

Improving the Performance of Deep Quantum Optimization Algorithms with Continuous Gate Sets


Nathan Lacroix^{1,*}, Christoph Hellings¹, Christian Kraglund Andersen¹, Agustin Di Paolo,²
Ants Remm¹, Stefania Lazar,¹ Sebastian Krinner¹, Graham J. Norris,¹ Mihai Gabureac,¹
Johannes Heinsoo¹, Alexandre Blais,^{2,3} Christopher Eichler,¹ and Andreas Wallraff^{1,4}

¹*Department of Physics, ETH Zurich, Zurich CH-8093, Switzerland*

²*Institut Quantique and Département de Physique, Université de Sherbrooke, Sherbrooke, Québec J1K2R1, Canada*

³*Canadian Institute for Advanced Research, Toronto, Ontario, Canada*

⁴*Quantum Center, ETH Zurich, Zurich 8093, Switzerland*

 (Received 20 May 2020; accepted 23 September 2020; published 20 October 2020)

Variational quantum algorithms are believed to be promising for solving computationally hard problems on noisy intermediate-scale quantum (NISQ) systems. Gaining computational power from these algorithms critically relies on the mitigation of errors during their execution, which for coherence-limited operations is achievable by reducing the gate count. Here, we demonstrate an improvement of up to a factor of 3 in algorithmic performance for the quantum approximate optimization algorithm (QAOA) as measured by the success probability, by implementing a continuous hardware-efficient gate set using superconducting quantum circuits. This gate set allows us to perform the phase separation step in QAOA with a single physical gate for each pair of qubits instead of decomposing it into two CZ gates and single-qubit gates. With this reduced number of physical gates, which scales with the number of layers employed in the algorithm, we experimentally investigate the circuit-depth-dependent performance of QAOA applied to exact-cover problem instances mapped onto three and seven qubits, using up to a total of 399 operations and up to nine layers. Our results demonstrate that the use of continuous gate sets may be a key component in extending the impact of near-term quantum computers.

DOI: [10.1103/PRXQuantum.1.020304](https://doi.org/10.1103/PRXQuantum.1.020304)

I. INTRODUCTION

Quantum computers have the potential to outperform classical computers on a range of computational problems such as prime factoring [1] and quantum chemistry [2]. Although many of these applications will require quantum error correction [3] to provide a quantum advantage, there is an increasing interest in exploring quantum applications on noisy intermediate-scale quantum (NISQ) devices [4] available in the near term. Recent experiments have demonstrated a computational advantage of quantum computers [5], explored many-body physics [6,7], and simulated small-scale quantum chemistry problems [8–10]. Moreover, there is a significant interest in solving

optimization problems on quantum computers, in particular with the quantum approximate optimization algorithm (QAOA) [11–13]. This variational algorithm has been used to study a range of discrete [11,14–16] and continuous [17] optimization problems, and may have applications for unstructured search [18]. While there is currently no proof that it can provide an asymptotic quantum advantage, QAOA is an emerging approach for benchmarking quantum devices and is a candidate for demonstrating a practical quantum speed up on near-term NISQ devices.

To find an approximate solution to a combinatorial problem with QAOA, a problem Hamiltonian is formulated, whose ground state corresponds to the solution of the combinatorial problem. To approximate this ground state, a quantum computer prepares an ansatz state with a parameterized gate sequence, whose parameters are iteratively updated by a classical optimizer. The gate sequence consists of layers, each characterized by two variational parameters, γ_q and β_q , see Fig. 1. The number of layers, p , sets the depth of the algorithm and QAOA can reach the global optimum of any cost function for $p \rightarrow \infty$ [11]. It is therefore expected that the computational power of

*lna@ethz.ch

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

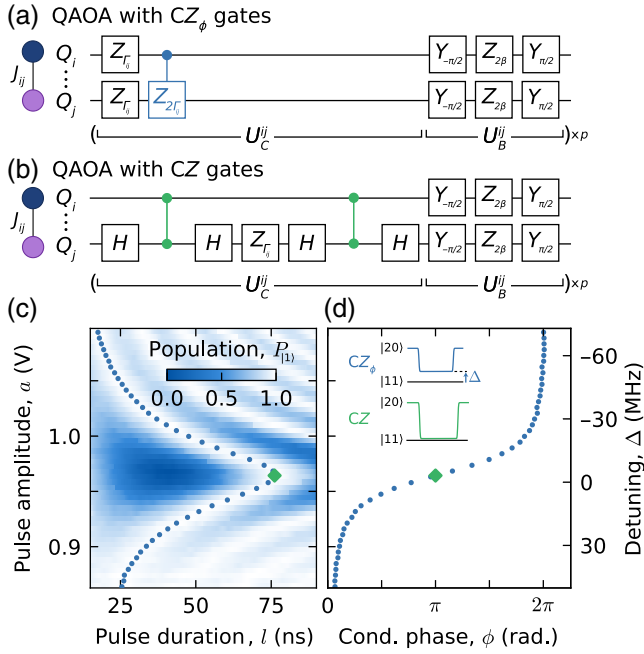


FIG. 1. (a) Quantum circuit of a layer q of QAOA for the two-qubit subspace $|Q_i Q_j\rangle$, using the controlled arbitrary-phase gate (blue) to rotate the $|11\rangle$ state by an angle $2\Gamma_{ij}$ where $\Gamma_{ij} = 2\gamma_q J_{ij}$. (b) A QAOA layer with the phase-separation unitary U_C^{ij} decomposed into CZ gates (green) and additional Hadamard gates and single-qubit Z -gates. (c) Excited-state population P_{11} of the higher-frequency qubit $Q_i = A_1$ brought in interaction with the lower-frequency qubit $Q_j = B_2$ via a flux pulse (all qubits are operated at their maximal frequency). We perform a two-dimensional sweep of flux-pulse amplitude a and flux-pulse duration l , and indicate the maximum population recovery with blue dots. P_{11} is extracted using a three-level readout (see Appendix A). (d) Conditional phase for the dots indicated in (c). The green diamond corresponds to the CZ gate. The right axis indicates the detuning between $|11\rangle$ and $|20\rangle$. The inset depicts the shift in frequency of the bare $|20\rangle$ state relative to the bare $|11\rangle$ state during the flux pulse for the CZ_ϕ gate and the CZ gate, see Appendix B for more details.

QAOA increases with p . In practice, however, the number of layers that can be executed reliably on near-term quantum computers is limited due to finite gate errors induced by relaxation, dephasing, and pulse imperfections [14,19].

Small-scale implementations of QAOA, while restricted to solving problems that can also be efficiently solved on classical computers, provide crucial insights into the feasibility and challenges related to the execution of the algorithm on NISQ devices. Previous studies of QAOA with superconducting qubits [12,13,19–21], photonics [22], and trapped ions [23] highlight the applicability of QAOA on a range of platforms and illustrate the breadth of problems that can be addressed with QAOA. The work presented in Ref. [12] studied the MaxCut problem, which is the canonical problem for QAOA [11], with up to

19 qubits, Ref. [20] studied a channel decoding problem, Ref. [23] searched the lowest-energy eigenstate of all-to-all connected Ising models with up to 40 qubits, and Ref. [21] considered an exact-cover problem with two qubits. Many of these experiments consider problems that can be solved with shallow QAOA circuits ($p = 1$ or 2). However, these examples may not be representative of the broad range of problems that can be addressed with QAOA. Indeed, studies of all-to-all connected Ising models show that deep circuits may be needed [13].

When implementing quantum algorithms on a quantum device, it is common to decompose the gate sequence into a discrete set of gates available on the hardware. To improve performance, recent experiments have explored continuous gate sets motivated by applications in quantum simulations [24,25], quantum chemistry [26,27], and for QAOA using XY interactions [28]. It was also recently suggested that QAOA can be implemented using always-on ZZ interactions [29].

In this work, we benchmark QAOA with a continuous hardware-efficient gate set. We present a controlled arbitrary-phase gate (CZ_ϕ gate), which allows the execution of each QAOA layer with only one two-qubit gate per ZZ term in problem Hamiltonians formulated as Ising models, see Fig. 1(a). We demonstrate how our gate set shortens the QAOA sequence and, thus, leads to better performance for a fixed QAOA depth compared with a decomposed implementation of the algorithm with a discrete gate set. In particular, we demonstrate with two concrete examples that the reduction in gate-sequence duration outweighs errors that are potentially introduced by implementing the continuous gate set, such as errors originating from the required interpolation of parameters. Taking advantage of this gain in performance, we investigate the tradeoff between experimental noise, which favors shallow circuits, and increasing the number of layers, which is needed to solve complex problem instances.

II. IMPLEMENTATION

The objective function of many NP-complete discrete optimization problems can be mapped to an Ising Hamiltonian [30,31],

$$\hat{C} = \sum_{i<j} J_{ij} Z_i Z_j + \sum_{i=1}^n h_i Z_i, \quad (1)$$

where Z_i is the Pauli- Z operator for spin i . QAOA can find the ground state of this Hamiltonian by minimizing the expectation value of \hat{C} for the ansatz state $|\vec{\gamma}, \vec{\beta}\rangle$, where $\vec{\gamma} = (\gamma_1, \dots, \gamma_p)$, $\vec{\beta} = (\beta_1, \dots, \beta_p)$ are variational parameters. In particular, the quantum circuit preparing $|\vec{\gamma}, \vec{\beta}\rangle$ consists of p layers each containing a phase-separation operator $U_C = e^{-i\gamma_q \hat{C}}$ and a mixing operator

$U_B = e^{-i\beta_q \hat{B}}$, where $\hat{B} = \sum_i X_i$, with $q = 1, \dots, p$ [11]. Since all terms of \hat{C} commute, we can implement each term $U_C^{ij} = e^{-i(\Gamma_{ij}/2)Z_i Z_j}$ separately, where $\Gamma_{ij}/2 = \gamma_q J_{ij}$ is a continuous parameter.

A common approach is to decompose U_C^{ij} into a gate sequence consisting of two conditional phase rotations of π , i.e., standard CZ gates, combined with several single-qubit gates [12,21]. We present such a decomposition in Fig. 1(b), where the dependence on the continuous parameters Γ_{ij} is introduced via an arbitrary-angle single-qubit Z-rotation. An alternative approach is to use a single controlled arbitrary-phase gate (CZ $_\phi$ gate), which can add any desired phase factor $e^{-i\phi}$ to the $|11\rangle$ state. This gate naturally applies the angle $\phi = 2\Gamma_{ij}$ and, together with two single-qubit Z-rotations of angle Γ_{ij} , realizes the unitary U_C^{ij} , see Fig. 1(a).

In QAOA, the number of unitaries U_C^{ij} grows linearly with the number of two-qubit terms in \hat{C} and with the number of QAOA layers, p . Thus, it is essential that each U_C^{ij} is implemented with high fidelity. The direct implementation we present in this work significantly reduces both the physical gate count and the sequence duration. Thus, this approach is expected to find correct solutions to complex problems with higher probability.

We run QAOA on a quantum device with seven superconducting transmon qubits, see Appendix A for device parameters and a false-colored micrograph of the device. The qubits are pairwise connected as illustrated in Fig. 2(a). Single-qubit X and Y rotations are implemented

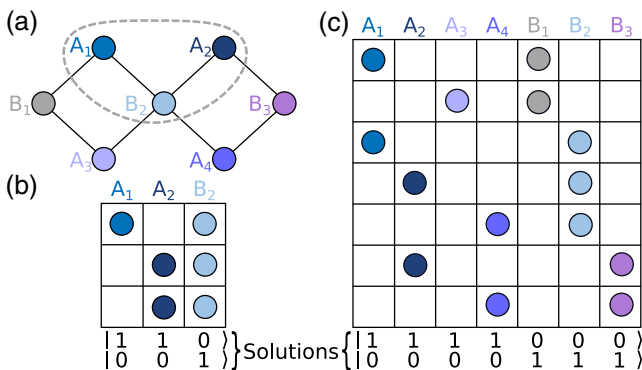


FIG. 2. (a) Hardware connectivity graph of the quantum device. Dots correspond to qubits and edges indicate between which pairs of qubits two-qubit gates can be realized. The gray dashed line indicates the subset of qubits used for the three-qubit problem instance depicted in (b). (b) Visual representation of the incidence matrix K (dots indicating entries $K_{\ell i} = 1$) for a chosen three-qubit exact-cover problem instance. The labels above the columns (and the colors) indicate which physical qubits are used to represent the corresponding subset. The two solution states are indicated below the grid. (c) Visual representation of the incidence matrix for a chosen seven-qubit problem instance.

with microwave pulses, while Z rotations are performed as virtual gates [32], which take zero time as they are implemented through a redefinition of the reference frame. To realize CZ gates, we use a standard approach relying on a flux pulse, which shifts the transition frequency of one of the qubits to bring the $|11\rangle$ state of a pair of coupled qubits in resonance with the noncomputational $|20\rangle$ state [33–35]. The resulting hybridization leads to a coherent population oscillation between the two states. The frequency detuning between the $|11\rangle$ and $|20\rangle$ states, Δ , is 0 during the gate, and after an interaction of the duration corresponding to one oscillation period, the population returns to the $|11\rangle$ state with an added phase of π , see green diamond in Figs. 1(c) and 1(d). We generalize the CZ gate to a CZ $_\phi$ gate on our device by exploiting near-resonant interactions of the $|11\rangle$ and $|20\rangle$ states [24], i.e., $\Delta \neq 0$, to acquire conditional phase angles ranging from 0 to 2π , see Fig. 1(d). We vary Δ by sweeping the flux-pulse amplitude and simultaneously adapting the pulse duration to maximize population recovery in the computational subspace, see blue dots in Fig. 1(c). Details about the gate implementation are provided in Appendix B.

We compare the performance of both approaches on two example instances of the NP-complete exact-cover problem [30]. The aim of exact cover is to decide whether it is possible to cover all elements in a set S *exactly once* by an appropriate selection of subsets $\{V_i\}$ from a given collection of subsets V . In the example visualized in Fig. 2(b), each row corresponds to an element of a three-element set S , while each column corresponds to a subset V_i out of three given subsets, see Appendix C for details. The dots visualize which elements (rows) are included in a subset (column). In this picture, the task is to find a selection of columns such that each row is covered by exactly one dot. This condition is fulfilled by two solutions: selecting the first two columns or selecting the last column. In a mathematical formulation of the exact-cover problem (see Appendix C), the grid in Fig. 2(b) corresponds to a visual representation of an incidence matrix K , where a dot in row ℓ and column i indicates an entry $K_{\ell i} = 1$ while empty cells in the grid indicate entries equal to 0. When mapping an instance of an exact-cover problem to an Ising Hamiltonian [31,36], the two-qubit coupling terms J_{ij} and single-qubit terms h_i are extracted from this incidence matrix, see Eq. (C3) and Eq. (C4), respectively. Furthermore, the i th qubit encodes whether a subset V_i is selected or not, see Appendix C. In the visualization in Fig. 2(b), the qubit that represents a subset is indicated by the label above the column and by the color used for the dots. Figure 2(c) shows an example of a larger instance of exact cover with seven subsets, requiring seven qubits.

To focus on the comparison between the two methods for realizing the two-qubit unitaries U_C^{ij} , these two problem instances are chosen such that the resulting Ising Hamiltonians respect the hardware connectivity graph of

our device, see Fig. 2(a), and that all single-qubit terms vanish, i.e., $h_i = 0$. The three-qubit problem instance depicted in Fig. 2(b) yields an Ising Hamiltonian with $J_{A_1B_2} = 0.5$ and $J_{A_2B_2} = 1$. In the basis $|A_1A_2B_2\rangle$, the two possible selections of columns covering all rows, namely $\mathcal{A} = \{A_1, A_2\}$ and $\mathcal{B} = \{B_2\}$, are encoded with the states $|110\rangle$ and $|001\rangle$, respectively, where a 1 in position i indicates that the i th column of K is included in the selection of subsets. For the seven-qubit problem instance of Fig. 2(c), we have $J_{A_3B_2} = 0$ and $J_{ij} = 0.5$ for all other physically connected qubit pairs. This instance also possesses two solutions, $\mathcal{A} = \{A_1, A_2, A_3, A_4\}$ and $\mathcal{B} = \{B_1, B_2, B_3\}$, corresponding to the states $|1111000\rangle$ and $|0000111\rangle$, respectively, using the basis $|A_1A_2A_3A_4B_1B_2B_3\rangle$. Note that we have labeled the qubits in Fig. 2(a) such that the solutions always correspond to either selecting the qubits labeled with A or the qubits labeled with B , see Appendix C.

Using the direct implementation instead of the decomposed one reduces the number of operations per layer from 42 to 15 for the three-qubit problem instance and from 98 to 42 for the seven-qubit problem instance. Consequently, the seven-qubit problem instance, which requires deep QAOA circuits, yields a circuit comprising 175 operations at $p = 4$ with the direct implementation compared with 399 operations with the decomposed implementation, see Appendix D for details.

III. PERFORMANCE OF QAOA

A single-layer QAOA implementation ($p = 1$) is a useful intermediate benchmark towards implementing multi-layer QAOA circuits since there are only two variational parameters $\vec{\gamma} = (\gamma_1)$ and $\vec{\beta} = (\beta_1)$, hereafter referred to as γ and β for ease of notation, which allows us to map out the full optimization landscape experimentally. As further discussed in Appendix E, when $p = 1$, the cost-function landscape is $\pi/2$ periodic in β for a problem without single-qubit terms. Moreover, since all eigenvalues of \hat{C} are odd multiples of $\frac{1}{2}$, the landscape is 2π periodic in γ . Finally, the landscape is always point-symmetric around the center point of a period. We can thus reduce our considerations to $\gamma \in [0, \pi)$ and $\beta \in [0, \pi/2)$. For each pair of parameters, we prepare the state $|\gamma, \beta\rangle$ 10 000 times, see Appendix D for the full pulse sequence, and we evaluate the cost function $C(\gamma, \beta) = \langle \gamma, \beta | \hat{C} | \gamma, \beta \rangle$. We use a three-level readout scheme discussed in Appendix A, which allows us to discard the measurement outcomes with leakage outside of the computational space, see Appendix F. In the context of QAOA, discarding leakage events corresponds to reducing the effective number of shots available for evaluating the cost function by rejecting outcomes that are not valid bit-strings. In this regard, leakage is different from other undetectable errors, for which such a postselection cannot be done.

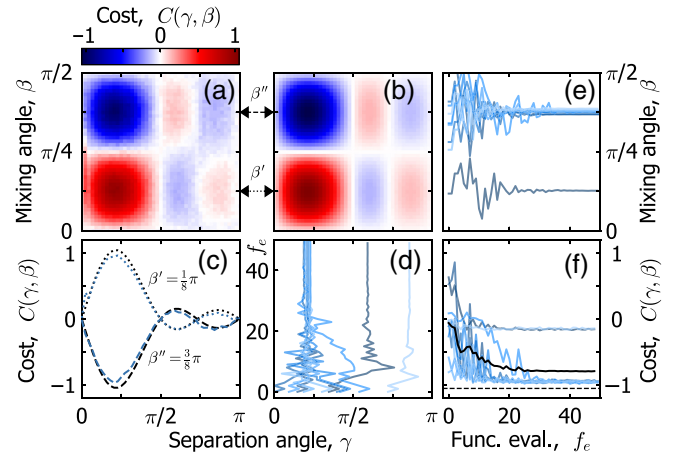


FIG. 3. Cost function evaluated for $p = 1$ on the three-qubit problem instance using CZ_ϕ gates. (a) Cost-function landscape as a function of variational parameters. (b) Cost-function landscape obtained from noise-free simulations. (c) Experimental evaluation (blue) and simulation (black) of the cost function for two horizontal line cuts of (a),(b), with $\beta' = \pi/8$ (dotted lines) and $\beta' = 3\pi/8$ (dashed lines), respectively. (d),(e) Ten convergence traces of the separation angle and the mixing angle, respectively, for end-to-end optimization starting from random parameter initialization. (f) Average energy (solid black line) and individual convergence traces (faded blue lines) of the energy corresponding to parameters shown in (d),(e).

We observe that the resulting cost-function landscapes, see Fig. 3, are odd functions of β with a line symmetry axis at $\beta = \pi/4$, see Appendix E. The locations of all extrema in the measured landscape, see Fig. 3(a), are in good agreement with noise-free simulations, see Fig. 3(b), which suggests that the coherent errors are small in our implementation. Errors due to decoherence mostly affect the contrast of the landscapes [37], see Fig. 3(c) and Appendix G. The distortions of the local extrema located at $\gamma > \pi/2$ are attributed to the residual ZZ coupling between the qubits [38], which we confirm with master-equation simulations, see Appendix G.

By embedding the evaluation of $C(\gamma, \beta)$ measured on the quantum device into a classical optimizer, we demonstrate that the landscape is suitable as a cost function for closed-loop optimization. The simple, gradient-free Nelder-Mead optimization method finds the optimal parameters for most random initialization parameters, see Figs. 3(d)–3(f), however, some convergence traces get trapped in local minima. Note that in this single-layer implementation, the cost never reaches the ground-state energy $C_{GS} = -1.5$, neither in the measurement nor in the noise-free simulation, which indicates that QAOA circuits of larger depth are indeed required for this problem.

To obtain better approximate solutions to the combinatorial problem instance, we execute QAOA circuits with additional layers and study the effect of the depth p on the

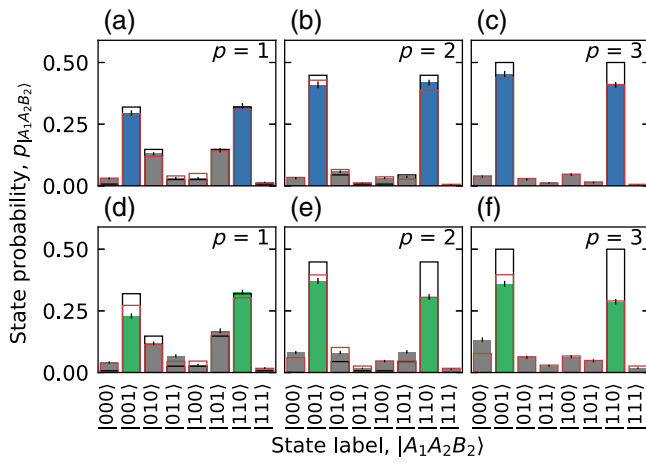


FIG. 4. Output state probability distribution for the three-qubit problem instance implemented with controlled arbitrary phase gates (a)–(c), and decomposed using CZ gates (d)–(f). States are measured at optimal parameters for depths of $p = 1$ (a),(d), $p = 2$ (b),(e), and $p = 3$ (c),(f). The filled bars correspond to the measured state probabilities in which we highlight the problem solutions in blue (direct implementation) and green (decomposed implementation), respectively. The black wireframes are the expected QAOA outcome from noise-free simulations and the red wireframes are from master-equation simulations. We use 15000 individual measurement runs to estimate each probability distribution. The black vertical markers indicate a bootstrap 99.7% confidence interval [39] for each state, estimated from the bootstrap standard deviation of 100 resampled datasets of the original single-shot measurements.

output state distribution. To investigate the performance of the quantum part of QAOA rather than the performance of the classical optimizer, we initialize the algorithm with optimal parameters obtained from noise-free simulations. We then optimize these parameters locally to correct for small coherent errors, and we estimate the resulting state distribution as a function of depth from 15 000 single-shot measurements, see Fig. 4 for the three-qubit case. Three layers are required in noise-free simulations (black wireframes) to fully concentrate the probability distribution on the two solution states $|110\rangle$ and $|001\rangle$ corresponding to the selection of subsets \mathcal{A} and \mathcal{B} , respectively.

We quantify the experimental outcomes using the classical fidelity [40] between the output state probability distribution arising from the measurements, P , and from noise-free simulations, \tilde{P} ,

$$\mathcal{F}(P, \tilde{P}) = \left(\sum_i \sqrt{P_i} \sqrt{\tilde{P}_i} \right)^2, \quad (2)$$

where P_i and \tilde{P}_i correspond to the probabilities of the i th basis state in the Hilbert space. Note that $0 \leq \mathcal{F}(P, \tilde{P}) \leq 1$ with $\mathcal{F}(P, \tilde{P}) = 1$ if and only if $P = \tilde{P}$. The state distributions of the implementation using CZ_ϕ gates, see filled bars

in Figs. 4(a)–4(c), have fidelities of 98.93 %, 95.93 %, and 86.20 % for $p = 1, 2$, and 3 respectively, with respect to the corresponding distribution obtained with noise-free simulations. As expected, the reduction of the fidelity with the number of layers p illustrates the accumulation of errors in circuits of increasing depth. However, the concentration of probability on solution states as p increases is stronger than the detrimental effect of the additional errors, such that overall, the probability of measuring a solution increases with p . By contrast, in the implementation using CZ gates, see Figs. 4(d)–4(f), the concentration of probability on solution states only compensates the additional errors for $p = 2$ while the errors outweigh the gain of an additional layer for $p = 3$. This is also reflected by lower fidelities of 96.37 %, 87.43 %, and 64.52 % for $p = 1, 2$, and 3 respectively, and is explained by the fact that decoherence and residual ZZ coupling accumulate over the longer gate sequence.

Master-equation simulations (red wireframes) yield fidelities of 98.3%, 94.9%, and 85.0% for the direct implementation and 97.5%, 88.6%, and 68.2% for the decomposed implementation, which is in excellent agreement with the measured distributions, see also Appendix G for details. We confirm from these simulations that decoherence is the main limitation in this experiment while residual ZZ coupling causes additional errors, in particular for the decomposed implementation.

The landscape and state probability distributions of the seven-qubit problem instance presented in Appendix H lead to a similar conclusion, i.e., the direct implementation exploiting CZ_ϕ gates is in better agreement with noise-free simulations than the decomposed implementation.

It is expected that additional layers in QAOA circuits increase the number of reachable states, thereby leading to better approximate solutions in the absence of experimental noise. To gain further insight into the tradeoff between the extended reachable state space and the additional noise resulting from increased depth, we determine the enhancement of the success probability provided by the output state distribution over a uniform state distribution as a function of p for both problem instances, see Fig. 5. We define the enhancement as P_s/P_u , where P_s is the success probability, i.e., the sum of the probabilities of all solution states, and $P_u = 2/(2^N - 1)$ is the probability of sampling a solution from a uniform probability distribution over all possible states. Note that we exclude the state $|0\rangle^{\otimes N}$, which is never a solution in the context of exact cover. We indicate the sequence duration for both implementations with additional axes in Fig. 5, where sequence duration is defined as the time between the start of the initialization pulse and the start of the readout.

In the three-qubit case, Fig. 5(a), the direct (decomposed) implementation shows a maximal enhancement of success probability of 3 (2.4) at $p = 3$ ($p = 2$). The direct implementation (blue dots) provides a higher enhancement

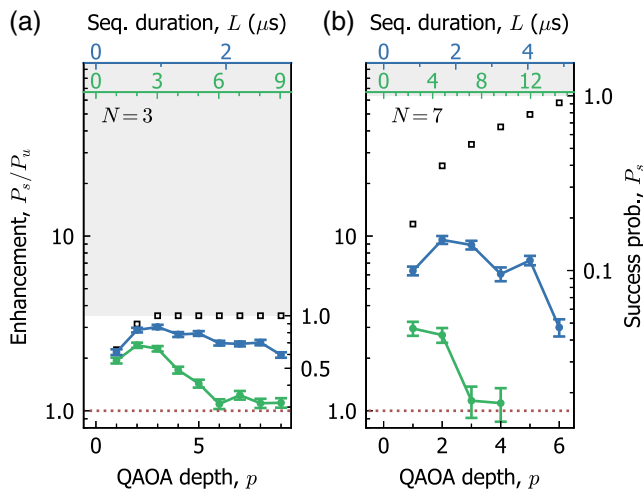


FIG. 5. Performance of QAOA for (a) three qubits and (b) seven qubits. The blue points are implemented with the direct controlled arbitrary-phase gate and the green points are implemented with the gate sequences decomposed into CZ gates. The black squares indicate the highest success probabilities found with noise-free simulations. The top axes indicate the sequence duration for the direct implementation and the decomposed implementation in green and blue, respectively. The gray areas indicate success probabilities above 1. The success probabilities are estimated from 15 000 (three-qubit problem instance) and 60 000 (seven-qubit problem instance) single-shot measurements. The error bars indicate a bootstrap 99.7% confidence interval.

than the decomposed implementation (green dots) because the sequence duration is shorter for a fixed p , and the problem instance requires at least $p = 3$ to reach maximal enhancement in an ideal setting (black squares).

When the problem increases in complexity, the number of layers required to reach maximal enhancement in a noise-free scenario also increases. For the seven-qubit instance, we find that $p = 6$ is required to reach a success probability above 90%, see Fig. 5(b). Consequently, the ability to execute more layers in shorter time provides an even more pronounced advantage. In particular, for the direct implementation, we find that increasing p from 1 to 2 increases the enhancement of the success probability to 9.5. However, when further increasing to $p = 3$, the extended reachable state space does not compensate the additional noise arising from the increased sequence duration. For the decomposed version, going beyond $p = 1$ does not provide any benefits. Thus, for the seven-qubit problem we only benefit from adding layers when taking advantage of the directly implemented CZ_ϕ gates, which improves the performance by a factor of 3 compared with the decomposed implementation.

Finally, to emphasize that the limitations for deeper circuits are directly related to the increased sequence duration rather than the depth itself, we notice that for a fixed

sequence duration of $L = 5 \mu s$, both implementations of the seven-qubit instance show similar enhancement of success probability despite being of depth $p = 6$ (direct) and $p = 2$ (decomposed).

IV. DISCUSSION

In this work, we show that controlled arbitrary phase gates (CZ_ϕ gates) enable a significant reduction of the number of physical gates required to implement QAOA circuits of any depth on quantum hardware. We demonstrate the advantage of this approach by comparing it with a standard QAOA decomposition on two problem instances of the exact-cover problem, with three and seven qubits, respectively. Despite a more demanding calibration scheme requiring interpolation of gate parameters, CZ_ϕ gates in QAOA circuits systematically outperform the decomposed alternative for a fixed depth and are able to benefit from the extended reachable state space of more layers.

We foresee an even more pronounced advantage for larger-scale combinatorial optimization problems because the number of layers required to solve problems with QAOA is expected to scale with the number of qubits involved in the experiment [41,42], in particular for dense problem graphs. In addition, the number of physical two-qubit gates saved within each layer also scales with the number of two-qubit terms in the cost Hamiltonian. Our results demonstrate that hardware-efficient gate sets are key components in extending the impact of near-term quantum applications, which may become even more relevant when solving problem instances that do not match the connectivity of the hardware. For example, it has recently been observed that the need for swap gates can significantly reduce the performance of a QAOA implementation if a decomposed implementation of swap gates is used [13]. A direct, hardware-efficient implementation combining a controlled arbitrary phase gate and a swap gate [27] may therefore be another key component to improve the performance in these cases, and should be considered in future algorithmic implementations on NISQ devices.

ACKNOWLEDGMENTS

The authors are grateful for valuable feedback on the manuscript by B.R. Johnson and D. Schuster and for valuable discussions with A. Choquette-Poitevin, P. Vikstål, and M. Collodo. The authors acknowledge S. Storz, F. Swiadek, and T. Zellweger for their contributions to the measurement setup.

The authors acknowledge financial support by the EU Flagship on Quantum Technology H2020-FETFLAG-2018-03 Project No. 820363 OpenSuperQ, by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via

the U.S. Army Research Office Grant No. W911NF-16-1-0071, by the National Centre of Competence in Research Quantum Science and Technology (NCCR QSIT), a research instrument of the Swiss National Science Foundation (SNSF), by the SNFS R'equip grant 206021-170731 and by ETH Zurich. This work was undertaken thanks in part to funding from NSERC, Canada First Research Excellence Fund and ARO W911NF-18-1-0411. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government.

APPENDIX A: EXPERIMENTAL SETUP AND DEVICE PARAMETERS

The experiments described in this paper are performed in a cryogenic setup [43,44], the wiring scheme of which is summarized in Fig. 6. Each qubit is controlled by a flux line for frequency tuning, which enables two-qubit gates, and a microwave drive line for realizing single-qubit gates. The pulses are generated with arbitrary waveform generators (AWGs). The drive pulses are generated at an intermediate frequency of 100 MHz and up-converted to microwave frequencies. Multiplexed readout is performed

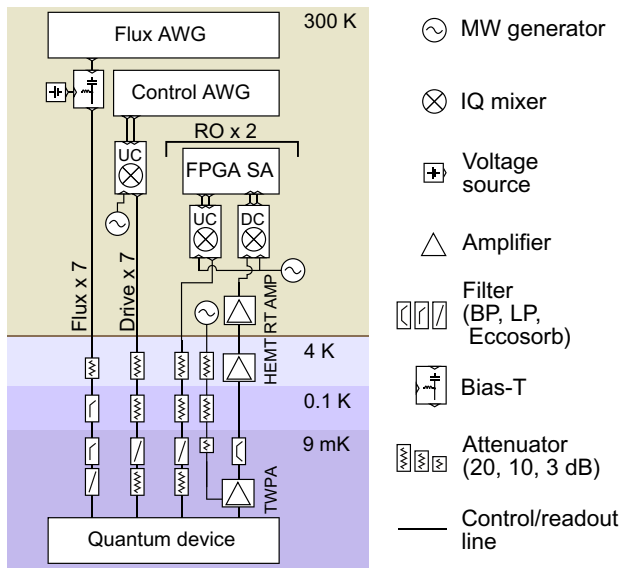


FIG. 6. Experimental setup. The experiment is controlled by AWGs whose signals are routed to the quantum device through a series of bandpass filters (BP), lowpass filters (LP), and Eccosorb filters. The flux pulses are combined with a voltage source using a bias-T. The IQ signal from the control AWG is up-converted (UC) to a microwave signal using an IQ mixer. The readout signal is generated by the FPGA SA, and the output from the quantum device is amplified by a chain of amplifiers before being down-converted (DC) and analyzed by the FPGA SA.

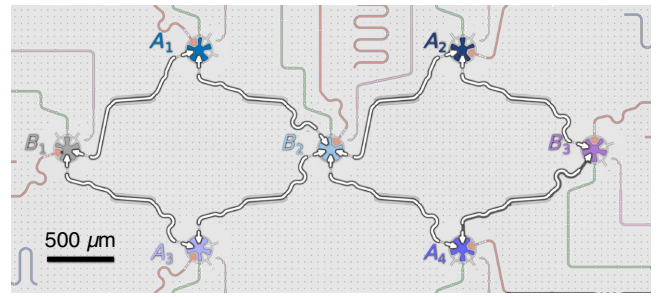


FIG. 7. Optical micrograph of the device. Each qubit (colors corresponding to the colors in Fig. 2) is connected to neighboring qubits by coupling resonators (white). Each qubit is connected to a readout resonator (red), a flux line (green), and a drive line (pink).

via two feedlines [44,45] with the readout pulses generated by an FPGA-based signal analyzer (FPGA SA). The measurement signals at the output ports of the sample are first amplified with a wide-bandwidth near-quantum-limited traveling-wave parametric amplifier (TWPA) [46], then with a HEMT amplifier and finally with low-noise, room-temperature amplifiers (RT AMP) [44]. Thereafter, the signals are down-converted and processed using the weighted integration units of the FPGA SAs.

The quantum device [44] shown in Fig. 7 is fabricated on a high-resistivity intrinsic silicon substrate. Photolithography and reactive ion etching are used to define resonators, signal lines and qubit structures in a 150-nm-thin niobium film sputtered onto the substrate. We also add air bridges to the device to establish a well-defined ground plane and for cross overs in signal lines. The Al/AIOx/Al Josephson junctions of the transmon qubits are fabricated using electron-beam lithography and shadow evaporation.

The parameters of the device listed in Table I are measured using standard spectroscopy and time-domain methods. All qubits are operated at their maximum frequency. We characterize the ability to identify the correct qubit state as well as the second excited state of each transmon qubit. In particular, we use two weighted-integration units per qubit to distinguish $|0\rangle$ from $|1\rangle$ and $|1\rangle$ from $|2\rangle$ [47,48], see Fig. 8 for an example data set for qubit A_1 . The weighted time-trace integration is performed in real time and yields a two-dimensional data point for each single-shot measurement. A trimodal Gaussian mixture model, whose parameters are obtained with maximum-likelihood estimation, is then used to classify the resulting integrated traces, for details see Ref. [47,48]. The correct readout state assignment probabilities reported in Table I are obtained with heralding of the ground state. Thermal populations are estimated by comparing correct state assignment probabilities with and without ground-state heralding.

TABLE I. Measured parameters of the seven qubits.

	A1	A2	A3	A4	B1	B2	B3
Qubit frequency, $\omega_q/2\pi$ (GHz)	5.462	5.684	4.077	4.195	4.825	4.920	5.165
Lifetime, T_1 (μ s)	12.9	6.7	24.5	21.1	16.7	14.8	14.3
Ramsey decay time, T_2^* (μ s)	18.1	12.2	7.4	4.6	27.2	24.5	13.3
Readout frequency, $\omega_r/2\pi$ (GHz)	6.611	6.836	5.832	6.063	6.255	6.042	6.300
Readout linewidth, $\kappa_{\text{eff}}/2\pi$ (MHz)	7.5	10.6	6.0	7.2	17.3	10.9	11.0
Dispersive shift, $\chi/2\pi$ (MHz)	-2.5	-2.5	-0.75	-1.0	-1.25	-2.4	-2.0
Thermal population, P_{th} (%)	0.04	0.01	0.2	0.8	0.3	0.04	0.2
$ 0\rangle$ readout assignment prob. (%)	99.98	99.97	96.54	96.85	99.47	99.92	99.97
$ 1\rangle$ readout assignment prob. (%)	98.06	96.19	90.39	88.94	97.70	94.45	98.20
$ 2\rangle$ readout assignment prob. (%)	96.63	89.68	78.90	80.86	95.18	94.89	96.96

To characterize the gate performance, we perform randomized benchmarking on all qubits to find the error per single-qubit Clifford, see Fig. 9. For the two-qubit gates, we only characterize with a fixed conditional phase of π such that the two-qubit gate is in the Clifford group and we measure the error per gate from interleaved randomized benchmarking, see infidelities next to lines indicating the coupling elements in Fig. 9.

We also characterize the residual ZZ coupling, α_{ij} , between pairs of coupled qubits, as the frequency shift of qubit i when qubit j is in the excited state. We verify that $\alpha_{ij} = \alpha_{ji}$. The experimentally determined residual ZZ couplings are shown below each error rate in Fig. 9.

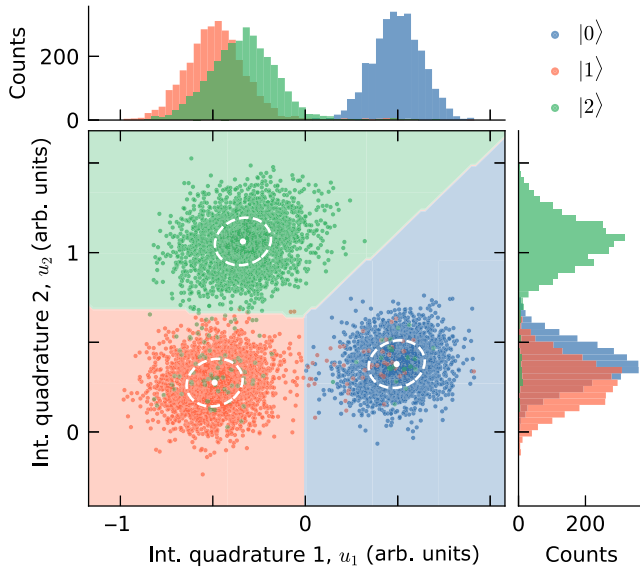


FIG. 8. Three-level single-shot readout characterization of qubit A_1 with ground-state heralding. We display the first 3000 of the 20000 single-shot measurements for each state. The blue (red) (green) shaded area correspond to the region of the integrated quadrature plane in which a measured data point is assigned to the $|0\rangle$ ($|1\rangle$) ($|2\rangle$) state. The white dots and dashed white lines correspond to the mean and the 1σ -confidence ellipse of each mode of the mixture model.

APPENDIX B: CONTROLLED ARBITRARY PHASE GATE

The CZ_ϕ gate is a unitary, which adds a desired phase ϕ to the $|11\rangle$ state in the two-qubit subspace,

$$U(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix}. \quad (\text{B1})$$

We realize this unitary by exploiting a near-resonant interaction of the $|11\rangle$ and $|20\rangle$ states with frequency detuning Δ , see Fig. 10(a) for an energy-level diagram. This exchange-type interaction with coupling rate J leads to a harmonic population oscillation between the two states. For an interaction time τ_g corresponding to one full period of the oscillation, the population of the $|11\rangle$ state cycles through the $|20\rangle$ state back to the $|11\rangle$ state, and additionally accumulates a phase dependent on Δ ,

$$\phi = \pi \left(1 + \frac{\Delta}{\sqrt{4J^2 + \Delta^2}} \right). \quad (\text{B2})$$

We can target any phase in the range $[0, 2\pi)$ by choosing an appropriate detuning. On our device, we control the

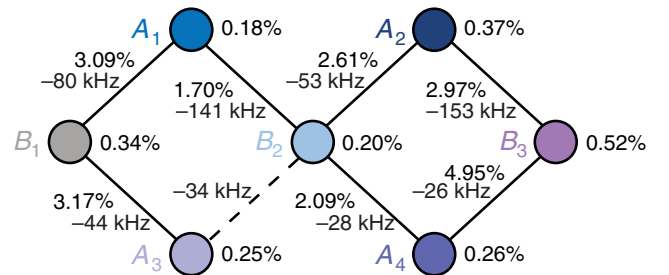


FIG. 9. Single-qubit error per gate (vertices) and two-qubit error per gate (edges) in percent, measured with randomized benchmarking. We indicate the residual ZZ coupling between each pair of coupled qubits below the corresponding two-qubit gate infidelity. The gate between A_3 and B_2 is not needed for the problem instances considered in this work (dashed line).

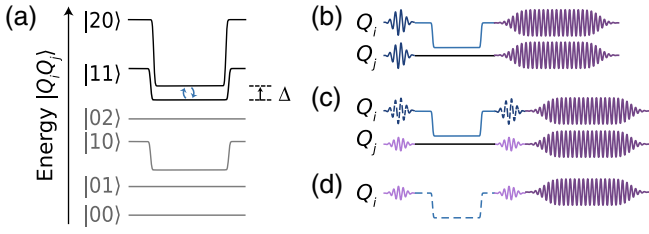


FIG. 10. Energy-level diagram and pulse schemes used for the calibration of the CZ_ϕ gate. (a) Time evolution of the energy levels of the qutrit pair $|Q_i Q_j\rangle$ when a flux pulse is applied to Q_i (not drawn to scale). The energy levels directly involved in the gate are shown in black and the interaction between the $|11\rangle$ and $|20\rangle$ states is indicated with two blue arrows. (b)–(d) Pulse scheme used to measure the Chevron pattern, the conditional phase and the dynamic phase, respectively. The π pulses, $\pi/2$ pulses, flux pulses, and readout pulses are shown in dark blue, pink, light blue, and purple, respectively. A dashed line indicates that the measurement is repeated once with, and once without the corresponding pulse.

detuning and interaction time of the CZ_ϕ gate with a flux pulse of amplitude a and duration l , respectively.

We calibrate the CZ_ϕ gate by adjusting a , l and the corresponding dynamic phases acquired by the flux-biased qubit(s).

First, we prepare a given qubit pair in the $|11\rangle$ state, apply a flux pulse to the higher-frequency qubit, and subsequently measure $P_{|1\rangle}$, the excited-state population of the higher-frequency qubit, see Fig. 10(b). Varying the flux-pulse amplitude and duration yields a characteristic pattern in the population, commonly referred to as a Chevron, owing to its characteristic shape, see Fig. 1(c). For each flux-pulse amplitude a , we fit a cosine to the harmonic oscillation of $P_{|1\rangle}$ as a function of pulse duration to find the pulse duration l corresponding to a full period τ_g in population exchange between the $|11\rangle$ and $|20\rangle$ states. We linearly interpolate between the pairs (a, l) [blue dots in Fig. 1(c)] to define a continuous function $l(a)$.

Then, using the pulse sequence depicted in Fig. 10(c), we excite the higher-frequency qubit to the $|1\rangle$ state with a π pulse, apply a flux pulse, and return the qubit to the ground state with another π pulse. A Ramsey experiment is performed on the lower-frequency qubit by applying $\pi/2$ pulses before and after the flux pulse and sweeping the phase of the second $\pi/2$ pulse. The acquired conditional phase ϕ is determined relative to a measurement without applying π pulses to the higher-frequency qubit. The conditional phase results from the total ZZ interaction, which naturally includes residual ZZ coupling for the duration of the gate [38].

By performing this measurement for 45 flux-pulse amplitudes a near resonance of the $|11\rangle$ and $|20\rangle$ states, in each case using the previously determined corresponding pulse duration $l(a)$, we obtain the pairs (ϕ, a) shown

as blue dots in Fig. 1(d). Note that on our device, we acquire phase from 0 to -2π , but reverse the sign in Fig. 1(d) for convenience. A continuous function $a(\phi)$ yielding the required amplitude for a target conditional phase $\phi \in [0, 2\pi)$ is obtained by linear interpolation.

With another Ramsey experiment (analogous to the one described above), we determine the acquired dynamic phase ϕ_D of the flux-biased qubit [34] as a function of the flux-pulse amplitude a and duration $l(a)$ relative to a measurement without applying a flux pulse, see Fig. 10(d) for the pulse scheme. The number of calibration points (a, ϕ_D) is chosen such that we can unwrap $\phi_D(a)$ without ambiguity. We interpolate between the unwrapped data points with cubic splines to obtain a continuous function $\phi_D(a)$. We compensate for this single-qubit phase shift $\phi_D(a)$ using a virtual Z gate applied after the CZ_ϕ gate.

Gates between selected qubit pairs require additional flux pulses on neighboring qubits to avoid undesired interactions, see Appendix D. The dynamic phase acquired by neighboring qubits depends on the flux-pulse duration $l(a)$. We perform a Ramsey experiment to characterize this dependency simultaneously with the Ramsey experiment for the qubit directly involved in the gate.

Using this calibration procedure, a CZ_ϕ gate with a desired conditional phase ϕ is implemented by calculating the amplitude $a(\phi)$, the corresponding duration $l[a(\phi)]$ of the flux pulse, and the dynamic phase correction(s) $\phi_D[a(\phi)]$ for the flux-biased qubit(s).

We automate the calibration procedure to require human interaction only to verify the quality of the fits. The approach is thus scalable to larger devices. Note that gate architectures allowing the acquisition of conditional phase as a linear function of the flux-pulse duration could further simplify and speed up the calibration procedure [49].

We characterize the deviation $\delta\phi$ from the target conditional phase ϕ for each pair of qubits, averaged over 30 evenly distributed values of ϕ in the range $[0, 2\pi)$, see Table II. Note that residual ZZ coupling makes it challenging to reach phases near 0. This challenge occurring at the boundaries of the phase interval may be avoided using CZ_ϕ gates, which can span a conditional phase interval larger than 2π [24,49]. We also list the average measured

TABLE II. Average phase deviation from target, $\delta\phi$, and average leakage, λ , for the seven CZ_ϕ gates used in the experiment.

Gate	$\delta\phi$ (deg.)	λ (%)
A_1-B_1	0.62 ± 1.21	0.92 ± 0.49
A_1-B_2	0.23 ± 1.12	0.54 ± 0.29
A_2-B_2	0.53 ± 1.26	0.21 ± 0.19
A_2-B_3	0.00 ± 1.30	0.10 ± 0.08
B_1-A_3	0.81 ± 1.08	0.24 ± 0.53
B_2-A_4	0.22 ± 1.84	3.50 ± 1.97
B_3-A_4	-0.43 ± 1.81	4.68 ± 3.19

leakage, λ , extracted with three-level readout during the characterization of the phase deviations. Five of the seven gates show an average leakage smaller than 1%. The gates between qubits B_2 - A_4 and B_3 - A_4 show average leakage of up to 5%, which we attribute to uncompensated flux-pulse distortions at small time scales.

APPENDIX C: EXACT COVER TO ISING

The exact-cover problem is mathematically formulated as follows [30]. Given a collection of subsets $V = \{V_i\}_{i \in 1, \dots, n}$ with $V_i \subseteq S$, the task is to verify whether there exists a set of indices $I \subseteq \{1, \dots, n\}$ such that $\{V_i\}_{i \in I}$ forms a partition of S , i.e., the sets in $\{V_i\}_{i \in I}$ are disjoint and their union equals S . This is the case if

$$0 = \min_{(b_1, \dots, b_n) \in \{0,1\}^n} \sum_{\ell} \left(1 - \sum_i K_{\ell i} b_i \right)^2, \quad (\text{C1})$$

where the element $K_{\ell i}$ of the incidence matrix K is 1 if the ℓ th element of S is contained in subset V_i and 0 otherwise. A bit value $b_i = 1$ indicates that the subset V_i is selected. Note that, while some authors (see, e.g., Ref. [50]) use the transposed convention in which the i th row of K corresponds to the subset V_i , the convention used here, in which the i th column corresponds to the subset V_i , leads to a clearer labeling in the visual representations of K in Figs. 2(b) and 2(c). Using spins $z_i \in \{\pm 1\}$ instead of bits $b_i = (z_i + 1)/2$, multiplying out, and dropping additive constants, the optimization problem can be formulated as [31,36,51]

$$\min_{(z_1, \dots, z_n) \in \{\pm 1\}^n} \sum_{i < j} J_{ij} z_j z_i + \sum_i h_i z_i, \quad (\text{C2})$$

where

$$J_{ij} = \sum_{\ell} \frac{K_{\ell i} K_{\ell j}}{2} \quad (\text{C3})$$

$$h_i = \sum_{\ell} K_{\ell i} \left(-1 + \frac{1}{2} \sum_j K_{\ell j} \right). \quad (\text{C4})$$

Solving this optimization problem is equivalent to finding the ground-state energy of the Ising Hamiltonian, see Eq. (1), with J_{ij} and h_i values given by Eq. (C3) and Eq. (C4), respectively. In the visual representations of the incidence matrices depicted in Figs. 2(b) and 2(c), the bullets represent the entries with $K_{\ell i} = 1$ while empty cells correspond to $K_{\ell i} = 0$. By substituting these values of $K_{\ell i}$ into the above equations, we see that $h_i = 0$ for all i in both problem instances, and we obtain the values of J_{ij} given in Sec. II.

To run QAOA without requiring swaps, we need all-to-all physical connectivity between qubits that occur jointly in any row of the incidence matrix K . For the physical connectivity graph shown in Fig. 2(a), this means that each row can contain only up to two nonzero entries. The positive sign in Eq. (C3) reveals that the spins corresponding to a row with two nonzero entries have an antiferromagnetic coupling. This is in line with the exact-cover constraint, which requires that exactly one of them is selected in a valid solution, but not both. Thus, any problem instance that does not require swaps on our device and that does not decompose into a set of isolated subgraphs must correspond to a lattice of antiferromagnetically coupled spins. Then, either all qubits labeled with A or all qubits labeled with B have to be in an excited state in a valid solution. In the presence of a row (or rows) with a single nonzero entry, some external field term(s) h_i of the Ising Hamiltonian become(s) nonzero and the solution that fulfills the exact-cover condition also for this row (these rows) is favored. Otherwise, both solutions are valid, which is the case for the problem instances considered in our experiments.

APPENDIX D: QAOA GATE SEQUENCES

Figure 11(a) shows the pulse sequence generated by the AWGs for a single layer of QAOA in the seven-qubit instance using the direct implementation of CZ_{ϕ} gates. Since the duration of the flux pulse depends on the required phase, see Figs. 1(c) and 1(d), the pulse sequence is shown

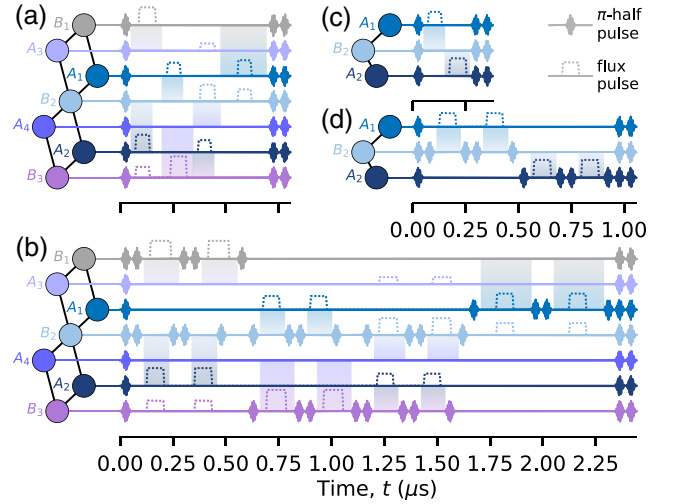


FIG. 11. Pulse sequences for implementing QAOA with $p = 1$. The shaded area around flux pulses illustrates which interaction they implement and how long buffer times before and after the flux pulse are chosen. (a) Direct implementation of the seven-qubit instance. (b) Decomposed implementation of the seven-qubit instance. (c) Direct implementation of the three-qubit instance. (d) Decomposed implementation of the three-qubit instance.

TABLE III. Number of two-qubit gates (first row), single-qubit microwave-driven gates (second row), and single-qubit virtual gates (third row) in the implemented QAOA sequences.

No. of layers, p	1	2	3	4	5	6	7	8	9
3 qb direct	2	4	6	8	10	12	14	16	18
	9	15	21	27	33	39	45	51	57
	7	14	21	28	35	42	49	56	63
3 qb decomposed	4	8	12	16	20	24	28	32	36
	17	31	45	59	73	87	101	115	129
	13	26	39	52	65	78	91	104	117
7 qb direct	7	14	21	28	35	42			
	21	35	49	63	77	91			
	21	42	63	84	105	126			
7 qb decomposed	14	28	42	56					
	49	91	133	175					
	42	84	126	168					

for a representative flux-pulse duration (close to the average) that we obtain for $\gamma = \pi/5$. After an initial $\pi/2$ pulse on each qubit to prepare a $|+\rangle^{\otimes 7}$ state, the phase-separation operator U_C of the first QAOA layer starts with two parallel CZ_ϕ gates corresponding to the couplings $J_{A_3B_1}$ and $J_{A_2B_2}$, while qubit B_3 is detuned by an additional flux pulse to avoid an unwanted interaction when the ef-transition frequency of A_2 crosses the parking frequency of qubit B_3 . Since the additional $Z_{\Gamma_{ij}}$ rotations, see Fig. 1(a), are implemented as virtual gates [32] through a redefinition of the reference frame, they are not shown in the pulse sequence. After the last round of flux pulses, the final two $\pi/2$ pulses for each qubit (plus a virtual gate between them) implement the mixing operator $U_B = e^{-i\beta B}$, where we have decomposed each term $e^{-i\beta X_i}$ as shown in Fig. 1(a). After the end of the shown pulse sequence, we perform qubit readout.

In the significantly longer pulse sequence shown in Fig. 11(b), the controlled arbitrary phase gates are decomposed as described in Fig. 1(b). Each Hadamard gate is implemented by a $\pi/2$ pulse and a Z_π rotation via a virtual gate, and the $Z_{\Gamma_{ij}}$ in the center of the gate decomposition is another virtual gate. Pulse sequences for the direct and the decomposed implementation of the three-qubit problem instance are shown in Figs. 11(c) and 11(d), where analogous explanations apply.

To implement additional layers, the pulses between the end of the initialization pulses and the start of the readout are repeated $p - 1$ times. For the configurations considered in Fig. 5, this leads to the gate counts shown in Table III.

APPENDIX E: PROPERTIES OF QAOA LANDSCAPES

Following Ref. [11], the parameter γ_q can be restricted to $[0, 2\pi)$ if the problem Hamiltonian \hat{C} has integer eigenvalues, while β_q can always be restricted to $[0, \pi)$. In this

appendix, we discuss further periodicity and symmetry properties of the QAOA cost function, which enable us to reduce the parameter space and better understand the cost-function landscapes we observe. To this end, we consider the cost of a p -layer QAOA circuit,

$$C(\vec{\gamma}', \vec{\beta}') = \langle + | U(\vec{\gamma}', \vec{\beta}')^\dagger \hat{C} U(\vec{\gamma}', \vec{\beta}') | + \rangle \quad (\text{E1})$$

in which $U(\vec{\gamma}', \vec{\beta}') = \prod_q e^{-i\beta'_q \hat{B}} e^{-i\gamma'_q \hat{C}}$ is the p -layer QAOA unitary with $\vec{\gamma}' = (\gamma'_1, \dots, \gamma'_p)$ and $\vec{\beta}' = (\beta'_1, \dots, \beta'_p)$.

If the eigenvalues of \hat{C} are integer multiples of α , then by setting $\gamma'_q = \gamma_q + 2\pi/\alpha$ in Eq. (E1) and noting that $e^{\pm i(2\pi/\alpha)\hat{C}} = I$ is the identity, we find that $C(\vec{\gamma}', \vec{\beta}')$ is $(2\pi/\alpha)$ periodic in γ_q .

In addition, if all eigenvalues of \hat{C} are odd multiples of α , we have $e^{\pm i(\pi/\alpha)\hat{C}} = -I$, where the minus sign is a global phase, so that the cost is (π/α) periodic. For both problem instances considered in this work, the eigenvalues of \hat{C} are odd multiples of $1/2$, so that the landscapes are 2π periodic.

Inserting $\beta'_q = \beta_q + \pi$ into Eq. (E1) and noting that $e^{\pm i\pi \hat{B}} = \prod_i e^{\pm i\pi X_i} = \prod_i (-I)$ yields the π periodicity in β_q mentioned in Ref. [11]. Moreover, since $e^{-i(\pi/2)\hat{B}} = \prod_i e^{-i(\pi/2)X_i}$ corresponds to an X_π rotation of all qubits, setting $\beta'_p = \beta_p + (\pi/2)$ in the last layer p corresponds to flipping the sign of all spins before estimating the energy of \hat{C} . If the Ising Hamiltonian \hat{C} does not contain single-qubit terms ($h_i = 0$ for all i), this sign flip does not change the energy, and the cost landscape is $\pi/2$ periodic in β_p . As this applies to the examples considered in this paper, we measure the landscapes for $p = 1$ only up to $\beta = \beta_1 = \pi/2$.

By simultaneously setting $\gamma'_q = -\gamma_q$ and $\beta'_q = -\beta_q$ in Eq. (E1), and noting that \hat{C} , \hat{B} , and $|+\rangle$ are real valued, we have $C(-\vec{\gamma}', -\vec{\beta}') = [C(\vec{\gamma}', \vec{\beta}')]^\dagger = C(\vec{\gamma}', \vec{\beta}')$. Therefore,

the cost landscape is point symmetric with respect to the origin, which implies that it is also point symmetric with respect to the center point of a period. When measuring a landscape, we can thus restrict either β or γ to half a period without losing information about the landscape. In the examples shown in this paper, we restrict γ to half a period, i.e., to the interval $[0, \pi)$.

Finally, when choosing $\gamma'_q = -\gamma_q$ and $\beta'_q = \beta_q$ in Eq. (E1), we obtain

$$-C(-\vec{\gamma}, \vec{\beta}) = \langle + | U'(\vec{\gamma}, \vec{\beta})^\dagger (-\hat{C}) U'(\vec{\gamma}, \vec{\beta}) | + \rangle, \quad (\text{E2})$$

where $U'(\vec{\gamma}, \vec{\beta}) = \prod_q e^{-i\beta_q \hat{B}} e^{-i\gamma_q (-\hat{C})}$. This is equivalent to the QAOA cost function for a problem Hamiltonian $\hat{C}' = -\hat{C}$. Thus, in cases for which running QAOA with \hat{C} and with $-\hat{C}$ leads to the same landscape, the landscape is an odd function of $\vec{\gamma}$. In particular, this occurs for both problem instances considered in this work.

Due to the point symmetry observed above, the landscape is also an odd function of $\vec{\beta}$ if it is an odd function of $\vec{\gamma}$. In the landscape plots for $p = 1$, this manifests as line symmetries (with a change of the sign of the energy) about both coordinate axes and with respect to the center line of each period. Within the chosen range of β , we observe this type of symmetry with respect to the horizontal line $\beta = \pi/4$.

APPENDIX F: POSTSELECTION

For both QAOA implementations, we discard all measured states containing at least one leakage event. We show the percentage of single-shot measurements we keep as a function of the number of layers in the top half of Table IV. We estimate the corresponding average leakage per gate as $\lambda_{\text{post}} \approx 1 - P_{\text{post}}^{1/n_g}$, where P_{post} is the fraction of data left after postselection and n_g is the number of two-qubit gates in the sequence.

We compare the λ_{post} with the mean of the measured average leakage, $\tilde{\lambda}_{3\text{qb}} = 0.38\%$ and $\tilde{\lambda}_{7\text{qb}} = 1.46\%$ defined as the average of the values reported in Table II for the two and seven gates used in the gate sequence, respectively. In

both cases, the mean of the measured average leakage is in decent agreement with the λ_{post} reported in Table IV.

APPENDIX G: MASTER-EQUATION SIMULATIONS

We model the dynamics of our system by a master equation given by

$$\dot{\rho} = -\frac{i}{\hbar} [H(t), \rho] + \sum_k \left[\hat{c}_k \rho \hat{c}_k^\dagger - \frac{1}{2} (\hat{c}_k^\dagger \hat{c}_k \rho + \rho \hat{c}_k^\dagger \hat{c}_k) \right], \quad (\text{G1})$$

where ρ is the density matrix describing the system at time t and $H(t)$ is the Hamiltonian, the time dependence of which models the applied gate sequence. Single-qubit Y gates are simulated for each qubit with a 50-ns-long DRAG pulse [52]. Arbitrary rotations along the Y axis are implemented by adapting the pulse amplitude accordingly, while keeping the gate time constant. We adjust the phase of the DRAG pulse to account for the virtual Z gates [53] and to simulate single-qubit X gates. Two-qubit gates between qubits i and j are simulated using the interaction Hamiltonian $H(t) = c(t) |11\rangle_{ij} \langle 11|$, where $c(t)$ is constant for a time τ_g corresponding to the gate duration. The amplitude $c(t)$ is adjusted such that the interaction implements the targeted CZ_ϕ gate. This model neither considers the dynamical phases that arise from the flux tuning of the qubits nor the possible leakage to noncomputational states during the gate. The collapse operators \hat{c}_k model incoherent processes. We solve the master equation numerically [54] in the rotating frame of qubits. Incoherent errors are described by Lindblad terms in Eq. (G1) with

$$\hat{c}_{T_{1,i}} = \sqrt{\frac{1}{T_{1,i}}} \sigma_{-i}, \quad (\text{G2})$$

$$\hat{c}_{T_{\phi,i}} = \sqrt{\frac{1}{2} \left(\frac{1}{T_{2,i}} - \frac{1}{2T_{1,i}} \right)} \sigma_{z,i}, \quad (\text{G3})$$

TABLE IV. (First four rows) Percentage of data kept after discarding all measured states containing at least one leakage event. (Bottom four rows) Corresponding average leakage per two-qubit gate in percent.

No. of layers, p	1	2	3	4	5	6	7	8	9
3 qb direct	99.2	97.1	97.9	97.1	96.4	97.1	96.6	94.1	93.6
3 qb decomposed	98.6	98.3	96.4	92.1	95.8	94.0	94.8	94.4	93.5
7 qb direct	87.1	87.7	80.6	55.9	79.2	77.5			
7 qb decomposed	80.2	82.4	61.6	62.1					
3 qb direct	0.4	0.7	0.3	0.4	0.4	0.2	0.2	0.4	0.4
3 qb decomposed	0.4	0.2	0.3	0.5	0.2	0.3	0.2	0.2	0.2
7 qb direct	1.9	0.9	1.0	2.1	0.7	0.6			
7 qb decomposed	1.6	0.7	1.1	0.8					

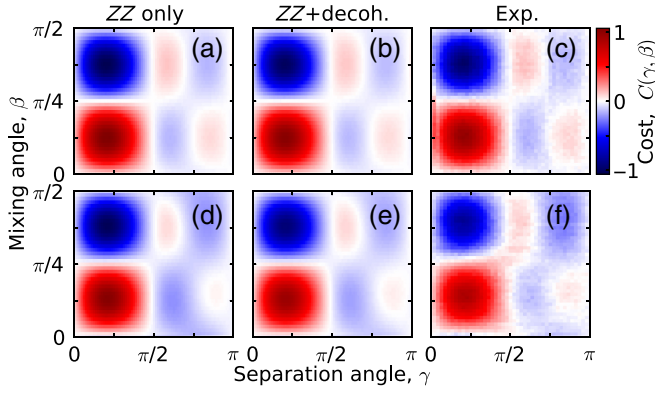


FIG. 12. Simulated and experimental cost-function landscapes of the three-qubit problem instance for $p = 1$. The direct implementation is displayed in (a)–(c). The decomposed version is shown in (d)–(f). The master-equation simulations are performed including errors from residual ZZ coupling only (a),(d), and with both residual ZZ coupling and decoherence (b),(e). The experimental data is shown in (c),(f).

where $T_{1,i}$ and $T_{2,i}$ are the lifetime and decoherence time (Ramsey decay time) of qubit i as listed in Table I. Since we do not explicitly apply echo pulses in the experimental sequences, we use the Ramsey decay time in the simulations rather than the dephasing times extracted from a Hahn-echo experiment.

In addition to the incoherent errors introduced by the Lindblad terms, it is important to also consider the impact of coherent errors on the algorithm. In our experiment, the main source of coherent errors is residual ZZ coupling between neighboring qubits [38]. To model this coupling in the numerical simulations, we include the Hamiltonian

$$H_{ZZ}/\hbar = \sum_{(i,j)} \alpha_{ij} |11\rangle_{ij} \langle 11| \quad (\text{G4})$$

where the sum is over connected pairs of qubits with the residual ZZ couplings listed in Appendix A. We notice from simulations of the full QAOA circuit that the residual ZZ couplings give rise to the distortions observed in the cost landscapes, see Fig. 12. The main effect of decoherence is to reduce the overall contrast of the landscape. In particular, for the direct implementation we find a minimum energy of -1.04 , -0.99 , and -0.98 for simulations including only residual ZZ couplings, for simulations including both residual ZZ couplings and decoherence, and for experiments, respectively. In comparison, the minimum energy in noise-free simulations is -1.06 , see Fig. 3(b).

APPENDIX H: ADDITIONAL QAOA MEASUREMENTS

We use a single-layer QAOA circuit with CZ_ϕ gates to measure the cost-function landscape of the seven-qubit problem instance, see Appendix D for the full pulse

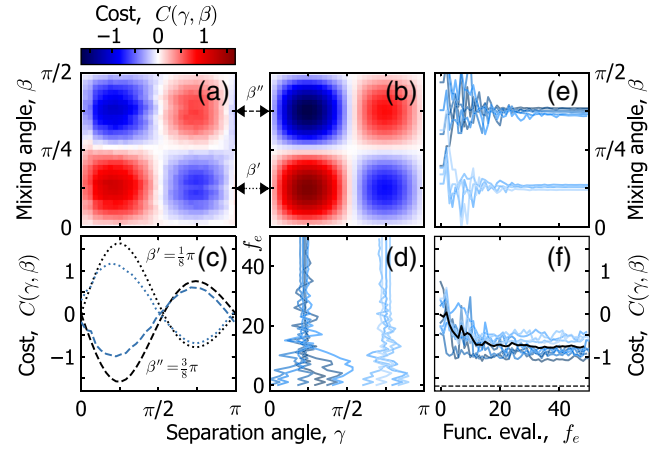


FIG. 13. Cost function evaluated for $p = 1$ on the seven-qubit problem instance using CZ_ϕ gates. (a) Cost-function landscape as a function of variational parameters. (b) Cost-function landscape obtained from noise-free simulations. (c) Experimental evaluation (blue) and simulation (black) of the cost function for two horizontal line cuts of (a),(b), with $\beta' = \pi/8$ (dotted lines) and $\beta' = 3\pi/8$ (dashed lines), respectively. (d),(e) Ten convergence traces of the separation angle and the mixing angle, respectively, for end-to-end optimization starting from random parameter initialization. (f) Average energy (solid black line) and individual convergence traces (faded blue lines) of the energy corresponding to parameters shown in (d),(e).

sequence. The measured landscape, see Fig. 13(a), is in good qualitative agreement with noise-free simulations, see Fig. 13(b). Due to decoherence, the absolute values of the global extrema are smaller than in noise-free simulations, see Fig. 13(c). Starting from random initialization, the convergence traces of the separating angle, the mixing angle and the corresponding cost are displayed in Figs. 13(d)–13(f), respectively.

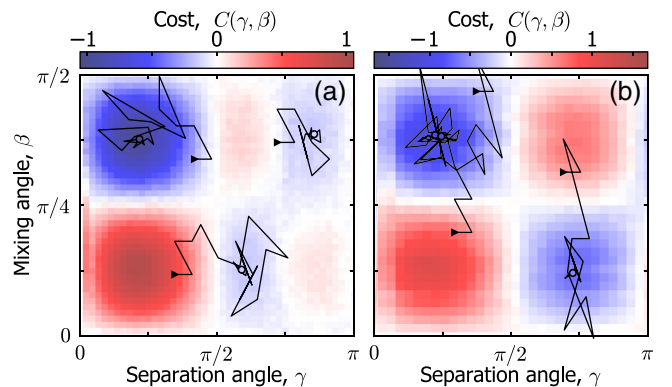


FIG. 14. Overlay of three selected convergence traces on the cost-function landscape of the (a) three-qubit and (b) seven-qubit problem instance. The random initialization is shown with a black triangle and the final location of each trace is indicated by a round marker whose color reflects the cost of the last function evaluation.

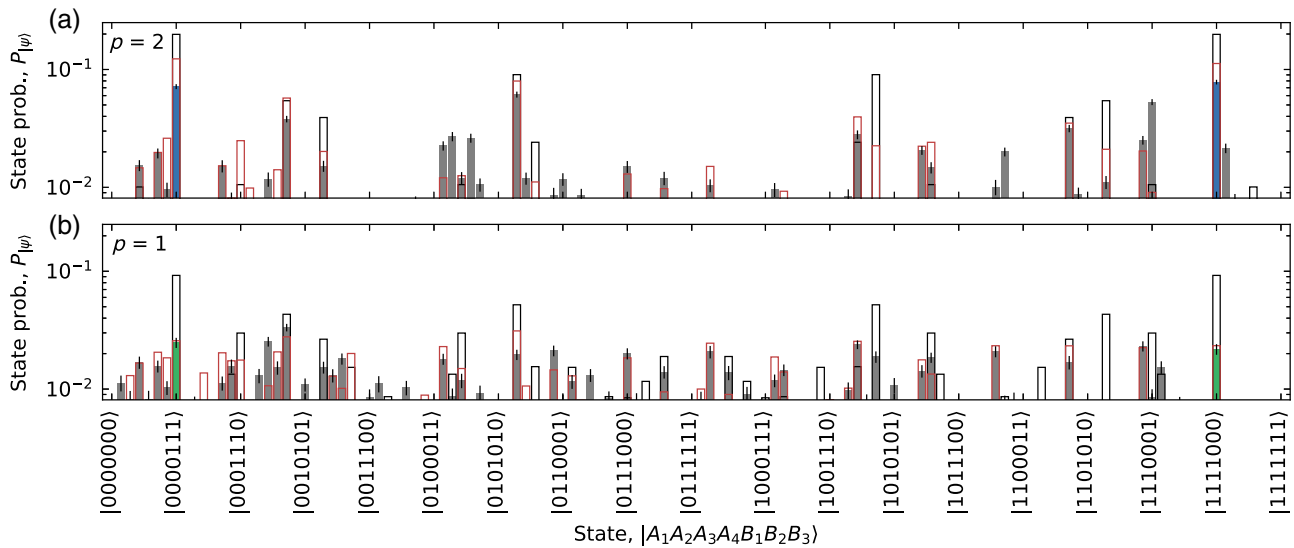


FIG. 15. Output state probability distribution for the seven-qubit problem instance implemented with CZ_ϕ gates (a) and decomposed using CZ gates (b). States are measured at optimal parameters for depth of $p = 2$ for (a) and $p = 1$ for (b). The filled bars correspond to the measured problem solutions, while the black (red) wireframes are the expected QAOA outcome from noise-free (master-equation) simulations. We use 60 000 individual measurement runs to estimate each distribution and indicate bootstrap 99.7% confidence intervals with black vertical markers.

In addition, we illustrate the convergence of a subset of traces to global and local minima by overlaying them on the cost-function landscape for the three-qubit and the seven-qubit problem instance, see Figs. 14(a) and 14(b), respectively.

The output state distributions at optimal parameters for the direct and decomposed implementation of CZ_ϕ gates are shown for the seven-qubit problem instance in Figs. 15(a) and 15(b), respectively. We display the distributions yielding highest success probability for each implementation, i.e., $p = 2$ for the direct implementation and $p = 1$ for the decomposed implementation. For the direct implementation, the two most likely measured states are $|1111000\rangle$ and $|0000111\rangle$, corresponding to the respective selections of subsets \mathcal{A} and \mathcal{B} forming exact covers of the considered problem instance. Conversely, the solution states are not the two most likely measured states for the decomposed implementation. For both implementations, the measured data matches well with expectations from master-equation simulations (red wireframe).

[1] P. W. Shor, in *Proceedings, 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Santa Fe, 1994), 124–134.

[2] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P. D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).

[3] Daniel A. Lidar and Todd A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, 2013).

[4] John Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).

[5] Frank Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).

[6] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, *Nature* **551**, 601 (2017).

[7] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin, Probing many-body dynamics on a 51-atom quantum simulator, *Nature* **551**, 579 (2017).

[8] P. J. J. O’Malley *et al.*, Scalable Quantum Simulation of Molecular Energies, *Phys. Rev. X* **6**, 031007 (2016).

[9] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).

[10] F. Arute *et al.*, Hartree-Fock on a superconducting qubit quantum computer, [arXiv:2004.04174](https://arxiv.org/abs/2004.04174) (2020).

[11] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann, A quantum approximate optimization algorithm, [arXiv:1411.4028](https://arxiv.org/abs/1411.4028) (2014).

[12] J. S. Otterbach *et al.*, Unsupervised machine learning on a hybrid quantum computer, [arXiv:1712.05771](https://arxiv.org/abs/1712.05771) (2017).

[13] F. Arute *et al.*, Quantum approximate optimization of non-planar graph problems on a planar superconducting processor, [arXiv:2004.04197](https://arxiv.org/abs/2004.04197) (2020).

- [14] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin, Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices, *Phys. Rev. X* **10**, 021067 (2020).
- [15] 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, *Algorithms* **12**, 34:1 (2019).
- [16] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou, The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size, [arXiv:1910.08187](https://arxiv.org/abs/1910.08187) (2019).
- [17] Guillaume Verdon, Juan Miguel Arrazola, Kamil Brádler, and Nathan Killoran, A quantum approximate optimization algorithm for continuous problems, [arXiv:1902.00409](https://arxiv.org/abs/1902.00409) (2019).
- [18] Zhang Jiang, Eleanor G. Rieffel, and Zhihui Wang, Near-optimal quantum circuit for Grover’s unstructured search using a transverse field, *Phys. Rev. A* **95**, 062317 (2017).
- [19] M. Alam, A. Ash-Saki, and S. Ghosh, Analysis of quantum approximate optimization algorithm under realistic noise in superconducting qubits, [arXiv:1907.09631](https://arxiv.org/abs/1907.09631) (2019).
- [20] Toshiaki Matsumine, Toshiaki Koike-Akino, and Ye Wang, in *Proc. 2019 IEEE Int. Symp. Inf. Theory (ISIT)* (IEEE, Paris, 2019), p. 2574.
- [21] A. Bengtsson, P. Vikstål, C. Warren, M. Svensson, X. Gu, A. F. Kockum, P. Krantz, C. Križan, D. Shiri, I. Svensson, G. Tancredi, G. Johansson, P. Delsing, G. Ferrini, and J. Bylander, Quantum approximate optimization of the exact-cover problem on a superconducting quantum processor, [arXiv:1912.10495](https://arxiv.org/abs/1912.10495) (2019).
- [22] Xiaogang Qiang, Xiaoqi Zhou, Jianwei Wang, Callum M. Wilkes, Thomas Loke, Sean O’Gara, Laurent Kling, Graham D. Marshall, Raffaele Santagati, Timothy C. Ralph, Jingbo B. Wang, Jeremy L. O’Brien, Mark G. Thompson, and Jonathan C. F. Matthews, Large-scale silicon quantum photonics implementing arbitrary two-qubit processing, *Nat. Photonics* **12**, 534 (2018).
- [23] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe, Quantum approximate optimization with a trapped-ion quantum simulator, [arXiv:1906.02700](https://arxiv.org/abs/1906.02700) (2019).
- [24] R. Barends *et al.*, Digital quantum simulation of fermionic models with a superconducting circuit, *Nat. Commun.* **6**, 7654 (2015).
- [25] P. Roushan *et al.*, Chiral ground-state currents of interacting photons in a synthetic magnetic field, *Nat. Phys.* **13**, 146 (2016), Advance online publication.
- [26] M. Ganzhorn, D. J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer, *Phys. Rev. Appl.* **11**, 044092 (2019).
- [27] B. Foxen *et al.*, Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms, [arXiv:200108343](https://arxiv.org/abs/200108343) (2020).
- [28] D. M. Abrams, N. Didier, B. R. Johnson, M. P. da Silva, and C. A. Ryan, Implementation of the xy interaction family with calibration of a single pulse, [arXiv:2005.12026](https://arxiv.org/abs/2005.12026) (2019).
- [29] D. Headley, T. Muller, A. Martin, E. Solano, M. Sanz, and F. K. Wilhelm, Approximating the quantum approximate optimisation algorithm, [arXiv:2002.12215](https://arxiv.org/abs/2002.12215) (2020).
- [30] Richard M. Karp, in *Complexity of Computer Computations: Proceedings of a Symposium on the Complexity of Computer Computations, Held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department*, edited by Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger (Springer, Boston, MA, USA, 1972), p. 85.
- [31] Andrew Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5:1 (2014).
- [32] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient z-gates for quantum computing, [arXiv:1612.00858](https://arxiv.org/abs/1612.00858) (2016).
- [33] Frederick W. Strauch, Philip R. Johnson, Alex J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Quantum Logic Gates for Coupled Superconducting Phase Qubits, *Phys. Rev. Lett.* **91**, 167005 (2003).
- [34] L. DiCarlo, J. M. Chow, J. M. Gambetta, Lev S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Demonstration of two-qubit algorithms with a superconducting quantum processor, *Nature* **460**, 240 (2009).
- [35] R. Barends *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, *Nature* **508**, 500 (2014).
- [36] P. Vikstål, M. Grönkvist, M. Svensson, M. Andersson, G. Johansson, and G. Ferrini, Applying the quantum approximate optimization algorithm to the tail assignment problem, [arXiv:1912.10499](https://arxiv.org/abs/1912.10499) (2019).
- [37] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo, Effects of quantum noise on quantum approximate optimization algorithm, [arXiv:1909.02196v2](https://arxiv.org/abs/1909.02196v2) (2019).
- [38] S. Krinner, S. Lazar, A. Remm, C. K. Andersen, N. Lacroix, G. J. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, Benchmarking Coherent Errors in Controlled-Phase Gates due to Spectator Qubits, *Phys. Rev. Appl.* **14**, 024042 (2020).
- [39] A. C. Davison and D. V. Hinkley, in *Bootstrap Methods and their Application*, *Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge University Press, Cambridge, 1997), p. 1169.
- [40] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distribution, *Bull. Calcutta Math. Soc.* **35**, 99 (1943).
- [41] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang, Obstacles to state preparation and variational optimization from symmetry protection, [arXiv:1910.08980](https://arxiv.org/abs/1910.08980) (2019).
- [42] E. Farhi, D. Gamarnik, and S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: A typical case, [arXiv:2004.09002](https://arxiv.org/abs/2004.09002) (2020).

- [43] S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinsoo, R. Keller, J. Lütolf, C. Eichler, and A. Wallraff, Engineering cryogenic setups for 100-qubit scale superconducting circuit systems, *EPJ Quantum Technol.* **6**, 2 (2019).
- [44] Christian Kraglund Andersen, Ants Remm, Stefania Lazar, Sebastian Krinner, Nathan Lacroix, Graham J. Norris, Mihai Gabureac, Christopher Eichler, and Andreas Wallraff, Repeated quantum error detection in a surface code, *Nat. Phys.* **16**, 875 (2020).
- [45] Johannes Heinsoo, Christian Kraglund Andersen, Ants Remm, Sebastian Krinner, Theodore Walter, Yves Salathé, Simone Gasparinetti, Jean-Claude Besse, Anton Potočnik, Andreas Wallraff, and Christopher Eichler, Rapid High-Fidelity Multiplexed Readout of Superconducting Qubits, *Phys. Rev. Appl.* **10**, 034040 (2018).
- [46] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovskiy, X. Zhang, W. D. Oliver, and I. Siddiqi, A near-quantum-limited Josephson traveling-wave parametric amplifier, *Science* **350**, 307 (2015).
- [47] P. Magnard, P. Kurpiers, B. Royer, T. Walter, J.-C. Besse, S. Gasparinetti, M. Pechal, J. Heinsoo, S. Storz, A. Blais, and A. Wallraff, Fast and Unconditional All-Microwave Reset of a Superconducting Qubit, *Phys. Rev. Lett.* **121**, 060502 (2018).
- [48] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J.-C. Besse, S. Gasparinetti, A. Blais, and A. Wallraff, Deterministic quantum state transfer and remote entanglement using microwave photons, *Nature* **558**, 264 (2018).
- [49] Michele C. Collodo, Johannes Herrmann, Nathan Lacroix, Christian Kraglund Andersen, Ants Remm, Stefania Lazar, Jean-Claude Besse, Theo Walter, Andreas Wallraff, and Christopher Eichler, Implementation of conditional-phase gates based on tunable ZZ-interactions, [arXiv:2005.08863](https://arxiv.org/abs/2005.08863) (2020).
- [50] Donald E. Knuth, Dancing links, [arXiv:cs/0011047](https://arxiv.org/abs/cs/0011047) (2000).
- [51] Vicky Choi, Adiabatic quantum algorithms for the NP-complete maximum-weight independent set, exact cover and 3SAT problems, [arXiv:1004.2226](https://arxiv.org/abs/1004.2226) (2010).
- [52] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits, *Phys. Rev. Lett.* **103**, 110501 (2009).
- [53] David C. McKay, Christopher J. Wood, Sarah Sheldon, Jerry M. Chow, and Jay M. Gambetta, Efficient z gates for quantum computing, *Phys. Rev. A* **96**, 022330 (2017).
- [54] J. R. Johansson, P. D. Nation, and Franco Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **184**, 1234 (2013).