

Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device

Xue, X.; Watson, T. F.; Helsen, J.; Ward, D. R.; Savage, D. E.; Lagally, M. G.; Coppersmith, S. N.; Eriksson, M. A.; Wehner, S.; Vandersypen, L. M.K.

DOI

[10.1103/PhysRevX.9.021011](https://doi.org/10.1103/PhysRevX.9.021011)

Publication date

2019

Document Version

Final published version

Published in

Physical Review X

Citation (APA)

Xue, X., Watson, T. F., Helsen, J., Ward, D. R., Savage, D. E., Lagally, M. G., Coppersmith, S. N., Eriksson, M. A., Wehner, S., & Vandersypen, L. M. K. (2019). Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device. *Physical Review X*, 9(2), Article 021011. <https://doi.org/10.1103/PhysRevX.9.021011>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device

X. Xue,^{1,2} T. F. Watson,^{1,2} J. Helsen,¹ D. R. Ward,³ D. E. Savage,³ M. G. Lagally,³ S. N. Coppersmith,^{3,†}
M. A. Eriksson,³ S. Wehner,¹ and L. M. K. Vandersypen^{1,2,*}

¹*QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, Netherlands*

²*Kavli Institute of Nanoscience, Delft University of Technology, 2600 GA Delft, Netherlands*

³*University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*



(Received 28 November 2018; published 18 April 2019)

We report the first complete characterization of single-qubit and two-qubit gate fidelities in silicon-based spin qubits, including cross talk and error correlations between the two qubits. To do so, we use a combination of standard randomized benchmarking and a recently introduced method called character randomized benchmarking, which allows for more reliable estimates of the two-qubit fidelity in this system, here giving a 92% fidelity estimate for the controlled-Z gate. Interestingly, with character randomized benchmarking, the two-qubit gate fidelity can be obtained by studying the additional decay induced by interleaving the two-qubit gate in a reference sequence of single-qubit gates only. This work sets the stage for further improvements in all the relevant gate fidelities in silicon spin qubits beyond the error threshold for fault-tolerant quantum computation.

DOI: [10.1103/PhysRevX.9.021011](https://doi.org/10.1103/PhysRevX.9.021011)

Subject Areas: Mesoscopics, Nanophysics,
Quantum Information

I. INTRODUCTION

With steady progress towards practical quantum computers, it becomes increasingly important to efficiently characterize the relevant quantum gates. Quantum process tomography [1–3] provides a way to reconstruct a complete mathematical description of any quantum process, but has several drawbacks. The resources required increase exponentially with qubit number and the procedure cannot distinguish pure gate errors from state preparation and measurement (SPAM) errors, making it difficult to reliably extract small gate error rates. Randomized benchmarking (RB) was introduced as a convenient alternative [4–7]. It estimates the gate fidelity as a concise and relevant metric, requires fewer resources, is more robust against SPAM errors, and works well even for low gate error rates.

Various randomized benchmarking methods have been investigated to extract fidelities and errors in different scenarios. In standard randomized benchmarking, sequences of increasing numbers of random Clifford operations are applied to one or more qubits [5,6] (for a single qubit,

the Clifford gates are precisely the gates that rotate states along the \hat{x} , \hat{y} , or \hat{z} axis on the Bloch sphere to each other). Then, loosely speaking, the average Clifford gate fidelity is extracted from how rapidly the final state diverges from the ideally expected state as a function of the number of random Clifford operations. For multiqubit systems, the degree of cross talk can be characterized by simultaneous randomized benchmarking, in which random single-qubit Clifford operations are applied simultaneously to different qubits [8].

In practice we will often be more interested in the fidelity of a particular quantum gate than the average fidelity of a gate set. To characterize this fidelity, we can make use of interleaved randomized benchmarking, in which a sequence of random Clifford gates is interleaved by a particular quantum gate. The fidelity of the interleaved gate can then be estimated by comparing the deviation rate from the ideal state with that obtained from standard randomized benchmarking [9], called the reference fidelity in this context. It must be noted, however, that interleaved randomized benchmarking does not provide an exact characterization of the fidelity of the interleaving gate but rather gives an estimate together with upper and lower bounds. These upper and lower bounds, which are essentially due to the fact that the fidelity of the composition of two gates is not necessarily equal to the product of the fidelities of the two gates, grow closer together with increasing reference fidelity [9,10].

A major drawback of these traditional randomized benchmarking methods is that the number of native gates that needs to be executed in sequence to implement a Clifford operation can rapidly increase with the qubit number. For example, it takes on average 1.5

*Corresponding author.

l.m.k.vandersypen@tudelft.nl

†Present address: School of Physics, University of New South Wales, Sydney NSW 2052, Australia.

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

controlled-phase (CPHASE) gates and 8.25 single-qubit gates to implement a two-qubit Clifford gate [11]. This in turn puts higher demands on the coherence time, which is still a challenge for near-term devices, and leads to rather loose bounds on the gate fidelity inferred from interleaved randomized benchmarking [9,10]. Therefore, in early work characterizing two-qubit gate fidelities for superconducting qubits, the effect of the two-qubit gate projected in single-qubit space was reported instead of the actual two-qubit gate fidelity [12,13]. For semiconductor spin qubits, even though two-qubit Bell states have been prepared [14–17] and simple quantum algorithms were implemented on two silicon spin qubits [15], the implementation issues of conventional randomized benchmarking have long stood in the way of quantifying the two-qubit gate fidelity. These limitations can be overcome by using native gates that compile efficiently [17], which is, however, not always possible. Another option, which we pursue here, is to use a new method called character randomized benchmarking (CRB), which we developed in Ref. [18]. CRB allows us to extract a two-qubit gate fidelity by interleaving the two-qubit gate in a reference sequence consisting of a small number of single-qubit Clifford gates only. This results in lower compilation overhead and thus lower demand on coherence times. Moreover, CRB will generically lead to higher reference fidelities and thus sharper bounds on the fidelity of the interleaved gate. As a final benefit, CRB provides detailed information on separate decay channels and error correlations.

Here we supplement standard randomized benchmarking with character randomized benchmarking for a comprehensive study of all the relevant gate fidelities of two electron spin qubits in silicon quantum dots, including the single-qubit and two-qubit gate fidelity as well as the effect of cross talk and correlated errors on single-qubit gate fidelities. This work is of strong interest since silicon spin qubits are highly scalable, owing to their compact size (<100 nm pitch), coherence times up to tens of milliseconds, and their ability to leverage existing semiconductor technology [19,20].

II. DEVICE AND QUBIT OPERATION

Figure 1 shows a schematic of the device, a double quantum dot defined electrostatically in a 12-nm-thick Si/SiGe quantum well, 37 nm below the semiconductor surface. The device is cooled to ~ 20 mK in a dilution refrigerator. By applying positive voltages on the accumulation gate, a two-dimensional electron gas is formed in the quantum well. Negative voltages are applied to the depletion gates in such a way that two single electrons are confined in a double well potential [15]. A 617-mT magnetic field is applied in the plane of the quantum well. Two qubits, $Q1$ and $Q2$, are encoded in the Zeeman split states of the two electrons, where spin-up represents $|1\rangle$ and spin-down represents $|0\rangle$.

Single-qubit rotations rely on electric dipole spin resonance (EDSR), making use of the transverse magnetic field gradient from three cobalt micromagnets fabricated on top

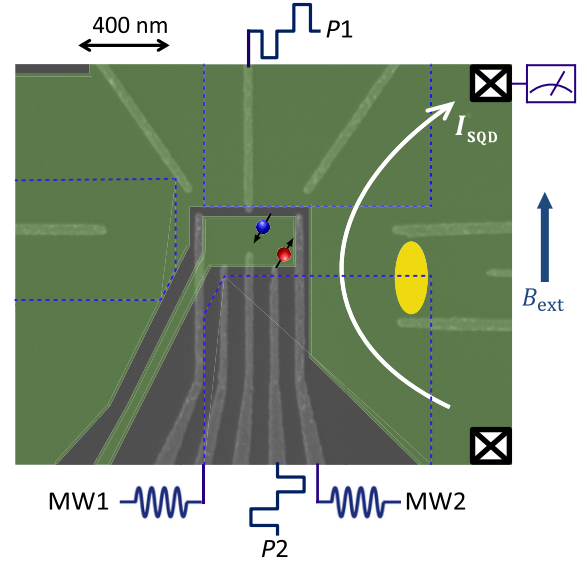


FIG. 1. Device schematic. A double quantum dot is formed in the Si/SiGe quantum well, where two spin qubits $Q1$ (blue spin) and $Q2$ (red spin) are defined. The green-shaded areas show the locations of the accumulation gates on top of the double dot and the reservoir. The blue dashed lines indicate the positions of three Co micromagnets, which form a magnetic field gradient along the qubit region. MW1 and MW2 are connected to two vector microwave sources to perform EDSR for single-qubit gates. The yellow ellipse shows the position of a larger sensing quantum dot (SQD) which is used as a charge sensor for single-shot readout. Plunger gates $P1$ and $P2$ are used to pulse to different positions in the charge stability diagram as needed for initialization, manipulation, and readout, as well as for pulsing the detuning for controlling the two-qubit gate.

of the gate stack [21]. The longitudinal magnetic field gradient leads to well-separated spin resonance frequencies of 18.34 and 19.72 GHz for $Q1$ and $Q2$, respectively. The rotation axis in the \hat{x} - \hat{y} plane is set by the phase of the on-resonance microwave drive, while rotations around the \hat{z} axis are implemented by changing the rotating reference frame in software [22].

We use the CPHASE gate as the native two-qubit gate. An exchange interaction $J(\varepsilon)$ is switched on by pulsing the detuning ε (electrochemical potential difference) between the two quantum dots, such that the respective electron wave functions overlap. Because of the large difference in qubit energy splittings, the flip-flop terms in the exchange Hamiltonian are ineffective and an Ising interaction remains [15,16,23,24]. The resulting time evolution operator in the standard $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ basis is given by

$$U_J(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{iJ(\varepsilon)t/2\hbar} & 0 & 0 \\ 0 & 0 & e^{iJ(\varepsilon)t/2\hbar} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (1)$$

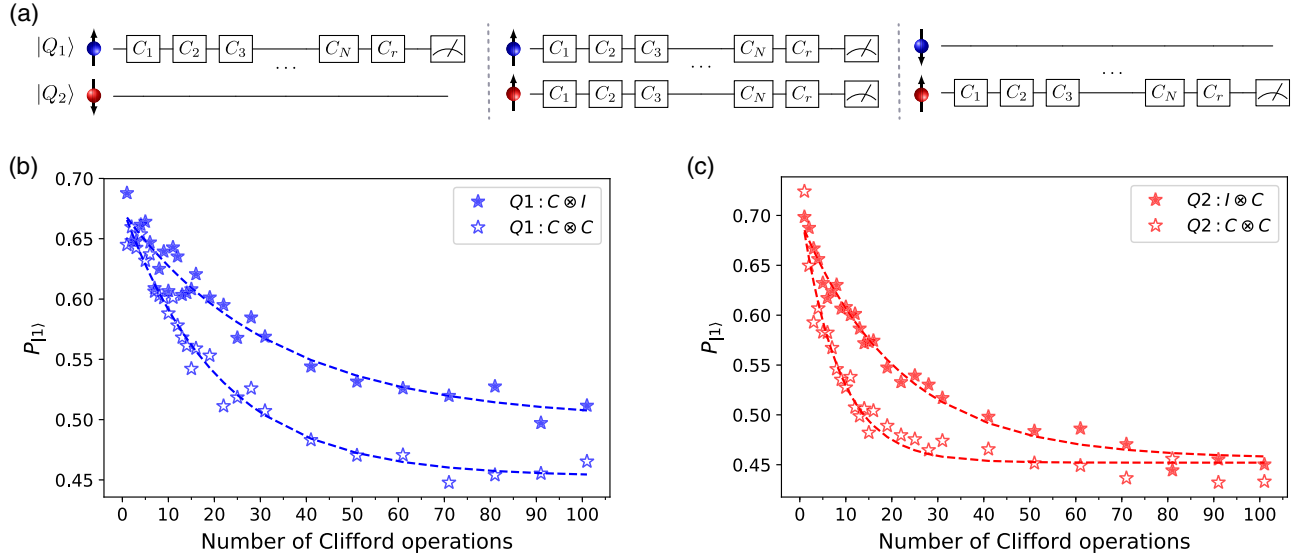


FIG. 2. Individual and simultaneous standard randomized benchmarking. (a) Circuit diagrams for individual single-qubit RB on $Q1$ (left) and $Q2$ (right), and simultaneous single-qubit RB (middle). (b) Probability for obtaining outcome 1 upon measurement in the $\sigma_z \otimes I$ basis as a function of the number of single-qubit Clifford operations. For the blue solid stars, $Q2$ is idle while a Clifford operation is applied to $Q1$ ($C \otimes I$). For the blue open stars, random Clifford operations are applied to $Q1$ and $Q2$ simultaneously ($C \otimes C$). For each data point, we sample 32 different random sequences, which are each repeated 100 times. Dashed lines are fit to the data with a single exponential. A constant offset of -0.06 is added to the open data points in order to compensate for a change in readout fidelities between the two data sets, making comparison of the two traces easier. Without SPAM errors, the data points would decay from 1 to 0.5. (c) Analogous single-qubit RB data for $Q2$, with $Q1$ idle (red solid stars) and subject to random Clifford operations (red open stars). A constant offset of -0.05 is added to the open data points. Throughout, single-qubit Clifford operations are generated by the native gate set $\{I, X(\pi), Y(\pm\pi), X(\pm\pi/2), Y(\pm\pi/2)\}$.

Choosing $t = \pi\hbar/J(\epsilon)$ and adding single-qubit \hat{z} rotations on both qubits, we obtain a controlled-Z (CZ) operator

$$Z_1\left(-\frac{\pi}{2}\right)Z_2\left(-\frac{\pi}{2}\right)U_J\left(\frac{\pi\hbar}{J(\epsilon)}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (2)$$

with $Z_i(\theta)$ a \hat{z} rotation of qubit i over an angle θ .

Spin initialization and single-shot readout of $Q2$ are realized by energy-selective tunneling [25]. $Q1$ is initialized to its ground spin state by fast spin relaxation at a hot spot [26]. For readout, the state of $Q1$ is mapped onto $Q2$ using a conditional π rotation [15,24], which enables extracting the state of $Q1$ by measuring $Q2$. Further details on the measurement setup are provided in Appendix A.

III. INDIVIDUAL AND SIMULTANEOUS RANDOMIZED BENCHMARKING

In standard randomized benchmarking, sequences of random multiqubit Clifford operations are applied to a number of target qubits, followed by a final Clifford operation that, in the absence of errors, maps the qubits' state back to the initial state. A “survival probability” is then extracted by measuring the system in a basis

containing the initial state. Repeating this procedure for many random sequences of Clifford gates and averaging over the resulting survival probabilities yields an average survival probability. This in effect symmetrizes the noise process in the system such that the qubits are effectively subject to a depolarizing channel. The average survival probability P then decays exponentially with the number of Clifford operations m , under broad assumptions [27–29]. By fitting the decay curve to

$$P = A\alpha^m + B, \quad (3)$$

where only A and B depend on the state preparation and measurement, the average fidelity of a Clifford operation can be extracted in terms of the depolarizing parameter α as

$$F_{\text{av}} = 1 - (1 - \alpha)\frac{d-1}{d}, \quad (4)$$

where $d = 2^N$ and N is the number of qubits.

In the present two-qubit system, we first perform standard RB on each individual qubit (filled data points in Fig. 2), finding $F_{\text{av}} = 98.50 \pm 0.05\%$ for $Q1$ and $F_{\text{av}} = 97.72 \pm 0.03\%$ for $Q2$ (all uncertainties are standard deviations). By dividing the error rate over the average number of single-qubit gates needed for a Clifford operation, we extract average single-qubit gate fidelities of $99.20 \pm 0.03\%$ for $Q1$ and $98.79 \pm 0.02\%$ for $Q2$.

In practice, single-qubit gates are normally applied simultaneously to multiple qubits during quantum computation. Driving one qubit can lead to errors on the other qubit gates due to cross talk effects such as an undesired excitation, residual qubit coupling, or heating effects. In order to assess the effects of cross talk, we next perform single-qubit RB while simultaneously applying random Clifford operations to the other qubit (Fig. 2, open data points). Following Ref. [8], we denote the corresponding depolarizing parameter for qubit i while performing random gates on qubit j as α_{ij} . In contrast to standard RB which is insensitive to SPAM errors, we have to assume here that operations on one qubit do not affect the readout fidelity of the other qubit [8]. Comparing with individual single-qubit randomized benchmarking results, we find that simultaneous RB decreases the average single-qubit gate fidelity for $Q1$ by 0.44% to $98.76 \pm 0.02\%$ while the fidelity for $Q2$ decreases by 1.85% to $96.94 \pm 0.05\%$. Since the difference in qubit frequencies of 1.38 GHz is almost 3 orders of magnitude larger than the Rabi frequencies (~ 2 MHz), this cross talk is not due to limited addressability. Furthermore, the cross talk on $Q2$ persists when the drive on $Q1$ is applied off-resonantly; hence, it is an effect of the excitation and not a result of performing random gates on $Q1$. Attempting to understand how the excitation leads to undesired cross talk, we performed detailed additional studies (see Ref. [15] and Appendix F), ruling out a number of other possible sources of cross talk, including the ac Stark effect and residual coupling between the qubits. In addition, we observed an anomalous frequency shift on $Q2$ while driving $Q1$ [15], although larger frequency shifts did not correlate with a faster Rabi decay of $Q2$. Increasing the driving power did not yield a higher quality factor of the Rabi oscillations, as reported in Ref. [30], which indicates that heating could contribute to their decay. Whether heating from driving $Q1$ also reduces the fidelity of $Q2$ needs to be confirmed by further studies. Finally, cross talk in the experimental setup is likely to be symmetric, so the observed asymmetry indicates that the microscopic details of the quantum dots must also play a role.

IV. TWO-QUBIT RANDOMIZED BENCHMARKING

To characterize two-qubit gate fidelities, the Clifford group is expanded to a four-dimensional Hilbert space. We first implement standard two-qubit RB, sampling Clifford operations from the 11 520 elements in the two-qubit Clifford group. Each two-qubit Clifford operation is compiled from single-qubit rotations and the two-qubit CZ gate, requiring on average 8.25 single-qubit rotations around \hat{x} or \hat{y} and 1.5 CZ gates. The measured probability to successfully recover the initial state is shown in Fig. 3. From a fit to the data using Eq. (3) and applying Eq. (4), we extract an average two-qubit Clifford fidelity F_{av} of $82.10 \pm 2.75\%$.

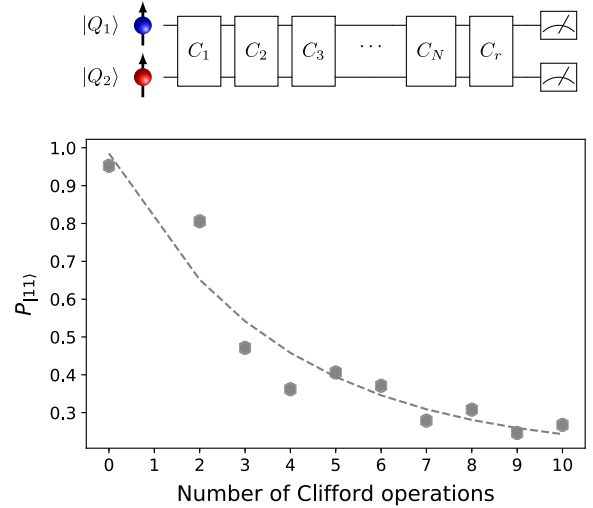


FIG. 3. Two-qubit Clifford randomized benchmarking. Probability for obtaining outcome 11 upon measurement in the $\sigma_z \otimes \sigma_z$ basis, starting from the initial state $|11\rangle$, as a function of the number of two-qubit Clifford operations. As the native gate set, we use $\{I, X(\pi), Y(\pm\pi), X(\pm\pi/2), Y(\pm\pi/2), CZ\}$. The elements of the two-qubit Clifford group fall in four classes of operations: the parallel single-qubit Clifford class, the CNOT-like class, the ISWAP-like class, and the SWAP-like class. They are compiled by single-qubit gates plus 0, 1, 2, and 3 CZ gates respectively. For each data point, we sample 30 random sequences, which are each repeated 100 times. The dashed line is a fit to the data with a single exponential. The data point for a single Clifford operation is missing because we absorb the recovery Clifford gate into the last random Clifford gate, in order to minimize dephasing effects. Here the probability has been normalized to remove initialization and readout errors.

The large number of native gates needed to implement a single two-qubit Clifford gate leads to a fast saturation of the decay, within about eight Clifford operations, leading to a large uncertainty on the two-qubit Clifford fidelity estimate. In addition, this fast saturation makes the randomized benchmarking data vulnerable to deviations from a single exponential decay due to different error processes being associated to each applied gate, possibly biasing the fidelity estimate [29,31,32]. Importantly, interleaving a specific gate in a fast decaying reference sequence also yields a rather unreliable estimate of the interleaved gate fidelity. In the present case, interleaving a CZ gate in the reference sequence of two-qubit Clifford operations is not a viable strategy to extract the CZ gate fidelity. Furthermore, the compilation of Clifford gates into two different types of native gates—single-qubit gates and the CZ gate—makes it impossible to confidently extract the fidelity of any of the native gates, such as the CZ gate, by itself. This is different from a recent experiment on silicon spin qubits where only a single physical native gate was used, the conditional rotation, in which case the error per Clifford operation can be divided by the average number of conditional rotations per Clifford operation for estimating the error per conditional rotation [17].

As a first step to obtain quantitative information on the CZ gate fidelity, we implement a simplified version of interleaved RB, which provides the fidelities of the two-qubit gate projected in various single-qubit subspaces, as was done earlier for superconducting transmon qubits [12] and hybrid gatemon qubits [13]. In this protocol, the CZ gate is interleaved in a reference sequence of single-qubit Clifford operations. When applying a CZ gate, we can (arbitrarily) consider one qubit the control qubit and the other the target qubit. When the control qubit is $|1\rangle$, the target qubit ideally undergoes a π rotation around the \hat{z} axis. With the control in $|0\rangle$, the target qubit ideally remains fixed (identity operation). Therefore, projected in the subspace corresponding to the target qubit, this protocol interleaves either a $Z(\pi)$ rotation or the identity operation in a single-qubit RB reference sequence applied to the target qubit. The decay of the return probability for interleaved RB is also expected to follow Eq. (3). The fidelity of the interleaved gate is then found from the depolarizing parameter α for the interleaved and reference sequence, as

$$F_{\text{gate}} = 1 - \left(1 - \frac{\alpha_{\text{interleaved}}}{\alpha_{\text{reference}}}\right) \frac{d-1}{d}. \quad (5)$$

From the experimental data, we find CZ fidelities projected in single-qubit space of 91%–95%, depending on which qubit acts as the control qubit for the CZ, and which eigenstate it is in (see Appendix E).

V. CHARACTER RANDOMIZED BENCHMARKING

In order to properly characterize the two-qubit CZ fidelity, we experimentally demonstrate a new approach to RB called character randomized benchmarking [18]. CRB is a powerful generic method that extends randomized benchmarking in a rigorous manner, making it possible to extract average fidelities from groups beyond the multi-qubit Clifford group while keeping the advantages of standard RB such as resistance to SPAM errors. The generality of CRB allows one to start from (a subset of) the natives gates of a particular device and then design a RB experiment tailored to that set. This can strongly reduce compilation overhead and gate-dependent noise, a known nuisance factor in standard RB [29,31,32]. Moreover, since the accuracy of interleaved randomized benchmarking depends on the fidelity of the reference gates [9,10], performing (through CRB) interleaved RB with a reference group generated by high-fidelity gates can significantly improve the utility of interleaved RB.

Character randomized benchmarking requires us to average over two groups (the second one usually being a subgroup of the first). The first group is the “benchmark group.” It is for the gates in this group that CRB yields the average fidelity. The second group is the “character group.”

CRB works by performing standard randomized benchmarking using the benchmark group but augments this by adding a random gate from the character group before each RB gate sequence. By averaging over this extra random gate, but weighting the average by a special function known from representation theory as a character function, it guarantees that the average over random sequences can always be fitted to a single exponential decay, even when the benchmark group is not the multiqubit Clifford group, and even in the presence of SPAM errors.

Guided by the need for high reference fidelities, we choose for our implementation of CRB the benchmark group to be the parallel single-qubit Clifford group ($C \otimes C$, the same as in standard simultaneous single-qubit RB) and the two-qubit Pauli group as the character group (see Ref. [18] for more information on why this is a good choice for the character group). It is nontrivial that the $C \otimes C$ group allows us to get information on two-qubit gates, since parallel single-qubit Clifford operations cannot fully depolarize the noise in the full two-qubit Hilbert space. In fact, for simultaneous single-qubit RB there are three depolarizing channels, each acting in a different subspace of the Hilbert space of density matrices, spanned by $I \otimes \sigma_i$, $\sigma_i \otimes I$, and $\sigma_i \otimes \sigma_i$, with I the identity operator and σ_i one of the Pauli operators. The three decay channels are reflected in the recovery probability for the final state, which is now described by

$$P_{C \otimes C} = A_1 \alpha_{1|2}^m + A_2 \alpha_{2|1}^m + A_{12} \alpha_{12}^m + B, \quad (6)$$

where $\alpha_{i|j}$ is again the depolarizing parameter for qubit i while simultaneously applying random Clifford operations to qubit j , and α_{12} is the depolarizing parameter for the two-qubit parity ($\{|00\rangle, |11\rangle\}$ versus $\{|01\rangle, |10\rangle\}$). We note that if the errors acting on both qubits are uncorrelated, then $\alpha_{12} = \alpha_{1|2} \alpha_{2|1}$ [8]. The question now is how to separate the three decays. Fitting the data using a sum of three exponentials will be very imprecise. Existing approaches combine the decay of specific combinations of the probabilities of obtaining 00, 01, 10, and 11 upon measurement, but suffer from SPAM errors [8]. As discussed above, CRB offers a clean procedure for extracting the individual decay rates that is immune to SPAM errors and does not incur additional overhead.

Concretely, CRB here proceeds as follows. (1) The two-qubit system is initialized to $|00\rangle$, then (2) one random Pauli operator on each qubit is applied to prepare the system in a state $|\phi_1 \phi_2\rangle$ (one of $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$), followed by (3) a random sequence of simultaneously applied single-qubit Clifford operators. In practice, the random Pauli operator is absorbed in the first Clifford operation, making the Pauli gates effectively noise-free. A final Clifford operation is applied which ideally returns the system to the state $|\phi_1 \phi_2\rangle$ and, finally, (4) both qubits are measured. Note that, contrary to standard RB, CRB does not

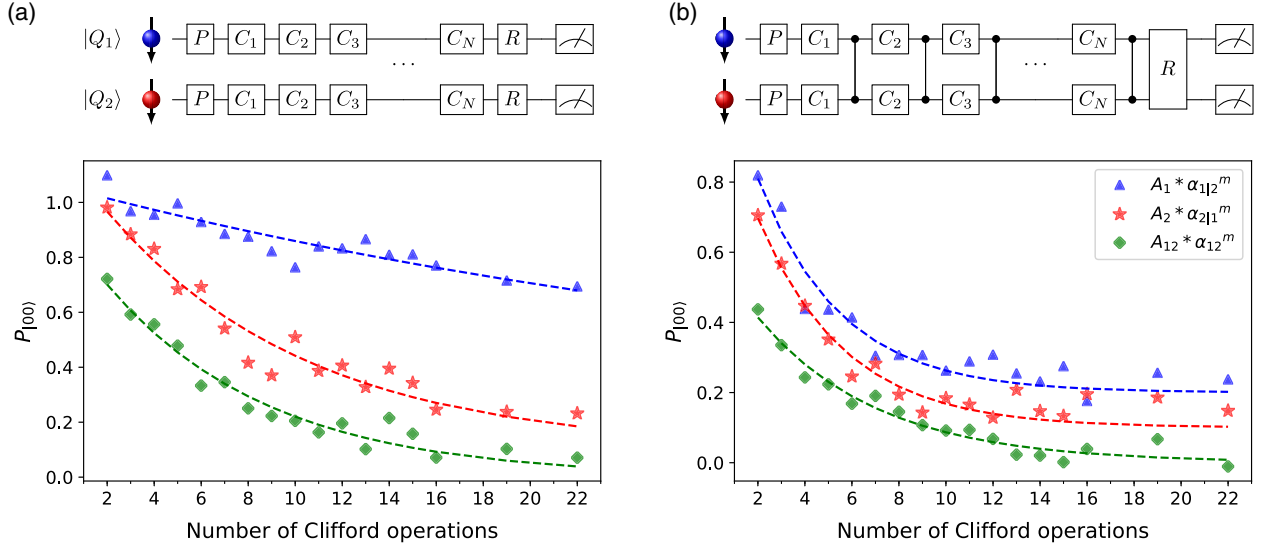


FIG. 4. Character randomized benchmarking. (a) Reference CRB experiment. The probabilities P_1 (blue triangles), P_2 (red stars), and P_3 (green diamonds), obtained starting from the initial state $|00\rangle$ followed by a Pauli operation, as a function of the number of subsequent single-qubit Clifford operations simultaneously applied to both qubits (see the schematic of the pulse sequence, note that the Clifford gates applied on Q_1 and Q_2 are sampled independently and are thus generally different). As the native gate set, we use $\{I, X(\pi), Z(\pm\pi), X(\pm\pi/2), Z(\pm\pi/2), CZ\}$. For each of the 16 Pauli operators, we apply 40 different random sequences, each with 20 repetitions. The dashed lines are fits to the data with a single exponential. Without SPAM errors, the data points would decay from 1 to 0. (b) Interleaved CRB experiment. This experiment is performed in an analogous way to the reference CRB experiment, but with a two-qubit CZ gate interleaved after each Clifford pair, as seen in the schematic of the pulse sequence. In both panels, the traces are offset by an increment of 0.1 for clarity and the probabilities have been normalized to remove initialization and readout errors.

always measure the probability of returning to the initial state. Each random sequence is repeated to collect statistics on the probability $P_{\phi_1\phi_2}$ of obtaining measurement outcome 00 when starting from $|\phi_1\phi_2\rangle$ (note that each $P_{\phi_1\phi_2}$ averages over four Pauli operations). We combine these probabilities according to their character (see Appendix B for more details) to obtain three fitting parameters:

$$\begin{aligned} P_1 &= P_{00} - P_{01} + P_{10} - P_{11}, \\ P_2 &= P_{00} + P_{01} - P_{10} - P_{11}, \\ P_3 &= P_{00} - P_{01} - P_{10} + P_{11}. \end{aligned} \quad (7)$$

Each of these three fitting parameters is expected to decay as a single exponential, isolating one of the decay channels in Eq. (6):

$$\begin{aligned} P_1 &= A_1 \alpha_{12}^m, \\ P_2 &= A_2 \alpha_{21}^m, \\ P_3 &= A_{12} \alpha_{12}^m. \end{aligned} \quad (8)$$

Note that there is no constant offset B . This is also a feature of CRB. The three experimentally measured probabilities are shown in Fig. 4(a). These contain a lot of useful information, including not only the separate depolarizing parameters but also the averaged CRB reference fidelity and information on error correlations. The blue (red) curve shows the decay in the subspace corresponding to Q_1 (Q_2),

spanned by $\sigma_i \otimes I$ ($I \otimes \sigma_i$). The green curve shows the decay in the subspace spanned by $\sigma_i \otimes \sigma_j$. This decay can be interpreted as the parity decay. The fitted depolarizing parameters are $\alpha_{12} = 0.9738 \pm 0.0008$, $\alpha_{21} = 0.8902 \pm 0.0020$, and $\alpha_{12} = 0.8652 \pm 0.0022$.

The average CRB depolarizing parameter can be found from the separate depolarizing parameters as

$$P = \frac{3}{15} \alpha_{12} + \frac{3}{15} \alpha_{21} + \frac{9}{15} \alpha_{12}, \quad (9)$$

where the weights are proportional to the dimension of the corresponding subspaces of the 16-dimensional Hilbert space of two-qubit density matrices. We obtain a reference CRB fidelity of $91.9 \pm 0.1\%$, which represents the fidelity of two simultaneous single-qubit Clifford operators ($C \otimes C$) in the full two-qubit space.

Finally, from the three depolarizing parameters in Eq. (6), we can infer to what extent errors occur independently on each qubit or exhibit correlations between the two qubits. The fact that $\alpha_{12} - \alpha_{12}\alpha_{21} = -0.0017 \pm 0.0031$ indicates that the errors are essentially independent.

Next we perform the interleaved version of CRB, for which we insert a CZ gate after each single-qubit Clifford pair. Figure 4(b) shows the three corresponding experimentally measured decays. The fitting parameters we extract now reflect the combined errors from a single-qubit Clifford pair followed by a CZ gate. The fitted

depolarizing parameters are $\alpha_{1|2} = 0.7522 \pm 0.0060$, $\alpha_{2|1} = 0.7623 \pm 0.0053$, and $\alpha_{12} = 0.8226 \pm 0.0030$. As can be expected, the three decays lie closer together than those for reference CRB: not only does the additional CZ gate contribute directly to all three decays, it also mixes the three subspaces. From the depolarizing parameters in interleaved and reference CRB measurement, we use Eq. (5) to isolate the fidelity of the CZ gate, now in two-qubit space as desired, yielding $92.0 \pm 0.5\%$.

The dominant errors in the CZ gate arise from nuclear spin noise and charge noise. In natural silicon, the abundance of Si^{29} atoms is about 4.7%, and the Si^{29} nuclear spins dephase the electron spin states due to the hyperfine interaction [19]. Charge noise modulates the overlap of the two electron wave functions, and thus also the two-qubit coupling strength. In the present device, we could not obtain a sufficiently strong coupling J at the symmetry point where the coupling strength is to first order insensitive to the detuning of the double dot potential [33,34]; hence, charge noise directly (to first order) affects the two-qubit coupling strength.

VI. CONCLUSIONS

Character randomized benchmarking provides a new method to effectively characterize multiqubit behavior. It combines the advantages of simultaneous randomized benchmarking and interleaved randomized benchmarking, and gives tighter bounds on the fidelity number than standard interleaved randomized benchmarking due to its simpler compilation. CRB is useful in a wide variety of settings, far beyond the particular case studied here. The general approach to exploiting CRB is to start from a set of native gates that can be implemented easily and with high fidelity, and to construct a suitable reference sequence based on this set. The decay for the reference sequence contains any number of exponentials, which can be separated without suffering from SPAM errors and which provide relevant additional information, in the present case on the fidelity of simultaneously applied gates, cross talk, and on noise correlations. Comparison with interleaved CRB allows one to extract the fidelity of specific gates of interest.

We perform the first comprehensive study of the single-qubit, simultaneous single-qubit, and two-qubit gate fidelities for semiconductor qubits, where the use of CRB, which allows for a compact reference sequence, was essential for extracting a reliable two-qubit gate fidelity. Summarizing, independent single-qubit gate fidelities are around 99% in this system, these drop to 98.8% for qubit 1 and to 96.9% for qubit 2 when simultaneously performing random gates on the other qubit, and the two-qubit CZ fidelity is around 92%. We expect that by working in an isotopically purified $\text{Si}^{28}/\text{SiGe}$ substrate and performing the two-qubit gate at the symmetry point, a CZ gate fidelity above the fault-tolerant threshold ($>99\%$) can be reached.

A recent report on the fidelity of controlled rotations in Si/SiO₂ quantum dots already comes close to this threshold [17]. With further improvements in charge noise levels, two-qubit gate fidelities above 99.9% are in reach.

All the data and the analysis scripts are uploaded to Zenodo: [35].

ACKNOWLEDGMENTS

This research was sponsored by the Army Research Office (ARO) under Grants No. W911NF-17-1-0274 and No. W911NF-12-1-0607. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ARO or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein. Development and maintenance of the growth facilities used for fabricating samples is supported by DOE (DE-FG02-03ER46028). We acknowledge the use of facilities supported by NSF through the University of Wisconsin-Madison MRSEC (DMR-1121288). J.H. and S.W. are supported by a NWO VIDI Grant (S.W.), an ERC Starting Grant (S.W.), and the NWO Zwaartekracht QSC. We acknowledge useful discussions with the members of the Vandersypen group, and technical assistance by R. Schouten and R. Vermeulen.

APPENDIX A: MEASUREMENT SETUP

The measurement setup is the same as the one used in Ref. [15]. We summarize here a few key points. The gates $P1$ and $P2$ are connected to arbitrary waveform generators (AWG, Tektronix 5014C) via coaxial cables. Applying dc voltage pulses to these two gates moves the system through different positions in the charge stability diagram for initialization, operation, and readout. Voltage pulses applied to these two gates are used to pulse the detuning between the two quantum dots, thereby turning on and off the controlled-phase gate. Gates MW1 and MW2 are connected to vector microwave sources (Keysight E8267D) for achieving EDSR. Each microwave source has two I/Q input channels, connected to two channels on the master AWG, which controls the clock of the entire system and triggers all the other instruments. The frequency, phase, and duration of the microwave bursts are thus controlled by I/Q modulation. In addition, we use pulse modulation to obtain a larger on/off ratio of the microwave bursts than is possible using I/Q modulation only. A digitizer card (Spectrum M4i.44) installed inside the measurement computer is used to record the current traces of the sensing quantum dot at a sampling rate ~ 60 kHz. Each time trace is converted into a single bit value (0 or 1) by the measurement computer using

threshold detection. The average over many repetitions gives us the spin-down and spin-up probabilities (0 and 1).

APPENDIX B: MATHEMATICAL BACKGROUND OF CRB

Character randomized benchmarking is a generic method for performing randomized benchmarking with finite groups other than the multiqubit Clifford group. As mentioned in the main text, CRB requires the user to specify two finite groups: the benchmark group and the character group. In this work we chose the benchmark group to be the simultaneous single-qubit Clifford group on two qubits and the character group to be the two-qubit Pauli group. Standard RB and CRB rely on the framework of representation theory. Representation theory gives us powerful tools to interpret averages over group elements, which form a core part of the randomized benchmarking protocol. In particular, we can use a result called Schur's lemma. In the context of randomized benchmarking, Schur's lemma states that, assuming for simplicity that every gate is subject to an identical noise map \mathcal{E} , the average noisy RB operator \mathcal{M} (here the average is taken over the group elements) is of the form

$$\mathcal{M} := \sum_{\substack{G_1, \dots, G_m \\ \in \mathcal{C} \otimes \mathcal{C}}} G_{inv} \mathcal{E} G_m \cdots \mathcal{E} G_1$$

$$= \begin{bmatrix} 1 & & & & \\ & \alpha_{1|2} \mathbb{I}_{1|2} & & & \\ & & \alpha_{2|1} \mathbb{I}_{2|1} & & \\ & & & \alpha_{12} \mathbb{I}_{12} & \\ & & & & 1 \end{bmatrix}^m,$$

where we are describing all quantum channels in the Pauli transfer matrix picture, i.e., $\mathcal{M}_{i,j} = \text{Tr}[\sigma_i \mathcal{M}(\sigma_j)]/2$, where σ_i, σ_j are Pauli matrices. One can think of the matrix entry $\mathcal{M}_{i,j}$ as describing how much the noise map \mathcal{M} maps the generalized Bloch sphere axis labeled σ_j to the one labeled σ_i . The submatrices $\mathbb{I}_{1|2}$, $\mathbb{I}_{2|1}$, and \mathbb{I}_{12} of the matrix \mathcal{M} are defined as the identity matrix on the sets of two-qubit Pauli operators of the form $\{\sigma_i \otimes I\}$, $\{I \otimes \sigma_i\}$, and $\{\sigma_i \otimes \sigma_j\}$, respectively. We would like to estimate the numbers $\alpha_{1|2}$, $\alpha_{2|1}$, and α_{12} individually in a way that does not depend on state preparation and measurement. To do this CRB adds an extra average over another group called the character group, which we choose to be the two-qubit Pauli group.

This average is weighted by a so-called character function. This average over the Pauli group projects any initial state onto a single axis of the Bloch sphere. Which axis is projected on depends on the character function used for the weights. By selecting the correct Bloch sphere axes, we can single out the individual blocks of the matrix \mathcal{M} . In order to isolate the parameter $\alpha_{1|2}$ we choose to project onto the Bloch sphere axis associated to $\sigma_z \otimes I$. Concretely this means that the character averaged RB operator \mathcal{M} becomes

$$\sum_{\sigma \in P_2} \chi_{\sigma_z I}(\sigma) \sum_{\substack{G_1, \dots, G_m \\ \in \mathcal{C} \otimes \mathcal{C}}} G_{inv} \mathcal{E} G_m \cdots \mathcal{E} G_1 \sigma = \mathcal{M}^m \mathcal{P}_{ZI},$$

where the function $\chi_{\sigma_z I}(\sigma)$ is given in the first row of Table I and the matrix $\mathcal{P}_{\sigma_z I}$ has all zero entries except on the diagonal entry corresponding to the Pauli $\sigma_z \otimes I$. By matrix multiplication we see that $\mathcal{M}^m \mathcal{P}_{\sigma_z I} = \alpha_{1|2}^m \mathcal{P}_{\sigma_z I}$. This means that the average measured survival probability in CRB, with input state ρ and measurement operator Q , is of the form

$$\sum_{\sigma \in P_2} \chi_{ZI}(\sigma) \sum_{\substack{G_1, \dots, G_m \\ \in \mathcal{C} \otimes \mathcal{C}}} \text{Tr}[Q G_{inv} \mathcal{E} G_m \cdots \mathcal{E} G_1(\sigma(\rho))] = A \alpha_{1|2}^m,$$

where A is a function of Q and ρ . Similarly, we can obtain estimates $\alpha_{2|1}$ and α_{12} by constructing projectors onto the Pauli operators $I \otimes \sigma_z$ and $\sigma_z \otimes \sigma_z$, respectively. The character functions for these projectors are given in rows two and three of Table I, respectively.

As noted in the main text, CRB is a generic procedure, which can be used beyond its application in this paper. Another notable example of where we suspect CRB can offer a benefit is when the device native gates are not single-qubit gates but rather two-qubit gates, as happens in Ref. [17]. In this case, compiling multiqubit Clifford gates is very cumbersome. In the theoretical RB literature benchmarking groups are discussed that are more suitable to this scenario, such as the CNOT-dihedral group (for native CNOT gates) [36] and the real Clifford group (for native CPHASE gates) [37]. Both of these groups lead to benchmarking data that mix two exponential decays, but using the CRB approach these can be fitted individually in a reliable manner (in both cases the Pauli group is a good choice for character group, see the example in Ref. [18] for more information).

TABLE I. Values for the character function $\chi_P(\sigma)$ for $P \in \{(\sigma_z \otimes I), (I \otimes \sigma_z), (\sigma_z \otimes \sigma_z)\}$.

$\sigma \backslash P$	I	$\sigma_z I$	$I \sigma_z$	$\sigma_z \sigma_z$	$\sigma_x I$	$I \sigma_x$	$\sigma_x \sigma_x$	$\sigma_y I$	$I \sigma_y$	$\sigma_y \sigma_y$	$\sigma_z \sigma_x$	$\sigma_x \sigma_z$	$\sigma_z \sigma_y$	$\sigma_y \sigma_z$	$\sigma_x \sigma_y$	$\sigma_y \sigma_x$
$\sigma_z I$	1	1	1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	-1
$I \sigma_z$	1	1	1	1	1	-1	-1	1	-1	-1	-1	1	-1	1	-1	-1
$\sigma_z \sigma_z$	1	1	1	1	-1	-1	1	-1	-1	1	-1	-1	-1	-1	1	1

TABLE II. Compilation of the single-qubit Clifford group with X/Y rotations and X/Z rotations. Here, $(-)K$ and $(-)K^2$ denote $K(\pm\pi/2)$ and $K(\pm\pi)$ gates ($K = X, Y, Z$), respectively.

Class	X/Y generation	X/Z generation
Pauli	I	$X, -X$
	X^2	X^2
	Y^2	$-Z, X^2, Z$
	Y^2, X^2	X, Z^2, X
$2\pi/3$	X, Y	$X, -Z, X, Z$
	$X, -Y$	$X, Z, X, -Z$
	$-X, Y$	$-X, -Z, X, Z$
	$-X, -Y$	$-X, Z, X, -Z$
	Y, X	$-Z, X, Z, X$
	$Y, -X$	$-Z, X, Z, -X$
	$-Y, X$	$Z, X, -Z, X$
	$-Y, -X$	$Z, X, -Z, -X$
$\pi/2$	X	$-Z, X, Z, X, -Z$
	$-X$	$Z, -X, -Z, -X, Z$
	Y	$X, Z, -X$
	$-Y$	$X, -Z, -X$
	$-X, Y, X$	$-X, Z^2, -X, -Z$
	$-X, -Y, X$	$-X, -Z^2, -X, Z$
Hadamard	X^2, Y	$X, -Z, X$
	$X^2, -Y$	X, Z, X
	Y^2, X	$-Z, X, Z, X, Z$
	$Y^2, -X$	$-Z, X, Z, -X, -Z$
	X, Y, X	X^2, Z
	$-X, Y, -X$	$-X^2, -Z$

APPENDIX C: EXPERIMENTAL DETAILS FOR CRB

The single-qubit Clifford group is commonly generated by the gate set $\{I, X(\pi), Y(\pm\pi), X(\pm\pi/2), Y(\pm\pi/2)\}$. In our experiment, we perform Z rotations by changing a qubit's reference frame in software [22], which makes Z rotations error-free. To benefit from this, we generate the single-qubit Clifford group by the gate set $\{I, X(\pi), Z(\pm\pi), X(\pm\pi/2), Z(\pm\pi/2)\}$ instead, as shown in Table II. Furthermore, we keep the Rabi frequency the same for all the X rotations; thus a $X(\pi)$ gate has twice the duration of a $X(\pi/2)$ gate. Combined with using X - Z compilation, we can keep the duration for all the 24 Clifford operations the same thereby avoiding any unnecessary idle time which would quickly dephase the qubits.

APPENDIX D: COMPARISON OF STANDARD AND CHARACTER INTERLEAVED TWO-QUBIT RB

Although it often goes unmentioned, the estimate for the fidelity of an interleaved gate given in Eq. (5) is

only exact when the qubit noise is exactly depolarizing. In the presence of other types of noise (such as dephasing or calibration errors), this number gives only upper and lower bounds on the fidelity of the interleaved gate. First upper and lower bounds were given in Ref. [9], and recently optimal upper and lower bounds were given in Ref. [10]. These bounds depend strongly on the fidelity of the gates in the reference sequence; in particular, they scale as $O(\sqrt{1 - \alpha_{\text{ref}}})$, where α_{ref} is the reference RB decay constant. This means that our implementation of CRB, which uses only single-qubit gates for the reference experiment, has a significant advantage over standard two-qubit interleaved RB also in this respect. We can illustrate this advantage by considering a hypothetical standard two-qubit interleaved experiment with interleaved CZ gate. Recall from Eq. (3) that standard two-qubit RB (here considered as a reference experiment) yielded a reference fidelity of 82% and thus a depolarizing parameter of $\alpha_{2,\text{ref}} = 0.73$ (suppressing uncertainty for the sake of this exercise). Assuming an interleaved CZ fidelity of 92% (which is what we extracted from the CRB experiment) and assuming that the error on a hypothetical interleaved two-qubit RB experiment scales multiplicatively (optimistic given the possibility of calibration errors), we estimate that a hypothetical two-qubit interleaved RB experiment would have a depolarizing parameter of $\alpha_{2,\text{int}}$. Using the optimal bounds calculated in Ref. [10], this would mean we can guarantee only that the fidelity of the interleaved gate lies in the range [0.58, 1]. From the CRB experiment we can, however, guarantee that the fidelity of the interleaved gate lies in the range [0.69, 1], a significant improvement even in the absolute worst case scenario discussed in Ref. [10]. We would also like to note that the bounds given in Refs. [9,10] significantly overestimate the range of possible interleaved gate fidelities if more is known about the noise process. If, for instance, the noise on the reference gates is assumed to be dominated by stochastic errors (as opposed to coherent errors due to miscalibration), then the upper and lower bounds can be made significantly tighter. This coincides with experimental consensus that interleaved RB generally gives good estimates of the interleaved gate fidelity. However, since single-qubit gates will typically suffer less from calibration errors than compiled two-qubit gates, we argue that interleaved CRB will yield sharper upper and lower bounds on the interleaved gate fidelity than standard interleaved RB when more is known about the noise process.

APPENDIX E: INTERLEAVED RB PROJECTED IN SINGLE-QUBIT SPACE

Figure 5 shows experimental results for the experiment discussed in the main text where a CZ gate is interleaved in a standard single-qubit RB sequence

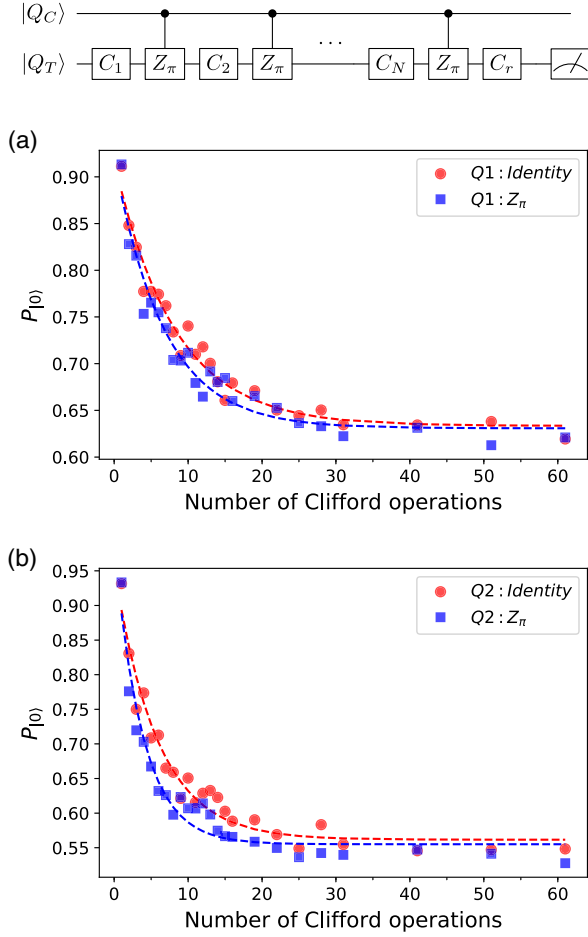


FIG. 5. Interleaved randomized benchmarking projected in single-qubit space. (a) Probability for obtaining outcome 0 upon measurement in the $\sigma_z \otimes I$ basis as a function of the number of single-qubit Clifford operations, interleaved with the CZ operation. For the red circles (blue squares), Q_2 is in $|0\rangle$ ($|1\rangle$) so Q_1 is expected to undergo the identity operation [a $Z(\pi)$ rotation]. For each data point, we sample 30 different random sequences for each Clifford number, which are each repeated 100 times. Dashed lines are fits to the data with a single exponential. A constant offset of $+0.045$ is added to the blue data points. (b) Analogous data for Q_2 .

applied to one qubit, while the other qubit is in either $|0\rangle$ or $|1\rangle$. This experiment provides the CZ fidelity projected in single-qubit space [12,13], summarized in Table III for the four possible cases.

TABLE III. CZ fidelities for different target qubits (Q_T) and different states of the control qubit (Q_C).

Q_C	Q_C state	Q_T	Operation	Fidelity
Q_1	$ 0\rangle$	Q_2	I	$94.62 \pm 0.24\%$
Q_1	$ 1\rangle$	Q_2	$Z(\pi)$	$90.79 \pm 0.38\%$
Q_2	$ 0\rangle$	Q_1	I	$95.50 \pm 0.20\%$
Q_2	$ 1\rangle$	Q_1	$Z(\pi)$	$94.38 \pm 0.25\%$

APPENDIX F: CROSS TALK

We here provide more information on the cross talk effects that occur on one qubit when applying a microwave drive to the other (see also Ref. [15] and the Supplemental Material therein). First, when we perform spectroscopy on Q_2 while driving Q_1 , we find that the frequency of Q_2 shifts by of the order of 2 MHz (depending on the power applied to Q_1). We compensate for this known frequency shift by shifting the drive frequency applied to Q_2 when we simultaneously drive Q_1 . We note that a frequency shift by a known amount is not expected to contribute to decoherence. However, Fig. 6 shows Rabi oscillations for both qubits in the absence and presence of an excitation to the other qubit. Clearly, when simultaneously driving Rabi oscillations on both qubits, we find a faster decay on Q_2 compared to driving Q_2 by itself. The effect of simultaneous driving on Q_1 is less pronounced. This is consistent with the observed effects of simultaneous driving on the measured single-qubit gate fidelities reported in the main text. The cross talk effect on Q_2 persists when the drive on Q_1 is applied off-resonantly or when dot 1 is emptied. We do note that the microwave power used to drive Q_1 (~ 20 dBm) is substantially higher than that

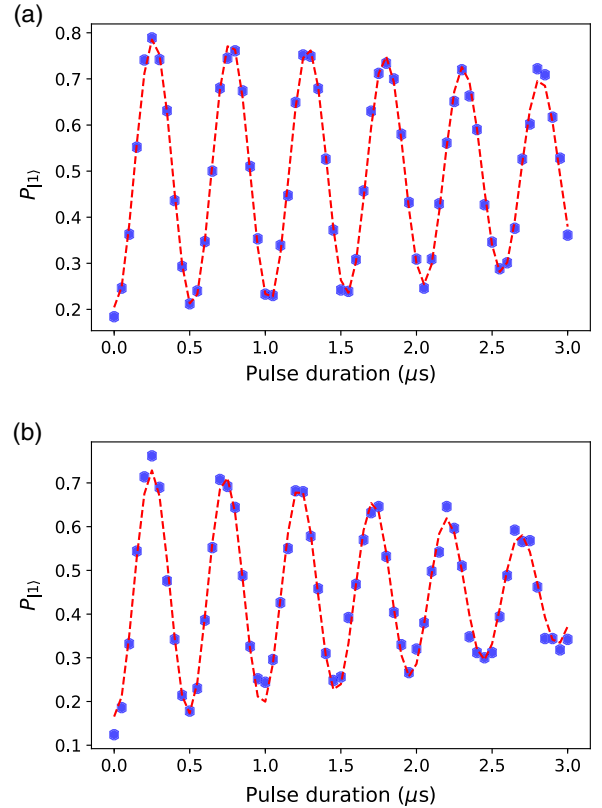


FIG. 6. Rabi oscillations of Q_2 . (a) Probability that measurement of Q_2 returns spin-up ($|1\rangle$) as a function of the duration of the resonant microwave burst driving the qubit. (b) Analogous data for Q_2 when Q_1 is simultaneously being driven.

used for $Q2$ (~ 8 dBm). This difference is needed to compensate for the tighter confining potential of dot 1 compared to dot 2.

-
- [1] I. L. Chuang and M. A. Nielsen, *Prescription for Experimental Determination of the Dynamics of a Quantum Black Box*, *J. Mod. Opt.* **44**, 2455 (1997).
- [2] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, *Quantum Process Tomography of a Controlled-NOT Gate*, *Phys. Rev. Lett.* **93**, 080502 (2004).
- [3] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, *Self-Consistent Quantum Process Tomography*, *Phys. Rev. A* **87**, 062119 (2013).
- [4] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, *Symmetrized Characterization of Noisy Quantum Processes*, *Science* **317**, 1893 (2007).
- [5] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Randomized Benchmarking of Quantum Gates*, *Phys. Rev. A* **77**, 012307 (2008).
- [6] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, *Randomized Benchmarking of Multiqubit Gates*, *Phys. Rev. Lett.* **108**, 260503 (2012).
- [7] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit*, *Phys. Rev. Lett.* **102**, 090502 (2009).
- [8] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Characterization of Addressability by Simultaneous Randomized Benchmarking*, *Phys. Rev. Lett.* **109**, 240504 (2012).
- [9] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking*, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [10] A. C. Dugas, J. J. Wallman, and J. Emerson, *Bounding the Average Gate Fidelity of Composite Channels using the Unitarity*, [arXiv:1610.05296](https://arxiv.org/abs/1610.05296).
- [11] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, *Process Verification of Two-Qubit Quantum Gates by Randomized Benchmarking*, *Phys. Rev. A* **87**, 030301(R) (2013).
- [12] Yu. Chen *et al.*, *Qubit Architecture with High Coherence and Fast Tunable Coupling*, *Phys. Rev. Lett.* **113**, 220502 (2014).
- [13] L. Casparis, T. W. Larsen, M. S. Olsen, F. Kueemmeth, P. Krogstrup, J. Nygård, K. D. Petersson, and C. M. Marcus, *Gatemon Benchmarking and Two-Qubit Operations*, *Phys. Rev. Lett.* **116**, 150505 (2016).
- [14] M. D. Shulman, O. E. Dial, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, *Demonstration of Entanglement of Electrostatically Coupled Singlet-Triplet Qubits*, *Science* **336**, 202 (2012).
- [15] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, *A Programmable Two-Qubit Quantum Processor in Silicon*, *Nature (London)* **555**, 633 (2018).
- [16] D. M. Zajac, A. J. Sigillito, M. Russ, F. Borjans, J. M. Taylor, G. Burkard, and J. R. Petta, *Resonantly Driven CNOT Gate for Electron Spins*, *Science* **359**, 439 (2018).
- [17] W. Huang, C. H. Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, A. Morello, A. Laucht, and A. S. Dzurak, *Fidelity Benchmarks for Two-Qubit Gates in Silicon*, [arXiv:1805.05027](https://arxiv.org/abs/1805.05027).
- [18] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, *A New Class of Efficient Randomized Benchmarking Protocols*, [arXiv:1806.02048](https://arxiv.org/abs/1806.02048).
- [19] F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, N. Coppersmith, and M. A. Eriksson, *Silicon Quantum Electronics*, *Rev. Mod. Phys.* **85**, 961 (2013).
- [20] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, and M. Veldhorst, *Interfacing Spin Qubits in Quantum Dots and Donors—Hot, Dense, and Coherent*, *npj Quantum Inf.* **3**, 34 (2017).
- [21] M. Pioro-Ladrière, T. Obata, Y. Tokura, Y.-S. Shin, T. Kubo, K. Yoshida, T. Taniyama, and S. Tarucha, *Electrically Driven Single-Electron Spin Resonance in a Slanting Zeeman Field*, *Nat. Phys.* **4**, 776 (2008).
- [22] L. M. K. Vandersypen and I. L. Chuang, *NMR Techniques for Quantum Control and Computation*, *Rev. Mod. Phys.* **76**, 1037 (2005).
- [23] T. Meunier, V. E. Calado, and L. M. K. Vandersypen, *Efficient Controlled-Phase Gate for Single-Spin Qubits in Quantum Dots*, *Phys. Rev. B* **83**, 121403(R) (2011).
- [24] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, *A Two-Qubit Logic Gate in Silicon*, *Nature (London)* **526**, 410 (2015).
- [25] J. M. Elzerman, R. Hanson, L. H. W. van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, *Single-Shot Read-Out of an Individual Electron Spin in a Quantum Dot*, *Nature (London)* **430**, 431 (2004).
- [26] V. Srinivasa, K. C. Nowack, M. Shafiei, L. M. K. Vandersypen, and J. M. Taylor, *Simultaneous Spin-Charge Relaxation in Double Quantum Dots*, *Phys. Rev. Lett.* **110**, 196803 (2013).
- [27] E. Magesan, J. M. Gambetta, and J. Emerson, *Scalable and Robust Randomized Benchmarking of Quantum Processes*, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [28] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing Quantum Gates via Randomized Benchmarking*, *Phys. Rev. A* **85**, 042311 (2012).

- [29] J.J. Wallman, *Randomized Benchmarking with Gate-Dependent Noise*, *Quantum* **2**, 47 (2018).
- [30] K. Takeda, J. Kamioka, T. Otsuka, J. Yoneda, T. Nakajima, M.R. Delbecq, S. Amaha, G. Allison, T. Kadera, S. Oda, and S. Tarucha, *A Fault-Tolerant Addressable Spin Qubit in a Natural Silicon Quantum Dot*, *Sci. Adv.* **2**, e1600694 (2016).
- [31] A. Carignan-Dugas, K. Boone, J.J. Wallman, and J. Emerson, *From Randomized Benchmarking Experiments to Gate-Set Circuit Fidelity: How to Interpret Randomized Benchmarking Decay Parameters*, *New J. Phys.* **20**, 092001 (2018).
- [32] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, *What Randomized Benchmarking Actually Measures*, *Phys. Rev. Lett.* **119**, 130502 (2017).
- [33] F. Martins, F.K. Malinowski, P.D. Nissen, E. Barnes, S. Fallahi, G.C. Gardner, M.J. Manfra, C.M. Marcus, and F. Kuemmeth, *Noise Suppression Using Symmetric Exchange Gates in Spin Qubits*, *Phys. Rev. Lett.* **116**, 116801 (2016).
- [34] M.D. Reed, B.M. Maune, R.W. Andrews, M.G. Borselli, K. Eng, M.P. Jura, A.A. Kiselev, T.D. Ladd, S.T. Merkel, I. Milosavljevic, E.J. Pritchett, M.T. Rakher, R.S. Ross, A.E. Schmitz, A. Smith, J.A. Wright, M.F. Gyure, and A.T. Hunter, *Reduced Sensitivity to Charge Noise in Semiconductor Spin Qubits via Symmetric Operation*, *Phys. Rev. Lett.* **116**, 110402 (2016).
- [35] <https://zenodo.org/record/2613798#.XJ0Czi-ZOuV>.
- [36] A.W. Cross, E. Magesan, L.S. Bishop, J.A. Smolin, and J.M. Gambetta, *Scalable Randomised Benchmarking of Non-Clifford Gates*, *npj Quantum Inf.* **2**, 16012 (2016).
- [37] A.K. Hashagen, S.T. Flammia, D. Gross, and J.J. Wallman, *Real Randomized Benchmarking*, *Quantum* **2**, 85 (2018).