Implementing two-qubit gates at the quantum speed limit

Joel Howard,[1,2] Alexander Lidiak,[1] Casey Jameson,[1] Bora Basyildiz,[3] Kyle Clark,[1] Tongyu Zhao,[4] Mustafa Bal,[4,5] Junling Long,[4] David P. Pappas,[6,2] Meenakshi Singh,[1,*] and Zhexuan Gong[1,†]

[1]*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*
[2]*Rigetti Computing, 775 Heinz Avenue, Berkeley, California 94701, USA*
[3]*Department of Computer Science, Colorado School of Mines, Golden, Colorado 80401, USA*
[4]*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*
[5]*Superconducting Quantum Materials and Systems Center, Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[6]*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

(Received 25 June 2022; revised 21 July 2023; accepted 22 October 2023; published 1 December 2023)

The speed of elementary quantum gates, particularly two-qubit gates, ultimately sets the limit on the speed at which quantum circuits can operate. In this work, we experimentally demonstrate commonly used two-qubit gates at nearly the fastest possible speed allowed by the physical interaction strength between two superconducting transmon qubits. We achieve this quantum speed limit by implementing experimental gates designed using a machine-learning-inspired optimal control method. Importantly, our method only requires the single-qubit drive strength to be moderately larger than the interaction strength to achieve an arbitrary two-qubit gate close to its analytical speed limit with high fidelity. Thus the method is applicable to a variety of platforms, including those with comparable single-qubit and two-qubit gate speeds, or those with always-on interactions. We expect our method to offer significant speedups for non-native two-qubit gates that are typically achieved with a long sequence of single-qubit and native two-qubit gates.

DOI: 10.1103/PhysRevResearch.5.043194

I. INTRODUCTION

Increasing the speed of elementary quantum gates boosts the "clock speed" of a quantum computer. For noisy, intermediate-scale quantum computers [1] with finite coherence times [2,3], speeding up single- and two-qubit quantum gates also increases the circuit depth needed for solving useful computational problems [4,5]. In most experimental platforms, single-qubit gates are achieved via electromagnetic fields that drive individual qubit transitions. The maximum speed of these gates is often limited by the strength of the driving fields [6,7]. However, a two-qubit entangling gate, necessary for universal quantum gates, can only operate at a speed proportional to the interaction strength between the qubits [8–10], which is typically weaker than available single-qubit drive strengths and cannot be easily increased.

Assuming a limited interaction strength, one can analytically obtain the maximum speed for any particular two-qubit gate in the limit of arbitrarily fast single-qubit gates [11,12]. In practice, all single-qubit gates have finite speeds, and in platforms such as superconducting qubits, single-qubit gates may not be much faster than two-qubit gates due to limited anharmonicity [13–15]. The speed limits of two-qubit gates in such scenarios have not been studied. Therefore we seek to both theoretically and experimentally investigate these practical speed limits, which are not only relevant to the optimal design of quantum gates and quantum circuits, but also directly related to the speed limits of entanglement generation, a fundamental topic of high interest in quantum information theory, condensed-matter physics, and black-hole physics [16–20].

In this paper we report a method for designing two-qubit gates that are speed optimized, and we implement the gates experimentally using superconducting transmon qubits. We find that the protocol for achieving the fastest two-qubit gates in Refs. [11,12] can be far from optimal with a finite single-qubit gate time. Our method differs from previous protocols [11,12,21] in that we apply single-qubit drives simultaneously with the two-qubit interaction, a crucial strategy for speed optimization. We optimize the pulse shapes of the single-qubit drives using a method that combines the well-known GRAPE algorithm [22,23] with state-of-art machine learning techniques. Importantly, our method works as long as there exists a physical interaction between two qubits and one can drive each qubit with controllable pulse shape and phase. This applies to almost any platform suitable for quantum computing. Furthermore, we experimentally demonstrate that our method can achieve maximally entangling two-qubit gates, such as theCNOT gate, close to their analytical speed limits found in Refs. [11,12] with modest single-qubit drive strengths and up to 98.3% average gate fidelity determined from quantum

*msingh@mines.edu
†gong@mines.edu

process tomography. This largely eliminates the impractical assumption of infinitely fast single-qubit gates. The same applies to the SWAP gate we implemented, which is crucial for remote quantum gates but notoriously hard to achieve with a short gate sequence [24]. We have also implemented a non-Clifford gate, the $\sqrt{\text{SWAP}}$ gate, close to its quantum speed limit at 97.0% average gate fidelity.

We emphasize that the above-mentioned gates we implemented are all non-native gates, meaning that they cannot be achieved by evolving the static interaction Hamiltonian without single-qubit gates or time-dependent drives. These non-native gates are crucial for efficient implementation of many useful quantum circuits [25–27]. With our method, one could even realize an arbitrary two-qubit gate close to its theoretical speed limit. The conventional method of using a universal gate set to achieve a general SU(4) unitary would instead require a sequence of up to three two-qubit gates and twelve single-qubit gates [28], which is not only much slower in practice, but also likely of lower fidelity due to error accumulation.

Although certain two-qubit gates with 99.5% fidelity or higher have already been achieved with superconducting qubits [13–15], the lower fidelity gates reported here are largely due to limitations of our experimental hardware, as the theoretical gate fidelities associated with our demonstrated gates are all above 99.9%. The main purpose of this work is to demonstrate that the theoretical two-qubit gate speed limits can be largely achieved using a general optimal control method with minimal hardware requirement. Our speed-optimized gates achieve the same level of fidelity as that achieved previously on the same hardware with only fidelity optimization [21], showing that our gate-speed optimization does not rely on sacrificing gate fidelity. Moreover, our method is scalable to a large number of qubits, provided that the two-qubit interactions can be switched on and off (such as via tunable couplers [29]), since we can perform the speed optimization for each two-qubit gate independently. With most of the research on quantum gates focusing on improving gate fidelities, our work represents an orthogonal direction that aims to improve the fidelity of the whole quantum circuit [30] by optimizing the speed of any two-qubit gates.

II. EXPERIMENTAL SETUP

Our experimental platform consists of strongly coupled fixed-frequency superconducting transmon qubits with static capacitive couplings in a hanger readout geometry [31]. An intrinsic silicon substrate is used on which aluminum oxide tunnel junctions are fabricated via an overlap technique [32]. The remaining circuit components are made of niobium. The full chip design and corresponding circuit model is shown in Fig. 1. The two transmon qubit transition frequencies are 5.10 GHz, 5.26 GHz, anharmonicities are −270 MHz, −320 MHz, T_1 decay times are 40 μs, 21 μs and T_2^* decay times are 12 μs, 10 μs, respectively. In the rotating frame of the two qubit frequencies and assuming $\hbar = 1$ from now on, the static Hamiltonian of the two qubits can be written

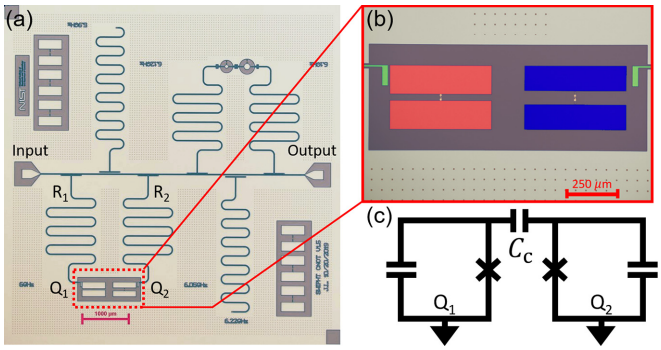

FIG. 1. (a) Optical micrograph of the experimental chip including qubits, readout resonators, test Josephson junctions, and test resonators. (b) Zoomed-in view of the two floating qubits. Each qubit consists of two identical pads (red for the left qubit and blue for the right qubit) and a Josephson junction connecting the two pads. Each qubit is coupled to its own readout resonator (blue). (c) Grounded circuit model of the capacitively coupled qubits.

as [21,33]

$$H_0 = g\left(\sigma_1^z + \sigma_2^z + \sigma_1^z\sigma_2^z\right), \tag{1}$$

where $g \approx 2\pi \times 1.75$ MHz represents a fixed Ising coupling strength between the qubits. To interact with the qubits, we deliver two microwave drives—resonant with each qubit's transition frequency—simultaneously through the feedline. Each of the drive fields contain two adjustable quadratures (X or Y) and can be described by the drive Hamiltonian:

$$H_1(t) = \sum_{\gamma=x,y}\sum_{i=1,2}\Omega_i^\gamma(t)\tilde{\sigma}_i^\gamma, \tag{2}$$

where $\Omega_i^{x,y}(t)$ denotes the Rabi frequency of the drive resonant with qubit i's transition in the X or Y quadrature at time t. For perfect single-qubit drives, $\tilde{\sigma}_i^\gamma = \sigma_i^\gamma$. However, due to the strong Ising coupling between the two qubits in H_0, the drive strength on one qubit is dependent on the other qubit's state, resulting in $\tilde{\sigma}_1^\gamma = \sigma^\gamma \otimes (|0\rangle\langle 0| + r_2|1\rangle\langle 1|)$ and $\tilde{\sigma}_2^\gamma = (|0\rangle\langle 0| + r_1|1\rangle\langle 1|) \otimes \sigma^\gamma$, with $r_1 \approx 1.1$ and $r_2 \approx 0.7$ for our current chip. We note that with a weaker coupling strength or with a tunable coupler [34,35], both r_1 and r_2 can be made closer to or equal to 1.

III. ANALYTICAL SPEED LIMIT

In the limit of arbitrarily strong single-qubit drives, i.e., $\Omega_{\max} \equiv \max|\Omega_{1,2}^{x,y}(t)| \to \infty$, one can derive an analytical speed limit for any target two-qubit unitary with the above-mentioned static Hamiltonian H_0 and control Hamiltonian H_1. Note that in this limit, the speed limit is only well defined when $r_1 = r_2 = 1$, since otherwise H_1 will lead to arbitrarily strong interactions. As detailed in [11], any two-qubit target unitary U can always be decomposed as

$$U = (U_1 \otimes U_2)U_d(V_1 \otimes V_2), \quad U_d = e^{-i\sum_\gamma \lambda_\gamma \sigma^\gamma \otimes \sigma^\gamma}, \tag{3}$$

where U_1, V_1 (U_2, V_2) are some single-qubit gates on the first (second) qubit, $\gamma = x, y, z$, and $\lambda_\gamma \in [-\frac{\pi}{4}, \frac{\pi}{4}]$. $\{\lambda_x, \lambda_y, \lambda_z\}$ form a canonical vector that uniquely specifies any given two-qubit gate U up to single-qubit rotations, and their exact

values can be obtained with the knowledge of U based on Ref. [12]. To obtain the analytical speed limit, we assume that single-qubit gates are arbitrarily fast and thus take a negligible amount of time. The total gate time for implementing U therefore reduces to the time spent on realizing U_d, which is responsible for any entanglement generation. Based on H_0, together with instantaneous single-qubit rotations, U_d can be realized with a minimum time of

$$T_{\min} = \frac{|\lambda_x| + |\lambda_y| + |\lambda_z|}{g} = \begin{cases} \pi/(4g) & \text{CNOT} \\ 3\pi/(4g) & \text{SWAP} \\ 3\pi/(8g) & \sqrt{\text{SWAP}} \end{cases}. \quad (4)$$

Equation (4) is the analytical speed limit for a two-qubit gate U. Its values for the CNOT, SWAP, and $\sqrt{\text{SWAP}}$ gate are shown above and derived in Appendix B.

IV. OPTIMAL CONTROL METHOD

In practice, single-qubit gate speeds are limited by finite drive strengths and the analytical speed limit in Eq. (4) does not apply. To our best knowledge, no analytical speed limit has been found with finite single-qubit gate time. In this case we can no longer rely on the decomposition in Eq. (3) to reduce the problem to just finding the minimum time in implementing U_d. The true time-optimal protocol may not feature the structure of such decomposition and is challenging to find analytically for a general target gate. In fact, if we still follow Eq. (3) for realizing a target two-qubit gate, the resulting gate time can be much longer than $T_{\min}$. To realize universal single-qubit gates needed in Eq. (3), we use a two-axis gate (TAG) protocol first developed in Ref. [21], which employs an analytically obtained, three-segment drive pulse for $\Omega_{1,2}^{x,y}(t)$ to exactly cancel the effects of static interaction for any values of r_1 and r_2. A single-qubit gate implemented this way has a gate time of at least $\pi/(2g)$ [33]. Apart from the native controlled-Z (CZ) gate that can be directly realized via an evolution of H_0 over a time $t = \pi/(4g) \approx 71.4$ ns [21,36,37], any two-qubit gate design that involves the use of single-qubit gate(s) realized via TAG requires a gate time of at least $T_{\min} + \pi/(2g)$ (since single-qubit gates cannot shorten $T_{\min}$) and is thus far from optimal.

Consequently, to approach the analytical speed limit with finite $\Omega_{\max}$ (or finite single-qubit gate time), we adopt an alternative approach that avoids the use of any single-qubit gate and generate the target two-qubit gate directly. Specifically, we directly optimize the pulse shapes $\Omega_{1,2}^{x,y}(t)$ in our control Hamiltonian $H_1(t)$ in order to minimize the gate time for achieving a certain target gate with sufficiently high fidelity. For a given set of pulse shape functions $\Omega_{1,2}^{x,y}(t)$, we numerically find the evolution operator

$$\mathcal{U} = \mathcal{T} e^{-i\int_0^T [H_0 + H_1(t)]dt}, \quad (5)$$

where $\mathcal{T}$ denotes the time-ordered integral and T denotes the total evolution time. Note that $\mathcal{U}$ can achieve any two-qubit gate with properly engineered pulse shapes, as the Hamiltonian and the commutators of the Hamiltonian at different times span all SU(4) generators.

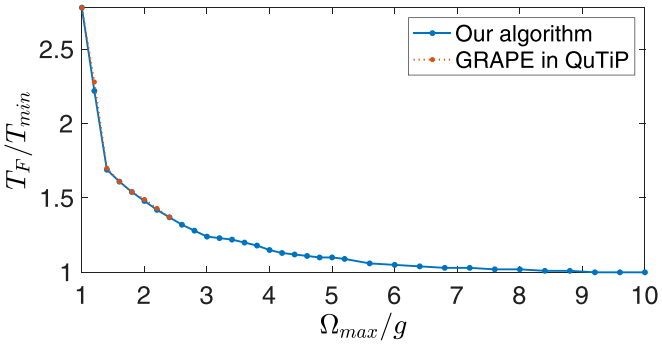

FIG. 2. The minimum time T_F (in units of $T_{\min}$) it takes to achieve a CNOT gate of $F > 99\%$ as a function of $\Omega_{\max}$ (in units of g) using either our optimization algorithm (blue) or the GRAPE algorithm in QUTIP (red). Both algorithms use 16 segments of the drive pulses and 200 random restarts. The GRAPE algorithm in QUTIP fails to reach $F > 99\%$ for larger $\Omega_{\max}$ values, where the gate time approaches its theoretical limit $T_{\min}$.

Next we calculate the average gate fidelity between the target unitary U and the evolution operator $\mathcal{U}$ using [38]

$$F = \frac{1}{5} + \frac{1}{20}\sum_j \text{Tr}(U U_j U^\dagger \mathcal{U} U_j \mathcal{U}^\dagger), \quad (6)$$

where $U_j \in \{\sigma^\gamma \otimes \sigma^{\gamma'}\}$ and $\sigma^\gamma \in \{\sigma^x, \sigma^y, \sigma^z, I\}$. For efficient numerical optimization, we will assume $\Omega_i^\gamma(t)$ is an M-segment piecewise function, i.e., $\Omega_i^\gamma(t) = \Omega_{i,m}^\gamma$ for $t \in [\frac{m-1}{M}T, \frac{m}{M}T]$, where m indexes the segments sequentially. Our goal is to maximize F over all possible values of $\{\Omega_{i,m}^\gamma\}$ subjected to the constraints $|\Omega_{i,m}^\gamma| \leqslant \Omega_{\max}$ for a given time T. The numerical speed limit T_F is then defined as the minimum T that can achieve $F > 1 - \epsilon$, where ϵ is the infidelity we can tolerate (set to 1% in the following).

Since F is a highly nonlinear function of $\{\Omega_{i,m}^\gamma\}$, simple numerical optimization methods will not work well in finding the global maximum of F. Here we develop a method that combines the standard GRAPE algorithm [22] with state-of-the-art machine learning techniques. Using the backward propagation method in the widely used machine learning library PyTorch [39], we calculate the gradients of F over each pulse parameter $\Omega_{i,m}^\gamma$ automatically. We then perform a stochastic gradient descent (SGD) algorithm with the Nesterov momentum method [40] to maximize F over the pulse parameters. To avoid obtaining only a local maximum for F, we repeat each gradient descent process with 200 random seeds used for both initialization and SGD, and then select the global maximum among all repetitions. Further increasing the number of random seeds does not lead to noticeable improvement in maximizing F, showing that the optimization has converged.

To benchmark our numerical optimization method, we choose the target gate to be the CNOT gate and find the above-mentioned numerical speed limit T_F for $F = 99\%$ as a function of $\Omega_{\max}$. We set $M = 16$, which allows the calculation to be done within a few hours on a small high-performance computing (HPC) cluster, and larger M does not lead to noticeable improvements. As shown in Fig. 2, we clearly see that as $\Omega_{\max}/g$ increases, T_F approaches the

analytical speed limit $T_{\min}$, indicating that the optimization succeeded in reaching the theoretical speed limit. Importantly, the maximum single-qubit drive strength $\Omega_{\max}$ does not need to be significantly larger than the interaction strength g to get close to the analytical speed limit. For example, setting $\Omega_{\max} = 3g$ already gives us a minimum gate time of $1.24T_{\min}$ with $F > 99\%$. We also compare our method with the standard GRAPE algorithm in the widely used QUTIP software [41]. With the same number of iterations and random initializations, the GRAPE algorithm in QUTIP can closely match our optimization results for $\Omega_{\max}/g < 2.5$. However, it fails to achieve $F > 99\%$ for $\Omega_{\max} > 2.5g$, and the same happens even with double the number of iterations or random initializations. This is likely because the algorithm struggles at escaping local minima due to a larger parameter space for a larger value of $\Omega_{\max}$.

In Appendix C we show that our optimization method also works well for different interaction Hamiltonians, such as the flip-flop interaction common in superconducting qubit systems. With such interaction, the speed optimization can significantly speed up CNOT and CZ gates, since they can operate as fast as the native i SWAP gate. In contrast, most existing experiments with flip-flop interacting Hamiltonians report much slower CZ gates compared to i SWAP gates [13,14,42].

V. EXPERIMENTAL RESULTS

We now proceed to demonstrate the speed limits of the two-qubit CNOT, SWAP, and $\sqrt{\text{SWAP}}$ gates experimentally. The procedure for this is as follows. First, for each gate the total evolution time T is varied from 0 to $\gtrsim T_{\min}$ in 20 steps, and the optimized pulse sequence is obtained numerically for each value of T. Next, this pulse sequence is applied to the transmon qubits experimentally by modulating the microwave drive signals. Finally, the average gate fidelity F is measured at time T by performing a quantum process tomography (QPT) [43]. Our QPT involves applying 36 different prerotations to an initial state with both qubits in the state $|0\rangle$, applying the optimized pulse sequence for time T, and then measuring nine different Pauli operators (see Appendix D for details), resulting in 324 different experimental protocols, each of which is further repeated 500 times to ensure low statistical errors. After correcting the state preparation and measurement (SPAM) errors as well as performing a maximum likelihood estimation to ensure a completely positive and trace-preserving quantum map (see Appendix D for details), the QPT allows us to find a Pauli transfer matrix [44] for the corresponding quantum process, which can be further used to infer F (Appendix D). This process allows us to find the value of T above which we can get sufficiently high gate fidelity. Such T is the experimental speed limit for the target gate.

There are several experimental limitations in this procedure. First, as strong microwave drives can heat up the superconducting qubits and cause decoherence, we only send microwave pulses of at most $2\pi \times 6$ MHz in Rabi frequency, roughly three times the coupling strength g. But as we have shown in Fig. 2, this limitation should not prevent us from getting close to the analytical speed limit. A more noticeable limitation is that we can only generate smoothly varying pulse shapes that approximate the segmented (and thus discontinuous) pulse shapes used in the numerical optimization. As the number of segments M increases, this approximation deteriorates while the gate speed increases (and eventually converges). For our setup, we choose $M = 4$ for the experiment as a sweet spot for balancing the error and speed. We note that this limitation can be addressed by numerically optimizing smooth pulse shape functions (such as a train of Gaussian envelopes), although such optimization is more resource intensive. Finally, with $r_1, r_2 \neq 1$ experimentally, our single-qubit drives will induce a small amount of extra interaction that would in principle allow us to go above the analytical speed limit for sufficiently large $\Omega_{\max}$. Our numerical optimizer accounts for this artifact. The amount of speedup over the scenario of $r_{1,2} = 1$ varies for different target gates.

Our experimental results are shown in Fig. 3. The measured gate fidelity F (red curves) closely matches the one obtained from the numerical simulation of the experiment with no error (blue curves). The deviations between the two grow as the gate fidelity gets close to 1 for reasons we discuss in the next section. For the CNOT gate, we were able to achieve $F \approx 96.5\%$ experimentally with a gate time of $T = 93.7$ ns $\approx 1.32T_{\min}$ [Fig. 3(a)]. We emphasize that this outperforms the CNOT gate implemented using the SWIPHT protocol [45] performed on the same hardware ($F \approx 94.6\%$ for a gate time of $1.87T_{\min}$ [21]), which is protocol designed specially for our hardware. The highest fidelity we achieve is $F \approx 98.3\%$ at time $T \approx 1.84T_{\min}$.

For the SWAP and $\sqrt{\text{SWAP}}$ gates, the extra interactions caused by nonunity r_1 and r_2 values have a more noticeable effect in speeding up the gates. For the SWAP gate [Fig. 3(b)], we obtain an experimental gate fidelity of $F \approx 95.9\%$ at $T = 216$ ns $\approx 1.01T_{\min}$, where theoretically $F \approx 99.997\%$. A SWAP gate with such a short time is hard to achieve via a gate sequence using a typical universal gate set, making our method particularly useful given the importance of SWAP gates in many quantum algorithms [24]. For the $\sqrt{\text{SWAP}}$ gate [Fig. 3(c)], we obtain an experimental gate fidelity of $F \approx 97.0\%$ at $T = 126$ ns $\approx 1.18T_{\min}$, with $F \approx 99.999\%$ in theory.

For all gates, the demonstrated experimental speed limits are reasonably close to the analytical speed limits. We note that the fidelities achieved here are lower than state-of-the-art due to limitations of the hardware platform (discussed in the next section) and not due to the optimal control algorithm. Even without optimal control, the fidelities obtained on this setup are close to or lower than what we are getting here [21].

VI. ERROR ANALYSIS

We have calculated the fidelity between the experimental process and the exact time evolution operator $\mathcal{U}$ in Eq. (5) for each point in Fig. 3, which is in general >95% (see Fig. 4) As seen from Fig. 3, the experimental errors get larger at large values of T. This is possibly due to the following reasons.

First, the qubits decohere as time increases. This is evidenced by our measurement of a dark evolution (i.e., with the drive Hamiltonian H_1 turned off) process fidelity that

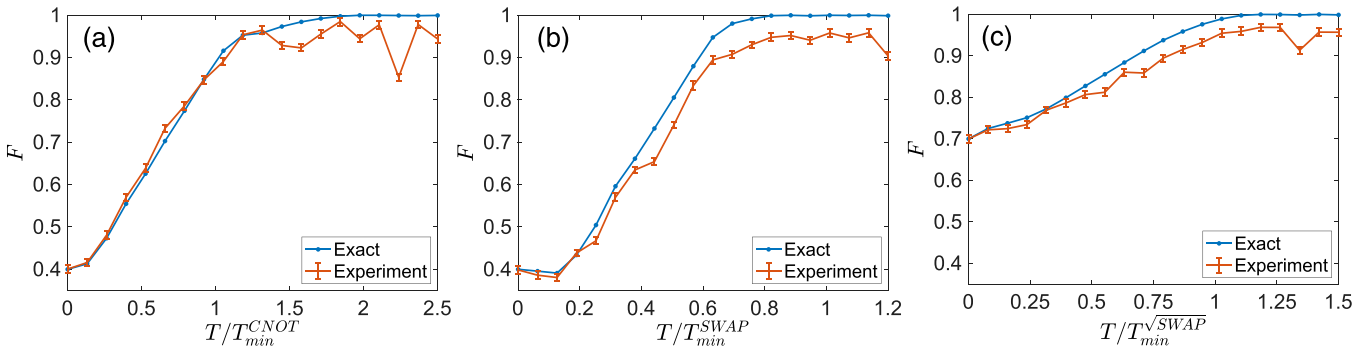

FIG. 3. Experimental measurements of the average gate fidelity F using optimized four-segment drive pulses, with the target gate being (a) CNOT, (b) SWAP, and (c) $\sqrt{\text{SWAP}}$. The red curves represent experimental measurements, while the blue curves represent the exact numerical calculation of F without considering any experimental error. $\Omega_{\max} = 6$ MHz for the CNOT gate, and $\Omega_{\max} = 5$MHz for SWAP and $\sqrt{\text{SWAP}}$ gate. The error bars represent a upper bound on the statistical error of the mean for 500 repeated measurements at each point.

drops from ≈99.3% to ≈96.3% from $T = 0$ to $T = 3\pi/(4g)$ (the theoretical minimum time for the SWAP gate), as shown in Fig. 4. This large loss in fidelity is unrelated to optimal control or errors in the control pulses, and likely results from finite T_1 time of the qubits, measurement errors, and low-fidelity (about 98% on average) single-qubit gates used in our QPT [21,33]. With better hardware designs, these errors can be largely eliminated. For example, single-qubit gates with >99.9% fidelities have already been achieved with superconducting qubits [46].

Second, when T is large enough to allow the numerically optimized F to approach 1, imperfect calibration or fluctuations on the microwave drive amplitudes or phases tend to create a larger discrepancy between the experiment and the theory, as we have discussed in detail in Appendix E. This can account for up to 0.1% loss in fidelity for ≈1% deviations in pulse shapes.

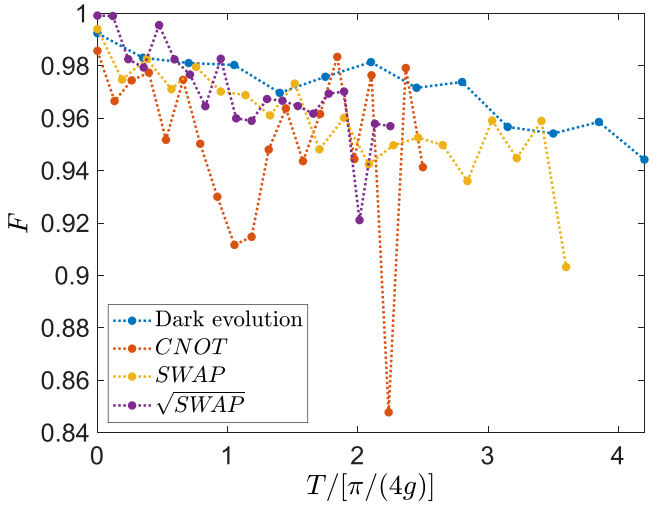

FIG. 4. Fidelity F between the experimental quantum process (characterized by the QPT) and the corresponding exact time evolution operator in Eq. (5) using the optimized pulse shapes for a given gate time T with the target gate being CNOT, SWAP, or $\sqrt{\text{SWAP}}$. The blue curve represents the fidelity between the experimental evolution without the drives (i.e., dark evolution) and the ideal evolution operator of e^{-iH_0T}.

Finally, there are also systematic errors coming from the leakage to higher excited states (in particular, the $|2\rangle$ state for each transmon), cross talk between the drives of each qubit, rotating-wave approximations, and the deviation of the experimental pulse shapes from the ideal square waves used in our numerical optimization. We can characterize these errors with the following Hamiltonian in the laboratory frame for two qutrits:

$$H(t) = \sum_m E_m |m\rangle\langle m| + \frac{1}{2}\sum_{i=1,2}\sum_{m\neq m'} (d_{m,m'}\mathcal{E}_i(t)|m\rangle\langle m'|e^{-i\omega_i t} + \text{H.c.}), \quad (7)$$

where $m \in \{00, 01, 10, 11, 02, 20, 12, 21, 22\}$ labels the energy eigenstates of the static Hamiltonian, and $d_{m,m'}$ represents the dipole moment for the transition between states $|m\rangle$ and $|m'\rangle$. $\mathcal{E}_1(t)$ and $\mathcal{E}_2(t)$ denote the electric fields of the two microwave drives we applied at frequencies $\omega_1 = E_{10} - E_{00}$ and $\omega_2 = E_{01} - E_{00}$, respectively. Their values are set by the actual experimentally applied electric fields that follow the pulse shapes from our optimization method but have finite rising/lowering edges between different pulse segments. The Hamiltonian in Eq. (7) then fully models the leakage outside the qubit subspace, the cross talk between the two drive fields, and realistic pulse shapes without rotating-wave approximations. We then calculate the fidelity between the exact evolution operator of this Hamiltonian and the one in Eq. (5) (which was used for Figs. 3 and 4). As shown in Fig. 5, these errors only add up to about 0.3% infidelity on average. And since these errors are explicitly modelled by the Hamiltonian in Eq. (7), we can further minimize their impact to gate fidelities using the same optimization method we built. However, this is beyond the scope of this work, as our experiment hardly benefits from such effort due to other error sources being dominating.

VII. CONCLUSION AND OUTLOOK

There are primarily three advancements made in this work. First, we have studied the speed limits for two-qubit gates under realistic experimental conditions and shown that these limits are close to the analytical speed limits derived

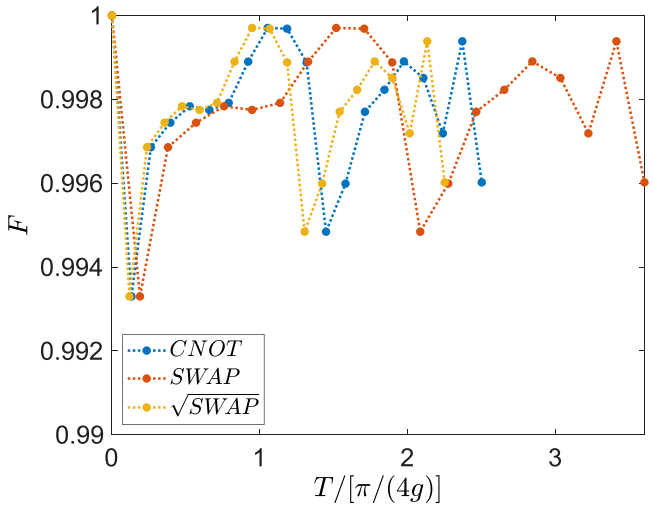

FIG. 5. Average fidelity F between the evolution operators calculated using the Hamiltonian in Eq. (7) and using Eq. (5) for the speed-optimized CNOT, SWAP, and $\sqrt{\text{SWAP}}$ gates shown in Fig. 3.

under ideal conditions. Second, we have developed an optimal control algorithm to generate a realistic pulse sequence to achieve these speed limits. Our algorithm performs better than a standard GRAPE algorithm and can be used to design speed-optimized two-qubit gates in a variety of quantum computing platforms with different types of interactions. It can offer significant speedups for non-native two-qubit gates, especially when single-qubit gate times are not negligible. Finally, we have experimentally demonstrated the quantum speed limits for various two-qubit gates using superconducting qubits.

We have also carefully characterized the error sources for our experimental gates. Most of the gate errors come from characterization/calibration errors, imperfect measurements, qubit decoherence, and low-fidelity single-qubit gates. While the strong drive pulses in the optimal control could lead to more leakage and cross-talk errors, we show that these errors only add up to about 0.3% infidelity on average, and they can be further mitigated by optimizing a more accurate Hamiltonian. It is also worth pointing out that by optimizing the speed of two-qubit gates, errors from qubit decoherence will be suppressed. We therefore expect our method to be able to improve the fidelity of the whole quantum circuit.

An important future direction is to generalize this work to a multiqubit scenario where additional qubits are used to speed up a two-qubit gate. Previous work has shown that significant scaling speedups may be obtained in performing remote quantum gates or preparing useful many-body entangled states [16,20] with long-range interacting qubits. However, questions regarding the speed limit of entangling gates when interactions are strongly long-ranged are still largely open [17]. Such interactions play important roles in quantum information scrambling [18] and the development of fully-connected quantum computers [47]. Another interesting direction is to study the speed limit of entangling gates when higher excited states outside the qubit subspace are utilized [48], where experimental and analytical results are both lacking.

ACKNOWLEDGMENTS

This work is jointly supervised by M.S. and Z.X.G. We thank NIST Boulder for hosting the experiment and the HPC center at Colorado School of Mines for providing computational resources. We acknowledge funding support from the NSF RAISE-TAQS program CCF-1839232, the NSF Triplets program DMR-1747426, the NSF NRT program DGE-2125899, and the W. M. Keck Foundation.

APPENDIX A: EXPERIMENTAL HARDWARE, CALIBRATION, AND CHARACTERIZATION

Our experimental device is operated at 10 mK in an Bluefors LD dilution refrigerator. Full schematics of the experimental setup are shown in Fig. 6. All qubit drive and readout microwave tones are delivered via the feedline, which has an output amplification chain of a Raytheon BBN Josephson parametric amplifier (JPA) preamp at base, high-electron-mobility transistor (HEMT) amplifier at 4 K, and a high-gain room-temperature amplifier.

Each experimental cycle consists of a state initialization, a time evolution under the engineered Hamiltonian flanked by process tomography rotations [43] and followed by a heterodyne state readout (see Fig. 6). The state initialization occurs by waiting 500 μs $\approx 12T_1$ between two experimental cycles, which is long enough to guarantee that each qubit is in the $|0\rangle$ state. All gates consist of microwave tones from a Holzworth HS9008B pulse shaped by a BBN arbitrary pulse sequencer (APS) quadrature modulation scheme. Readout consists of a simultaneous 2-μs probe (Agilent N5183Ms) of the two readout resonators to detect shifts in their frequencies due to their respective qubit states. The I/Q components of the readout signal shift are extracted via down conversion and a digital lock-in routine with a reference tone. They are then used to identify the two-qubit states as $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ via a classification algorithm using support vector machines. Total state preparation and measurement errors, quantified by the basis-state preparation confusion matrix [43], stayed under 5%, with the errors dominated by readout errors associated with qubit state relaxation during the measurement.

The computational subspace spectrum of the transmons was determined via a combination of spectroscopy (directly probing excitations with a 10-μs square pulse) and Ramsey experiments (driving 2 MHz off-resonant, running a typical Ramsey sequence, and noting the deviation of the fitted frequency from 2 MHz) [49]. The strength of the drive fields on the qubits was inferred from the frequency of Rabi oscillations of the excited-state population incurred by driving at uniform strength for a linearly increasing duration. The linearity of the pulse-shaping quadrature channels on the pulse sequencer was characterized by measuring the Rabi oscillation frequency resulting from a sweep over pulse amplitudes, analyzing it via Fourier filtering [33], and correcting for it at the software level.

APPENDIX B: OBTAINING ANALYTICAL SPEED LIMITS

We provide details on how to obtain the analytical speed limit T_{min} defined in Eq. (4) of the main text. Given a two-qubit unitary operator U, the key step is to find the decomposition of

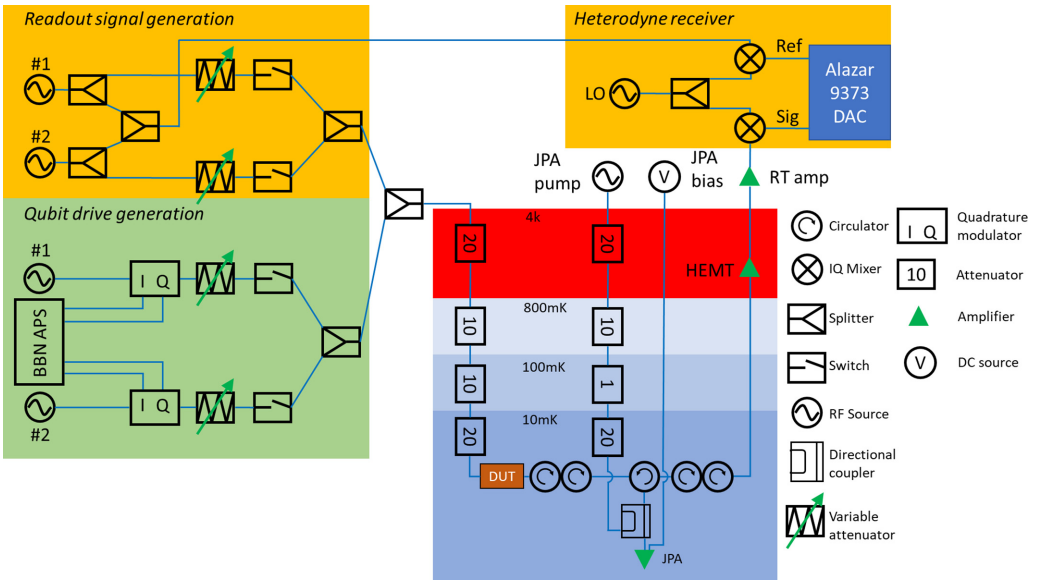

FIG. 6. Our experimental setup composed of qubit drives and heterodyne state readout. Each qubit drive (green section) is shaped via quadrature modulation by a BBN APS1 and Polyphase Microwave AM4080A. Readout (orange) consists of digitally locking in the signal passing through the device with a reference and extracting I/Q shifts used to classify the ground/excited states.

U into $(U_1 \otimes U_2)U_d(V_1 \otimes V_2)$, where $U_d = e^{-i\sum_{\gamma=x,y,z}\lambda_\gamma \sigma^\gamma \otimes \sigma^\gamma}$ with $\lambda_{x,y,z} \in [-\frac{\pi}{4}, \frac{\pi}{4}]$, and U_1, V_1 (U_2, V_2) are some single-qubit gates on the first (second) qubit. This decomposition is nontrivial, and the detailed procedure can be found in Ref. [11]. Here we provide the results of the decomposition for the three target gates we studied in Table I. The values of $\lambda_{x,y,z}$ directly lead to $T_{\min}$ values for the three target gates shown in Eq. (4) of the main text. Note that an overall phase difference is tolerated for the decomposition of U.

APPENDIX C: OPTIMIZATION FOR DIFFERENT STATIC HAMILTONIANS

To demonstrate the universality of our optimal control method, here we apply it to two different static Hamiltonians commonly seen in superconducting qubit platforms [13,25,35,50,51]: a flip-flop (also known as XY) Hamiltonian H_0^{XY} and an XXZ Hamiltonian H_0^{XXZ} that contains both flip-flop interaction and ZZ (Ising-type) interaction:

$$H_0^{\mathrm{XY}} = g\left(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y\right),$$
$$H_0^{\mathrm{XXZ}} = g\left(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y + \eta\sigma_1^z\sigma_2^z\right). \tag{C1}$$

As an example, we set the parameter $\eta = 1/2$ in the following, and similar results are expected for different η values. We now perform our gate-speed optimization described in Sec. IV with our experimental static Hamiltonian H_0 in Eq. (1) replaced by the above H_0^{XY} or H_0^{XXZ}, and the target gate being either SWAP or CNOT.

TABLE I. Detailed decompositions of theCNOT, SWAP, and $\sqrt{\mathrm{SWAP}}$ gates based on Eq. (3).

	CNOT	SWAP	$\sqrt{\mathrm{SWAP}}$
U_1	$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$
U_2	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$
V_1	$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & i\\ -1 & i\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$
V_2	$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$
λ_x	$\frac{\pi}{4}$	$\frac{\pi}{4}$	$\frac{\pi}{8}$
λ_y	0	$\frac{\pi}{4}$	$-\frac{\pi}{8}$
λ_z	0	$\frac{\pi}{4}$	$-\frac{\pi}{8}$

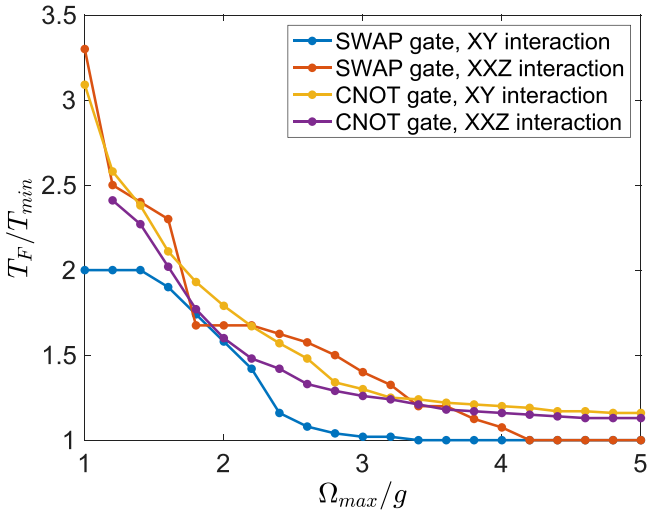

FIG. 7. Minimum gate times at different maximum drive strengths achieved by our optimization method for an XY or XXZ interacting Hamiltonian shown in Eq. (C1), with the target gate being SWAP or CNOT. Every point here has average gate fidelity $F > 99.99\%$.

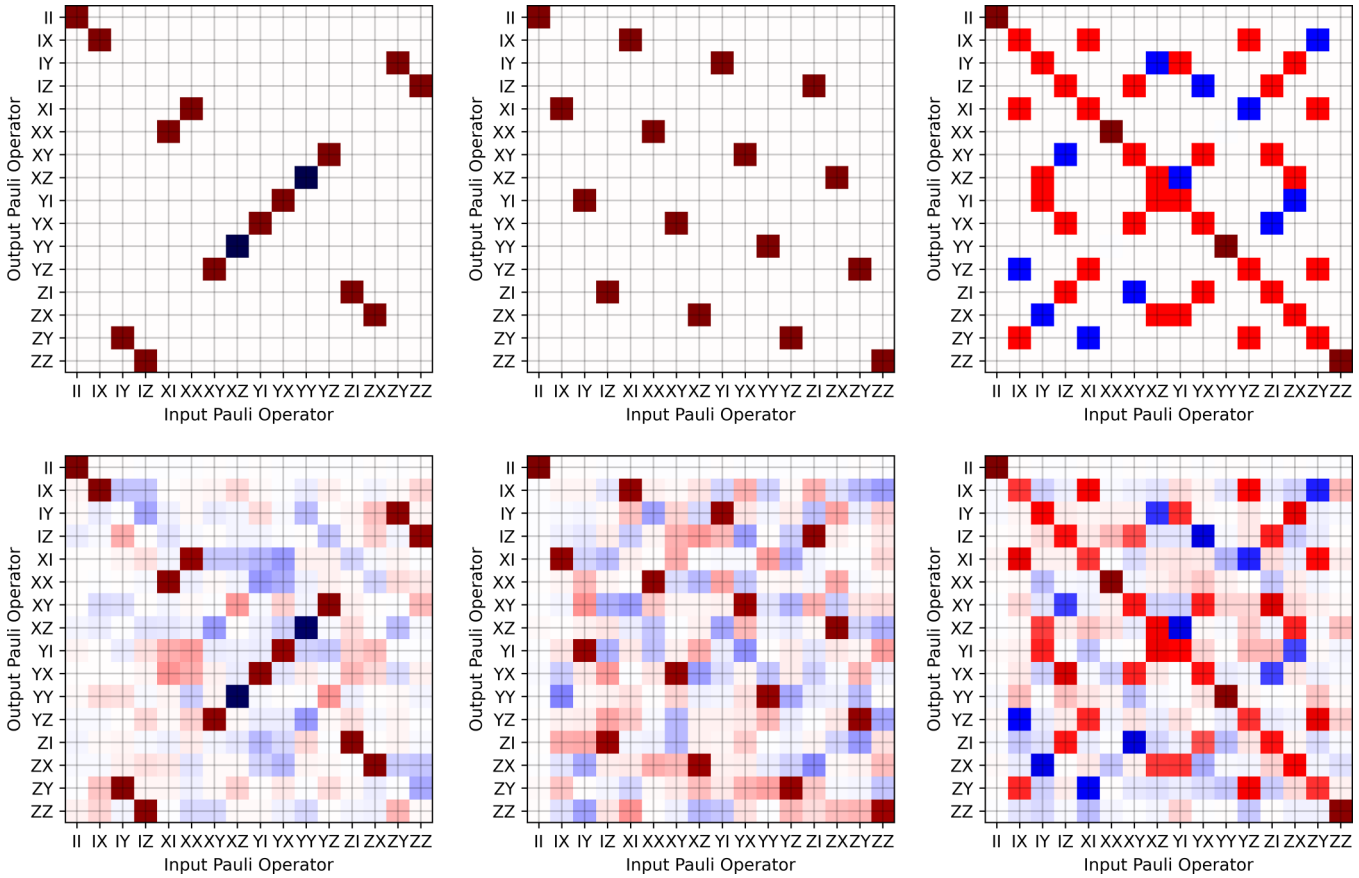

FIG. 8. Quantum process tomography based on the Pauli transfer matrix for the three target gates we performed experimentally. Top row from left to right: Pauli transfer matrices for an ideal CNOT gate, SWAP gate, and $\sqrt{\text{SWAP}}$ gate. Bottom row from left to right: Examples of SPAM-error-corrected Pauli transfer matrices obtained experimentally for the CNOT, SWAP, and $\sqrt{\text{SWAP}}$ gates, with average gate fidelities being 95.6%, 93.1%, and 95.7%, respectively.

In Fig. 7 we show the minimum gate time T (in units of the corresponding analytical speed limit T_{min}) as a function of the maximum drive strength Ω_{max} (in units of the interaction strength g), similar to Fig. 2. Note that for the CNOT gate (as well as the CZ gate), $T_{\text{min}} = \pi/(4g)$ holds for H_0, H_0^{XY}, and H_0^{XXZ}. But for the SWAP gate, $T_{\text{min}} = 3\pi/(8g)$ for H_0^{XY} and $T_{\text{min}} = 3\pi/(10g)$ for H_0^{XXZ}. In other words, the XY or XXZ interaction can generate a SWAP gate faster than the Ising interaction used in our experiment at the same strength.

We have also set a higher fidelity threshold of $F > 99.99\%$ to better show the potentials of our optimization method in Fig. 7, while the number of pulse segments, random seeds, and iterations remain identical to those in Fig. 2 (for our experiment such a high-fidelity threshold is unnecessary due to experimental error sources). We find that for both H_0^{XY} and H_0^{XXZ}, one can reach $T \approx T_{\text{min}}$ with $\Omega_{\text{max}}/g < 4$. For the CNOT gate, one needs larger drive strengths to reach the theoretical speed limit, but at moderate drive strengths ($\Omega_{\text{max}} \approx 3g$ as in our experiment), one can still get close to the theoretical speed limit ($T \approx 1.2T_{\text{min}}$).

APPENDIX D: PROCESS TOMOGRAPHY AND SPAM ERROR CORRECTION

We perform a two-qubit quantum process tomography (QPT) via a standard protocol [43,44] based on the measurement of a Pauli transfer matrix $\mathcal{R}$ defined as

$$\mathcal{R}_{i,j} \equiv \text{Tr}[P_i \mathcal{E}(P_j)], \quad P_i \in \{I, X, Y, Z\} \otimes \{I, X, Y, Z\}, \quad \text{(D1)}$$

where $\mathcal{E}(\cdot)$ denotes a quantum map of the process being measured.

For an ideal two-qubit gate U, we can calculate the 16×16 Pauli transfer matrix $\mathcal{R}$ numerically using Eq. (D1) (see Fig. 8 for examples). To measure $\mathcal{R}$ experimentally, we need to perform single-qubit rotations before and after the quantum process [52]. Here we apply nine postrotations $R_k \in \{I, X_{-\pi/2}, Y_{\pi/2}\}^{\otimes 2}$ and 36 prerotations $R_l \in \{Y_{\pi/2}, Y_{-\pi/2}, X_{-\pi/2}, X_{\pi/2}, I, X_\pi\}^{\otimes 2}$ [44], for a total of 324 different sequences, each of which is repeated 500 times to suppress statistical noise in the measurement. The single-qubit rotations here are achieved via the "two-axis gate" protocol [33], and they are fined tuned to ensure single-qubit gate fidelities up to 99.1% [21].

Each single experiment returns a measurement outcome as one of the four two-qubit basis states $|j\rangle \in \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. We then group the outcomes of all 162 000 experiments using a tensor n_{jkl} which counts the number of measurement outcome states $|j\rangle$ for the kth postrotation and lth prerotation. To take into account possible measurement errors, we separately measure a 4×4 confusion matrix $\mathcal{P}$, where $\mathcal{P}_{i,j}$ denotes the probability of obtaining a

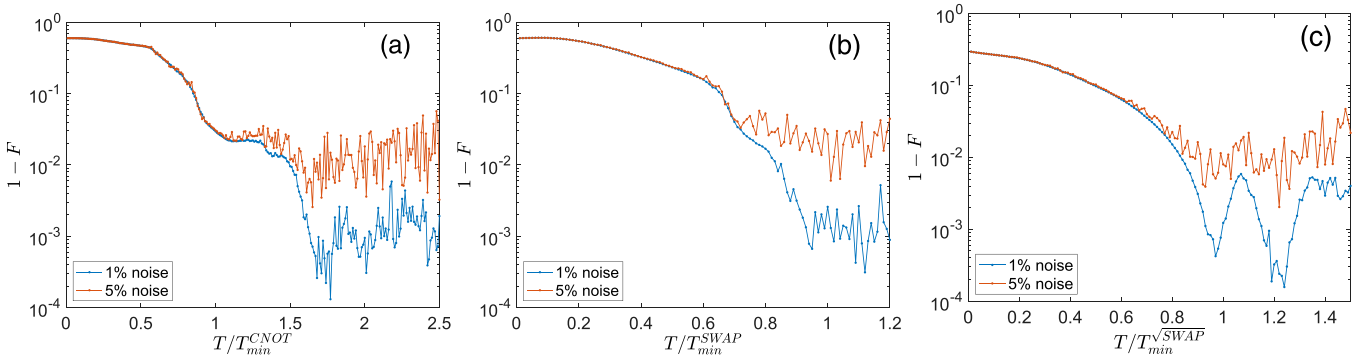

FIG. 9. Average gate infidelity $1-F$ calculated using the optimized pulse shapes in Fig. 3 of the main text but with random Gaussian noise added to each pulse shape parameter $\Omega_{i,m}^{\gamma}$ (see main text) with the target gate being CNOT (a), SWAP (b), and $\sqrt{\text{SWAP}}$ (c). The blue (red) curves correspond to a standard deviation of the Gaussian noise at $0.01\Omega_{\max}$ ($0.05\Omega_{\max}$).

measurement outcome as the state $|i\rangle$ if we initialize the two qubits in state $|j\rangle$.

To infer the Pauli transfer matrix $\mathcal{R}$ from the measurement outcome tensor n_{jkl}, we further invoke a maximum likelihood estimation (MLE) method with the following log-likelihood function of $\mathcal{R}$ [43]:

$$\log L(\mathcal{R}) = \sum_{j,k,l} n_{jkl} \log\left(\sum_{m,n=0}^{15} B_{jklmn}\mathcal{R}_{mn}\right), \tag{D2}$$

$$B_{jklmn} = \sum_{m',n'=0}^{15}\sum_{j'=0}^{3} \times \mathcal{P}_{jj'}\langle j'|P_{m'}|j'\rangle \mathcal{R}_k)_{m'm}(\mathcal{R}_l)_{nn'}\text{Tr}(P_{n'}\rho_0), \tag{D3}$$

where $\mathcal{R}_k$ and $\mathcal{R}_l$ are the Pauli transfer matrices for the above-defined postrotation unitary R_k and prerotation unitary R_l, respectively. $\rho_0 = |00\rangle\langle 00|$ is our initial state, and $P_{m'}$, $P_{n'}$ are defined in Eq. (D1). The experimental Pauli transfer matrix $\mathcal{R}$ is then obtained by maximizing $\log L(\mathcal{R})$ under the constraint that $\mathcal{R}$ represents a trace-preserving and completely positive quantum map [43,44].

Before we perform QPT for the speed-optimized two-qubit gates, we first perform the above QPT procedure for a zero-time evolution to obtain a Pauli transfer matrix $\mathcal{R}_I^{\text{exp}}$, which without state preparation and measurement (SPAM) errors should represent an identity quantum map. The measured $\mathcal{R}_I^{\text{exp}}$ can be used to correct the SPAM errors for a quantum process we perform with nonzero time evolution, whose measured Pauli transfer matrix is denoted by $\mathcal{R}_U^{\text{exp}}$ for a target unitary U.

To achieve SPAM error correction, we first obtain the process matrix χ_I^{exp} and χ_U^{exp} for the corresponding Pauli transfer matrix $\mathcal{R}_I^{\text{exp}}$ and $\mathcal{R}_U^{\text{exp}}$, respectively [by inverting Eq. (D5) below] [52]. The process matrix for a quantum map $\mathcal{E}$ is defined via $\mathcal{E}(\rho) = \sum_{m,n}\chi_{mn}P_m\rho P_n$ with P_m, P_n defined in Eq. (D1). Second, we obtain the SPAM-error-corrected process matrix $\chi_U^{\text{corrected}}$ for the target process using

$$\chi_U^{\text{corrected}} = T^{-1}\left(T\chi_U^{\text{exp}}T^{\dagger} - V\chi_I^{\text{exp}}V^{\dagger} + \chi_I^{\text{exp}}\right)(T^{\dagger})^{-1}, \tag{D4}$$

where $T_{mn} = \text{Tr}(P_mP_nU^{\dagger})/4$ and $V_{mn} = \text{Tr}(P_mP_n)/4$ [53]. Finally, we convert the error-corrected process matrix $\chi_U^{\text{corrected}}$ to the error-corrected Pauli transfer matrix $\mathcal{R}_U^{\text{corrected}}$ via

$$\mathcal{R}_{ij} = \sum_{m,n=0}^{15}\chi_{mn}\text{Tr}P_iP_mP_jP_n. \tag{D5}$$

Examples of such error-corrected Pauli transfer matrix $\mathcal{R}_U^{\text{corrected}}$ for U being the CNOT, SWAP, or $\sqrt{\text{SWAP}}$ gate are shown in Fig. 8. Finally, $\mathcal{R}_U^{\text{corrected}}$ allows us to find the average gate fidelity F to the target gate U via [38]

$$F = \frac{4\text{Tr}\left(\mathcal{R}_U^{\text{corrected}}\mathcal{R}_U^{\text{ideal}}\right)+1}{5}, \tag{D6}$$

where $\mathcal{R}_U^{\text{ideal}}$ represents the Pauli transfer matrix of the ideal target gate U.

APPENDIX E: ADDITIONAL ERROR ANALYSIS

One error source shared among all our data points is the statistical error due to quantum measurements. To quantify this error, we simulate additional measurements by adding a Gaussian-distributed random noise with zero mean and unity standard deviation on each Pauli operator measured during our QPT [44]. This allows us to set an upper bound on the statistical error of the mean that would be obtained on reperforming the full experiment with all other error sources held fixed. As shown in Fig. 3 of the main text, this statistical error on the measured F is less than 1% in all cases.

We have also numerically simulated the effects of imperfect calibration or noises on the optimized pulse shapes (either for amplitudes or phases) by adding random perturbations to each optimized pulse parameter $\Omega_{i,m}^{\gamma}$ (see main text). We expect such perturbations to be present in our experimental setup with magnitudes of a few percent of $\Omega_{\max}$. The simulated average gate fidelity F is shown in Fig. 9, where all other parameters are identical to the exact F curves in Fig. 3 of the main text. We see that our optimization method is robust to small amount (1%) of noises on the pulse shapes, where the fidelity can still exceed 99.9% for gate time close to $T_{\min}$. For larger noises (5%), the infidelity caused by errors on the drive pulses can be around $1\%-2\%$, but such large noises are rare in most experimental platforms.

[1] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[3] A. P. M. Place *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, Nat. Commun. **12**, 1779 (2021).

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[5] M. H. Devoret and R. J. Schoelkopf, Superconducting circuits for quantum information: An outlook, Science **339**, 1169 (2013).

[6] M. R. Frey, Quantum speed limits—Primer, perspectives, and potential future directions, Quant. Inf. Proc. **15**, 3919 (2016).

[7] N. Malossi, M. G. Bason, M. Viteau, E. Arimondo, D. Ciampini, R. Mannella, and O. Morsch, Quantum driving of a two level system: Quantum speed limit and superadiabatic protocols – an experimental investigation, J. Phys.: Conf. Ser. **442**, 012062 (2013).

[8] L.-M. Duan, Scaling ion trap quantum computation through fast quantum gates, Phys. Rev. Lett. **93**, 100502 (2004).

[9] V. M. Schäfer, C. J. Ballance, K. Thirumalai, L. J. Stephenson, T. G. Ballance, A. M. Steane, and D. M. Lucas, Fast quantum logic gates with trapped-ion qubits, Nature (London) **555**, 75 (2018).

[10] A. M. Steane, G. Imreh, J. P. Home, and D. Leibfried, Pulsed force sequences for fast phase-insensitive quantum gates in trapped ions, New J. Phys. **16**, 053049 (2014).

[11] B. Kraus and J. I. Cirac, Optimal creation of entanglement using a two-qubit gate, Phys. Rev. A **63**, 062309 (2001).

[12] G. Vidal, K. Hammerer, and J. I. Cirac, Interaction cost of nonlocal gates, Phys. Rev. Lett. **88**, 237902 (2002).

[13] A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, Demonstration of a high-fidelity CNOT gate for fixed-frequency transmons with engineered zz suppression, Phys. Rev. Lett. **127**, 130501 (2021).

[14] I. N. Moskalenko, I. A. Simakov, N. N. Abramov, A. A. Grigorev, D. O. Moskalev, A. A. Pishchimova, N. S. Smirnov, E. V. Zikiy, I. A. Rodionov, and I. S. Besedin, High fidelity two-qubit gates on fluxoniums using a tunable coupler, npj Quantum Inf. **8**, 130 (2022).

[15] J. Stehlik, D. M. Zajac, D. L. Underwood, T. Phung, J. Blair, S. Carnevale, D. Klaus, G. A. Keefe, A. Carniol, M. Kumph, M. Steffen, and O. E. Dial, Tunable coupling architecture for fixed-frequency transmon superconducting qubits, Phys. Rev. Lett. **127**, 080505 (2021).

[16] Z. Eldredge, Z.-X. Gong, J. T. Young, A. H. Moosavian, M. Foss-Feig, and A. V. Gorshkov, Fast quantum state transfer and entanglement renormalization using long-range interactions, Phys. Rev. Lett. **119**, 170503 (2017).

[17] A. Y. Guo, M. C. Tran, A. M. Childs, A. V. Gorshkov, and Z.-X. Gong, Signaling and scrambling with strongly long-range interactions, Phys. Rev. A **102**, 010401(R) (2020).

[18] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, Towards the fast scrambling conjecture, J. High Energy Phys. 04 (2013) 022.

[19] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Commun. Math. Phys. **28**, 251 (1972).

[20] M. C. Tran, C.-F. Chen, A. Ehrenberg, A. Y. Guo, A. Deshpande, Y. Hong, Z.-X. Gong, A. V. Gorshkov, and A. Lucas, Hierarchy of linear light cones with long-range interactions, Phys. Rev. X **10**, 031009 (2020).

[21] J. Long, T. Zhao, M. Bal, R. Zhao, G. S. Barron, H.-s. Ku, J. A. Howard, X. Wu, C. R. H. McRae, X.-H. Deng, G. J. Ribeill, M. Singh, T. A. Ohki, E. Barnes, S. E. Economou, and D. P. Pappas, A universal quantum gate set for transmon qubits with strong zz interactions, arXiv:2103.12305.

[22] N. Khaneja *et al.*, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, J. Magn. Reson. **172**, 296 (2005).

[23] J. Werschnik and E. K. U. Gross, Quantum optimal control theory, J. Phys. B **40**, R175 (2007).

[24] IBM, The open science prize: Solve for swap gates and graph states, 2021, https://research.ibm.com/blog/open-science-prize.

[25] D. M. Abrams, N. Didier, B. R. Johnson, M. P. da Silva, and C. A. Ryan, Implementation of XY entangling gates with a single calibrated pulse, Nat. Electron. **3**, 744 (2020).

[26] 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).

[27] 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).

[28] F. Vatan and C. Williams, Optimal quantum circuits for general two-qubit gates, Phys. Rev. A **69**, 032315 (2004).

[29] Y.-X. Xi Liu, L. F. Wei, J. S. Tsai, and F. Nori, Controllable coupling between flux qubits, Phys. Rev. Lett. **96**, 067003 (2006).

[30] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell *et al.*, Demonstration of quantum volume 64 on a superconducting quantum computing system, Quantum Sci. Technol. **6**, 025020 (2021).

[31] K. O'Brien, V. Ramasesh, A. Dove, J. M. Kreikebaum, J. Colless, and I. Siddiqi, Design, characterization, and multiplexed single-shot readout of a multi-qubit circuit for quantum simulation, in *Quantum Information and Measurement (QIM) 2017, Paris France* (Optica Publishing Group, 2017), p. QF6A.4.

[32] X. Wu, J. L. Long, H. S. Ku, R. E. Lake, M. Bal, and D. P. Pappas, Overlap junctions for high coherence superconducting qubits, Appl. Phys. Lett. **111**, 032602 (2017).

[33] J. Long, Superconducting quantum circuits for quantum information processing, Ph.D. thesis, University of Colorado, Boulder, 2020.

[34] P. Mundada, G. Zhang, T. Hazard, and A. Houck, Suppression of qubit crosstalk in a tunable coupling superconducting circuit, Phys. Rev. Appl. **12**, 054023 (2019).

[35] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Tunable coupling

scheme for implementing high-fidelity two-qubit gates, Phys. Rev. Appl. **10**, 054062 (2018).

[36] G. S. Barron, F. A. Calderon-Vargas, J. Long, D. P. Pappas, and S. E. Economou, Microwave-based arbitrary CPHASE gates for transmon qubits, Phys. Rev. B **101**, 054508 (2020).

[37] M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Lazar, J.-C. Besse, T. Walter, A. Wallraff, and C. Eichler, Implementation of conditional phase gates based on tunable zz interactions, Phys. Rev. Lett. **125**, 240502 (2020).

[38] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).

[39] A. Paszke *et al.*, *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (Curran Associates Inc., Red Hook, NY, 2019).

[40] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, In Dokl. Akad. Nauk SSSR **269**, 543 (1983).

[41] J. Johansson, P. Nation, and F. Nori, QUTIP: An open-source python framework for the dynamics of open quantum systems, Comput. Phys. Commun. **183**, 1760 (2012).

[42] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim, A. Melville, B. M. Niedzielski, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Realization of high-fidelity CZ and zz-free ISWAP gates with a tunable coupler, Phys. Rev. X **11**, 021058 (2021).

[43] Grove, Histogram based tomography, 2016, https://grove-docs.readthedocs.io/en/latest/tomography.html.

[44] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits, Phys. Rev. Lett. **109**, 060501 (2012).

[45] S. E. Economou and E. Barnes, Analytical approach to swift nonleaky entangling gates in superconducting qubits, Phys. Rev. B **91**, 161405(R) (2015).

[46] R. Manenti, E. A. Sete, A. Q. Chen, S. Kulshreshtha, J.-H. Yeh, F. Oruc, A. Bestwick, M. Field, K. Jackson, and S. Poletto, Full control of superconducting qubits with combined on-chip microwave and flux lines, Appl. Phys. Lett. **119**, 144001 (2021).

[47] 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, Proc. Natl. Acad. Sci. **114**, 3305 (2017).

[48] S. Ashhab, F. Yoshihara, T. Fuse, N. Yamamoto, A. Lupascu, and K. Semba, Speed limits for two-qubit gates with weakly anharmonic qubits, Phys. Rev. A **105**, 042614 (2022).

[49] N. F. Ramsey, A molecular beam resonance method with separated oscillating fields, Phys. Rev. **78**, 695 (1950).

[50] S. A. Caldwell *et al.*, Parametrically activated entangling gates using transmon qubits, Phys. Rev. Appl. **10**, 034050 (2018).

[51] X. Li, Y. Ma, J. Han, T. Chen, Y. Xu, W. Cai, H. Wang, Y.-P. Song, Z.-Y. Xue, Z.-Q. Yin, and L. Sun, Perfect quantum state transfer in a superconducting qubit chain with parametrically tunable couplings, Phys. Rev. Appl. **10**, 054009 (2018).

[52] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921.

[53] X. Y. Han, T. Q. Cai, X. G. Li, Y. K. Wu, Y. W. Ma, Y. L. Ma, J. H. Wang, H. Y. Zhang, Y. P. Song, and L. M. Duan, Error analysis in suppression of unwanted qubit interactions for a parametric gate in a tunable superconducting circuit, Phys. Rev. A **102**, 022619 (2020).