoding including the considered enumeration of the qubits. *Right column:* The construction of the $\widetilde{\mathrm{CP}}$ gates for binary encoding. Construction for both $\mathfrak{q} = \log_2 d$ and $\mathfrak{q} \neq \log_2 d$ case is shown (based on Ref. [62]) together with the decomposition of the multi-controlled $\mathrm{C}^{\mathfrak{q}-1}(U)$ gate for $\mathfrak{q} = 4$ and $\mathfrak{q} = 3$ (based on Ref. [115]). The unitaries V satisfy $V^2 = U$ and $V^4 = U$ in the decomposition of $\mathrm{C}^2(U)$ and $\mathrm{C}^3(U)$ respectively [115].

Geometry	Encoding		$n_{\widetilde{SWAP}}^{\text{inter}}$	$[\widetilde{SWAP}] =$ $= \mathfrak{q}^2[SWAP_2]$ $= 3 \times \mathfrak{q}^2[CX]$	$n_{SWAP_2}^{\text{intra}}[SWAP_2] =$ $= 3\mathfrak{q}(\mathfrak{q}-1)[SWAP_2]$ $= 3 \times 3\mathfrak{q}(\mathfrak{q}-1)[CX]$	$[C^{\mathfrak{q}-1}(U)]$	$\mathcal{C}_{\text{tot}}$
1D	binary	$\mathfrak{q}=2$	$\frac{(N-1)(N-2)}{2}$	3×4	3×6	$1[CX] = 1$	$\lvert E \rvert \times 35 - O(N) \times 12$
		$\mathfrak{q}=3$		3×9	3×18	$5[CX] + 2[SWAP_2] = 11$	$\lvert E \rvert \times 98 - O(N) \times 27$
		$\mathfrak{q}=4$		3×16	3×36	$13[CX] + 12[SWAP_2] = 49$	$\lvert E \rvert \times 213 - O(N) \times 48$
	qudit	$\mathfrak{q}=2$		$[SWAP_d] = 3(d-1)$		$[CP] = 2(d-1)$	
				9		6	$\lvert E \rvert \times \mathbf{15} - O(N) \times 9$
		$\mathfrak{q}=3$		21		14	$\lvert E \rvert \times \mathbf{35} - O(N) \times 21$
		$\mathfrak{q}=4$		45		30	$\lvert E \rvert \times \mathbf{75} - O(N) \times 45$
2D	binary		$n_{\widetilde{SWAP}}^{\text{inter}}$	$[\widetilde{SWAP}]$	$n_{SWAP_2}^{\text{intra}}[SWAP_2]$	$[C^{\mathfrak{q}-1}(U)]$	$\mathcal{C}_{\text{tot}}$
		$\mathfrak{q}=2$	$O(N)$	$3 \times 3\frac{1}{3}^*$	$3 \times 4^*$	$1[CX] = 1$	$\lvert E \rvert \times 17^* + O(N) \times 10^*$
		$\mathfrak{q}=3$		$3 \times 8^*$	$3 \times 6^*$	$5[CX] = 5$	$\lvert E \rvert \times 29^* + O(N) \times 24^*$
		$\mathfrak{q}=4$		$3 \times 9\frac{1}{3}^*$	$3 \times 12\frac{2}{3}^*$	$13[CX] + 2[SWAP_2] = 19$	$\lvert E \rvert \times 65^* + O(N) \times 28^*$
	qudit	$\mathfrak{q}=2$		$[SWAP_d] = 3(d-1)$		$[CP] = 2(d-1)$	
				9		6	$\lvert E \rvert \times \mathbf{6} + O(N) \times 9$
		$\mathfrak{q}=3$		21		14	$\lvert E \rvert \times \mathbf{14} + O(N) \times 21$
		$\mathfrak{q}=4$		45		30	$\lvert E \rvert \times \mathbf{30} + O(N) \times 45$

TABLE II. Gate count as per Eqs. (39) for qudits and qubit binary encoding for 1D (blue shaded lines) and 2D (red shaded lines) processor geometries. The cost of the $\widetilde{CP}$ gate is evaluated using the circuits from Ref. [62] as well as the standard decomposition of multi-controlled qubit gates shown in the right column of Table I. The cost of the $\widetilde{SWAP}$ gate is evaluated according to the qubit arrangements shown in the left column of Table I. In 2D, it is obtained as a weighted average over the neighbours (for $\mathfrak{q} = 2, 3, 4$, each e-dit has four, three and two neighbours to which it is connected by one leg and two, three and four neighbours to which it is connected by three legs). This might result in a fractional value of $[\widetilde{SWAP}]$ and gate counts stemming from such weighted average are denoted by a star. The total gate count $\mathcal{C}_{\text{tot}}$ of the algorithm is indicated in the last column in purple. For the qubit encoding, $\mathcal{C}_{\text{tot}}$ also includes the contribution $2\mathfrak{q}[CX]$ from (40) not listed in the Table. For complete graphs considered here, the dominant contribution to $\mathcal{C}_{\text{tot}}$ is coming from the number of edges $\lvert E \rvert = N(N-1)/2 = O(N^2)$. This is highlighted in bold for qudits in the last column. It is apparent from $\mathcal{C}_{\text{tot}}$ that the qudit encoding is superior to the (best-case scenario $\mathfrak{q} = \log_2 d$) binary one in all but the 2D $\mathfrak{q} = 2$ cases.

A. Gate count and comparison to qubits

As stated in the introduction, the use of qudits in general offers a resource-efficient alternative to qubit encodings and for certain problems, such as correlation clustering, also a natural physical platform for their implementation. To proceed with the comparison between qudits and qubits, we first specify the qubit encoding. In what follows, we denote by tilde a qudit-like gate acting on the effective qudit encoded in qubits. We also term such an effective qudit an *e-dit* to distinguish it from the physical qudit.

Effective controlled-phase gates $\widetilde{CP}$ equivalent to (24) have been proposed in [62]. The Ref. [62] studied the MAX-k-cut using QAOA and considered two possible encodings, a one-hot and a binary one. The one-hot encoding seems to produce smaller 2-gate count when $\mathfrak{q} \neq \log_2 d$ and larger when $\mathfrak{q} = \log_2 d$, the precise numbers depending on the graph topology [62]. Here

$$\mathfrak{q} = \lceil \log_2 d \rceil, \tag{38}$$

is the number of qubits necessary to contain the qudit Hilbert space as subspace such that the two become equal when $\mathfrak{q} = \log_2 d$. On the other hand, the one-hot encoding is more resource extensive than the binary encoding and we therefore choose the binary one. In the scheme of Ref. [62] $\mathfrak{q} = \log_2 d$ and $\mathfrak{q} \neq \log_2 d$ correspond to two different realizations of $\widetilde{CP}$, which we list in Table I for the reader's convenience.

Next, we have to specify the processor topology. The versatility of neutral atom platforms allows in principle for arbitrary arrangements of the atoms, which can be exploited for efficient encoding of the graph instance at hand. For specificity, we perform the gate count on one of the examples of main interest, the complete graph, for which $\lvert E \rvert = N(N-1)/2$. Motivated by ongoing experiments [46–52], we consider a simple 1D chain (with open boundaries) and a 2D regular lattice. In 2D we consider a triangular lattice, which provides the densest packing of atoms. The 2-gate count is given by

$$\text{qubits}: \mathcal{C}_{\text{tot}} = n_{\widetilde{SWAP}}^{\text{inter}}[\widetilde{SWAP}] + \lvert E \rvert[\widetilde{CP}] \tag{39a}$$

$$\text{qudits}: \mathcal{C}_{\text{tot}} = n_{SWAP_d}^{\text{inter}}[SWAP_d] + \lvert E \rvert[CP], \tag{39b}$$

where $n_{\widetilde{SWAP}}^{\text{inter}}$, $n_{SWAP_d}^{\text{inter}}$ count the number of SWAPs between the e-dits and qudits respectively such that each vertex has been a neighbour of every other vertex at least once. In principle we can perform the gate count for any d. We note, that due to the $\widetilde{CP}$ gate structure for $\mathfrak{q} \neq \log_2 d$, the total count is relatively

much higher than for $\mathfrak{q} = \log_2 d$ (Table 1 of [62] gives $[\widetilde{\text{CP}}] = $2,70,$\underline{6}$,206,142,78,$\underline{14}$ for $d = $2,3,$\underline{4}$,5,6,7,$\underline{8}$, where the values with underline correspond to $\mathfrak{q} = \log_2 d$). By contrast, since the gate structure for qudits is the same for all d and since our main goal here is to compare the gate count for qudits and qubits, we focus only on the $\mathfrak{q} = \log_2 d$ case. The reason for this is that for a given ratio $\mathcal{C}_{\text{tot}}^{\text{qudits}}/\mathcal{C}_{\text{tot}}^{\text{qubits}}$ of gate counts for qudits and qubits for $\mathfrak{q} = \log_2 d$, this ratio will be only smaller when $\mathfrak{q} \neq \log_2 d$. In this case, the cost of $\widetilde{\text{CP}}$ can be further decomposed as

$$[\widetilde{\text{CP}}] = n_{\text{SWAP}_2}^{\text{intra}}[\text{SWAP}_2] + 2\mathfrak{q}[\text{CX}] + [\text{C}^{\mathfrak{q}-1}(U)], \quad (40)$$

where $n_{\text{SWAP}_2}^{\text{intra}}$ counts the number of SWAPs within the e-dit and $\text{C}^{\mathfrak{q}-1}(U)$ is the multicontrolled unitary performed within the target e-dit, cf. Table I. The unitary U in $\text{C}^{\mathfrak{q}-1}(U)$ is a single qubit gate and its particular form is not relevant for the counting (cf. Ref. [62] for details).

1D. Let us first describe the situation in 1D. Here, $n_{\widetilde{\text{SWAP}}}^{\text{inter}} = n_{\text{SWAP}_d}^{\text{inter}}$ indicates the number of SWAPs between qudits (or e-dits) such that each qudit (or e-dit) has been a neighbour of all the others. Here we shall use a rather natural choice of SWAP sequences described in [116, 117]. It consists of a repeated application of two layers of SWAPs, one performing SWAP operations on qudit (e-dit) pairs $1-2, 3-4, \ldots$ followed by the other on pairs $2-3, 4-5, \ldots$. This yields $n_{\widetilde{\text{SWAP}}}^{\text{inter}} = n_{\text{SWAP}_d}^{\text{inter}} = (N-1)(N-2)/2$, see also Appendix E. Furthermore we also have $|E| = N(N-1)/2$ such that Eq. (39b) together with Eq. (29b) and Eq. (36) yield

$$\mathcal{C}_{\text{tot}} = \frac{(5N-6)(N-1)}{2}(d-1) = |E| \times 5(d-1) - O(N). \quad (41)$$

To analyse the qubit case, let us start with the analysis of the $\text{C}^{\mathfrak{q}-1}(U)$ gate. Ref. [118] describes a systematic construction of $\text{C}^{\mathfrak{q}}(U)$ for arbitrary $\mathfrak{q}$ proposing two schemes for such a construction, one scaling exponentially and the other one polynomially with $\mathfrak{q}$. While the polynomial scaling is clearly favourable for large $\mathfrak{q}$, the actual gate count favours the exponential scheme for $\mathfrak{q} \leq 4$ considered here. We thus consider the decomposition for $\text{C}^2(U)$ and $\text{C}^3(U)$ shown in Table I, proposed in the early works on quantum information [115, 119]. The considered ordering within the e-dit shown in Table I leads to the following counts as function of $\mathfrak{q}$

$$[\widetilde{\text{SWAP}}] = \mathfrak{q}^2[\text{SWAP}_2] \quad (42a)$$

$$n_{\text{SWAP}_2}^{\text{intra}} = 3\mathfrak{q}(\mathfrak{q}-1). \quad (42b)$$

The total gate count for qudits and e-dits and their break down as per (39) is summarized in Table II.

2D. For qudits, the cost (39b) carries over to two dimensions. For qubits, the situation is more involved and we shall analyse only the specific cases $\mathfrak{q} = 2, 3, 4$. In Table I we show a possible arrangement of the e-dits including the qubit labelling within the e-dit. We note that the effective lattice geometry composed of the e-dits retains the topology of the triangular lattice in that each e-dit has six nearest neighbours and consequently $n_{\text{SWAP}_2}^{\text{inter}} = n_{\text{SWAP}_d}^{\text{inter}}$ as in the 1D case. However the difference with qudits is that the e-dit lattice is "anisotropic", namely for $\mathfrak{q} = 2, 3, 4$, each e-dit has four, three and two neighbours to which it is connected by one leg and two, three and four neighbours to which it is connected by three legs respectively.

We also note that the proposed tilings implementing the e-dits are not necessarily unique.

In order to determine $n_{\text{SWAP}}^{\text{inter}}$, one can apply a generalization of the alternating SWAP sequence from the 1D case, cf. [117], which yields a scaling $O(N)$ for the number of SWAPs between the qudits (or e-dits). Since $|E| = N(N-1)/2 = O(N^2)$, this gives a subleading contribution to the gate count and we do not elaborate on the precise sequence further.

We are thus left with evaluating the cost of $[\widetilde{\text{SWAP}}]$ for the e-dit SWAP and $[\widetilde{\text{CP}}]$ which takes into account the respective e-dit and processor geometries. The summary of the costs for 2D is given in Table II.

As a result, we find the expected outcome, namely that in all considered cases the gate count is lower for qudit encoding than for qubit binary encoding, even for the best-case scenario $\mathfrak{q} = \log_2 d$ of the latter.

Here we have evaluated the gate count considering the realization of the SWAP gates via the Rydberg interactions. A remark is that the gate count of a sequence of consecutive SWAPs can be further reduced by considering so-called bridge gates, leading however only to a modest improvement by a factor ≈ 1.5 [120]. In the context of neutral atoms in optical tweezers, it would be interesting to exploit a strength of these platforms and perform the SWAP by physically exchanging the atoms, which for distances $\sim 5\,\mu$m can be done on the timescale of $\sim 50\,\mu$s [121].

VIII.   ERRORS

In this section we consider an error model used in Ref. [122] in the theoretical analysis of a Rydberg quantum computer and we discuss the implications of the errors for the algorithm performance. Importantly, the error model of Ref. [122] can be cast in the *unitary* evolution framework used in our work. We comment on the actual experimental errors and how they relate to the considered error model in Appendix F.

Unitary error model. Let us consider a set of d^2 single qudit unitaries $\mathbb{U} \equiv \{(\Sigma^X)^r(\Sigma^Z)^s\}$, where $r, s = 0, \ldots, d-1$.

Here $\Sigma^X = \Sigma^x + (\Sigma^x)^\dagger$, Σ^x is given by Eq. (12) and

$$\Sigma^Z = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \lambda^{d-1} \end{bmatrix} \tag{43}$$

with $\lambda = \exp(i2\pi/d)$ is the generalized Pauli Z. For a qubit, $d = 2$, this reduces to $\mathbb{U} = \{\mathbb{1}, \Sigma^X, \Sigma^Z, \Sigma^X \Sigma^Z\}$. Motivated by the experimental considerations, namely the fact that the errors are dominated by the 2-gate ones, cf. Appendix F, the model consists of applying an identity with probability $1 - p_2$ or a unitary $U \in \mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}$, with probability $p_2/(|\mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}|) = p_2/(|\mathbb{U}^{\otimes 2}| - 1)$ on each pair of qudits after each cost unitary $U_C = e^{-iH_C}$, cf. Eq. (10). Put formally, for $\rho = U_C \rho' U_C^\dagger$

$$\rho \to \rho \text{ with prob. } 1 - p_2$$

$$\rho \to U\rho U^\dagger, \ U \in \mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}, \text{ with prob. } \frac{p_2}{|\mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}|} \tag{44}$$

We consider the same data sets for complete graphs as have been used in Section V for $N \in \{3, 4, 5\}$, and use the *optimized* values for γ, β and d obtained in the noise-free setting. This way makes it possible to discard the classical optimization loop, saving computation time, and allows us to focus on the performance degradation as a result of the randomization of the state vector. It has been argued in Ref. [123] that noise generated by dephasing, bit flip, and depolarizing channels tends to flatten (on average) the parameter space energy landscape without changing its structure. Since the error model (44) is a qudit generalization of these types of channels, we expect the γ, β obtained in the noise-free setting to be optimal also in the noisy setting.

The results are shown in Figure 8. For p_2 small ($\lesssim 10^{-3}$), we see that the performance is hardly affected by the noise. Once p_2 increases, we enter the regime where performance quickly degrades until we reach the performance of the completely randomized state. Whilst the performance at $p_2 = 1$ is considerably smaller than in the noise-free setting, the approximation ratio achieved on average is still relatively large. This is due to the fact that d is already pre-determined: instances with $d = 1$ are not affected by the noise and still maintain a large approximation ratio. We define the threshold noise $p_{2,\mathrm{Th}}$ (threshold amount of 2-gates g_{Th}) as the noise level (amount of 2-gates) for which the QAOA has lost half of its performance as compared to random guessing, on average for all instances.

By determining the values of $p_{2,\mathrm{Th}}$ from the data in Figures 8a-c, and using that the amount of two-qubit gates is simply $pN(N-1)/2$ (with p the QAOA depth), we plot g_{Th} as a function of p_2 in Figure 8d. We observe that our data are compatible with a linear dependence for g_{Th} of the form

$$g_{\mathrm{Th}} = \frac{\kappa}{p_2}, \quad \kappa = 0.84. \tag{45}$$

This naive model for the noise shows scaling similar to that of Ref. [124] – here the authors showed that a bound on the circuit depth scales inversely proportional to the quantum gate error – but now in the amount of operations instead of circuit depth.

It is interesting to interpret the results of Fig. 8d in terms of the achievable system sizes and required hardware gate operations. Denoting the error probability of an elementary CX-like hardware gate as p_{CX} we get for the total success probability (no error) after application of $|E| = g$ gates on the $|E|$ edges

$$p_{\mathrm{success}} = (1 - p_2)^{|E|}$$
$$= (1 - [\mathrm{CP}]p_{\mathrm{CX}})^{|E|}(1 - [\mathrm{SWAP}_d]p_{\mathrm{CX}})^{O(N)}$$
$$\approx (1 - [\mathrm{CP}]p_{\mathrm{CX}})^{|E|},$$

where in the last relation we used that $p_2, p_{\mathrm{CX}} \ll 1$, $|E| \gg N$ while $[\mathrm{CP}] \approx [\mathrm{SWAP}_d]$. This allows us to identify $p_2 \approx [\mathrm{CP}]p_{\mathrm{CX}}$. Using the result of Table II, considering $d = 8$ for specificity, and using $|E| = N(N-1)/2$, we find that for $p_2 = (10^{-2}, 10^{-3}, 10^{-4})$, $p_{\mathrm{CX}} \approx (10^{-3}, 10^{-4}, 10^{-5})$ resulting in $N \approx 13, 41, 130$ [125].

We thus see that the inclusion of the errors strongly limits the scaling of the algorithm to large problem instances. We note that similar limitations of the QAOA due to experimental errors have been discussed recently in [126] and [124]. In particular, it has been argued in Refs. [124, 126] that hardware-native graph instances, in case of Ref. [126] a simple square lattice with nearest-neighbour interactions, are less prone to errors and potentially allow to scale up to problem sizes large enough to achieve quantum advantage. This is due to significant simplification of the quantum circuit, which avoids a number of extra compilation steps (such as SWAP gates). The major downside is however that a quantum computer hardware is, typically, not application or instance specific.

In this respect, the neutral atom based platforms seem to be particularly interesting as they allow for implementing graph topologies beyond simple planar ones by exploiting the atom arrangements in three dimensions and the long-range interactions. Nevertheless, as is apparent from the results in this section, the robustness to the errors is directly related to the graph topology. Specifically, for the complete graph considered here, the errors strongly limit the scaling even if it is native to the hardware, see also the discussion in Sec. IX.

IX. CONCLUSIONS AND OUTLOOK

In this work we have addressed the problem of solving the correlation clustering using QAOA and a qudit quantum computer. We have specifically considered a neutral atom based architecture, which has the potential to offer up to ~ 1000 qudits in the near future [49, 50, 60]. Here the gates are realized through the interaction of atoms
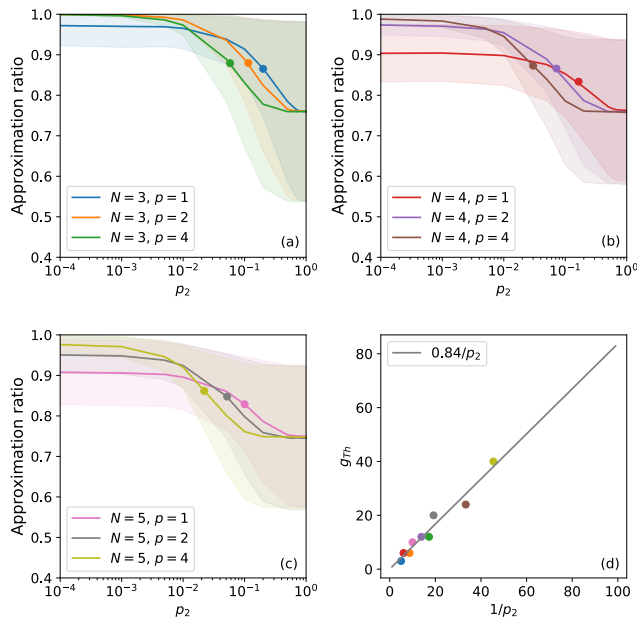
FIG. 8. **(a)-(c)** Numerical results for the QAOA with a error channel given by (44) for $N = 3, 4, 5$, respectively. The plots are the average approximation ratios over all instances in the complete graph data sets and the shaded areas correspond to one standard deviation. The filled circles data points correspond to the threshold noise for which the QAOA has lost half of its performance over random guessing. **(d)** Linear fit for $g_{\mathrm{Th}} \propto 1/p_2$, individual colours correspond to the respective data points where $p_2 = p_{2,\mathrm{Th}}$ in (a), (b) and (c).

in a highly excited electronic state, a so-called Rydberg state. Considering specifically the element ^{87}Sr we have identified a suitable level structure for the qudit, which in turn allowed us to design the gates for implementing the QAOA on the quantum processor. It is worth emphasizing that while we have considered correlation clustering, the discussed qudit gates can be used for the closely related MAX-k-CUT and MAX k-VERTEX COVER problems, cf. [62] and [63, 64, 114] respectively.

We assessed the algorithm's performance by numerical simulations including various optimization strategies, namely restarts, optimizing the initial points and looping over the cluster number. Focusing specifically on complete and Erdős-Rényi graphs of up to 7 vertices and clusters we found that in all studied cases the QAOA with depth $p \geq 2$ provides approximation ratios above the Swamy bound 0.7666 [67], corresponding to the best known classical strategy (based on SDP) with performance guarantees for MAXAGREE.

While this result is encouraging, the inclusion of errors suggests that it is challenging for the QAOA to outperform classical algorithms, cf. Sec. III, at least for complete graphs on near-term noisy quantum devices. This is in agreement with related results reported recently in Ref. [124] and Ref. [126], where various graph

instances were considered to compare the performance of the QAOA using a superconducting quantum chip, including the effect of the errors.

In this respect, the neutral atom based platforms seem to be particularly interesting and our work indicate possible directions in the design of experiments for benchmarking the QAOA on correlation clustering instances. Not only do neutral atoms allow to assemble arbitrary structures in both 2D and 3D [46–48, 121] (a PTAS does exist for planar graphs [127], but one could still investigate the rate of convergence), but they also allow for native long-range interactions, known to yield an advantage over quantum processors featuring only nearest-neighbour ones [112]. Such long-range interactions in turn allow for implementation of *non-planar* graphs [126] as native geometry using even a planar arrangement of atoms and at the same time avoid the need for additional SWAP gates. One should however keep in mind that for complete graphs, the errors strongly limit the scaling even when this graph topology is native to the hardware, cf. Sec. VIII. This suggests, together with the results of Refs. [124, 126], that the robustness of (low depth) QAOA increases with the decrease of the graph degree. Additional improvement in the performance might be also achieved when using multiqudit gates, which can further reduce the circuit depth [128].

It would be highly interesting to address systematically the above listed scenarios, which we leave for future work. It should be also emphasized that while achieving a practical quantum advantage using the QAOA remains a challenge, constructing a qudit quantum processor is a task worth pursuing - it constitutes an exquisite tool for applications beyond the QAOA, allowing for instance for the realization of a plethora of condensed matter models such as the $d-$state Potts and other SU(d) spin systems [129–144].

Author contributions. JW and FSp have performed the analysis and numerical simulations of the algorithm. AR has supported the analysis of the numerical simu-

[1] E. Farhi, J. Goldstone, and S. Gutmann, A Quantum Approximate Optimization Algorithm, arXiv e-prints (2014), arXiv:1411.4028 [quant-ph].

[2] Z. Jiang, E. G. Rieffel, and Z. Wang, Near-optimal quantum circuit for grover's unstructured search using a transverse field, Phys. Rev. A **95**, 062317 (2017).

[3] S. Lloyd, Quantum approximate optimization is computationally universal, arXiv e-prints (2018), arXiv:1812.11075 [quant-ph].

[4] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás, On the universality of the quantum approximate optimization algorithm, Quantum Information Processing **19**, 291 (2020), arXiv:1909.03123 [quant-ph].

[5] E. Farhi and A. W. Harrow, Quantum Supremacy through the Quantum Approximate Optimization Algorithm, arXiv e-prints (2016), arXiv:1602.07674 [quant-ph].

[6] V. Akshay, H. Philathong, M. E. S. Morales, and J. D. Biamonte, Reachability Deficits in Quantum Approximate Optimization, Phys. Rev. Lett. **124**, 090504 (2020), arXiv:1906.11259 [quant-ph].

[7] M. B. Hastings, Classical and Quantum Bounded Depth Approximation Algorithms, arXiv e-prints (2019), arXiv:1905.07047 [quant-ph].

[8] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, Obstacles to Variational Quantum Optimization from Symmetry Protection, Phys. Rev. Lett. **125**, 260505 (2020), arXiv:1910.08980 [quant-ph].

[9] K. Marwaha, Local classical MAX-CUT algorithm outperforms $p = 2$ QAOA on high-girth regular graphs, arXiv e-prints (2021), arXiv:2101.05513 [quant-ph].

[10] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size, arXiv e-prints (2019), arXiv:1910.08187 [quant-ph].

[11] M. M. Wauters, G. Bigan Mbeng, and G. E. Santoro, Polynomial scaling of QAOA for ground-state preparation of the fully-connected p-spin ferromagnet, arXiv e-prints (2020), arXiv:2003.07419 [quant-ph].

[12] J. Claes and W. van Dam, Instance Independence of Single Layer Quantum Approximate Optimization Algorithm on Mixed-Spin Models at Infinite Size, arXiv e-prints (2021), arXiv:2102.12043 [quant-ph].

[13] J. Wurtz and P. J. Love, Bounds on MAXCUT QAOA Performance for $p > 1$, arXiv e-prints (2020), arXiv:2010.11209.

[14] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, Quantum approximate optimization algorithm for maxcut: A fermionic view, Physical Review A **97**, 022304 (2018).

[15] G. G. Guerreschi and A. Y. Matsuura, Qaoa for maxcut requires hundreds of qubits for quantum speed-up, Scientific reports **9**, 1 (2019).

[16] E. Farhi, J. Goldstone, and S. Gutmann, A Quantum Approximate Optimization Algorithm Applied to a Bounded Occurrence Constraint Problem, arXiv e-prints (2014), arXiv:1412.6062 [quant-ph].

[17] N. Bansal, A. Blum, and S. Chawla, Correlation clustering., Mach. Learn. **56**, 89 (2004).

[18] O. Benjelloun, H. Garcia-Molina, D. Menestrina, Q. Su, S. E. Whang, and J. Widom, Swoosh: a generic approach to entity resolution, The VLDB Journal **18**, 255 (2009).

[19] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, Deepercut: A deeper, stronger, and faster multi-person pose estimation model, in *Computer Vision – ECCV 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling (Springer International Publishing, Cham, 2016) pp. 34–50.

[20] M. Charikar, V. Guruswami, and A. Wirth, Clustering with qualitative information, J. Comput. Syst. Sci. **71**, 360 (2005).

[21] N. Bent, H. Qassim, A. A. Tahir, D. Sych, G. Leuchs, L. L. Sánchez-Soto, E. Karimi, and R. W. Boyd, Experimental realization of quantum tomography of photonic qudits via symmetric informationally complete positive operator-valued measures, Phys. Rev. X **5**, 041006 (2015).

[22] M. Kues, C. Reimer, P. Roztocki, L. R. Cortés, S. Sciara, B. Wetzel, Y. Zhang, A. Cino, S. T. Chu, B. E. Little, *et al.*, On-chip generation of high-dimensional entangled quantum states and their coherent control, Nature **546**, 622 (2017).

[23] A. Babazadeh, M. Erhard, F. Wang, M. Malik, R. Nouroozi, M. Krenn, and A. Zeilinger, High-dimensional single-photon quantum gates: Concepts and experiments, Phys. Rev. Lett. **119**, 180510 (2017).

[24] N. T. Islam, C. C. W. Lim, C. Cahall, J. Kim, and D. J. Gauthier, Provably secure and high-rate quantum key distribution with time-bin qudits, Science advances **3**, e1701491 (2017).

[25] J.-i. Yoshikawa, M. Bergmann, P. van Loock, M. Fuwa, M. Okada, K. Takase, T. Toyama, K. Makino, S. Takeda, and A. Furusawa, Heralded creation of photonic qudits from parametric down-conversion using linear optics, Phys. Rev. A **97**, 053814 (2018).

[26] H.-H. Lu, J. M. Lukens, N. A. Peters, O. D. Odele, D. E. Leaird, A. M. Weiner, and P. Lougovski, Electro-optic frequency beam splitters and tritters for high-fidelity photonic quantum information processing, Phys. Rev. Lett. **120**, 030502 (2018).

[27] P. Imany, J. A. Jaramillo-Villegas, O. D. Odele, K. Han, D. E. Leaird, J. M. Lukens, P. Lougovski, M. Qi, and A. M. Weiner, 50-ghz-spaced comb of high-dimensional frequency-bin entangled photons from an on-chip silicon nitride microresonator, Optics express **26**, 1825 (2018).

[28] J. Wang, S. Paesani, Y. Ding, R. Santagati, P. Skrzypczyk, A. Salavrakos, J. Tura, R. Augusiak, L. Mančinska, D. Bacco, *et al.*, Multidimensional quantum entanglement with large-scale integrated optics, Science **360**, 285 (2018).

[29] C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. R. Cortés, Y. Zhang, B. Fischer, S. Loranger, R. Kashyap,

A. Cino, *et al.*, High-dimensional one-way quantum processing implemented on d-level cluster states, Nature Physics **15**, 148 (2019).

[30] C. Senko, P. Richerme, J. Smith, A. Lee, I. Cohen, A. Retzker, and C. Monroe, Realization of a quantum integer-spin chain with controllable interactions, Phys. Rev. X **5**, 021026 (2015).

[31] J. Randall, S. Weidt, E. D. Standing, K. Lake, S. C. Webster, D. F. Murgia, T. Navickas, K. Roth, and W. K. Hensinger, Efficient preparation and detection of microwave dressed-state qubits and qutrits with trapped ions, Phys. Rev. A **91**, 012322 (2015).

[32] F. M. Leupold, M. Malinowski, C. Zhang, V. Negnevitsky, A. Cabello, J. Alonso, and J. P. Home, Sustained state-independent quantum contextual correlations from a single ion, Phys. Rev. Lett. **120**, 180401 (2018).

[33] R. Hussain, G. Allodi, A. Chiesa, E. Garlatti, D. Mitcov, A. Konstantatos, K. S. Pedersen, R. De Renzi, S. Piligkos, and S. Carretta, Coherent manipulation of a molecular ln-based nuclear qudit coupled to an electron qubit, Journal of the American Chemical Society **140**, 9814 (2018).

[34] R. Sawant, J. A. Blackmore, P. D. Gregory, J. Mur-Petit, D. Jaksch, J. Aldegunde, J. M. Hutson, M. Tarbutt, and S. L. Cornish, Ultracold polar molecules as qudits, New Journal of Physics **22**, 013027 (2020).

[35] V. V. Albert, J. P. Covey, and J. Preskill, Robust encoding of a qubit in a molecule, Phys. Rev. X **10**, 031050 (2020).

[36] M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, E. Lucero, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, A. N. Cleland, *et al.*, Emulation of a quantum spin with a superconducting phase qudit, Science **325**, 722 (2009).

[37] E. Svetitsky, H. Suchowski, R. Resh, Y. Shalibo, J. M. Martinis, and N. Katz, Hidden two-qubit dynamics of a four-level josephson circuit, Nature communications **5**, 1 (2014).

[38] X. Tan, D.-W. Zhang, W. Zheng, X. Yang, S. Song, Z. Han, Y. Dong, Z. Wang, D. Lan, H. Yan, S.-L. Zhu, and Y. Yu, Experimental observation of tensor monopoles with a superconducting qudit, Phys. Rev. Lett. **126**, 017702 (2021).

[39] S. Dogra, K. Dorai, *et al.*, Determining the parity of a permutation using an experimental nmr qutrit, Physics Letters A **378**, 3452 (2014).

[40] Z. Gedik, I. A. Silva, B. Çakmak, G. Karpat, E. L. G. Vidoto, D. d. O. Soares-Pinto, E. DeAzevedo, and F. F. Fanchini, Computational speed-up with a single qudit, Scientific reports **5**, 1 (2015).

[41] F. Moro, A. J. Fielding, L. Turyanska, and A. Patanè, Realization of universal quantum gates with spin-qudits in colloidal quantum dots, Advanced Quantum Technologies **2**, 1900017 (2019).

[42] V. Soltamov, C. Kasper, A. Poshakinskiy, A. Anisimov, E. Mokhov, A. Sperlich, S. Tarasenko, P. Baranov, G. Astakhov, and V. Dyakonov, Excitation and coherent control of spin qudit modes in silicon carbide at room temperature, Nature communications **10**, 1 (2019).

[43] Y. Wang, Z. Hu, B. C. Sanders, and S. Kais, Qudits and high-dimensional quantum computing, Frontiers in Physics **8**, 479 (2020).

[44] M. Saffman, T. G. Walker, and K. Mølmer, Quantum

information with rydberg atoms, Rev. Mod. Phys. **82**, 2313 (2010).

[45] M. Saffman, Quantum computing with atomic qubits and rydberg interactions: progress and challenges, Journal of Physics B: Atomic, Molecular and Optical Physics **49**, 202001 (2016).

[46] D. Barredo, S. de Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys, An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays, Science **354**, 1021 (2016).

[47] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, Atom-by-atom assembly of defect-free one-dimensional cold atom arrays, Science **354**, 1024 (2016).

[48] D. Barredo, V. Lienhard, S. De Leseleuc, T. Lahaye, and A. Browaeys, Synthetic three-dimensional atomic structures assembled atom by atom, Nature **561**, 79 (2018).

[49] P. Scholl, M. Schuler, H. J. Williams, A. A. Eberharter, D. Barredo, K.-N. Schymik, V. Lienhard, L.-P. Henry, T. C. Lang, T. Lahaye, A. M. Läuchli, and A. Browaeys, Programmable quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms, arXiv e-prints (2020), arXiv:2012.12268 [quant-ph].

[50] G. Semeghini, H. Levine, A. Keesling, S. Ebadi, T. T. Wang, D. Bluvstein, R. Verresen, H. Pichler, M. Kalinowski, R. Samajdar, A. Omran, S. Sachdev, A. Vishwanath, M. Greiner, V. Vuletic, and M. D. Lukin, Probing Topological Spin Liquids on a Programmable Quantum Simulator, arXiv e-prints (2021), arXiv:2104.04119 [quant-ph].

[51] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Parallel implementation of high-fidelity multiqubit gates with neutral atoms, Phys. Rev. Lett. **123**, 170503 (2019).

[52] I. S. Madjarov, J. P. Covey, A. L. Shaw, J. Choi, A. Kale, A. Cooper, H. Pichler, V. Schkolnik, J. R. Williams, and M. Endres, High-fidelity entanglement and detection of alkaline-earth rydberg atoms, Nature Physics **16**, 857 (2020).

[53] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, Phys. Rev. X **10**, 021067 (2020).

[54] C. Dalyac, L. Henriet, E. Jeandel, W. Lechner, S. Perdrix, M. Porcheron, and M. Veshchezerova, Qualifying quantum approaches for hard industrial optimization problems. a case study in the field of smart-charging of electric vehicles, EPJ Quantum Technology **8** (2021).

[55] Pasqal, https://pasqal.io/.

[56] Atom computing, https://www.atom-computing.com/.

[57] Coldquanta, https://coldquanta.com/.

[58] Quera, https://www.quera-computing.com/.

[59] L. Henriet, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, Quantum computing with neutral atoms, Quantum **4**, 327 (2020).

[60] M. Morgado and S. Whitlock, Quantum simulation and computing with Rydberg-interacting qubits, AVS Quantum Science **3**, 023501 (2021), arXiv:2011.03031 [quant-ph].

[61] Y. Alexeev, D. Bacon, K. R. Brown, R. Calderbank, L. D. Carr, F. T. Chong, B. DeMarco, D. En-

glund, E. Farhi, B. Fefferman, *et al.*, Quantum computer systems for scientific discovery, arXiv preprint arXiv:1912.07577 (2019).

[62] F. G. Fuchs, H. Ø. Kolden, N. H. Aase, and G. Sartor, Efficient encoding of the weighted max k-cut on a quantum computer using qaoa, SN Computer Science **2**, 1 (2021).

[63] J. Cook, S. Eidenbenz, and A. Bärtschi, The quantum alternating operator ansatz on max-k vertex cover, Bulletin of the American Physical Society **65** (2020).

[64] A. Bärtschi and S. Eidenbenz, Grover mixers for qaoa: Shifting complexity from mixer design to state preparation, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2020) pp. 72–82.

[65] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli, and R. Biswas, From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz, arXiv e-prints (2017), arXiv:1709.03489.

[66] J. Tan, A Note on the Inapproximability of Correlation Clustering, arXiv e-prints (2007), arXiv:0704.2092 [cs.LG].

[67] C. Swamy, Correlation clustering: Maximizing agreements via semidefinite programming, in *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, Vol. 15 (2004) pp. 526–527.

[68] D. Emanuel and A. Fiat, Correlation clustering – minimizing disagreements on arbitrary weighted graphs, in *Algorithms - ESA 2003*, edited by G. Di Battista and U. Zwick (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003) pp. 208–220.

[69] W. M. Soon, H. T. Ng, and D. C. Y. Lim, A machine learning approach to coreference resolution of noun phrases, Computational Linguistics **27**, 521 (2001).

[70] V. Ng and C. Cardie, Improving machine learning approaches to coreference resolution, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 2002) pp. 104–111.

[71] N. Ailon, M. Charikar, and A. Newman, Aggregating inconsistent information: Ranking and clustering, J. ACM **55**, 10.1145/1411509.1411513 (2008).

[72] A. Gionis, H. Mannila, and P. Tsaparas, Clustering aggregation, ACM Trans. Knowl. Discov. Data **1**, 4–es (2007).

[73] A. Goder and V. Filkov, Consensus clustering algorithms: Comparison and refinement, in *Proceedings of the Meeting on Algorithm Engineering & Expermiments* (Society for Industrial and Applied Mathematics, USA, 2008) p. 109–117.

[74] S. Bagon and M. Galun, Large Scale Correlation Clustering Optimization, arXiv e-prints (2011), arXiv:1112.2903 [cs.CV].

[75] M. Elsner and W. Schudy, Bounding and comparing methods for correlation clustering beyond ilp, in *Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*, ILP '09 (Association for Computational Linguistics, USA, 2009) p. 19–27.

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

[77] C. Developers, Cirq (2021), See full list of authors on Github: https://github .com/quantumlib/Cirq/graphs/contributors.

[78] W. Lavrijsen, A. Tudor, J. Müller, C. Iancu, and W. de Jong, Classical Optimizers for Noisy Intermediate-Scale Quantum Devices, arXiv e-prints (2020), arXiv:2004.03004.

[79] Scipy documentation, `https://www.scipy.org`.

[80] R. Shaydulin, I. Safro, and J. Larson, Multistart Methods for Quantum Approximate Optimization, arXiv e-prints (2019), arXiv:1905.08768.

[81] F. G. S. L. Brandao, M. Broughton, E. Farhi, S. Gutmann, and H. Neven, For Fixed Control Parameters the Quantum Approximate Optimization Algorithm's Objective Function Value Concentrates for Typical Instances, arXiv e-prints (2018), arXiv:1812.04170.

[82] M. Streif and M. Leib, Training the Quantum Approximate Optimization Algorithm Without Access to a Quantum Processing Unit, Quantum Science and Technology **5**, 10.1088/2058-9565/ab8c2b (2020), arXiv:1908.08862.

[83] V. Akshay, D. Rabinovich, E. Campos, and J. Biamonte, Parameter Concentration in Quantum Approximate Optimization, arXiv e-prints (2021), arXiv:2103.11976 [quant-ph].

[84] L. Zhu, H. Lun Tang, G. S. Barron, F. A. Calderon-Vargas, N. J. Mayhall, E. Barnes, and S. E. Economou, An adaptive quantum approximate optimization algorithm for solving combinatorial problems on a quantum computer, arXiv e-prints (2020), arXiv:2005.10258 [quant-ph].

[85] Z. Wang, S. Hadfield, Z. Jiang, and E. G. Rieffel, Quantum approximate optimization algorithm for maxcut: A fermionic view, Phys. Rev. A **97**, 022304 (2018).

[86] E. Farhi, D. Gamarnik, and S. Gutmann, The Quantum Approximate Optimization Algorithm Needs to See the Whole Graph: A Typical Case, arXiv e-prints (2020), arXiv:2004.09002 [quant-ph].

[87] S. G. Porsev and A. Derevianko, Hyperfine quenching of the metastable $^{3}P_{0,2}$ states in divalent atoms, Phys. Rev. A **69**, 042506 (2004).

[88] B. DeSalvo, M. Yan, P. Mickelson, Y. Escobar, and T. Killian, Degenerate fermi gas of sr-87, Physical Review Letters **105**, 030402 (2010).

[89] S. Stellmer, R. Grimm, and F. Schreck, Production of quantum-degenerate strontium gases, Phys. Rev. A **87**, 013611 (2013).

[90] S. Stellmer, F. Schreck, and T. C. Killian, Degenerate quantum gases of strontium, in *Annual Review of Cold Atoms and Molecules* (World Scientific, 2014) pp. 1–80.

[91] T. Fukuhara, Y. Takasu, M. Kumakura, and Y. Takahashi, Degenerate fermi gases of ytterbium, Phys. Rev. Lett. **98**, 030401 (2007).

[92] S. Taie, Y. Takasu, S. Sugawa, R. Yamazaki, T. Tsujimoto, R. Murakami, and Y. Takahashi, Realization of a $SU(2) \times SU(6)$ system of fermions in a cold atomic gas, Phys. Rev. Lett. **105**, 190401 (2010).

[93] S. Sugawa, Y. Takasu, K. Enomoto, and Y. Takahashi, Ultracold ytterbium: Generation, many-body physics, and molecules, in *Annual Review of Cold Atoms and Molecules: Volume 1* (2013) pp. 3–51.

[94] O. Onishchenko, S. Pyatchenkov, A. Urech, C.-C. Chen, S. Bennetts, G. A. Siviloglou, and F. Schreck, Frequency

of the ultranarrow $^1S_0 - ^3P_2$ transition in ^{87}Sr, Phys. Rev. A **99**, 052503 (2019).

[95] A. Urech, Ph.D. Thesis, University of Amsterdam, In preparation.

[96] M. Safronova, private communication.

[97] I. H. Deutsch and P. S. Jessen, Quantum-state control in optical lattices, Phys. Rev. A **57**, 1972 (1998).

[98] M. M. Boyd, T. Zelevinsky, A. D. Ludlow, S. Blatt, T. Zanon-Willette, S. M. Foreman, and J. Ye, Nuclear spin effects in optical lattice clocks, Phys. Rev. A **76**, 022510 (2007).

[99] T. F. Gallagher, *Rydberg atoms*, 3 (Cambridge University Press, 2005).

[100] A. Browaeys and T. Lahaye, Many-body physics with individually controlled rydberg atoms, Nature Physics **16**, 132 (2020).

[101] We note that analogous approach has been adopted in Ref. [52], where the $^3P_0$ manifold has been used for a realization of a qubit state.

[102] For large d (i.e. close to $d_{\max} = 10$), achieving $c_{\ell,\ell'=\ell+1} = 1$, $\forall \ell$, might be challenging. This is because the ratio $\Omega_{0,1}/\Omega_{9,10}$ between the Rabi frequencies coupling the smallest and largest ℓ states is too large (or too small, depending on the orientation of the magnetic field) so that $c_{\ell,\ell'=\ell+1} = 1$, $\forall \ell$, implies either slow operation or further detrimental decoherence due to the off-resonant scattering of the intense mixer beams [95]. In this case the mixers could be implemented in a stepwise fashion by means of the Givens rotations [103].

[103] D. P. O'Leary, G. K. Brennen, and S. S. Bullock, Parallelism for quantum computation with qudits, Phys. Rev. A **74**, 032334 (2006).

[104] For a qubit, the cost of the symmetric gate Eq. (26) can be further reduced to $[\text{CP}] = d$, cf. Appendix D.

[105] J. C. Garcia-Escartin and P. Chamorro-Posada, A swap gate for qudits, Quantum information processing **12**, 3625 (2013).

[106] S. Balakrishnan, Various constructions of qudit swap gate, Physics Research International **2014** (2014).

[107] A. M. C. Stroud, Quantum fast fourier transform using multilevel atoms, journal of modern optics **49**, 2115 (2002).

[108] A. Urech, I. Knotterus, R. Spreeuw, and F. Schreck, In preparation.

[109] S. Saskin, J. T. Wilson, B. Grinkemeyer, and J. D. Thompson, Narrow-line cooling and imaging of ytterbium atoms in an optical tweezer array, Phys. Rev. Lett. **122**, 143002 (2019).

[110] J. P. Covey, I. S. Madjarov, A. Cooper, and M. Endres, 2000-times repeated imaging of strontium atoms in clock-magic tweezer arrays, Phys. Rev. Lett. **122**, 173201 (2019).

[111] I. S. Madjarov, *Entangling, Controlling, and Detecting Individual Strontium Atoms in Optical Tweezer Arrays*, Ph.D. thesis, California Institute of Technology (2021).

[112] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Experimental comparison of two quantum computing architectures, Proceedings of the National Academy of Sciences **114**, 3305 (2017).

[113] M. Saffman and K. Mølmer, Scaling the neutral-atom rydberg gate quantum computer by collective encoding in holmium atoms, Phys. Rev. A **78**, 012336 (2008).

[114] Y. Wang, S. Shevate, T. M. Wintermantel, M. Morgado,

G. Lochead, and S. Whitlock, Preparation of hundreds of microscopic atomic ensembles in optical tweezer arrays, npj Quantum Information **6**, 1 (2020).

[115] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Elementary gates for quantum computation, Phys. Rev. A **52**, 3457 (1995).

[116] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, Phys. Rev. Lett. **120**, 110501 (2018).

[117] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Low-depth quantum simulation of materials, Phys. Rev. X **8**, 011044 (2018).

[118] Y. Liu, G. L. Long, and Y. Sun, Analytic Constructions of General n-Qubit Controlled Gates, arXiv e-prints (2007), arXiv:0708.3274 [quant-ph].

[119] T. Sleator and H. Weinfurter, Realizable universal quantum logic gates, Phys. Rev. Lett. **74**, 4087 (1995).

[120] T. Itoko, R. Raymond, T. Imamichi, and A. Matsuo, Optimization of quantum circuit mapping using gate transformation and commutation, Integration **70**, 43 (2020).

[121] K.-N. Schymik, V. Lienhard, D. Barredo, P. Scholl, H. Williams, A. Browaeys, and T. Lahaye, Enhanced atom-by-atom assembly of arbitrary tweezer arrays, Phys. Rev. A **102**, 063107 (2020).

[122] J. M. Auger, S. Bergamini, and D. E. Browne, Blueprint for fault-tolerant quantum computation with rydberg atoms, Phys. Rev. A **96**, 052320 (2017).

[123] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, Effects of Quantum Noise on Quantum Approximate Optimization Algorithm, arXiv e-prints (2019), arXiv:1909.02196 [quant-ph].

[124] D. Stilck Franca and R. Garcia-Patron, Limitations of optimization algorithms on noisy quantum devices, arXiv e-prints (2020), arXiv:2009.05532 [quant-ph].

[125] A technical remark is that strictly speaking the formula (45) was verified when the number of clusters d saturates the number of vertices N. Keeping this in mind, the discussed achievable system sizes should thus be considered merely as an estimate.

[126] 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, Nature Physics **17**, 332 (2021).

[127] P. N. Klein, C. Mathieu, and H. Zhou, Correlation clustering and two-edge-connected augmentation for planar graphs, in *32nd International Symposium on Theoretical Aspects of Computer Science, STACS 2015, March 4-7, 2015, Garching, Germany* (2015) pp. 554–567.

[128] E. O. Kiktenko, A. S. Nikolaeva, P. Xu, G. V. Shlyapnikov, and A. K. Fedorov, Scalable quantum computing with qudits on a graph, Phys. Rev. A **101**, 022304 (2020).

[129] F. Y. Wu, The potts model, Rev. Mod. Phys. **54**, 235 (1982).

[130] P. Nataf and F. Mila, Exact diagonalization of heisenberg su (n) models, Physical review letters **113**, 127204 (2014).

[131] P. Nataf and F. Mila, Exact diagonalization of heisenberg su (n) chains in the fully symmetric and antisym-

metric representations, Physical Review B **93**, 155134 (2016).

[132] F. H. Kim, K. Penc, P. Nataf, and F. Mila, Linear flavor-wave theory for fully antisymmetric su (n) irreducible representations, Physical Review B **96**, 205142 (2017).

[133] P. Nataf and F. Mila, Density matrix renormalization group simulations of su (n) heisenberg chains using standard young tableaus: Fundamental representation and comparison with a finite-size bethe ansatz, Physical Review B **97**, 134420 (2018).

[134] J. Dufour, P. Nataf, and F. Mila, Variational monte carlo investigation of su (n) heisenberg chains, Physical Review B **91**, 174427 (2015).

[135] C. Romen and A. M. Läuchli, Structure of spin correlations in high-temperature su (n) quantum magnets, Physical Review Research **2**, 043009 (2020).

[136] P. Chen, Z.-L. Xue, I. McCulloch, M.-C. Chung, C.-C. Huang, and S.-K. Yip, Quantum critical spin-2 chain with emergent su (3) symmetry, Physical review letters **114**, 145301 (2015).

[137] J. D'Emidio, M. S. Block, and R. K. Kaul, Rényi entanglement entropy of critical su (n) spin chains, Physical Review B **92**, 054411 (2015).

[138] H. Song, M. Hermele, *et al.*, Mott insulators of ultracold fermionic alkaline earth atoms in three dimensions, Physical Review B **87**, 144423 (2013).

[139] P. Corboz, K. Penc, F. Mila, and A. M. Läuchli, Simplex solids in su(n) heisenberg models on the kagome and checkerboard lattices, Phys. Rev. B **86**, 041106 (2012).

[140] P. Corboz, M. Lajkó, K. Penc, F. Mila, and A. M. Läuchli, Competing states in the su(3) heisenberg model on the honeycomb lattice: Plaquette valence-bond crystal versus dimerized color-ordered state, Phys. Rev. B **87**, 195113 (2013).

[141] P. Nataf, M. Lajkó, P. Corboz, A. M. Läuchli, K. Penc, and F. Mila, Plaquette order in the su(6) heisenberg model on the honeycomb lattice, Phys. Rev. B **93**, 201113 (2016).

[142] P. Corboz, A. M. Läuchli, K. Penc, M. Troyer, and F. Mila, Simultaneous dimerization and su(4) symmetry breaking of 4-color fermions on the square lattice, Phys. Rev. Lett. **107**, 215301 (2011).

[143] B. Bauer, P. Corboz, A. M. Läuchli, L. Messio, K. Penc, M. Troyer, and F. Mila, Three-sublattice order in the su(3) heisenberg model on the square and triangular lattice, Phys. Rev. B **85**, 125116 (2012).

[144] A. Weichselbaum, S. Capponi, P. Lecheminant, A. M. Tsvelik, and A. M. Läuchli, Unified phase diagram of antiferromagnetic su (n) spin ladders, Physical Review B **98**, 085104 (2018).

[145] J. Weggemans, *Solving Correlation Clustering with the Quantum Approximate Optimisation Algorithm*, Master's thesis, University of Twente, The Netherlands (2020), available at `http://essay.utwente.nl/85484/`.

[146] Our derivation does not include the possibility that G is the complete $N = 4$ graph, which is not neccesary as it cannot be extended.

[147] H.-W. Jiao, F.-H. Wang, and Y.-Q. Chen, An effective branch and bound algorithm for minimax linear fractional programming, Journal of Applied Mathematics **2014**, 1 (2014).

[148] S. Caron, lpsolvers, `https://pypi.org/project/lpsolvers/` (2007).

[149] M. O. Scully and M. S. Zubairy, Quantum optics (1999).

[150] A. Steck, Quantum and atom optics.

[151] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, Low-depth quantum simulation of materials, Phys. Rev. X **8**, 011044 (2018).

[152] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, Phys. Rev. Lett. **120**, 110501 (2018).

[153] M. Saffman and T. G. Walker, Analysis of a quantum logic device based on dipole-dipole interactions of optically trapped rydberg atoms, Phys. Rev. A **72**, 022347 (2005).

[154] N. Šibalić, J. D. Pritchard, C. S. Adams, and K. J. Weatherill, Arc: An open-source library for calculating properties of alkali rydberg atoms, Computer Physics Communications **220**, 319 (2017).

[155] N. Šibalić, J. D. Pritchard, C. S. Adams, and K. J. Weatherill, Arc package, `https://arc-alkali-rydberg-calculator.readthedocs.io/en/latest/`.

[156] L. Henriet, Robustness to spontaneous emission of a variational quantum algorithm, Phys. Rev. A **101**, 012335 (2020).

Appendix A: Results optimizer study

We compare the performance of five different optimizers with off-the-shelf hyper-parameter settings using the vanilla QAOA: ImFil, SnobFit and BOBYQA taken from the scikit-quant package [78] and an implementation of COBYLA, used both as single optimizer and in conjunction with basin-hopping (BH), both from the SciPy optimization package [79]. We consider a single instance out of our correlation clustering data sets: the complete graph with $N = 4$ and all edge weights '$-$' such that the optimal solution corresponds to all nodes being put in different clusters. For each $p \in \{1, \ldots, 5\}$, we generate 25 random initial points in the respective parameter space. The maximum number of iterations for each optimizer is set to 500. Since basin-hopping has two different budgets, one for the basin-hopping steps and one for its local optimizer, we set the maximum number of iterations by using the number of evaluations the local optimizer (COBYLA) used in its individual run: the number of basin-hopping steps is the rounded ratio of the budget over this number. The results for state vector sampling (1000 samples) are given in Figure 9.
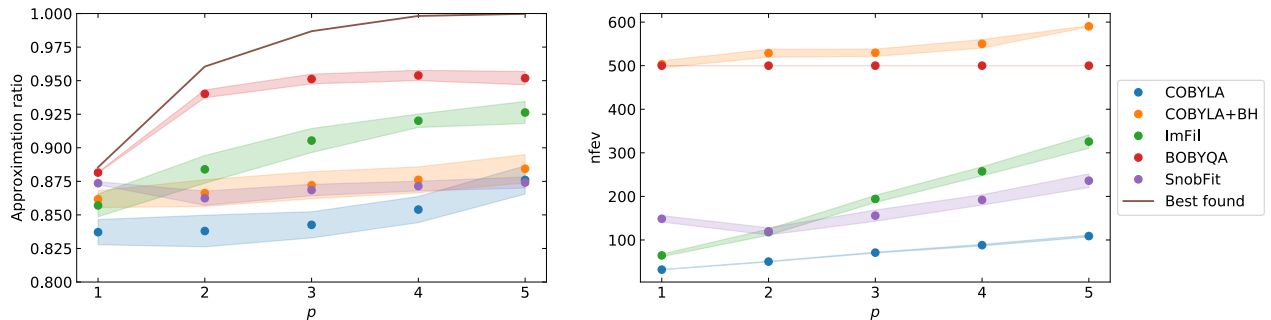
FIG. 9. State vector sampling with 1000 samples. Left: found approximation ratios for 25 random initial points using different optimizers. Right: total number of function evaluations 'nfev' as a function of p. The shaded area indicates the error in the mean, where the discrete points have been connected in order to improve the readability of the figure. 'Best found' indicates the best approximation ratio that was observed over all instances and optimizers.

We find that BOBYQA outperforms all other optimizers in terms of the achieved approximation ratios, but does always use up the maximum available amount of iterations. Unfortunately, the scikit-quant optimization package does not allow us to change the tolerance, which would allow for fairer comparison. As far as the ratio of performance to number of function evaluations is concerned, ImFil performs well. Even though adopting BOBYQA already potentially results in a large performance increase compared to using COBYLA, which was used to obtain our initial results, we still observe that a large performance difference exists between the best value found and the average performance over different optimizer runs. In addition, note that the standard deviation is relatively large as we plot the error in the mean in Fig. 9. This means that there is still a lot to be gained in the classical optimization step, the most natural being the use of good initial points, followed by hyper-parameter optimization. Good initial points will have a larger effect on the local optimization methods (in particular COBYLA and BOBYQA) compared to the global optimizers (e.g. ImFil). The concentration of initial points at $p = 1$ is studied in the next appendix.

Appendix B: Initial points study for $p = 1$

In this part of the appendix we give numerical evidence that supports the conclusions of the work by Brandão et al. [81]: initial points for different instances, belonging to similar problem classes, are concentrated. To investigate this effect at $p = 1$, we run the following: we start with initial points we obtained from the all-negative-weights graphs, for some fixed N and d, and then solve for all 50 instances in our correlation clustering data set, using COBYLA as the optimizer. The resulting optimal points are plotted in Fig. 10: Note how all points are in the neighbourhood of our initial points. In fact, the smallest rectangular area containing all points (indicated by the black rectangle) takes about 0.2% of the entire possible parameter space. The plot on the right zooms in on this rectangle, showing how the optimal points are located relative to each other. In this plot x_0^* is the initial point, which is for both $N = 4$ and $N = 5$ positioned at the boundary of the collection of points. This makes sense due to the structure of the graph it belongs to—the all-negative-weights graph is itself an extreme case of the correlation clustering problem (requiring all clusters to be used). This also indicates that other graphs might be more suitable to generate initial points.

For the same problem instance, Figure 11 shows the location of the points obtained for different N and d in parameter space ($p = 1$). We observe that $d = 2$ is somewhat of an outsider, but again all points fall in a rectangular area encompassing about 0.2% of the entire possible parameter space.

Appendix C: Approximation ratio bound on 3-regular graphs

A recent work by Wurtz and Love [13] shows a derivation of lower bounds for QAOA depths $p = 1$ and $p = 2$ (and a conjectured result at $p = 3$) on MAXCUT. In this appendix, we apply their techniques to the correlation clustering problem on 3-regular graphs: graphs for which every node has a fixed degree of 3. The goal is to find a lower bound for the approximation ratio α at $p = 1$ on 3-regular graphs, defined as

$$\alpha = \max_{d \in \{1, \ldots, d_{\max}\}} \frac{F_d(\gamma, \beta)}{C_{\max}}, \tag{C1}$$
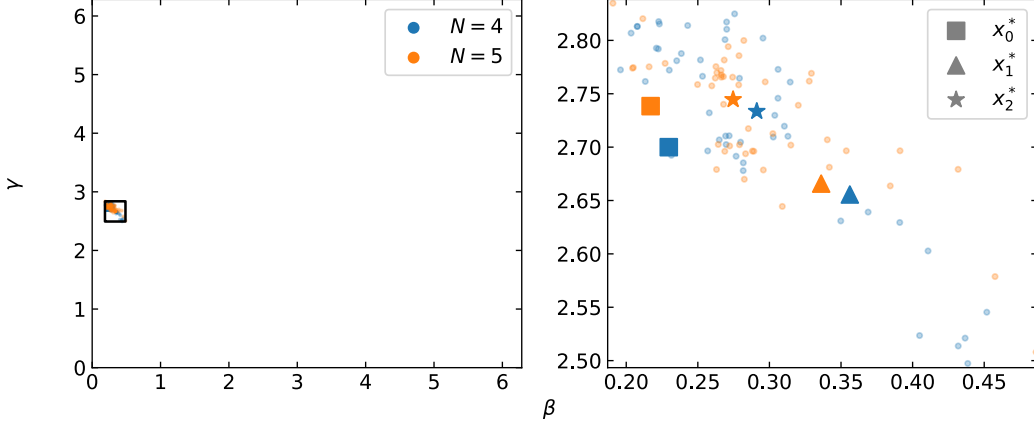
FIG. 10. Locations of points obtained through the use of COBYLA starting from initial point x_0^*, which corresponds to an optimized point for solving the all-negative-weights graph, for $N = 4$ and $N = 5$. The number of levels is set as $d = N$. Left: overview over the entire possible parameter space. Right: close-up to the smallest possible square area that contains all optimal points. The point x_1^* has the smallest maximum Euclidean distance to other points and x_2^* the smallest average Euclidean distance to all other points.
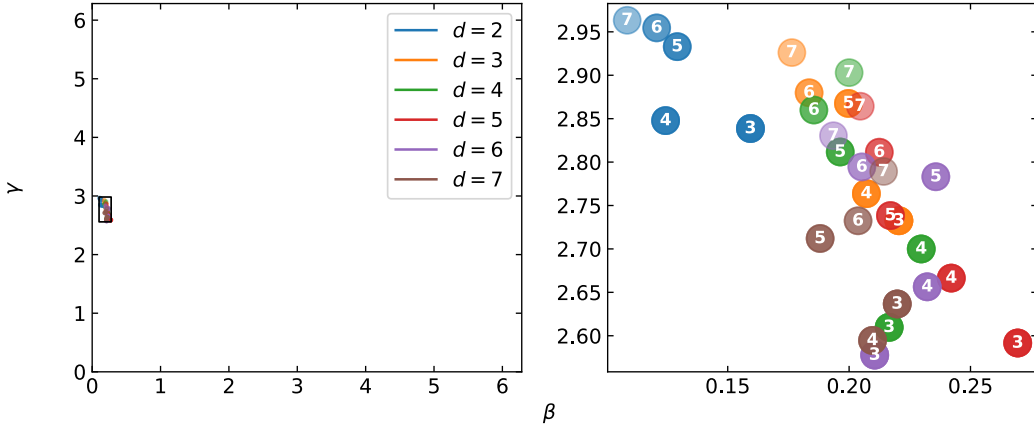

FIG. 11. Left: locations of optimal points obtained over the entire possible parameter space. Right: close-up of the smallest rectangle that contains all optimal points. The number inside the point indicates the number of nodes.

where $F_d(\gamma, \beta)$ is the expectation value of the state produced by the QAOA algorithm using d levels and $C_{\max}$ denotes the optimal objective function value for a single correlation clustering instance. Since only graphs with degree 3 are considered, we need at most 4 clusters. Therefore, we only consider $d \in \{1, 2, 3, 4\}$. Based on local optimization, we will conjecture that for 3-regular graphs $G = (V, E)$ initial points γ_d^*, β_d^* exist such that QAOA that loops over the clusters has an approximation ratio larger then 0.6699, even without the classical optimization loop[146]. By relaxing the problem into a linear program (LP) we can prove that this bound is at least 0.6367.

1. Problem setup and a lower bound for the energy

Consider an arbitrary 3-regular graph G of $N(G)$ nodes that is not the complete graph with $N = 4$. We identify for each edge $\langle i, j \rangle$ the sub-graph $G_{\langle i,j \rangle}^p$ induced by all neighbouring edges at most p steps away from $\langle i, j \rangle$. At $p = 1$ there are only three possible kinds of sub-graph structures as indicated in Fig. 12.
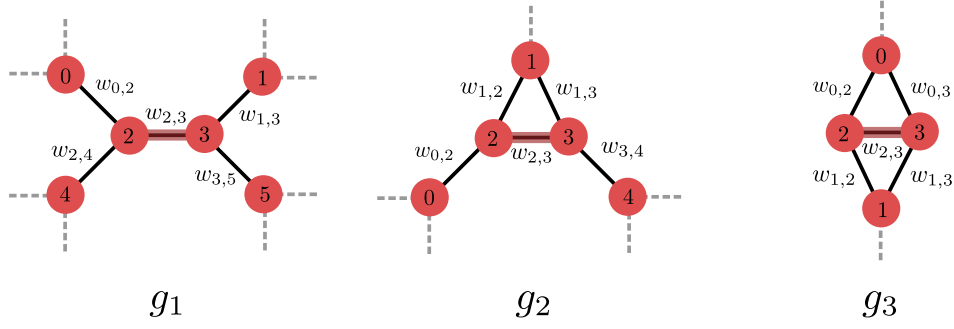
FIG. 12. The 3 types of sub-graphs for $p = 1$. The sub-graphs describe the environment of the highlighted edge, note how only neighbouring edges are included in the sub-graph. The dotted edges indicate edges outside the sub-graph.

Since all sub-graph types have 5 edges, we have a total of $3 \cdot 2^5 = 96$ possible sub-graphs when we only consider weights $w_{u,v} \in \{-1, +1\}$. However, by symmetry arguments we can reduce the total number of weighted sub-graphs we have to consider: if reordering the edge labels results in the same graph under any rotation, it will look the same to the QAOA. We define three sets of sub-graphs $g_i, i \in \{1, 2, 3\}$, representing all 3-regular sub-graph structures with 6, 5 and 4 nodes respectively (see Figure 12), such that for every sub-graph $\lambda \in g_i$ there exists no other graph $\lambda' \in g_i$ that is equivalent in the QAOA setting. Our total set of possible sub-graphs is then $S = g_1 \cup g_2 \cup g_3$. We now decompose our graph G into sub-graph environments $\lambda \in S$ for which the multiplicity of each λ is $N_\lambda(G)$. Since we have such a sub-graph environment for every edge, we must have that the sum over all $N_\lambda(G)$ is equal to the total number of edges $3N(G)/2$. For a single edge $\langle u, v \rangle$ with sub-graph environment λ, denoted as $\langle u, v \rangle \to \lambda$, the contribution to the expectation value is given by

$$f_{d, \langle u,v \rangle \to \lambda}(\beta_d, \gamma_d) = \langle \beta_d, \gamma_d | w_{\langle u,v \rangle} V_d | \beta_d, \gamma_d \rangle. \tag{C2}$$

Note that Eq. (C2) only contains terms that operate within λ. The expectation value of the algorithm with a maximum of k clusters at $p = 1$ for some β_d, γ_d is given by

$$F_d(\beta_d, \gamma_d) = \sum_\lambda N_\lambda(G) f_{d, \langle u,v \rangle \to \lambda}(\beta_d, \gamma_d). \tag{C3}$$

We must also have that this is always smaller than or equal to the expectation value using the best values of β_d, γ_d

$$F_{d,\max} = \max_{\beta_d, \gamma_d} F_d(\beta_d, \gamma_d) \geq \sum_\lambda N_\lambda(G) f_{d, \langle u,v \rangle \to \lambda}(\beta_d, \gamma_d), \tag{C4}$$

providing us with a lower bound on the expectation value of the algorithm $F_{d,\max}$ at a given d.

2. Upper bound on the number of agreements

We are now faced with the task of finding an upper bound on $C_{\max}$. A naive bound would be the total number of edges, but we can do better by considering the same argument Wurtz and Love used to determine an upper bound for MAXCUT [13]. Consider the graph $\mathcal{G}$, which is a collection of $N_\lambda(G)$ disconnected sub-graphs $G^p_{\langle i,j \rangle}$ for each edge in G. Since the largest sub-graph has only 6^4 feasible solutions, a brute-force method can be used to find the ratio between the optimal objective function value and the number of edges for every sub-graph λ, which we will call c_λ. Since all sub-graphs are isolated (i.e. not connected to each other), the global fraction of agreements to edges is equal to the average fraction over each sub-graph. However, we have several edges belonging to different disjoint sub-graphs in $\mathcal{G}$ that are actually the same edge in G. In this case, we can have that for both sub-graphs a clustering exists for which the edge contributes to the objective value, but that the required clustering is different in both sub-graphs. As a result, we have that the objective function value of G is bounded from above by

$$C_{\max} \leq \sum_\lambda N_\lambda(G) c_\lambda. \tag{C5}$$

3. Constructing the hardest graph

Since we have a lower bound for the algorithm's expectation value and an upper bound for the optimal objective function value, the approximation ratio α is bounded from below by

$$\alpha(G) \geq \max_d \frac{\sum_\lambda N_\lambda(G) f_{d,\lambda}(\gamma_d, \beta_d)}{\sum_\lambda N_\lambda(G) c_\lambda}. \tag{C6}$$

The worst case approximation ratio is then given by the hardest graph $G = G^*$, which corresponds to some combinations of $N_\lambda(G^*) \in \mathbb{Z}_0$ for all sub-graphs λ. However, not all combinations of $N_\lambda(G)$ correspond to a valid graph, as was already noted by Farhi et al. [1]. For now, we will only consider the structure of the graph and not take the feasibility of certain weight combinations into account. First, we note that for every edge in sub-graph g_3 there must be at least 4 other edges that have the environment corresponding to sub-graph g_2. Also, we have that the 'triangle' of g_2 and 'crossed square' of g_3 cannot share the same vertex, which means that the number of triangular edges and crossed square edges must be smaller than the number of nodes $N(G)$. Our final constraints are that all $N_\lambda(G)$ are non-negative integers and must sum up to $3N(G)/2$. We define $n_\lambda(G) \equiv N_\lambda(G)/N(G)$ such that we can relax the integer constraint in the limit of large $N(G)$, and obtain the following *Minimax Linear Fractional Program*:

$$\min_{n_\lambda(G):\lambda \in S} \quad \max_d \frac{\sum_\lambda n_\lambda(G) f_{d,\lambda}(\gamma_d, \beta_d)}{\sum_\lambda n_\lambda(G) c_\lambda}$$

$$\text{s.t.} \quad \sum_{n_\lambda(G):\lambda \in g_2} n_\lambda(G) - 4 \sum_{n_\lambda(G):\lambda \in g_3} n_\lambda(G) \geq 0$$

$$- \sum_{n_\lambda(G):\lambda \in g_2} n_\lambda(G) \geq -1 \tag{C7}$$

$$\sum_{n_\lambda(G):\lambda \in S} n_\lambda(G) = \frac{3}{2}$$

$$n_\lambda(G) \geq 0 \text{ for all } \lambda \in S$$

Equation (C7) is in the form of a generalized fractional program, which is not reducible to a linear program (LP) which can be solved efficiently. However, there exist linear relaxation bounding techniques that do allow for global optimization up to some error ϵ. We have performed initial experiments with one of those techniques [147], but were not able to achieve desirable results so far due to the difficulties in approximating our objective function. However, we can construct two related LP formulations that upper and lower bound the value of α. Let us define the feasible region $\mathcal{C}$ such that $x \in \mathcal{C}$ when it satisfies the constraints of (C7).

1. Take the number of edges as the upper bound instead of the fractional objectives (C5), i.e. we let $c_\lambda \to 1$ for all λ. Under this relaxation we can write (C7) as the following LP:

$$\min \quad \alpha$$

$$\text{s.t.} \quad \sum_\lambda n_\lambda(G) f_{d,\lambda}(\gamma_d, \beta_d) \leq \frac{3}{2}\alpha \quad \text{for all } d \tag{C8}$$

$$n_\lambda(G) \in \mathcal{C}, \alpha \in \mathbb{R}$$

The solution of (C8) is a lower bound to the actual bound, as sub-graphs that originally had $c_\lambda = 0.8$ now contribute too much to the upper bound of the optimal objective function value (C5).

2. Similarly, we can also assume that only sub-graphs for which a perfect clustering exists contribute to the construction of the most difficult graph. Define the set $S' = \{\lambda | \lambda \in S, c_\lambda = 1\}$ to be such a set. Since $S' \subset S$, the resulting solution provides an upper bound to the best value of α that we can find for $\lambda \in S$.

$$\min \quad \alpha$$

$$\text{s.t.} \quad \sum_\lambda n_\lambda(G) f_{d,\lambda}(\gamma_d, \beta_d) \leq \frac{3}{2}\alpha \quad \text{for all } d \tag{C9}$$

$$n_\lambda(G) \in \mathcal{C}, \alpha \in \mathbb{R}, \lambda \in S'$$

All LPs will be solved by CVXOPT contained in the package *lpsolvers* [148] for Python.

4. Iterative procedure for determining the bound

The minimax optimization problem gives us a method to determine the hardest instance G^*, given that we fix the parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4), \beta = (\beta_1, \beta_2, \beta_3, \beta_4)$. But how do we choose the values of these parameters? As one would normally do with QAOA, a classical optimization loop can be adopted. First, we choose some initial values for γ, β and calculate $f_{d,\lambda}(\gamma_d, \beta_d)$ for all sub-graph environments λ and all $d \in \{1, 2, 3, 4\}$. Next, we construct the hardest graph G^* by solving (C7), (C8) or (C9). Instead of minimizing over n_λ whilst keeping γ, β fixed, in the next step we now fix n_λ and try to maximize the objective over γ, β, i.e., we want to

$$\max_{\gamma_d, \beta_d} \quad \frac{\sum_\lambda n_\lambda(G) f_{d,\lambda}(\beta_d, \gamma_d)}{\sum_\lambda n_\lambda(G) c_\lambda} \tag{C10}$$
$$\text{s.t.} \quad \gamma_d, \beta_d \in [0, 2\pi),$$

for all $d \in \{1, 2, 3, 4\}$. For these new values of γ^*, β^*, there might exist some other graph G^{**} that is more difficult than the one we originally obtained. After we obtained G^{**} we can again try to find new parameters β^{**}, γ^{**}. This suggests the use of an iterative procedure for establishing the bound. We do not know whether this procedure converges, but it doesn't necessarily have to: any combination of γ, β has its own lower bound that holds specifically for these parameters, and can therefore be used as our result. Convergence would only suggest something about the tightness of this bound, guaranteed global convergence would mean that no better values of γ, β exist. In our iterative procedure, the best parameter combinations we found are listed in Table III:

	$d = 1$	$d = 2$	$d = 3$	$d = 4$
γ_d^*	$-$	2.857	2.773	2.682
β_d^*	$-$	0.4833	0.1310	0.1435

TABLE III. Parameter values for different d at which we were able to obtain the results of Theorem C.1 and Conjecture C.1. At $d = 1$ the algorithm has only one state and hence no parameters.

For the parameter combinations in Table III, (C8) can be solved to an arbitrary precision, which establishes the proof of the following theorem:

Theorem C.1. *For 3-regular graphs $G = (V, E)$, where G is not the complete $N = 4$ graph, initial points γ_d^*, β_d^* exist such that at $p = 1$ the QAOA that loops over the clusters gives a 0.6367-approximation algorithm, even without the classical optimization loop.*

However, as stated before, this bound is too strict as the objective function of G is overestimated. Using COBYLA as a local optimizer, we numerically observed that solving (C7) never resulted in a bound lower than 0.6699. This is equal to the upper bound we find by solving (C9) (also 0.6699). Since COBYLA does not guarantee a global minimum and (C9) only considers a subset of all possible graphs, we can only conjecture that this is the actual lower bound:

Conjecture C.1. *For 3-regular graphs $G = (V, E)$, where G is not the complete $N = 4$ graph, initial points γ_d^*, β_d^* exist such that at $p = 1$ the QAOA that loops over the clusters gives a 0.6699-approximation algorithm, even without the classical optimization loop.*

For this particular combination of sub-graphs belong to the hardest graph G^*, the classical optimization step actually does not make much of a difference: our results show that solving (C10) with $G = G^*$ results in

$$\max_{\gamma_d, \beta_d} \quad \max_d \frac{\sum_\lambda n_\lambda(\mathcal{G}^*) f_{d,\lambda}(\gamma_d, \beta_d)}{\sum_\lambda n_\lambda(\mathcal{G}^*) c_\lambda} \approx 0.674,$$

which is only a small improvement on the conjectured bound. This shows the quality of these values of γ^*, β^* (as well as the hardness of the graph G^*). Additionally, this also provides further indication that QAOA, when having access to good initial points, can also be used without the classical optimization step [81, 82].

5. Performance bounds for $p > 1$?

Unfortunately, we do not observe evidence for the existence of a trivial graph hierarchy as Wurtz and Love proved (and conjectured) for MAXCUT at $p \leq 2$ ($p > 2$) [13]. In fact, when we consider their large loop conjecture [13] for our problem, we find that for our used γ, β we obtain a worst-case approximation ratio of $\alpha = 0.693$, which is significantly
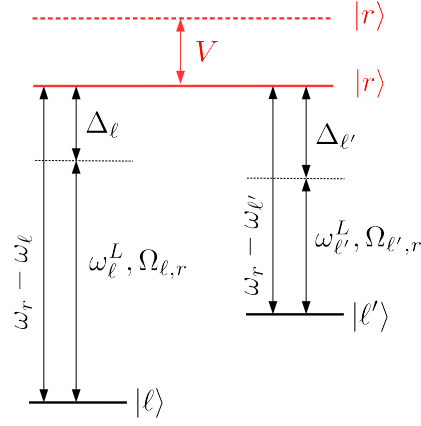
FIG. 13. Scheme of a driven three-level system in a so-called Λ-configuration. Two qudit states $|\ell\rangle, |\ell'\rangle$ of energies $\omega_\ell, \omega_{\ell'}$ are coupled with laser light of frequencies $\omega_\ell^L, \omega_{\ell'}^L$ and Rabi frequencies $\Omega_{\ell,r}, \Omega_{\ell',r}$ to a Rydberg state $|r\rangle$ of energy ω_r. The Rydberg state can be additionally shifted by energy $V(R)$ if an atom at distance R is excited to a Rydberg state (we recall we use $\hbar = 1$).

larger than the bounds in Theorem C.1 and Conjecture C.1. Therefore, we do not conjecture the structure of the most difficult graph at any p, which would ease the determination of lower bounds at larger p. If we are to use the same method as we used for $p = 1$, we will have to consider of the order of $123 \cdot 2^{13} \approx 10^6$ different sub-graphs (not taking symmetries into account). We can reduce this number by exploiting symmetries, but since determining the energy of the largest sub-graph (14 nodes) is computationally very expensive we do not attempt to determine bounds for $p > 1$. In fact, a back-of-the-envelope estimation to the amount of computing hours needed to execute such a computation – again not utilizing symmetries – shows that we would need on the order of 10 million computing hours.

Appendix D: Details of the experimental building blocks

1. Dynamics of a driven three-level system

Here we briefly review the standard derivation of the unitaries Eqs. (19),(20). Let us consider the level scheme depicted in Fig. 13. Two qudit states $|\ell\rangle, |\ell'\rangle$ of energies $\omega_\ell, \omega_{\ell'}$ are coupled with lasers of frequencies $\omega_\ell^L, \omega_{\ell'}^L$ and Rabi frequencies $\Omega_{\ell,r}, \Omega_{\ell',r}$ to a Rydberg state $|r\rangle$ of energy ω_r. Furthermore, we allow for the Rydberg state to be shifted by energy V if a nearby atom is in a Rydberg state ($V = V(R) = C_6/R^6$ for the atoms separated by a distance R and C_6 is the Van der Waals coefficient, cf. Sec. VI). Within the rotating wave approximation and in the frame where the levels $|\ell\rangle, |\ell'\rangle$ rotate at the laser frequencies $\omega_\ell, \omega_{\ell'}$ respectively [149, 150], the corresponding Hamiltonian reads

$$H = \sum_{j=\ell,\ell'} \Omega_{j,r} |j\rangle\langle r| + \text{H.c.} + \Delta_j |j\rangle\langle j| + V|r\rangle\langle r|. \tag{D1}$$

Here $\Delta_j = \omega_j^L - (\omega_r - \omega_j)$ is the single-photon detuning between the qudit state $j = \ell, \ell'$ and the Rydberg state $|r\rangle$ (note our sign convention, where $\Delta_j < 0$ refers to a red-detuned laser beam). Written in the $\{|\ell\rangle, |\ell'\rangle, |r\rangle\}$ basis, the Hamiltonian (D1) is

$$H = \begin{pmatrix} \Delta_\ell & 0 & \Omega_{\ell,r} \\ 0 & \Delta_{\ell'} & \Omega_{\ell',r} \\ \Omega_{\ell,r}^* & \Omega_{\ell',r}^* & V \end{pmatrix}. \tag{D2}$$

The associated unitary operator $U = \mathrm{e}^{-itH}$ can in principle be obtained analytically by diagonalizing (D2), leading to a cubic equation for the eigenvalues. The situation simplifies at two-photon resonance $\Delta_\ell = \Delta_{\ell'} = 0$. In the limit of large interaction, $V \to \infty$, we get

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mathrm{e}^{-iVt} \end{pmatrix}. \tag{D3}$$

Denoting $\Omega_0 = \Omega_{\ell,r}, \Omega_1 = \Omega_{\ell',r}$ for the ease of notation, the $|r\rangle\langle r|$ element of U for a general V reads

$$U_{rr} = e^{-\frac{1}{2}itV}\left(\cos\left(\frac{1}{2}t\sqrt{V^2 + 4\Omega^2}\right) - \frac{iV\sin\left(\frac{1}{2}t\sqrt{V^2 + 4\Omega^2}\right)}{\sqrt{V^2 + 4\Omega^2}}\right), \tag{D4}$$

where

$$\Omega = \sqrt{|\Omega_0|^2 + |\Omega_1|^2}. \tag{D5}$$

Requiring that the Rydberg population returns to zero at the end of the evolution, if it was initially unpopulated, is equivalent to $|U_{rr}| = 1$. This leads to the condition

$$t = \frac{2m\pi}{\sqrt{V^2 + 4\Omega^2}}, \ m \in \mathbb{N}. \tag{D6}$$

Specifically, in the absence of interaction ($V = 0$) and for m odd we get the expression Eq. (20)

$$U_{\ell,\ell'}^{3-\text{level}} = -\begin{pmatrix} \cos\frac{\theta}{2} & e^{i\varphi}\sin\frac{\theta}{2} & 0 \\ e^{-i\varphi}\sin\frac{\theta}{2} & -\cos\frac{\theta}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with θ and φ defined in Eqs. (21).

The unitary Eq. (19) is obtained in an analogous way when considering a two-level system with levels $|\ell\rangle, |\ell'\rangle$ coupled by a laser with Rabi frequency $\Omega_{\ell,\ell'}$ and described by the Hamiltonian

$$H = \Omega_{\ell,\ell'}|\ell\rangle\langle\ell'| + \text{H.c.} + \Delta|\ell'\rangle\langle\ell'|, \tag{D7}$$

where $U = e^{-itH}$ is again evaluated on resonance $\Delta = 0$.

2. Derivation of the qudit controlled-phase gates

In this section we describe a systematic construction of the qudit controlled-phase gates Eqs. (26) and (27).

a. On qudit controlled-phase gate Eq. (26)

Let us start with the gate (26) acting on a qubit, $d = 2$. First we note that the action of the unitary $\mathcal{U}_0^{(2|1)}$ corresponds to

$$\mathcal{U}_0^{(2|1)}|\ell\rangle|\ell'\rangle = \begin{cases} e^{i\frac{\gamma}{2}}|\ell\rangle|\ell'\rangle & \text{if } (\ell,\ell') = (1,0) \\ |\ell\rangle|\ell'\rangle & \text{otherwise.} \end{cases} \tag{D8}$$

It then follows that for a qubit, the controlled-phase gate expressed in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ reads

$$\text{CP}(\frac{\gamma}{2}) = \mathcal{U}_0^{(2|1)}\mathcal{U}_1^{(2|1)} = \text{diag}(1, e^{i\gamma/2}, e^{i\gamma/2}, 1) \tag{D9}$$

(in this case, an alternative sequence $\mathcal{U}_0^{(1|2)}\mathcal{U}_0^{(2|1)}$ [or $\mathcal{U}_0^{(2|1)}\mathcal{U}_0^{(1|2)}$] yields the same result). Furthermore, the cost of the controlled-phase gate (D9) is $[\text{CP}] = 2 = d$.

Motivated by the construction in Eq. (D9), we next wish to extend it to qutrits. Here the situation becomes slightly more involved. Observing that the gate $\mathcal{U}_\ell^{(q_t|q_c)}$ is always diagonal, i.e. introducing at most a phase for each element $|\ell\rangle|\ell'\rangle$ of the two-qutrit system, we introduce the following graphical notation for a representation of a general *diagonal* gate acting in the two-qutrit space

$$\mathcal{U} \rightarrow \boxed{\begin{array}{c} \boxed{22} \\ 02 \ 20 \qquad 21 \ 12 \\ \boxed{00} \ \begin{array}{c} 10 \\ 01 \end{array} \ \boxed{11} \end{array}} \quad , \tag{D10}$$

where the labels on the right-hand side denote the coordinates of the diagonal element $\langle \ell\ell' | \mathcal{U} | \ell\ell' \rangle$ of $\mathcal{U}$. Since $\mathcal{U}$ is diagonal with elements, which are either 1 or a pure phase $e^{i\varphi}$, we shall use the notation $\bullet$ for 1 and φ for the pure phase for easier readability. With the help of (D10), we can list the action of the possible six unitaries $\mathcal{U}_\ell^{(q_t | q_c)}$ acting on the two qutrits:

$$
\mathcal{U}_2^{(1|2)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \bullet & \gamma/2 & & \gamma/2 & \bullet \\ & \boxed{\bullet} & \bullet & \boxed{\bullet} & \\ & & \bullet & & \\ \hline \end{array}
\qquad
\mathcal{U}_2^{(2|1)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \gamma/2 & \bullet & & \bullet & \gamma/2 \\ & \boxed{\bullet} & \bullet & \boxed{\bullet} & \\ & & \bullet & & \\ \hline \end{array}
$$

$$
\mathcal{U}_1^{(1|2)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \bullet & \bullet & & \bullet & \gamma/2 \\ & & \gamma/2 & & \\ & \boxed{\bullet} & \bullet & \boxed{\bullet} & \\ \hline \end{array}
\qquad
\mathcal{U}_1^{(2|1)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \bullet & \bullet & & \gamma/2 & \bullet \\ & & \bullet & & \\ & \boxed{\bullet} & \gamma/2 & \boxed{\bullet} & \\ \hline \end{array}
$$

$$
\mathcal{U}_0^{(1|2)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \gamma/2 & \bullet & & \bullet & \bullet \\ & \boxed{\bullet} & \bullet & \boxed{\bullet} & \\ & & \gamma/2 & & \\ \hline \end{array}
\qquad
\mathcal{U}_0^{(2|1)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \bullet & \gamma/2 & & \bullet & \bullet \\ & & \gamma/2 & & \\ & \boxed{\bullet} & \bullet & \boxed{\bullet} & \\ \hline \end{array}
\qquad . \tag{D11}
$$

It then follows that the controlled-phase gate can be obtained by concatenation of all the gates in (D11)

$$
\mathrm{CP}(\gamma) = \mathcal{U}_2^{(1|2)} \mathcal{U}_2^{(2|1)} \mathcal{U}_1^{(1|2)} \mathcal{U}_1^{(2|1)} \mathcal{U}_0^{(1|2)} \mathcal{U}_0^{(2|1)} = \begin{array}{|ccccc|} \hline & & \bullet & & \\ \gamma & \gamma & & \gamma & \gamma \\ & & \gamma & & \\ & \boxed{\bullet} & \gamma & \boxed{\bullet} & \\ \hline \end{array} \quad , \tag{D12}
$$

which corresponds to Eq. (24) up to a global phase. It is then a straightforward exercise to verify that the manifestly symmetric form suggested by (D12) generalizes to higher d yielding the expression (26). In fact, it also applies to a qubit, yielding the cost $[\mathrm{CP}] = 4 = 2d$, which is twice as large as that of (D9). For the ease of exposition and compactness of notation, and motivated by the fact that we are mainly interested in applications with $d > 2$, we have kept the controlled-phase gate (26) also for qubits in the main text.

b. On qudit controlled-phase gate Eq. (27)

To understand the logic behind the construction of (27), it is instructive to work out a specific example and we shall consider the simplest case beyond qubit, the qutrit. To appreciate the role of the qudit CX gates in (27), we focus specifically on the equal-state elements $|\ell\rangle|\ell\rangle$ of the two-qutrit system. The evolution of these elements under the action of the controlled-phase gate (27) can be symbolically written as

$$
|22\rangle \xunderleftrightarrow[\mathrm{CX}_{1,2|\neg 1}^{(1|2)}]{\mathrm{CX}_{1,2|\neg 1}^{(1|2)}} |12\rangle \xunderleftrightarrow[\mathrm{CX}_{0,1|\neg 0}^{(1|2)}]{\mathrm{CX}_{0,1|\neg 0}^{(1|2)}} e^{i\gamma}|02\rangle \hookleftarrow P_0^{(1)}(\gamma)
$$

$$
|11\rangle \xLeftrightarrow{\qquad} |11\rangle \xLeftrightarrow{\qquad} e^{i\gamma}|01\rangle \hookleftarrow P_0^{(1)}(\gamma)
$$

$$
|00\rangle \xLeftrightarrow{\qquad} |00\rangle \xLeftrightarrow{\qquad} e^{i\gamma}|00\rangle \hookleftarrow P_0^{(1)}(\gamma). \tag{D13}
$$

Here, the CX gates (listed only in the first line to avoid cluttering) act in the direction indicated by the arrows. We thus see that the purpose of the CX gates is to bring the equal state elements to the elements of the form $|0\ell\rangle$, $\forall \ell$, i.e. $\mathrm{CX}_{0,1|\neg 0}^{(1|2)} \mathrm{CX}_{1,2|\neg 1}^{(1|2)} |\ell\ell\rangle \to |0\ell\rangle$, $\ell = 0, 1, 2$. Since there are at most three states of the form $|0\ell\rangle$, the states $|02\rangle, |01\rangle, |00\rangle$ on the right of (D13) exhaust all such states. The phase gate $P_0^{(1)}$ then imprints a phase γ to all these states, which is highlighted in red, and acting with the CX gates backwards yields the controlled-phase gate (27). Similarly to the construction in Sec. D 2 a above, it is straightforward to verify that the scheme (D13) holds also for higher d.

Appendix E: Simulating complete graphs using only nearest-neighbour interactions

For first experiments, it might be prudent to focus on problem instances where the problem graph matches the topology of the quantum computer well. Our simulations considered complete graphs, so for those it is natural to ask: How many swap operations are required to let a limited-interaction quantum computer execute the algorithm on a complete graph? Or, if the hardware allows swaps on distinct nearest-neighbour pairs of qudits to be performed in parallel, how many *layers* of swaps?

We can construct a simple sequence of swaps that achieves this task using not too many layers:

Proposition E.1. *There exists a sequence of swaps such that the complete graph can be simulated on a line graph in $n-2$ layers of swaps. The total swap count for this protocol is $\frac{(n-1)(n-2)}{2}$.*

Proof. First assume n is even (the argument follows the same structure if n is odd). The sequence will be generated by alternating the two following sets of swaps: $\pi = \{(1,2),(3,4),\ldots,(n-1,n)\}$, $\sigma = \{(2,3),(4,5),\ldots,(n-2,n-1)\}$. First observe what happens when applying π and σ to a qudit starting in an odd position i. If $i \neq n-1$, this qudit will move to $i+1$ because of application of π, and then to $i+2$ by σ. If $i = n-1$, the qudit will move to position n. Similarly, for a qudit starting in even position j will move to position $j-2$, except that the qudit at position 2 will move to position 1.

It is easily checked that after $n-2$ layers of swaps, all pairs of qudits have been nearest-neighbours at some stage of the process. ∎

For a more thorough analysis of swapping sequences for executing all-to-all interactions, see also Ref. [151] and Ref. [152], where variants of the previous construction are analysed in-depth.

We can easily see that the previous strategy is close to optimal, and that no strategy can exist that is much better: by a simple counting argument the number of layers of swaps required to enable all interactions of a complete graph on a line could not be much lower. This motivates us to look for the corresponding lower bound.

Proposition E.2. *Consider a quantum computer for which the graph of possible interactions is a line of n qudits. Then at least $\frac{1}{2}\binom{n}{2} - o(n)$ SWAP gates are required to enable all-to-all interactions. Additionally, at least $\frac{n}{2} - 1$ layers of SWAP gates are required.*

Proof. We can lower bound the number of necessary swaps by a simple counting argument: The complete graph has $\binom{n}{2}$ edges, so that number of interactions will be required. A line of n qubits has $n-1$ edges, which is how many interactions are possible before the first SWAP gate. Next, every SWAP gate gives two qubits a new neighbour, enabling at most two new interactions. The required swap count therefore is at least $\frac{1}{2}\left[\binom{n}{2} - (n-1)\right]$.

For the number of layers, note again between a layer of SWAP gates at most $n-1$ nearest-neighbour interactions are possible. This implies a lower bound of $\frac{\binom{n}{2}}{n-1} = \frac{n}{2}$ of layers, hence $\frac{n}{2} - 1$ layers of SWAP gates to be able to have all interactions. ∎

Appendix F: On experimental errors

The gate errors for Rydberg atoms have been analysed extensively, e.g. in [153], and verified experimentally, e.g. in [52]. The total error budget consists of various contributions including the collisions of the atoms with background gas, phase and intensity laser noise, light scattering from both the trapping and resonant lasers operating the processor, spurious level shifts from fluctuating electric and magnetic fields or atom motion in the tweezer traps. A detailed analysis of the individual contributions of these error sources goes beyond the scope of the present work. Given that the total 2-gate error dominates over the 1-gate one [153] and is one of the bottlenecks for scalability across different platforms, we consider only the 2-gate error. Specifically, we consider the error due to the spontaneous decay from the Rydberg levels.

Let us start with a brief overview of the situation and let us consider $\mathbb{P}$ as the qudit manifold (we recall we consider the excitation to a Rydberg state as a resonant two step process through the long-lived $\mathbb{P}$ manifold, cf. Sec. VI). For the corresponding Rydberg manifold $|\mathbb{R}y\rangle = |n^3S_1\rangle$, the Rydberg atom decays approximately with branching ratios $\approx$1:4 to the $(np)^3P_J$ and $(5p)^3P_J$ manifolds respectively, where $n > 5$. Out of the decay to $(5p)^3P_J$, it branches further with ratios $\approx$1:3:5 to the $J = 0, 1, 2$ manifolds and then further among the $F-$manifolds and the respective m_F states (allowed by the selection rules). Consequently, the probability of the Rydberg state to decay back to the qudit manifold (here $|^3P_2\rangle$) – but not necessarily to the same qudit state – is ≈ 0.4 [154, 155].
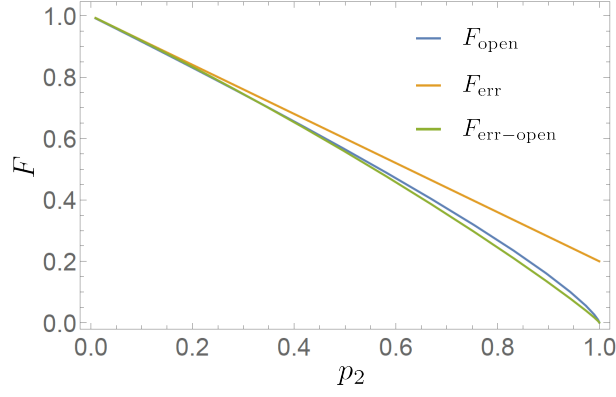
FIG. 14. Fidelities (F4), (F5) and (F6) [solid orange, blue and green lines] as a function of the 2-gate error p_2, cf. Eqs. (44) and (F7), for $|\psi_{\text{in}}\rangle = |+^2\rangle^{\otimes 2}$ and CP gate phase $\gamma = \pi/2$. The approximate agreement between F_{open} and F_{err} for small values of p_2 was obtained by setting $\eta \approx 1.5$ in (F7).

The complete dynamics can be described by the standard means of a master equation for amplitude decay including all the decay channels. However, since the probability to decay outside the qudit manifold $\gtrsim 0.6$, we consider a simplified dynamics of a pair of interacting qudits $q = 1, 2$ given by

$$\dot{\rho} = i[\rho, H] + \frac{\Gamma}{2} \sum_{q=1,2} \sum_{\ell=0}^{d-1} 2\sigma^{(q)-}_{\ell} \rho \sigma^{(q)+}_{\ell} - \{\sigma^{(q)+}_{\ell} \sigma^{(q)-}_{\ell}, \rho\} \tag{F1}$$

where $\sigma^{(q)+}_{\ell} = \left| r^{(q)}_{\ell} \right\rangle \langle \text{aux}|$, $\{\cdot, \cdot\}$ is the anticommutator and a single auxiliary level $|\text{aux}\rangle$ is used to model the decay at rate Γ outside the qudit manifold.

We seek to compare the effect of the realistic errors described by the master equation (F1) to the unitary error model introduced above. Let us denote

$$\rho_{\text{id}} \equiv \text{CP}|\psi_{\text{in}}\rangle\langle\psi_{\text{in}}|\text{CP}^{\dagger} \tag{F2}$$

the ideal state obtained upon an action of the qudit controlled-phase gate (24) on some pure state $|\psi_{\text{in}}\rangle$. We also define the fidelity between two quantum states with density matrices ρ, σ as usual,

$$F(\rho, \sigma) = \text{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]. \tag{F3}$$

Next, we define the average fidelity of a state corresponding to the error model (44) as

$$F_{\text{err}}(p_2) = (1 - p_2) + \frac{p_2}{|\mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}|} \sum_{U \in \mathbb{U}^{\otimes 2}\setminus\{\mathbb{1}\}} F\left(U\rho_{\text{id}}U^{\dagger}, \rho_{\text{id}}\right). \tag{F4}$$

Similarly, denoting the density matrix resulting from the open dynamics given by the master Eq. (F1) as ρ_{open}, we define the fidelity

$$F_{\text{open}} = F(\rho_{\text{open}}, \rho_{\text{id}}) \tag{F5}$$

and the average fidelity between ρ_{open} and the states generated by (44)

$$F_{\text{err}-\text{open}} = (1 - p_2)F(\rho_{\text{id}}, \rho_{\text{open}}) + \frac{p_2}{|\mathbb{U}^{\otimes 2} \setminus \{\mathbb{1}\}|} \sum_{U \in \mathbb{U}^{\otimes 2}\setminus\{\mathbb{1}\}} F\left(U\rho_{\text{id}}U^{\dagger}, \rho_{\text{open}}\right). \tag{F6}$$

The resulting fidelities depend on $|\psi_{\text{in}}\rangle$ and they in principle vary in the course of the algorithm. Furthermore, there is no direct unambiguous identification between the the error probability p_2 of the error model (44) and that of the open dynamics (F1). Denoting the gate time of the cost unitary (27) as t_{CP}, the decay probability of an excited atom $\propto 1 - e^{-\Gamma t_{\text{CP}}}$. This motivates a parametrization

$$p_2 = 1 - e^{-\eta\Gamma t_{\text{CP}}}, \tag{F7}$$

where we have introduced a phenomenological factor η.

For illustration, in Fig. 14 we show the fidelities (F4), (F5) and (F6) as a function of the error probability p_2, cf. Eq. (44), for the "isotropic" initial state $|\psi_{\rm in}\rangle = |+^2\rangle^{\otimes 2}$, $|+^d\rangle = 1/\sqrt{d}\sum_{\ell=0}^{d-1}|\ell\rangle$.

Denoting $\mathcal{P}_\mathbb{Q}$ the projector on the qudit subspace, $\mathbb{Q} = \mathbb{P}$, we have also verified that the state evolution subject to the open dynamics (F1) satisfies to a good accuracy $\mathcal{P}_\mathbb{Q}\rho_{\rm open}\mathcal{P}_\mathbb{Q} \approx {\rm e}^{-\Gamma t_{\rm CP}}\rho_{\rm id}$. This corresponds well to the naive guess that the resulting density matrix is just the rescaled ideal matrix $\rho_{\rm id}$ due to the decay with rate Γ outside the qudit subspace.

In summary, while the error model (44) seems to capture qualitatively the decrease of the state fidelity, it clearly cannot account for the dynamics outside the qudit subspace and more rigorous analysis of the errors in the context of the QAOA is desirable, see also [156] for related developments.