Article

Calibrating quantum gates up to 52 qubits in a superconducting processor

Check for updates

Daojin Fan[1,2,3,7], Guoding Liu [4,7], Shaowei Li[1,2,3,7], Ming Gong [1,2,3,5], Dachao Wu[1,2,3], Yiming Zhang[1,2,3], Chen Zha[1,2,3], Fusheng Chen[1,2,3], Sirui Cao[1,2,3], Yangsen Ye[1,2,3], Qingling Zhu[1,2,3], Chong Ying[1,2,3], Shaojun Guo[1,2,3], Haoran Qian[1,2,3], Yulin Wu[1,2,3], Hui Deng[1,2,3,5], Gang Wu[1,2,3,5,6], Cheng-Zhi Peng [1,2,3,5], Xiongfeng Ma [4] ✉, Xiaobo Zhu [1,2,3,5] ✉ & Jian-Wei Pan [1,2,3,5] ✉

Benchmarking large-scale quantum gates, typically involving multiple native two-qubit and single-qubit gates, is crucial in quantum computing. Global fidelity, encompassing information about inter-gate correlations, offers a comprehensive metric for evaluating and optimizing gate performance, unlike the fidelities of individual local native gates. In this work, utilizing the character-average benchmarking protocol implementable in a shallow circuit, we successfully benchmark gate fidelities up to 52 qubits. Notably, we achieved a fidelity of 63.09% ± 0.23% for a 44-qubit parallel CZ gate. Utilizing the global fidelity of the parallel CZ gate, we explore the correlations among local CZ gates by introducing an inter-gate correlation metric, enabling one to simultaneously quantify crosstalk error when benchmarking gate fidelity. Finally, we apply our methods in gate optimization. By leveraging global fidelity for optimization, we enhance the fidelity of a 6-qubit parallel CZ gate from 87.65% to 92.04% and decrease the gate correlation from 3.53% to 3.22%, compared to local gate fidelity-based optimization. The experimental results align well with our established composite noise model, incorporating depolarizing and *ZZ*-coupling noises, and provide valuable insight into further study and mitigation of correlated noise.

Realistic quantum computers suffer from noise, hindering themselves from demonstrating an advantage over their classical counterpart[1–5]. Quantum error correction is the key to reducing noise[6–10], yet its implementation hinges on the availability of high-fidelity quantum gates and weak correlations among different gate operations[11–13]. Thus, diagnosing the noise level and assessing the interaction within the quantum circuits is essential to improving their performance and, hence, realizing fault-tolerant quantum computing. Within a step of quantum circuits, single-qubit and two-qubit gates are parallelly implemented to reduce the circuit depth. Due to the potential crosstalk and residual coupling between the qubits[14], the local gate fidelities may be insufficient to give an overall evaluation of the global gate. In contrast, global gate fidelity contains information about correlations among different local gates, providing an overall performance metric.

As quantum hardware advances and platforms grow, large-scale quantum gate calibration becomes increasingly crucial. Researchers have made great efforts to develop practical benchmarking methods[15–17] and enlarge the gate calibration size. Currently, the most frequently used method to evaluate the quantum gate fidelity is randomized benchmarking (RB)[18–22], which enjoys the advantage of low complexity and robustness to state preparation and measurement errors. Nonetheless, the original RB protocol[20] is only realized up to three qubits[23] owing to the compiling problem of global Clifford gates. Several RB variants[24–26] were proposed circumventing compiling issues and significantly advanced the benchmarking size to 10 qubits for an individual gate via cycle benchmarking[25] and 27 qubits for a gate set via mirror RB[26].

Nonetheless, a gap exists between these achievements and the scale of state-of-the-art quantum computers. Currently, quantum computers have been realized with tens of or even hundreds of qubits in superconducting[27–29], ion trap[30,31], and neutral atom systems[32]. Calibrating larger gates is demanding for assessing and enhancing the performance of

[1]Hefei National Research Center for Physical Sciences at the Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei, China. [2]Shanghai Research Center for Quantum Science and CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai, China. [3]Hefei National Laboratory, University of Science and Technology of China, Hefei, China. [4]Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China. [5]Jinan Institute of Quantum Technology and Hefei National Laboratory Jinan Branch, Jinan, China. [6]University of Science and Technology of China, Shanghai Research Institute, Shanghai, China. [7]These authors contributed equally: Daojin Fan, Guoding Liu, Shaowei Li. ✉e-mail: xma@tsinghua.edu.cn; xbzhu16@ustc.edu.cn; pan@ustc.edu.cn

current quantum computers. Meanwhile, a noticeable challenge to achieving this goal is the low fidelity often associated with large-scale gates. Many RB protocols require repeated execution of the target gate. However, noise can prevent the gate from being repeated without significant signal loss, which can render the protocol ineffective. Resolving this issue requires enhancing benchmarking protocols to ensure their effectiveness in short-depth circuits and achieving the realization of higher-performance quantum gates.

In this work, to achieve large-scale gate calibration, we utilize the character-average benchmarking (CAB)[33] protocol. This method can evaluate the fidelity robust to state preparation and measurement errors for an individual Clifford gate up to a local unitary transformation and be scalable with respect to the system size. Compared to cycle benchmarking[25] that also aims to evaluate the individual gate fidelity, CAB requires shorter circuit depth and can tolerate higher gate errors and benchmarking larger gates. Additionally, CAB exhibits smaller statistical fluctuations than cycle benchmarking, with more details shown in the Supplemental Material (See Supplementary Material for more detailed experimental data of the comparison between character-average benchmarking and cycle benchmarking, the fluctuation analysis of the experimental results, and additional benchmarking and optimization results, which includes references[25,33]). Meanwhile, in experiments, we improve gate fidelity and reduce a significant portion of gate control errors via qubit frequency adjustment and pre-calibration of gate parameters, as elaborated in Methods. The high fidelity and nearly depolarizing noise of the quantum gates are crucial to our success in realizing large-scale CAB experiments.

Here, we consider two important types of gates for experimental benchmarking. One is a fully connected gate composed of two layers of CZ gates and two layers of local Clifford gates. This gate is a part of the brickwise architecture circuit with an efficient realization scheme on current devices and hence is favored in variational quantum algorithms[34]. It also plays an essential role in other quantum information processing tasks like simulating a nearest-neighbor interacting Hamiltonian evolution. We benchmark such gates up to 46 qubits and get a fidelity of 17.42% ± 0.45% dressed with local twirling gates.

We also benchmark the parallel CZ gate, composed of parallelly implemented local CZ gates, which excels in generating entanglement across multiple parties and is essential in preparing graph states and executing numerous quantum algorithms[35,36]. We benchmark such gates from 4 to 44 qubits. The average fidelity of a single local CZ gate is about 98% and does not decrease with the qubit number increase.

With the global fidelity of the parallel CZ gate, we characterize the correlation among the local CZ gates. This procedure can be done simultaneously when benchmarking global fidelity without extra experiments. The correlation data provides the interaction information within the circuit and helps to evaluate and optimize the gate performance. The correlation of the 44-qubit parallel CZ gate turns out to be weak and constantly positive. In contrast, the results of the 52-qubit parallel CZ gate present some negative values. The experimental results are well explained by our established composite noise model, which incorporates local depolarizing and ZZ-coupling noises. The correlation magnitude positively depends on the coupling strength. Interestingly, the correlation sign between two gates varies in two cases: staying positive in two-gate coupling yet turning negative in three-gate coupling when one gate strongly couples with a third one. The correlation results help identify coupling gates and advance the study of correlated noise.

Since an important application of fidelity benchmarking is gate optimization, we demonstrate optimization experiments for parallel CZ gates and compare outcomes when setting the target function as the parallel CZ gate fidelity and the individual local CZ gate fidelities. We observe better results of the former. This result validates the effectiveness of CAB and suggests that global fidelity is more effective in optimizing quantum circuit performance, originating from containing more correlation information. The optimization result is consistent with the ZZ-coupling noise model. When three gates couple, improving the fidelity of one local CZ gate may

reduce the fidelities of others. This antagonistic relationship demonstrates the limitations of local fidelities.

Results
Preliminary
Let us start with briefly revisiting the concept of gate fidelity. Any noisy quantum gate $\widetilde{U}$ can be treated as a composite of its noise channel Λ, and the ideal gate U, expressed as $\widetilde{U} = U \circ \Lambda$. In this work, the fidelity of gate U refers to the process fidelity of Λ, defined by,

$$F(\Lambda) = 2^{-3n} \sum_{P \in \mathsf{P}_n} \mathrm{tr}(P\Lambda(P)), \tag{1}$$

where n is the number of qubits and $\mathsf{P}_n = \{\mathbb{I}, X, Y, Z\}^{\otimes n}$ represents the n-qubit Pauli group. This process fidelity is a linear function of the average fidelity, a metric commonly employed in quantum gate benchmarking studies. The summation in Eq. (1), initially spanning 4^n terms, can be effectively restructured into 2^n terms as follows:

$$F(\Lambda) = 2^{-2n} \sum_{w \in \{0,1\}^{\otimes n}} 3^{|w|} \lambda_w;$$
$$\lambda_w = \frac{2^{-n} \sum_{\mathrm{pt}(P)=w} \mathrm{tr}(P\Lambda(P))}{3^{|w|}}, \tag{2}$$

where $\mathrm{pt}(P)$ denotes a bitstring that takes 0 on bit i if P acts as the identity $\mathbb{I}$ on qubit i and takes 1 if P acts non-trivially, and $|w|$ signifies the weight of the bitstring w. The quantity λ_w is termed the weighted quality parameter of Λ with weight $2^{-2n}3^{|w|}$, and below, we concisely refer to it as the quality parameter of Λ.

The core of CAB lies in assessing the quality parameter of the channel $\sqrt{U^{-1}\Lambda'_U U \Lambda_U}$ through the circuit depicted in Fig. 1(a), with Λ_U and Λ'_U being the Pauli-twirled noise channels of U and U^{-1}, respectively. This approach yields the CAB fidelity of U, closely approximating U's fidelity under physically reasonable conditions[33]. Hereafter, we refer to the CAB fidelity simply as the gate fidelity unless stated otherwise. While estimating all quality parameters for fidelity evaluation demands exponential resources, the number of quality parameters necessary for accurate fidelity estimation within acceptable error margins and confidence levels is independent of the qubit count, thereby ensuring the scalability of the protocol. The whole procedure is shown in Fig. 1 and elaborated in Methods. Note that the fidelity evaluated by this procedure is associated with the noise of U and that of the local twirling gates adjacent to U and U^{-1}, referred to as dressed fidelity. To isolate U's fidelity, one can employ interleaved RB techniques[22], comparing the dressed fidelity against the local twirling gate fidelity. Particularly, the fidelity of U is derived by ref.[22]

$$F = \frac{4^n F_{\mathrm{dress}} - 1}{4^n F_{\mathrm{twirl}} - 1}\left(1 - \frac{1}{4^n}\right) + \frac{1}{4^n}, \tag{3}$$

with F_{dress} and F_{twirl} the dressed and twirling gate fidelities, respectively. The local twirling gate fidelity itself is determined through CAB, with the identity operation as the target gate.

Fully connected gate and parallel CZ gate benchmarking
Our CAB experiments for the fully connected gate and the parallel CZ gate are conducted on a 54-qubit superconducting quantum computer. On a two-qubit system, the CZ gate is $\mathrm{CZ} = |0\rangle\langle0| \otimes \mathbb{I} + |1\rangle\langle1| \otimes Z$, and its parallel extension across multiple qubits is defined by $U = \bigotimes_{k=1}^{r} \mathrm{CZ}^{(i_k j_k)}$, where $\mathrm{CZ}^{(i_k j_k)}$ denotes the CZ gate acting on a specific qubit pair, (i_k, j_k), making U a $2r$-qubit gate. The fully connected gate comprises two layers of different parallel CZ gates intertwined with two layers of single-qubit gates, as shown in Fig. 2(a). One of the parallel CZ gates connects qubits 1 and 2, 3 and 4, ... and the other connects qubits 2 and 3, 4 and 5...
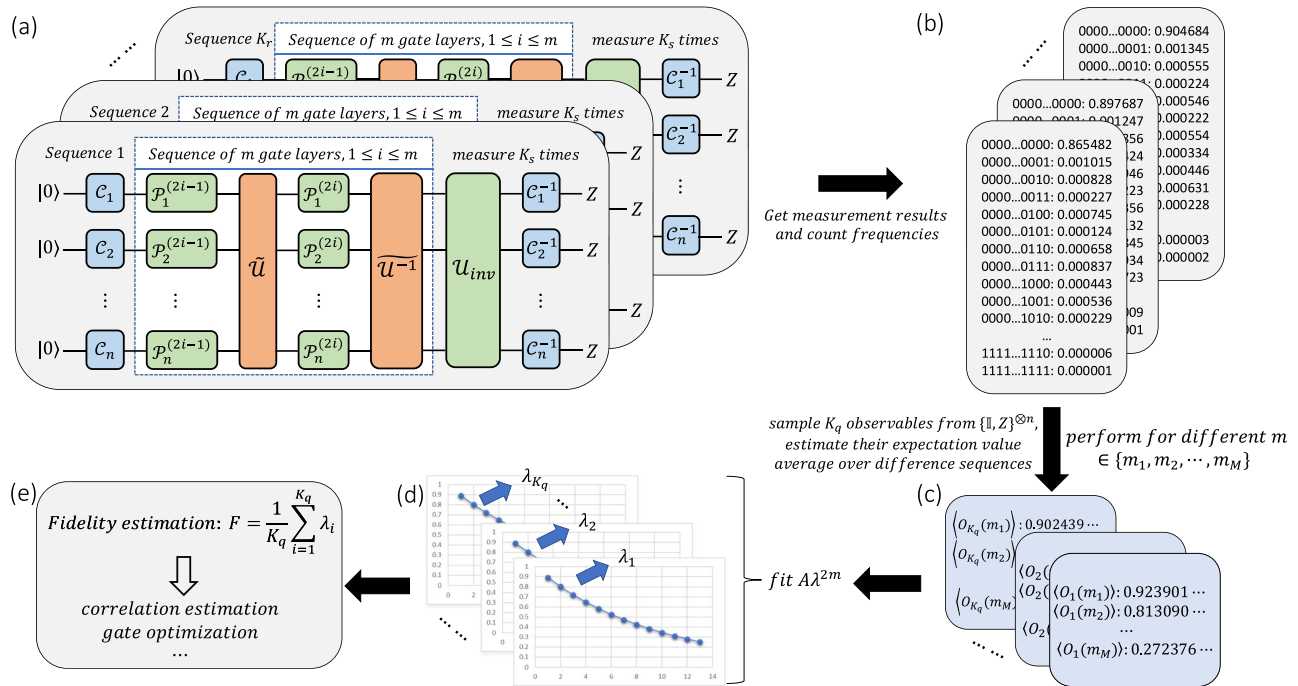
Fig. 1 | Illustration of the CAB procedure for assessing the fidelity of an n-qubit gate, U. a shows the circuit, beginning with the preparation of the state $|0\rangle^{\otimes n}$, followed by a random local Clifford gate $\bigotimes_{i=1}^{n} C_i$. Subsequently, $2m$ layers of random Pauli gates $\bigotimes_{j=1}^{n} P_j^{(i)}$ are interleaved with alternate sequences of U and U^{-1}. The inverse gate $U_{inv} = (\Pi_{i=1}^{m}(U^{-1}\bigotimes_{j=1}^{n} P_j^{(2i)} U\bigotimes_{j=1}^{n} P_j^{(2i-1)}))^{-1}$ and the inverse of the local Clifford gate $\bigotimes_{i=1}^{n} C_i^{-1}$ are applied thereafter. Finally, one applies computational-basis measurements and records the outcome. One needs to sample K_r random sequences, and for each sequence, one measures K_s times. The outcome statistics are counted for each sequence, as in (**b**). After that, one randomly chooses K_q observables $Z_w \in \{\mathbb{I}, Z\}^{\otimes n}$, where $w \in \{0,1\}^n$ and $pt(Z_w) = w$, with probability

$2^{-2n}3|w|$. One estimates the expectation values of these chosen observables, and the expectation values need to be averaged across K_r random sequences. The above procedure is repeated for different m from a circuit depth set, $\{m_1, m_2, \cdots, m_M\}$. For all m, the chosen observables have to be the same. **c** demonstrates the expectation values of different observables for different sequence lengths. The expectation value $O_w(m)$ is approximately proportional to λ_w^{2m}, which can be fit to $A\lambda^{2m}$ to determine quality parameter λ_w like (**d**). **e** shows the last step. The final fidelity estimation is the average of the fitting values, and subsequently, one can further evaluate the gate correlation and employ gate optimization. Practical experimental settings choose constant values for K_r, K_s, and K_q independent of the qubit count.
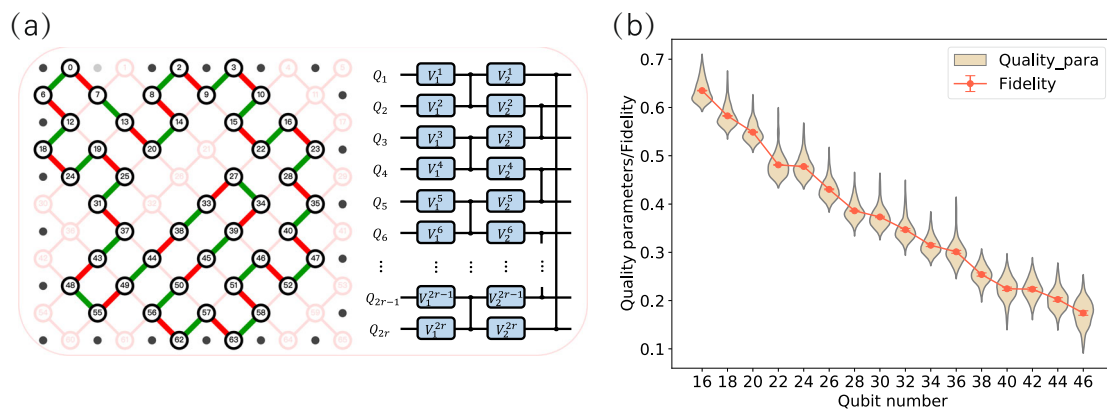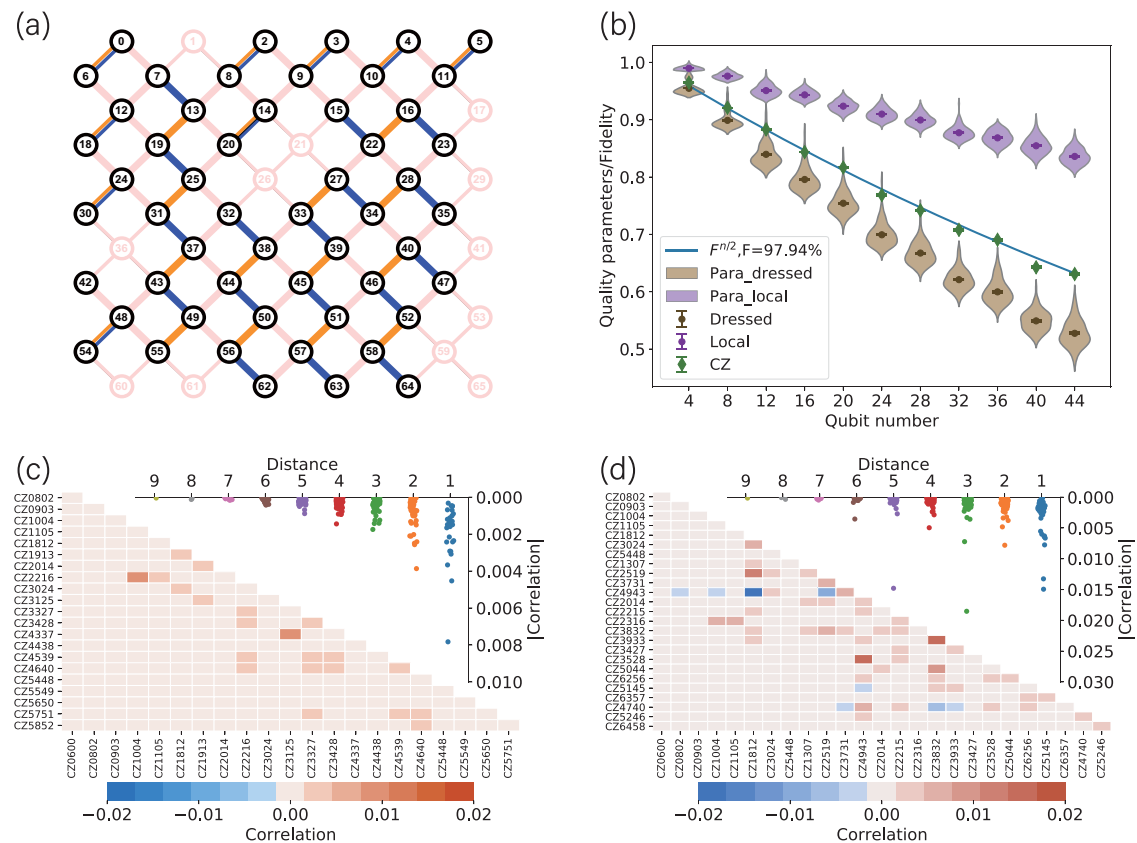

Fig. 2 | The topological diagram and corresponding fidelity of a fully connected gate. a The left figure demonstrates the qubits and CZ gates used to realize the fully connected gate with 46 qubits. We select a ring on the two-dimensional quantum processor, corresponding to a one-dimensional quantum system. The ring contains two patterns of parallel CZ gates, shown in red and green, respectively. The right figure shows the circuit structure for the fully connected quantum gate on $n = 2r$ qubits, where Q_i represents the i-th qubit. The gate

comprises two layers of single-qubit gates, $\bigotimes_{i=1}^{n} V_1^i$ and $\bigotimes_{i=1}^{n} V_2^i$, and two layers of different parallel CZ gates. **b** The quality parameters and the fidelities of the dressed gates, including the error from the local twirling gates and the target gates. The distributions of the noise channel quality parameters are shown with violin plots. Their mean values equal the gate fidelities, shown with a point with an error bar. The length of the error bar equals the standard error of the fidelity estimation.

Fully connected gate benchmarking. The single-qubit gates in the fully connected gate, like V_1^i and V_2^i in Fig. 2(a), are free to vary. In quantum information tasks like variational quantum algorithms, they are normally chosen according to specific problems. In our experiments, we randomly sample all the single-qubit gates from the Clifford group so that the global gate is a Clifford gate and can be benchmarked with CAB. It is worth

mentioning that cycle benchmarking requires repeatedly implementing the target gate U proportional to the gate order, defined as the smallest positive integer p to make $U^p = \mathbb{I}$. The fully connected gate generally has a rapidly increasing order with respect to the qubit number, and the order has already been thousands on average for 16 qubits, which we show in the Supplemental Material (See Supplementary Material for more

Fig. 3 | The topological diagram and corresponding experimental results of parallel CZ gate benchmarking. a Two parallel CZ gate patterns, colored orange (22 CZ gates) and blue (26 CZ gates) on a 54-qubit quantum processor, with available qubits shown in black. **b** Benchmarking fidelities for the parallel CZ gate within the orange pattern. The benchmarking is done progressively from 2 to 22 CZ gates. Brown and purple violin plots illustrate the noise channel quality parameters for the dressed and local twirling gates, respectively. The point with an error bar in each violin plot represents the mean value equal to the gate fidelity, and the length of the error bar equals the standard error of the fidelity estimation. Green diamond-shaped points denote individual fidelities of target gates, fitting well with the theoretical curve $F^{n/2}$, where n is the qubit number and $F = 97.94\%$. **c** A heatmap displays the pairwise correlation among CZ gates within the orange pattern, revealing weak positive correlations among several neighboring gates. We label the indexes of the CZ gates on the two axes. CZAB means the CZ gate between qubits A and B. For instance, CZ0802 is the CZ gate between qubits 2 and 8. The scatter diagram shows that the absolute value of the correlation decreases with the distance of the CZ gates and is mainly large between nearby CZ gates. The CZ gate distance is measured as the minimal line count connecting qubit pairs in the processor layout from (**a**). **d** The heatmap and scatter plot for the blue pattern, with the same color scale and interpretation as in (**c**), indicate more significant correlations compared to the orange pattern and show that large correlations also exist for two remote CZ gates.

detailed experimental data of the comparison between character-average benchmarking and cycle benchmarking, the fluctuation analysis of the experimental results, and additional benchmarking and optimization results, which includes references[25,33]). The long-depth circuit will lead to extremely noisy experimental results and cannot provide effective benchmarking. Instead, by using the target gate and its inverse, this gate can be benchmarked within a short-depth circuit using CAB.

We realized the fully connected gate with qubit numbers from 16 to 46 and benchmarked its dressed fidelity. The dressed fidelity of each target gate is estimated with circuit depths of {0, 1} through 50 circuit samples per circuit depth and performed 20,000 measurements per circuit. Theoretically, larger circuit depth differences can reduce fidelity estimation fluctuations, but practical constraints limit how deep the circuits can be. To reduce the impact of gate noise on the measurement results, we choose the circuit depths below 2.

Based on the measurement results, we estimated 100 quality parameters and averaged them to compute the gate fidelity. The results are shown in Fig. 2(b), ranging from 63.49% ± 0.07% to 17.42% ± 0.45% for qubit numbers from 16 to 46, with the full data available in the Supplemental Material (See Supplementary Material for more detailed experimental data of the comparison between character-average benchmarking and cycle benchmarking, the fluctuation analysis of the experimental results, and additional benchmarking and optimization results, which includes

references[25,33]). The quality parameters are distributed near the mean value, meaning the noise is close to a depolarizing noise. This feature is extremely useful in realizing CAB experiments in a low-fidelity region. If the noise is far from the depolarizing noise, like the unitary noise, and the fidelity is low, the quality parameters may be below 0. Since a quality parameter λ is obtained by fitting it to $A\lambda^{2m}$ as shown in Fig. 1, a negative quality parameter λ is indistinguishable from—λ and cannot be obtained correctly by the exponential fitting.

Parallel CZ gate benchmarking. The benchmarking for the parallel CZ gates contains two patterns, depicted in Fig. 3(a) using two distinct colors. We first evaluate the fidelities of the gates within the orange pattern, consisting of 22 pairs of CZ gates aligned in the same physical direction. This benchmarking was conducted progressively, starting with 2 pairs of CZ gates and incrementally including more gates up to the full set of 22 pairs. Subsequently, we evaluate the gate correlations within the orange and the blue patterns. The latter incorporates 26 pairs of CZ gates and engages almost the entire quantum processor.

The settings for benchmarking parallel CZ gates within the orange pattern are the same as that for the fully connected gate, except the circuit depths changed to {0, 2}. In this experiment, we benchmark both the dressed fidelity and the local twirling gate fidelity. The local twirling gate fidelity is obtained by changing the target gate U with the identity operation. We then
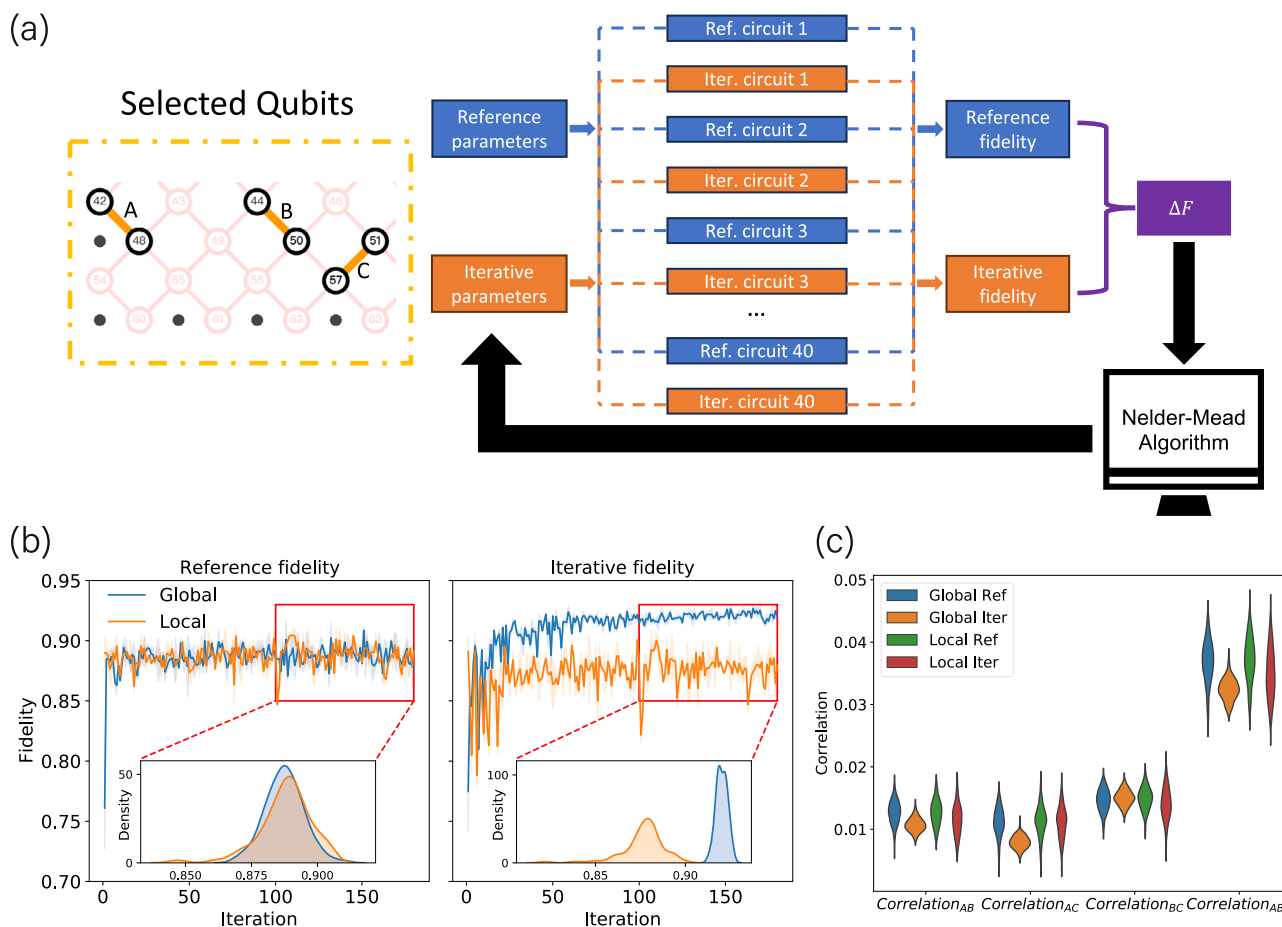
Fig. 4 | The experimental optimization process and corresponding results of parallel CZ gates. a The left figure shows the topology of 3-pair parallel CZ gates on the quantum processor. The three CZ gates are labeled with *A*, *B*, and *C*. During the optimization, we benchmark the fidelities of the target gate associated with two sets of parameters via CAB. The reference fidelity, fluctuating due to environmental factors, is based on reference parameters, and the iterative fidelity, mainly influenced by varying CZ gate parameters, corresponds to iterative parameters. The target function subtracts the reference fidelity from the iterative fidelity, thus mitigating the interference of environmental factors. The optimization algorithm is the Nelder-Mead algorithm. **b** The data of global fidelity during the optimization procedure. The line is the fidelity estimation, and the shadow above and below represents the value of one standard error away from the fidelity. The blue and orange lines correspond to the optimization results utilizing global and local CZ gate fidelities, respectively. The left and right figures show the reference and iterative fidelities, respectively. The small figure shows the probability density distribution of the fidelities within iterations 100–180, obtained by kernel density estimation. During this range, two reference fidelities remain stable and close, and iterative fidelities converge. Comparing the iterative fidelities in this range allows for a fair assessment, with the right figure indicating more effective fidelity improvements when global fidelity is used for optimization. **c** The distribution of gate correlations derived from reference and iterative fidelities within iterations 100–180 in (**b**). For a 3-pair parallel CZ gate, correlations are observed among all three gates ($Correlation_{ABC}$) and between each pair ($Correlation_{AB}$, $Correlation_{AC}$, and $Correlation_{BC}$). For instance, $Correlation_{AB}$ refers to the correlation between CZ gates *A* and *B*. The average values and standard deviations of correlations from reference fidelities are similar for both objective functions, indicating consistent environmental influences. However, for iterative fidelities, except $Correlation_{BC}$, employing global fidelity as the target function generally leads to improved outcomes.

use the interleaved technique[22] and Eq. (3) to isolate the pure parallel CZ gate fidelity.

In Fig. 3(b), we show the noise channel quality parameter distribution for the dressed parallel CZ gates and the local gates with violin plots. The mean value equals the fidelity and is shown with a point. The green diamond-shaped points represent the pure fidelities of the parallel CZ gates within the orange pattern. The largest one, or the 22-pair parallel CZ gate, possesses a fidelity of 63.09% ± 0.23%. These fidelities have been fit using the function $F^{n/2}$, where *n* is the qubit number. The fit aligns closely with our experimental data, suggesting a fidelity value of approximately 97.94% for a single CZ gate. This indicates that, in the orange pattern, the fidelity of individual CZ gates remains nearly constant and is not affected by an increase in qubit number. This observation implies that the crosstalk among these parallel CZ gates is either limited to short-range interactions or is remarkably minimal. Such a characteristic is critical for the implementation of quantum error correction. The detailed fidelity and standard error data are available in the Supplemental Material (See Supplementary Material for more detailed experimental data of the comparison between character-average benchmarking and cycle benchmarking, the fluctuation analysis of the experimental results, and additional benchmarking and optimization results, which includes references[25,33]).

Correlation benchmarking

Beyond gate fidelity, our analysis extends to examining correlations within parallel CZ gates. We define this correlation as the deviation of the parallel CZ gate fidelity from the product of the fidelities of its comprising local CZ gates. The concept of correlation is elaborated in Methods. Any nonzero correlation value emerges as an indicator of interactions among the local gates.

The correlations among every two CZ gates within the orange and blue patterns are visualized through heatmaps in Fig. 3(c) and (d), respectively. Additionally, we plot the correlation magnitudes as a function of the physical distance between CZ gates in the processor. For the orange pattern, we used the pure CZ gate fidelities from the parallel CZ gate benchmarking experiment to evaluate the correlation. The result of the blue pattern is

obtained by another experiment. The orange pattern exhibits a notable feature: correlations between CZ gates are consistently positive and pronounced only when the gates are closed. This observation suggests a negligible presence of long-range interactions in the implementation. However, the 26-pair parallel CZ gate within the blue pattern reveals different performance, with substantially higher correlations even for distant CZ gates, indicating the existence of long-range interactions within this configuration. The difference between the two patterns can be explained by whether a parallel calibration of CZ gates before benchmarking exists. Before experiments, gates in the same direction–such as all the two-qubit gates in the orange pattern–were calibrated in parallel using the method detailed in the "Parallel calibration of controlled-Z gate parameters with back probability" subsection in Methods. To maximize the qubit number of a pattern, the blue pattern is formed by combining gates in two directions. The results show that correlation benchmarking serves as an additional indicator of gate performance, alongside fidelity.

In Methods, we establish a noise model composed of depolarizing and ZZ-coupling noises to explain the correlation benchmarking results. The model first introduces a local depolarizing noise on each gate, followed by a unitary correlated noise. The Hamiltonian of the correlated noise is a summation of pairwise ZZ on each pair of CZ gates, where the coefficients depend on the coupling strengths. Our analysis focuses on the two-gate and three-gate coupling cases. A natural result is that the correlation value between two gates positively depends on their own coupling strength. Weak correlation implies weak coupling. Interestingly, we find that when only two gates couple with each other, the correlation is always positive. The correlation value is normally less than 0.001 for uncorrelated gates and can be on the order of 0.01 for correlated gates, which is consistent with the experimental data. Nonetheless, if one of the two gates couples with a third gate, the correlation between the original two gates can become negative. Particularly, when one gate is strongly coupled with the third gate, and the other gate is weakly coupled to it, the negative correlation becomes significant.

Applying the theoretical analysis to the experimental data, we observe that most correlation values are positive, corresponding to weak coupling or two-gate coupling cases. The results of the orange pattern can be fully explained by two-gate coupling. The pair of CZ2216 and CZ1004 and the pair of CZ4337 and CZ3125 exhibit the strongest couplings. The CZ gates in the two pairs are both nearest neighbors. For the blue pattern, negative correlation values are observed, indicating that the two-gate coupling model is not sufficient to explain the data. Take the pair of CZ4943 and CZ1812 as an example, the negative correlation value implies that besides the coupling between CZ4943 and CZ1812, there exists a third gate strongly coupled with CZ4943 or CZ1812. From the correlation data, we infer that a neighbor of CZ4943, like CZ3731, couples strongly with CZ4943 but not with CZ1812. The coupling relationship among these three CZ gates can lead to a negative correlation between CZ4943 and CZ1812. Note that in real experiments, couplings can involve more than three gates, making the origins of negative correlations more complicated than the three-gate model described here. We expect our results to inspire more explorations into the correlation and coupling among quantum gates.

Parallel CZ gate optimization

In addition to benchmarking, we conducted optimization on parallel CZ gates, employing two distinct approaches: optimizing with global fidelity and optimizing with individual local CZ gate fidelities. In both cases, the Nelder-Mead algorithm is utilized for optimization[37,38]. Note that the noise of the target gate is much larger than that of the local twirling gates, and the fidelities of the local twirling gates are stable. The noise of the target gate dominates the dressed fidelity, making this quantity sufficient for optimization. To reduce benchmarking time, we optimize using the dressed fidelity, avoiding additional benchmarking of local twirling gates and the interleaved procedure. To minimize the influence of other unstable factors on gate fidelity, we measure both the fidelity of the parameters being iterated (iterative fidelity) and the fidelity of the initial parameters as a reference (reference fidelity) throughout the optimization process. The optimization

target function is defined as the difference between these two fidelities. Before this optimization experiment, each local CZ gate was calibrated with the "fast calibration" and "parallel calibration" approaches shown in Methods.

In our experiments, the parallel CZ gate comprises $2n$ optimizable parameters with n the qubit number. We chose n as 4 and 6 so that the number of parameters is suitable for the Nelder-Mead algorithm to work. Optimizing gates with tens of qubits requires more scalable optimization algorithms. Figure 4 presents the optimization results for a parallel CZ gate comprising 3 local CZ gates on 6 qubits. The topology of the three CZ gates and the optimization procedure are shown in Fig. 4(a). The three gates are relatively close, which are more likely to correlate with each other. Meanwhile, the readout channels of these three gates are relatively stable compared to other gates, ensuring minimal influence of environmental noise on the optimization procedure.

The progression of fidelities during the optimization is depicted in Fig. 4(b), and the inter-gate correlations, calculated based on data from iterations 100-180, are illustrated in Fig. 4(c). This specific iteration range is chosen as it is the phase of the iterative parameter convergence and stable reference fidelities, indicating a reduced impact from other fluctuating factors. Further experimental details and results of a 4-qubit setup are available in the Supplemental Material (See Supplementary Material for more detailed experimental data of the comparison between character-average benchmarking and cycle benchmarking, the fluctuation analysis of the experimental results, and additional benchmarking and optimization results, which includes references[25,33]).

Below, we compare the optimization results employing global fidelity and the local gate fidelities with data from iterations 100-180. The fidelity is improved to 92.04% and 87.65%, and the correlation is reduced to 3.22% and 3.53% when using global fidelity and the local gate fidelities for optimization, respectively. It is clear that using global fidelity for optimization more effectively enhances fidelity and reduces correlation. Using local gate fidelities tends to yield inferior optimization outcomes, attributable to the lack of correlation information within the target function. In Methods, we use the ZZ-coupling model to explain the difference between global and local gate fidelities. When three gates are mutually coupled, optimizing one local fidelity may cause a decrease in the other two. This antagonistic relationship between local fidelities can trap the optimization in a local region. In contrast, global fidelity incorporates all correlation information, allowing the optimization to proceed monotonically. This finding underscores the significance of correlation in optimizing parallel gates and demonstrates the crucial advantage and essence of benchmarking large-scale quantum gates.

Discussion

In conclusion, we utilize CAB to conduct large-scale experiments of benchmarking the fully connected and parallel CZ gates. The benchmarking of gate correlation provides quantity to evaluate the quantum gate performance beyond gate fidelity, allowing the detection of long-range interaction, which may further be useful in studying many-body physics. Combined with the ZZ-coupling noise model, it is feasible to detect the coupling pattern of the parallel gates with correlation benchmarking. The established noise model also provides insight into further study of near-term quantum devices and quantum error correction.

The results highlight the crucial role of correlation in optimizing parallel quantum gates. Practically, one can decompose the circuit into multiple layers and optimize each layer with improved gate fidelity and reduced inter-gate correlation. This approach is more effective than optimizing each local gate individually, as evidenced by our optimization results. Meanwhile, one can first identify strongly coupled gates through correlation benchmarking and divide them into distinct groups. Gates within each group are strongly correlated, while groups themselves are weakly correlated. By optimizing each group individually, the overall performance of the entire layer can be enhanced. This approach simplifies the optimization process by focusing on the dominant gate correlations.

Box 1 | Procedure of character-average benchmarking

1. Choose a list of circuit depths $\{m_1, m_2, \cdots, m_M\}$, where M is the number of circuit depths.

2. Choose integers K_r and K_s as the number of random sequences and single-shot measurements, respectively.

3. For any $1 \leq j \leq M$, choose K_r random sequences $\mathcal{S}_\mathbf{j}^\mathbf{k} = \mathbf{C}^{-1}\mathbf{U_{inv}}\prod_{\mathbf{i}=1}^{\mathbf{m_j}}$ $(\mathbf{U}^{-1}\mathbf{P}^{(2i)}\mathbf{U}\mathbf{P}^{(2i-1)})\mathbf{C}$. Here, C is a local Clifford gate uniformly and randomly sampled from the n-qubit local Clifford group $C_1^{\otimes n}$, and $\forall\, 1 \leq i \leq 2m_j$, $P^{(l)}$ is uniformly and randomly sampled from the n-qubit Pauli group $P_\mathbf{n}$. The inverse gate $\mathbf{U_{inv}} = (\prod_{\mathbf{i}=1}^{\mathbf{m_j}}(\mathbf{U}^{-1}\mathbf{P}^{(2i)}\mathbf{U}\mathbf{P}^{(2i-1)}))^{-1}$ is a Pauli gate.

4. Prepare state $|0\rangle^{\otimes \mathbf{n}}$, implement each random sequence K_s times, and collect all Z-basis measurement results.

5. Independently sample K_q Z-basis observables, $\{O_i, 1 \leq i \leq K_q\}$, from $\{\mathbb{I}, \mathbf{Z}\}^{\otimes n}$. The sampling distribution is given by $2^{-2\mathbf{n}}3^{|O_i|}$, where $|O_i|$ is the weight of O_i, or the number of Z in O_i. For each O_i, compute $\text{tr}\widetilde{\mathbf{O}}_\mathbf{i}\mathcal{S}_\mathbf{j}^\mathbf{k}(\rho)$ with the measurement results from the previous step where ρ is the noisy version of $|0\rangle\langle 0|^{\otimes \mathbf{n}}$, and $\widetilde{\mathbf{O}}_\mathbf{i}$ is the noisy version of O_i. Then, average $\text{tr}\mathbf{O}_\mathbf{i}\mathcal{S}_\mathbf{j}^\mathbf{k}(\rho)$ over different random sequences for each circuit depth m_j and obtain $\mathbf{f_i(m_j)} = \frac{1}{\mathbf{K_r}}\sum_{\mathbf{k}=1}^{\mathbf{K_r}} \text{tr}\widetilde{\mathbf{O}}_\mathbf{i}\mathcal{S}_\mathbf{j}^\mathbf{k}(\rho)$, which is named survival probability.

6. For each O_i, fit $\{f_i(m_j), m_j\}$ to the function $\mathbf{f_i(m)} = \mathbf{A}\lambda_\mathbf{i}^{2\mathbf{m}}$ and obtain λ_i, which is a quality parameter of the noise channel. Then, the process fidelity of the target gate U is given by the average of $\{\lambda_i, 1 \leq i \leq K_q\}$,

$$\mathbf{F} = \frac{1}{\mathbf{K_q}}\sum_{\mathbf{i}=1}^{\mathbf{K_q}} \lambda_\mathbf{i}. \tag{4}$$

Our experimental methodology for benchmarking and optimizing large-scale gates applies to enhancing all Clifford circuits up to a local gauge transformation[33], including the essential case of quantum error-correcting code circuits[39]. Typical quantum encoding schemes like surface codes[9,10] or, more generally, quantum low-density-parity-check codes[40], involve tens or even hundreds of qubits. Optimizing such large batches involves managing many gate parameters, necessitating the development of scalable optimization algorithms rather than solely scalable benchmarking. In the future, advanced large-scale optimization algorithms, particularly gradient-free ones, will help to explore large-gate optimization, ultimately contributing to realizing universal fault-tolerant quantum computers.

Methods
Procedure of character-average benchmarking
Below in Box 1, we introduce the procedure of the CAB protocol when benchmarking an n-qubit Clifford gate corresponding to Fig. 1. For target gates as non-Clifford gates and more protocol details, one can refer to Ref. 33.

Note that the procedure in Box 1 differs from the original one in Ref. 33. The main modification lies in step 5. In Ref. 33, one does not sample observables but takes all observables from $\{\mathbb{I}, Z\}^{\otimes n}$, which requires $K_q = 2^n$. Here, we only need to set $K_q = O(-\epsilon^{-2}\log\delta)$ to ensure that Eq. (4) only differs from the fidelity estimated by traversing observables a small quantity, ϵ, with a high confidence level, $1 - \delta$. This can be seen from Hoeffding's inequality, which is shown below. Note that λ_i is limited in the region $[-1, 1]$.

$$\Pr(|F - \mathbb{E}_{\lambda_i}[F]| > \epsilon) \leq 2\exp(-\frac{K_q\epsilon^2}{2}). \tag{5}$$

Setting $2\exp(-\frac{K_q\epsilon^2}{2}) = \delta$, we get $K_q = 2\epsilon^{-2}(\log 2\delta^{-1})$. Note that K_q is irrelevant to the qubit number n. Thus, the complexity of the classical postprocessing is independent of n. The number of sampled sequences for fidelity estimation is also independent of n as proved in Ref. 33. Thus, the whole benchmarking protocol is scalable.

Correlation
Here, we introduce the formal definition of the correlation of a parallel gate. We consider a parallel gate, $U = \bigotimes_{i=1}^{g} U_i$, where U_i is more local or acts on fewer qubits than U. Normally, U_i is a one-local or two-local gate in experiments. That is, U_i only acts on one qubit or two qubits. Via CAB, one can simultaneously get the fidelity of U and the fidelities of U_i, denoted as $F(U)$ and $F(U_i)$, respectively. If there is no correlation among U_i, the noise channel of U can be expressed as $\Lambda = \bigotimes_{i=1}^{g}\Lambda_i$ where Λ_i is the individual

Table 1 | Summary of system parameters

Parameters	Mean	Median	SD
Qubit maximum frequency (GHz)	5.504	5.504	0.093
Qubit idle frequency (GHz)	5.411	5.403	0.104
Qubit anharmonicity (MHz)	−243	−242	4
T_1 at idle frequency (μs)	22.14	22.11	7.60
T_2^* at idle frequency (μs)	4.94	4.55	3.24
Readout $e_{|0\rangle}$ (%)	1.03	0.90	0.72
Readout $e_{|1\rangle}$ (%)	3.82	3.43	1.64
1Q XEB e_1 (%)	0.31	0.24	0.18
2Q-CZ XEB e_2 (%)	3.35	3.21	0.98

SD denotes the standard deviation.

noise of U_i. Then, the global fidelity, $F(U)$, would be equal to the product of local gate fidelities, $\prod_{i=1}^{g} F(U_i)$. In reality, the interaction among different gates would make the two values different. We define the following quantity to characterize the total correlation among $\{U_i, 1 \leq i \leq g\}$,

$$C_g(U, \{U_i, 1 \leq i \leq g\}) = \frac{F(U) - \prod_{i=1}^{g} F(U_i)}{\sqrt{F(U)\prod_{i=1}^{g} F(U_i)}}. \tag{6}$$

The denominator is a normalization factor. When the correlation is positive, the global fidelity is larger than the product fidelity, indicating that the correlation helps to increase the global fidelity. Since Eq. (6) is defined among g gates, we call it g-correlation. Except for g-correlation, one can also obtain j-correlation among each j gate in $\{U_i, 1 \leq i \leq g\}$ where $2 \leq j \leq g - 1$ for parallel gate $U = \bigotimes_{i=1}^{g} U_i$. Note that since $F(U)$ and $F(U_i)$ can be obtained simultaneously from the same experimental data, gate correlation can also be evaluated concurrently without any additional experimental effort.

Experimental platform
In this work, we utilized a processor with the same design as the *Zuchongzhi* 2.0 processor[41] and selected up to 54 qubits for our experiments. The basic performance of the processor is shown in Table 1, where the single-qubit gate error and the two-qubit CZ gate error with a median of 0.24% and 3.21% by cross-entropy benchmarking (XEB)[42–44]. Our scheme to realize a two-qubit CZ gate is implementing an all-microwave coupler with a fixed gate time of 110 ns[45]. The approach involves applying a microwave

Fig. 5 | Circuits for fast calibration of CZ gates. Q_1 and Q_2 represent the first qubit and the second qubit, respectively. C represents the coupler. $X_\beta = e^{-\frac{i\beta X}{2}}$ where X is the Pauli X operator. N normally takes 1, 3, 5, 7.
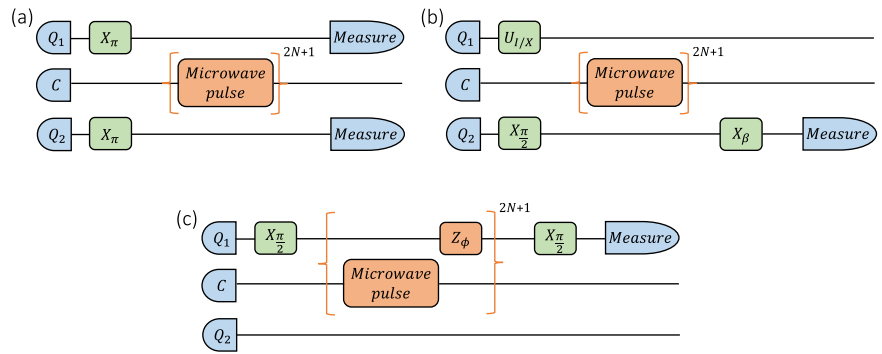
signal with an envelope $A(t)$ and a driving frequency of ω_t to the tunable coupler, with an extra flux given by $\Phi(t) = A(t) \cos(\omega_t t + \phi_0)$. When the driving frequency ω_t matches the energy difference between the $|11\rangle$ and $|02\rangle$ states, i.e., $\omega_t = \omega_{11} - \omega_{02}$, resonance takes place between these two states. Here, $|11\rangle$ is a computational-basis state with each qubit at state $|1\rangle$, and $|02\rangle$ is a state outside the computational subspace. However, due to the nonlinear relationship between the extra flux and the coupling strength, although $\Phi(t)$ is a good single-frequency signal when transformed into the coupling strength, the signal contains significant frequency components not only at ω_t, but also at $2\omega_t$ and $4\omega_t$. Therefore, when considering the frequency layout of qubits, it is necessary to avoid ω_t, $2\omega_t$, and $4\omega_t$ equal to either $\Delta_{01,10} = |\omega_{01} - \omega_{10}|$ or $\Delta_{11,20} = \omega_{11} - \omega_{20}$.

Before our experiments, we adjusted the qubit frequency and calibrated the parameters for each local gate to make the processor perform well. In the following subsections, we will elaborate on this procedure in detail.

Frequency conflict and frequency adjustment

The distribution of qubit frequencies is typically constrained within a range of 0-400 MHz due to magnetic flux noise and the bandwidth of the digital-to-analog converter. When arranging the qubit frequencies, the following factors need to be considered and balanced: (1) Energy relaxation time T_1 and dephasing time T_ϕ. (2) Spacing of frequencies between neighboring qubits and next-to-nearest neighboring qubits. (3) Two-qubit gate frequencies, as well as their second and fourth harmonic frequencies, and the frequency conflicts with $\Delta_{01,10} = |\omega_{01} - \omega_{10}|$ and $\Delta_{11,20} = \omega_{11} - \omega_{20}$. (4) Maximum frequencies for each qubit, which represent the available frequency range for each qubit. By defining the above factors as error functions and frequency domains, we can obtain a set of theoretically optimal frequencies. After adjusting the frequencies of all qubits to the optimized arrangement, a majority of single-qubit gates and two-qubit gates can achieve high fidelity through standard calibration. However, local fine-tuning is still required for poorly performing gates. Additionally, the performance of a quantum processor may deteriorate during certain time intervals due to long-term periodic frequency variations in two-level systems. This also requires fine-tuning of the corresponding qubits. Qubit frequency tuning is relatively frequent and tedious, as adjusting the frequency of one qubit requires re-calibrating two-qubit gates associated with it. Therefore, it is crucial to calibrate single-qubit gates and two-qubit gates efficiently in this process. On the other side, the success of subsequent benchmarking relies on an initial good adjustment of qubit frequencies, as a higher gate fidelity improves the benchmarking accuracy and stability. The configuration of qubit frequencies is the key to the success of our experiments.

Fast calibration of controlled-Z gates

After fine-tuning the qubit frequency, we need to recalibrate CZ gates related to it. The main parameters for calibrating the CZ gates are microwave frequency, microwave amplitude, and dynamic phase of the two relevant qubits. First, we roughly determine the microwave frequency and

amplitude through the circuit in Fig. 5(a) with N typically set to 0. Within this circuit, the two qubits Q_1 and Q_2 are initially set at $|0\rangle$. We first flip these two qubits by applying pulse X_π. Then, we apply the microwave pulse once, and after that, we measure the probability of two qubits returning to the $|11\rangle$ state. In the process of applying microwave pulses, $|11\rangle$ and $|02\rangle$ states will be exchanged, and we try to find the microwave pulse parameters to maximize the probability back to $|11\rangle$ for the ending state. Then, we fine-tune the microwave amplitude by implementing circuit (a) again. To amplify the errors caused by the parameters, we superimpose $2N + 1$ CZ gates, with $N > 0$ this time. As the conditional phase is relatively sensitive to the frequency, we fine-tune the microwave frequency through the circuit in Fig. 5(b). When the β of X_β changes, the probability of Q_1, or the first qubit, changes as follows:

$$P(\beta) = \frac{1 + \cos(\beta + \phi_I)}{2} \text{ for } U = I; \quad (7)$$

$$P(\beta) = \frac{1 + \cos(\beta + \phi_x)}{2} \text{ for } U = X. \quad (8)$$

By fitting $P(\beta)$ and obtaining ϕ_I and ϕ_x, we can get the conditional phase $\phi = \phi_x - \phi_I$. The optimal microwave frequency is the frequency at which $\phi = \pi$ is satisfied. We then repeat the circuit (a) again to calibrate the microwave amplitude further with a large N. The optimal microwave amplitude is the amplitude that makes the probability of the $|11\rangle$ state closest to 1. The dynamic phase of the two relevant qubits can be calibrated through the circuit in Fig. 5(c). We first change the dynamic phase Z_ϕ of Q_1 and find the point where the probability of the $|1\rangle$ state is closest to 1 to complete the dynamic phase compensation for Q_1. This process is repeated for Q_2 subsequently with $X_{\frac{\pi}{2}}$ and Z_ϕ applied at Q_2 in circuit (c).

Parallel calibration of controlled-Z gate parameters with back probability

When we calibrate CZ gate parameters by means of amplifying errors through the superposition of multiple layers of CZ gate circuits, we can attain a high level of fidelity for most CZ gates. To further refine the gate parameters, we resort to the Nelder-Mead algorithm to continue to search for CZ gate parameters. Particularly, for each local CZ gate, we input $|00\rangle$, implement a random sequence of two-qubit Clifford gates, and record the probability of the final state back to $|00\rangle$. Each two-qubit Clifford gate is decomposed into CZ and single-qubit gates in implementation. Additionally, we alternate running the same set of random circuits with the reference and iterative parameters. The difference between their outcomes is used as the target function of the Nelder-Mead algorithm to mitigate environmental influence. To quickly calibrate a set of CZ gates where no two gates share a common qubit, we execute the above procedure in parallel for each local CZ gate. Each gate is calibrated independently using its own back probability. This parallel calibration method typically yields better gate parameters than the parameter scanning approach described in the previous subsection.

Fidelity and correlation analysis of depolarizing and correlated noise

In this part, we establish a simple but physical noise model to explain the experimental results. Note that CAB faithfully evaluates the process fidelity of quantum gates[33]. We consider how the process fidelity behaves under a combination of local depolarizing noise and gate interaction. Particularly, for a unitary gate $U = \bigotimes_{i=1}^{g} U_i$, we consider the following noise model.

$$\Lambda = \Lambda_V \circ \bigotimes_i \Lambda_{p_i}, \qquad (9)$$

where Λ_{p_i} is a depolarizing noise on the i-th gate with parameter p_i such that

$$\Lambda_{p_i}(\rho) = p_i \rho + (1 - p_i) \frac{\mathbb{I}_i}{d_i}. \qquad (10)$$

Here, d_i is the dimension of gate U_i, and $\mathbb{I}_i$ is the identity operator on this subsystem. The noise Λ_V is a correlated noise among all local gates U_i, modeled as a unitary evolution:

$$\Lambda_V(\rho) = V \rho V^\dagger. \qquad (11)$$

The form of Λ incorporates both decoherence on individual gates and interactions between gates. The decoherence on each gate is set as a depolarizing type, which is a standard setting in studies of near-term quantum devices[3–5,28]. Different from previous works, we introduce an additional interaction term V, which makes the noise model more physical. The form of V depends on the implemented gate U, which will be specified later. When considering the action of V on a subsystem S, the remainder of the system is treated as a maximally mixed state or a thermalized state at infinite temperature. That is, for state ρ_S on subsystem S,

$$\Lambda_V|_S(\rho_S) = \mathrm{tr}_{\bar{S}} V \left(\rho_S \otimes \frac{\mathbb{I}_{\bar{S}}}{d_{\bar{S}}} \right) V^\dagger. \qquad (12)$$

Here, $\Lambda_V|_S$ is the restriction of Λ_V to S, $\bar{S}$ is the complementary subsystem of S, $d_{\bar{S}}$ is the dimension of $\bar{S}$, and $\mathbb{I}_{\bar{S}}$ is the identity operator on $\bar{S}$.

Since the gate $U = \bigotimes_{i=1}^{g} U_i$ comprises g gates $\{U_1, U_2, \cdots, U_g\}$, we use $[g] = \{1, 2, \cdots, g\}$ to denote the whole system. Given a subset $S \subseteq [g]$, we can represent a part of the gate U, $\bigotimes_{i \in S} U_i$, whose fidelity is given by $F(\Lambda_V|_S \circ \bigotimes_{i \in S} \Lambda_{p_i})$ and denoted as F_S. Note that the process fidelity has an expression $F(\Lambda) = \mathrm{tr}(|\Phi^+\rangle\langle\Phi^+| \Lambda(|\Phi^+\rangle\langle\Phi^+|))$ where $|\Phi^+\rangle$ is a maximally entangled state on two copies of the system. Through direct calculation, we have that

$$F_S = \sum_{L \subseteq S} p^L (1 - p)^{S \setminus L} \frac{d_L}{d d_S^2} \parallel \mathrm{tr}_L V \parallel_2^2, \qquad (13)$$

where L is a subset of S, $S \setminus L$ is the complementary set of L in S, d_L and d_S are the dimensions of subsystems L and S, respectively, and

$$p^L = \prod_{i \in L} p_i, (1 - p)^{S \setminus L} = \prod_{i \in S \setminus L} (1 - p_i), \parallel \mathrm{tr}_L V \parallel_2^2 = \mathrm{tr}_{\bar{L}}(\mathrm{tr}_L V \mathrm{tr}_L V^\dagger). \qquad (14)$$

Note that $\bar{L}$ is the complementary set of L in $[g]$. From Eq. (13), we can obtain the global fidelity and local gate fidelities, and hence evaluate the correlation and investigate how fidelities depend on noise parameters.

In our experiments, we mainly consider the parallel CZ gate $U = \bigotimes_{k=1}^{r} \mathrm{CZ}^{(i_k j_k)}$. Based on experimental observations, the correlated noise mainly arises from the ZZ coupling among qubits. Particularly, we consider a simplified noise model where only one qubit from each CZ gate couples with each other. Without loss of generality, we set this qubit as i_k. Meanwhile, the Hamiltonian of the correlated noise only contains two local terms while the strength between i_k and i_l is set as γ_{kl}. Thus, the correlated noise V is

$$V = e^{-i \sum_{1 \le k < l \le r} \gamma_{kl} Z_{i_k} Z_{i_l}}. \qquad (15)$$

The strength parameter is determined by $\gamma_{kl} = g_{kl} t$, where t is the evolution time, and g_{kl} is the coupling strength between two qubits. In our experiments, t corresponds to the two-qubit gate time, which is 110ns. For two physically isolated qubits, their coupling strength is typically less than 0.3 MHz, resulting in γ_{kl} less than 0.033 for two uncorrelated gates. For two coupled qubits, γ_{kl} can be on the order of 0.1.

Substituting Eq. (15) into Eq. (13) gives the global fidelity and local CZ gate fidelities. We provide the results when $r = 2$ and $r = 3$, which relates to our correlation benchmarking and gate optimization results. More general cases can be straightforwardly obtained using Eq. (13).

When $r = 2$, the fidelities of the two local CZ gates are

$$F_1 = p_1 \cos^2 \gamma_{12} + \frac{1 - p_1}{4}; \qquad (16)$$

$$F_2 = p_2 \cos^2 \gamma_{12} + \frac{1 - p_2}{4}. \qquad (17)$$

The global fidelity of the two CZ gates is

$$F_{[2]} = \left(p_1 p_2 + \frac{p_1(1 - p_2)}{4} + \frac{p_2(1 - p_1)}{4} \right) \cos^2 \gamma_{12} + \frac{(1 - p_1)(1 - p_2)}{16}. \qquad (18)$$

The above gives the correlation when only two CZ gates correlate with each other via the ZZ coupling:

$$\frac{F_{[2]} - F_1 F_2}{\sqrt{F_{[2]} F_1 F_2}} = \frac{p_1 p_2 \cos^2 \gamma_{12} \sin^2 \gamma_{12}}{\sqrt{F_{[2]} F_1 F_2}}$$
$$\overset{p_1, p_2 \to 1}{\approx} \sin \gamma_{12} \tan \gamma_{12}. \qquad (19)$$

In the limit that p_1 and p_2 are close to 1, $\sqrt{F_{[2]} F_1 F_2}$ is approximately $p_1 p_2 \cos^3 \gamma_{12}$. Then, the correlation is approximately $\sin \gamma_{12} \tan \gamma_{12}$. Given the value of γ_{12} as 0.033 and 0.1, the correlation values take 0.001 and 0.01, respectively. This is consistent with our experimental results.

When $r = 3$, we give the fidelities of the three local CZ gates, the fidelities of each pair of CZ gates, and the global fidelity.

$$F_1 = p_1 (\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13}) + \frac{1 - p_1}{16}; \qquad (20)$$

$$F_2 = p_2 (\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23}) + \frac{1 - p_2}{16}; \qquad (21)$$
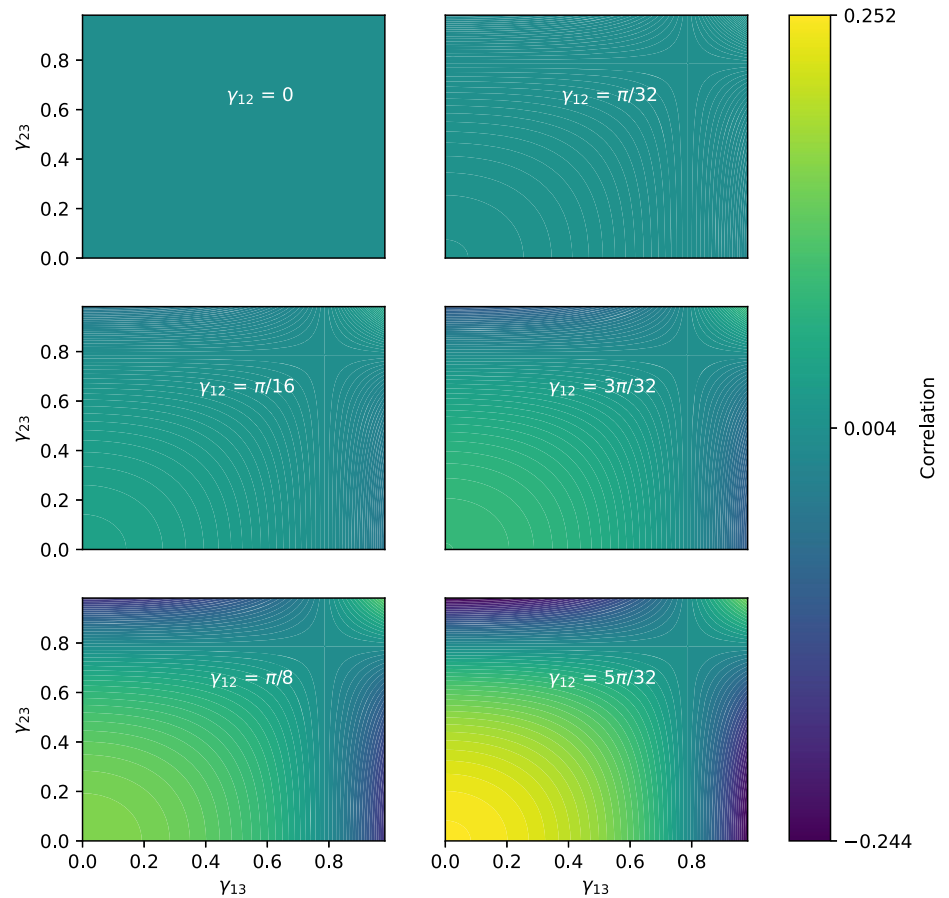
$$F_3 = p_3 (\cos^2 \gamma_{13} \cos^2 \gamma_{23} + \sin^2 \gamma_{13} \sin^2 \gamma_{23}) + \frac{1 - p_3}{16}; \qquad (22)$$

$$F_{12} = p_1 p_2 (\cos^2 \vec{\lambda} + \sin^2 \vec{\lambda}) + \frac{p_1(1 - p_2)}{16} (\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13})$$
$$+ \frac{(1 - p_1)p_2}{16} (\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23}) + \frac{(1 - p_1)(1 - p_2)}{256}; \qquad (23)$$

$$F_{13} = p_1 p_3 (\cos^2 \vec{\lambda} + \sin^2 \vec{\lambda}) + \frac{p_1(1 - p_3)}{16} (\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13})$$
$$+ \frac{(1 - p_1)p_3}{16} (\cos^2 \gamma_{13} \cos^2 \gamma_{23} + \sin^2 \gamma_{13} \sin^2 \gamma_{23}) + \frac{(1 - p_1)(1 - p_3)}{256}; \qquad (24)$$

$$F_{23} = p_2 p_3 (\cos^2 \vec{\lambda} + \sin^2 \vec{\lambda}) + \frac{p_2(1 - p_3)}{16} (\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23})$$
$$+ \frac{(1 - p_2)p_3}{16} (\cos^2 \gamma_{13} \cos^2 \gamma_{23} + \sin^2 \gamma_{13} \sin^2 \gamma_{23}) + \frac{(1 - p_2)(1 - p_3)}{256}; \qquad (25)$$

Fig. 6 | These graphs depict the relationship between different coupling strengths $\{\gamma_{12}, \gamma_{13}, \gamma_{23}\}$ **and correlation** F_{12} **within the context of three-gate coupling examples.** Particularly, we assign fixed values to γ_{12}, selecting from the set $\{0, \pi/32, \pi16, 3\pi/32, \pi/8, 5\pi/32\}$. The values of γ_{13} and γ_{23} range from 0 to $5\pi/16$.

$$F_{[3]} = (p_1 p_2 p_3 + \frac{p_1 p_2 (1-p_3)}{16} + \frac{p_1 p_3 (1-p_2)}{16} + \frac{p_2 p_3 (1-p_1)}{16})(\cos^2 \overrightarrow{\lambda} + \sin^2 \overrightarrow{\lambda})$$
$$+ \frac{p_1 (1-p_2)(1-p_3)}{256}(\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13})$$
$$+ \frac{p_2 (1-p_1)(1-p_3)}{256}(\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23})$$
$$+ \frac{p_3 (1-p_1)(1-p_2)}{256}(\cos^2 \gamma_{13} \cos^2 \gamma_{23} + \sin^2 \gamma_{13} \sin^2 \gamma_{23})$$
$$+ \frac{(1-p_1)(1-p_2)(1-p_3)}{4096}.$$

$$(26)$$

Here, $\cos \overrightarrow{\lambda} = \cos \gamma_{12} \cos \gamma_{13} \cos \gamma_{23}$ and $\sin \overrightarrow{\lambda} = \sin \gamma_{12} \sin \gamma_{13} \sin \gamma_{23}$. In this case, the correlation between the first and second local CZ gates is given by

$$\frac{F_{12} - F_1 F_2}{\sqrt{F_{12} F_1 F_2}} = \frac{p_1 p_2 ((\cos^2 \overrightarrow{\lambda} + \sin^2 \overrightarrow{\lambda}) - (\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13})(\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23}))}{\sqrt{F_{12} F_1 F_2}}$$
$$= \frac{p_1 p_2 \cos^2 \gamma_{12} \sin^2 \gamma_{12} (\cos^2 \gamma_{13} - \sin^2 \gamma_{13})(\cos^2 \gamma_{23} - \sin^2 \gamma_{23})}{\sqrt{F_{12} F_1 F_2}}$$
$$\overset{p_1 \cdot p_2 \to 1}{\approx} \frac{\cos^2 \gamma_{12} \sin^2 \gamma_{12} (\cos^2 \gamma_{13} - \sin^2 \gamma_{13})(\cos^2 \gamma_{23} - \sin^2 \gamma_{23})}{\sqrt{(\cos^2 \overrightarrow{\lambda} + \sin^2 \overrightarrow{\lambda})(\cos^2 \gamma_{12} \cos^2 \gamma_{13} + \sin^2 \gamma_{12} \sin^2 \gamma_{13})(\cos^2 \gamma_{12} \cos^2 \gamma_{23} + \sin^2 \gamma_{12} \sin^2 \gamma_{23})}}.$$

$$(27)$$

Note that when p_1 and p_2 are close to 1, the first term in the formula of fidelities dominates, and we get the above approximation. It is interesting that the sign of the correlation depends on $(\cos^2 \gamma_{13} - \sin^2 \gamma_{13})(\cos^2 \gamma_{23} - \sin^2 \gamma_{23})$. A negative correlation value implies that one of γ_{13} and γ_{23} is larger than $\pi/4$.

To further investigate how the correlation depends on parameters γ_{12}, γ_{13}, and γ_{23}, we fix γ_{12} with values in $\{0, \pi/32, \pi/16, 3\pi/32, \pi/8, 5\pi/32\}$ and depict the correlation with respect to γ_{13} and γ_{23}. The results are available in Fig. 6.

From the expressions in Eqs. (19) and (27), we observe a natural result that the correlation between two gates always increases with their own coupling strength. In the two-gate coupling case, the correlation is $\sin \gamma_{12} \tan \gamma_{12}$, which increases monotonically with γ_{12}. In the three-gate coupling case, the numerator of the correlation is proportional to $\cos^2 \gamma_{12} \sin^2 \gamma_{12}$, which also increases monotonically with γ_{12}. Thus, the correlation magnitude directly reflects the coupling strength.

Examining the sign of the correlation reveals distinct behaviors. If only two CZ gates are coupled, their correlation is always positive. However, in a three-gate coupling scenario, the situation changes. The correlation is still positive for a low-strength correlated noise when all parameters are less than $\pi/4$. Nonetheless, if two CZ gates exhibit strong ZZ coupling, one of the CZ gates will have a negative correlation with the third CZ gate. From Eq. (27) and Fig. 6, we can see that the negativity becomes particularly pronounced when one coupling is weak while the other is strong. This characteristic is useful for identifying the strongly coupled pair of gates.

Beyond correlation analysis, the dependence of local gate fidelities and global fidelity on the coupling strength helps explain the differences between these two types of fidelities in gate optimization. In the case of $r = 2$, this difference is small. Both kinds of fidelities are proportional to $\cos^2 \gamma_{12}$, and optimization naturally reduces γ_{12} to improve performance. Nonetheless, for the three-gate coupling model, the situation can be different. Each local fidelity depends only on two coupling parameters. When optimizing F_1, the coupling parameters γ_{12} and γ_{13} tend to be smaller to make F_1 higher and make the correlation weaker. However, given a fixed γ_{23}, the decrease of γ_{12} or γ_{13} can also decrease F_2 or F_3. For instance, $F_2 = p_2 (\cos^2 \gamma_{12}(\cos^2 \gamma_{23} - \sin^2 \gamma_{23}) + \sin^2 \gamma_{23}) + \frac{1-p_2}{16}$. When $\cos^2 \gamma_{23} - \sin^2 \gamma_{23} < 0$, the decrease of γ_{12} makes F_2 also decrease. This creates a competition between optimizing one local fidelity and another, often

leading to the optimization being trapped in a local region. In contrast, global fidelity incorporates all coupling parameters, allowing the optimization process to simultaneously reduce all coupling strengths.

References

1. Aharonov, D., Ben-Or, M., Impagliazzo, R. & Nisan, N. Limitations of noisy reversible computation. https://arxiv.org/abs/quant-ph/9611028 (1996).
2. Müller-Hermes, A., Stilck França, D. & Wolf, M. M. Relative entropy convergence for depolarizing channels. *J. Math. Phys.* **57**, 022202 (2016).
3. Stilck França, D. & García-Patrón, R. Limitations of optimization algorithms on noisy quantum devices. *Nat. Phys.* **17**, 1221–1227 (2021).
4. Aharonov, D., Gao, X., Landau, Z., Liu, Y. & Vazirani, U. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, 945-957 https://doi.org/10.1145/3564246.3585234. (Association for Computing Machinery, New York, NY, USA, 2023).
5. Yan, Y., Du, Z., Chen, J. & Ma, X. Limitations of noisy quantum devices in computational and entangling power. https://arxiv.org/abs/2306.02836 (2023).
6. Egan, L. et al. Fault-tolerant control of an error-corrected qubit. *Nature* **598**, 281–286 (2021).
7. Gong, M. et al. Experimental exploration of five-qubit quantum error-correcting code with superconducting qubits. *Natl Sci. Rev.* **9**, nwab011 (2021).
8. Postler, L. et al. Demonstration of fault-tolerant universal quantum gate operations. *Nature* **605**, 675–680 (2022).
9. Zhao, Y. et al. Realization of an error-correcting surface code with superconducting qubits. *Phys. Rev. Lett.* **129**, 030501 (2022).
10. Acharya, R. et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023).
11. Kitaev, A. Y. Quantum computations: algorithms and error correction. *Russian Math. Surv.* **52**, 1191 (1997).
12. Aharonov, D. & Ben-Or, M. Fault-tolerant quantum computation with constant error. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97,176-188 https://doi.org/10.1145/258533.258579 (Association for Computing Machinery, New York, NY, USA, 1997).
13. Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation: error models and thresholds. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **454**, 365–384 (1998).
14. Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
15. Chuang, I. L. & Nielsen, M. A. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455–2467 (1997).
16. Flammia, S. T. & Liu, Y.-K. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett.* **106**, 230501 (2011).
17. Emerson, J., Alicki, R. & Zyczkowski, K. Scalable noise estimation with random unitary operators. *J. Opt. B Quantum Semiclassical Opt.* **7**, S347–S352 (2005).
18. Emerson, J. et al. Symmetrized characterization of noisy quantum processes. *Science* **317**, 1893–1896 (2007).
19. Knill, E. et al. Randomized benchmarking of quantum gates. *Phys. Rev. A* **77**, 012307 (2008).
20. Magesan, E., Gambetta, J. M. & Emerson, J. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.* **106**, 180504 (2011).
21. Magesan, E., Gambetta, J. M. & Emerson, J. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A* **85**, 042311 (2012).
22. Magesan, E. et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.* **109**, 080505 (2012).
23. McKay, D. C., Sheldon, S., Smolin, J. A., Chow, J. M. & Gambetta, J. M. Three-qubit randomized benchmarking. *Phys. Rev. Lett.* **122**, 200502 (2019).
24. Proctor, T. J. et al. Direct randomized benchmarking for multiqubit devices. *Phys. Rev. Lett.* **123**, 030503 (2019).
25. Erhard, A. et al. Characterizing large-scale quantum computers via cycle benchmarking. *Nat. Commun.* **10**, 5347 (2019).
26. Hines, J. et al. Demonstrating scalable randomized benchmarking of universal gate sets. *Phys. Rev. X* **13**, 041030 (2023).
27. Zhu, Q. et al. Quantum computational advantage via 60-qubit 24-cycle random circuit sampling. *Sci. Bull.* **67**, 240–245 (2022).
28. Morvan, A. et al. Phase transitions in random circuit sampling. *Nature* **634**, 328–333 (2024).
29. McKay, D. C. et al. Benchmarking quantum processor performance at scale. https://arxiv.org/abs/2311.05933 (2023).
30. Moses, S. A. et al. A race-track trapped-ion quantum processor. *Phys. Rev. X* **13**, 041052 (2023).
31. Joshi, M. K. et al. Exploring large-scale entanglement in quantum simulation. *Nature* **624**, 539–544 (2023).
32. Bluvstein, D. et al. Logical quantum processor based on reconfigurable atom arrays. *Nature* **626**, 58–65 (2024).
33. Zhang, Y., Yu, W., Zeng, P., Liu, G. & Ma, X. Scalable fast benchmarking for individual quantum gates with local twirling. *Photon. Res.* **11**, 81–99 (2023).
34. Cerezo, M. et al. Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
35. Raussendorf, R., Browne, D. E. & Briegel, H. J. Measurement-based quantum computation on cluster states. *Phys. Rev. A* 68 https://doi.org/10.1103/PhysRevA.68.022312 (2003).
36. Hein, M., Eisert, J. & Briegel, H. J. Multiparty entanglement in graph states. *Phys. Rev. A* **69**, 062311 (2004).
37. Nelder, J. A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
38. Rol, M. A. et al. Restless tuneup of high-fidelity qubit gates. *Phys. Rev. Appl.* **7**, 041001 (2017).
39. Gottesman, D. Stabilizer codes and quantum error correction (California Institute of Technology, 1997).
40. Bravyi, S. et al. High-threshold and low-overhead fault-tolerant quantum memory. *Nature* **627**, 778–782 (2024).
41. Wu, Y. et al. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* **127**, 180501 (2021).
42. Barends, R. et al. Diabatic gates for frequency-tunable superconducting qubits. *Phys. Rev. Lett.* **123**, 210501 (2019).
43. Boixo, S. Characterizing Beyond-Classical Computation in Near-Term Devices. In APS March Meeting Abstracts, vol. 2017 of *APS Meeting Abstracts*, A19.005 (2017).
44. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
45. Xu, H. et al. Realization of adiabatic and diabatic cz gates in superconducting qubits coupled with a tunable coupler*. *Chin. Phys. B* **30**, 044212 (2021).

Acknowledgements

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41534-025-00983-5.

Correspondence and requests for materials should be addressed to Xiongfeng Ma, Xiaobo Zhu or Jian-Wei Pan.

Reprints and permissions information is available at http://www.nature.com/reprints